Sound for Film and Television

Tomlinson Holman

TMH QUALIFIED

Third Edition

Focal Press

Sound for Film and Television

Third Edition

This page intentionally left blank

Sound for Film and Television

Third Edition

Tomlinson Holman

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.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

Contents

This book is an introduction to the art and technique of sound for film and television. The focus in writing the book has been to span the gulf between typical film and television production textbooks, with perhaps too little emphasis on underlying principles, and design engineering books that have few practical applications, especially to the film and television world. The guiding principle for inclusion in the text is the usefulness of the topic to film or video makers.

The first three chapters provide background principles of use to anyone dealing with sound, especially sound accompanying a picture, and, by way of examples, demonstrate the utility of the principles put into practice. The rest of the book walks through the course of a production, from the pickup of sound by microphones on the set to the reproduction of sound in cinemas and homes at the end of the chain.

How to Read This Book

Depending on who you are, there are various approaches you can take to reading this book.

If you have to start on a set tomorrow morning, read Chapter 4, Capturing Sound, and Chapter 8, Sync, Sank, Sunk, tonight. These two chapters contain the most salient features that you have to know to get started. Of these, the sync chapter is the harder one, and you may have to call the postproduction house to know what to do, explaining to them what you are about to embark on—above all, be careful: cameras that say they are 24 P may in fact be 23.976 P. Be sure to have camera, mixer/recorder, and slate model numbers so that the post house can help you. If the production has not determined a postproduction house, your audio rental facility should be helpful.

Then, having mastered the material in these chapters, move on to the other chapters related to recording sound, 5, 6, and 7. You will find in them concepts that the background provided by Chapters 1, 2, and 3 will be helpful in explaining. From there, work linearly through the book from Chapter 9 through Chapter 13. Note that the Glossary at the end should help in defining terms.

For a university course in film sound, I start with the first background chapters and proceed forward straight through the book. I do skip some material in a starting course, which is here for completeness but is beyond the scope of early courses. We use the book at multiple levels in our program at the University of Southern California.

For the sake of completeness, some information has been included that may be tangential to end users. This information has been made separate from the main text by being indented and of smaller type.

Examples

No study of film and television sound would be complete without listening to a lot of film and television shows. This is practical today in classrooms and at home because with a decent home theater sound system, available for a few thousand dollars, the principles given in the text can be demonstrated. Here are some film examples that are especially useful.

- *Citizen Kane:* the scene in the living room at Xanadu in which Kane and his love interest interact, photographed with the great depth of field that was the innovation of Greg Toland for the picture. The sound in this scene can be contrasted with that in the next scene of Kane and his girlfriend in the back seat of a car. In the first scene, the sound is very reverberant, emphasizing the separation of the characters. In the second, the sound is very intimate, emphasizing the change of scene that has taken place. Orson Welles brought the techniques he had learned in radio to films. This is used to illustrate Chapter 1, Objective Sound, and the difference that attention to such factors can make.

- *Days of Heaven,* reel 1: from opening to arrival of the train at the destination. After an opening still photo montage accompanied by music, we come in on an industrial steel mill in which the sound of the machinery is so loud we often cannot hear the characters. A fight ensues between a worker and his boss, and the content of the argument is actually made stronger by the fact of our not being able to discern it. This illustrates frequency masking, a topic in Chapter 2. A train leaves for the country then, accompanied by music, and the question posed is: Do we hear the train or not and what does it matter if we do or don't? A voice-over narration illustrates the speech mode of perception, when it abruptly enters and demands our attention. The lyrical music accompanied by the train motion is a strong contrast with the sound that has come before, and is used in the vaudeville sense of "a little traveling music, please"—making it an effective montage. At the end of the scene there is a cross-fade

between the music and the reality sound that puts an end to the montage, punctuating the change in mood.

- *Das Boot,* reel 1: the entrance of the officers into the submarine compound until the first shot of the submarine on the open ocean illustrates many things. At first the submarine repair station interior is very noisy, and the actors have to raise their voices to be heard. Actually, the scene was almost certainly looped, therefore it was the direction to the actors that caused their voices to be raised, not the conditions under which they were recorded. Next, the officers come upon their boat, and despite the fact they are still in a space contiguous with the very noisy space, the noisy background gives way to a relatively quiet one, illustrating a subjective rather than a totally objective view. Then the submarine leaves the dock, accompanied by a brass band playing along in sync (an example of prerecorded, or at least postsynced, sound). The interior of the boat is established through the medium of telling a visitor where everything is. Sound is used to help establish each space in the interior: noise for the men's quarters, and a little music for the officer's, Morse code for the radio room, and mechanical noise for the control room. Next we come upon a door from behind which we hear a loud sound, then going through the door, we find it is the engine room with the very noisy engine causing the actors to speak loudly once again. The whole reel, up to the going to sea shot, is useful in the ways that sound is used to tell the story.

- *Cabaret* is useful for two principal purposes. The first is to show a scene that involved extensive preproduction preparation of a music recording, then filming, then using the prerecorded and possibly postrecorded materials after the picture was edited to synchronize to the perspective of the picture. The scene is of the Nazi boy singing "The sun on the meadow is summery warm . . ." until the main characters leave with the question, "Do you still think you will be able to control them?" What is illustrated here is very well controlled filmmaking, for we always hear what we expect to, that is, sound matched to picture, but over a bed of sound that stays more constant with the picture cuts. The second point of using *Cabaret* is that filmmaking is a powerful, but neutral, tool that can help move people to heights of folly. Whether the techniques taught here are used for good or ill is in the hands of the filmmaker.

- *Platoon* demonstrates a number of sound ideas in the final battle scene. Despite the fact that it is difficult to understand the characters while they are under fire, the effect of their utterances is bone chilling nevertheless. The absolutely essential lines needed for exposition are clearly exposed, with practically no competition from sound effects or music. On the other hand, there is one line that can be understood only by lip reading, because it is covered by an explosion. Still, the meaning is clear and the "dialog" can be understood because the words spoken are so right in the context of the scene.

Other films that I have found to be of enduring interest are listed in the Appendix II Filmography at the end of the book.

ACKNOWLEDGMENTS

Art Baum read the entire book in minute detail and provided great feedback. Martin Krieger also read the entire manuscript for clarity. Others read specific areas of their expertise, which is often beyond my own. Mark Schubin, of the live Metropolitan Opera broadcasts, and Paul Chapman, Senior Vice President of Technology at Fotokem, were particularly helpful on synchronization issues. The discussion of the Gestalt psychologists and psychoacoustics in Chapter 2 owes a great deal to Dr. Brian C. J. Moore's book *An Introduction to the Psychology of Hearing*. Other readers included colleagues Tom Abrams, Midge Costin, and Don Hall. Dr. Dominic Patawaran contributed to my well-being during the writing of this book.

DEDICATION

This work is dedicated to the hardworking men and women, often unsung, who perform feats of skill and amazing perseverance every day in the making of sound for film and video.

SOUND FOR FILM AND TELEVISION DEFINED

Sound for film and television is an aural experience constructed to support the story of a narrative, documentary, or commercial film or television program. Sound may tell the story directly, or it may be used indirectly to enhance the story. Although there are separate perceptual mechanisms for sound and picture, the sound may be integrated by the audience along with the picture into a complete whole, without differentiation. In such a state, the sound and picture together can become greater than the sum of the parts.

In most instances, film and television sound for entertainment and documentary programming is constructed in postproduction by professionals utilizing many pieces of sound mixed seamlessly together to create a complete whole. The sources used for the sound include recordings made during *principal photography* on sets or on location, sound effects libraries and customized recordings, and music, both composed for the film and from preexisting sources. Sound for film and television is thus a thoroughly constructed experience, usually meant to integrate many elements together seamlessly and not draw specific attention to itself.

The relative roles of picture and sound can change with regard to storytelling from scene to scene and moment to moment. A straight narrative picture will probably have dialog accompanying it, whereas a picture montage will often be accompanied by music, or at least manipulated sound effects, as the filmmaker varies the method of storytelling from time to time to add interest to the film and provide a moment for audiences to soak up the action, make scene transitions, and so forth.

Nearly everyone involved in the production of a film or television program affects, and is affected by, sound. Writers use sound elements in their storytelling, with suggestions in the script for what may be heard. Location scouts should note bad noise conditions at potential shooting sites because, although the camera can "pan off" an offending sign, there is no such effective way to eliminate airplanes flying over from the soundtrack—the "edges" of a sound frame are not hard like those of a picture frame. Directors need to be keenly aware of the potential for sound, for what they are getting on location and what can be substituted in postproduction, as sound is "50 percent of the experience" according to a leading filmmaker.

Cinematographers can plan lighting so that a sound boom is usable, with the result being potentially far better sound. Costumers can supply pouches built into clothing that can conceal microphones and can supply booties so that actors can wear them for low noise when their feet don't show. Grips, gaffers, and set dressers can make the set quiet and make operable items work silently. Often, the director need only utter the word to the crew that sound is important to him or her for all this to occur.

ROLES OF SOUND

Many kinds of sound have a direct storytelling role in filmmaking.[1] Dialog and narration tell the story, and narrative sound effects can be used in such a capacity, too, for example, to draw the attention of the characters to an off-screen event. Such direct narrative sound effects are often written into the script, because their use can influence when and where actors have to take some corresponding action.

Sound also has a subliminal role, working on its audience subconsciously. Whereas all viewers can tell the various objects in a picture apart—an actor, a table, the walls of a room—listeners barely ever perceive sound so analytically. They tend to take sound in as a whole, despite its actually being deliberately constructed from many pieces. Herein lies the key to an important storytelling power of sound: the inability of listeners to separate sound into ingredient parts can easily produce "a willing suspension of disbelief" in the audience, because they cannot separately discern the functions of the various sound elements. This fact can be manipulated by filmmakers to produce a route to emotional involvement in the material by the audience. The most direct example of this effect is often the film score. Heard in isolation, film scores[2] often do not make much musical sense; the music is deliberately written to enhance the mood of a scene and to underscore the action, not as a foreground activity, but a background one. The function of the music is to "tell" the audience

---

[1]This term is used instead of the clumsier, but more universal, "program making." What is meant here and henceforth when terms such as this are used is the general range of activities required to make a film, video, or television program.
[2]The actual score played with the film, not the corresponding music-only CD release.

how to feel, from moment to moment: soaring strings mean one thing, a single snare drum, another.

Another example of this kind of thing is the emotional sound equation that says that low frequencies represent a threat. Possibly this association has deep primordial roots, but if not, exposure to film sound certainly teaches listeners this lesson quickly. A distant thunderstorm played underneath an otherwise sunny scene indicates a sense of foreboding, or doom, as told by this equation. An interesting parallel is that the shark in *Jaws* is introduced by four low notes on an otherwise calm ocean, and there are many other such examples.

Sound plays a grammatical role in the process of filmmaking too. For instance, if sound remains constant before and after a picture cut, the indication being made to the audience is that, although the point of view may have changed, the scene has not shifted—we are in the same space as before. So sound provides a form of continuity or connective tissue for films. In particular, one type of sound represented several ways plays this part. *Presence* and *ambience* help to "sell" the continuity of a scene to the audience.

SOUND IS OFTEN "HYPERREAL"

Sound recordings for film and television are often an exaggeration of reality. One reason for this is that there is typically so much competing sound at any given moment that each sound that is recorded and must be heard has to be rather overemphatically stated, just to "read" through the clutter. Heard in isolation, the recordings seem silly, overhyped; but heard in context, they assume a more natural balance. The elements that often best illustrate this effect are called *Foley* sound effects. These are effects recorded while watching a picture, such as footsteps, and are often exaggerated compared to how they would be in reality, both in loudness and in intimacy. Although some of this exaggeration is due to the experience of practitioners finding that average sound playback systems obscure details, a good deal of the exaggeration still is desirable under the best playback conditions, simply because of the competition from other kinds of sound.

SOUND AND PICTURE

Sound often has an influence on picture, and vice versa. For instance, making picture edits along with downbeats in a musical score often makes the picture cuts seem very right. In *The Wonderful Horrible Life of Leni Riefenstahl,* we see Hitler's favorite filmmaker teaching us this lesson, for she cut the waving flags in the Nuremberg Nazi rally in *Triumph of the Will* into sync with the music, increasing the power of the scene to move people.

Scenes are different depending on how sound plays out in them. For example, "prelapping" a sound edit before a scene-changing picture edit[3] simply feels different from cutting both sound and picture simultaneously. The sense is heightened that the outgoing scene is over, and the story is driven ahead. Such a decision is not one taken at the end of the process in postproduction by a sound editor typically, but more often by the picture editor and director working together, because it has such a profound impact on storytelling. Thus involvement with sound is important not only to those who are labeled with sound-oriented credits, but also to the entire filmmaking process represented by directing and editing the film.

SOUND PERSONNEL

Sound-specific personnel on a given film or television job may range from one person, that being the camera person on a low-budget documentary with little postproduction, to quite large and differentiated crews as seen in the credits of theatrical motion pictures. In typical feature film production, a production sound recordist serves as head of a crew, who may add one or more boom operators and cable persons as needed to capture all the sound present. On television programs shot in the multicamera format, "filmed in Hollywood before a live studio audience," an even larger crew may be used to control multiple boom microphones, to plant microphones on the set, and to place radio microphones on actors, and then mix these sounds to a multitrack tape recorder. Either of these situations is called production sound recording.

Following in postproduction, picture editors cut the production soundtrack along with the picture, so that the story can be told throughout a film. They may add some additional sound in the way of principal sound effects and music, producing, often with the help of sound-specific editors, "temp mixes" useful in evaluating the current state of a film or video in postproduction. Without such sound, audiences, including even sophisticated professional ones, cannot adequately judge the program content, as they are distracted by such things as cutting to silence. By stimulating two senses, program material is subject to a heightened sensation on the part of the viewer/listener, which would not occur if either the picture or the sound stood alone. A case in point is one of an observer looking at an action scene silently, and then with ever increasing complexity of sound by adding each of the edited sound sources in turn. The universal perception of observers under these conditions is that the picture appears to run faster with more complex sound, despite the fact

---

[3]By cutting to the sound for the incoming scene before the outgoing picture changes.

that precisely the same time elapses for the silent and the sound presentations: the sound has had a profound influence on the perception of the picture.

When the picture has been edited, sound postproduction begins in earnest. Transfer operators take the production sound recordings and transfer them to an editable format such as into a digital audio workstation. Sound editors pick and place sound, drawing on production sound, sound effects libraries, and specially recorded effects, which are also all transferred to an editable format. From the edited soundtracks, various mixes are made by rerecording mixers (called dubbing mixers in England). Mixing may be accomplished in one or more steps, more generations becoming necessary as the number of soundtracks cut increases to such an extent that all the tracks cannot be handled at one time. The last stage of postproduction mixing prepares masters in a format compatible with the delivery medium, such as optical sound on film, or videotape.

THE TECHNICAL VERSUS THE AESTHETIC

Although it has a technical side, in the final analysis what is most important for film and television sound is what the listener hears, that is, what choices have been made throughout production and postproduction by the filmmakers. Often, thoughts are heard from producers and others such as, "Can't you just improve the sound by making it all digital?" In fact, this is a naive point of view, because, for instance, what is more important to production sound is the microphone technique, rather than the method of tape recording. Unwanted noise on the set is not reduced by digital recording and often causes problems, despite the method used to record the production sound.

When film sound started in the late 1920s, the processes to produce the soundtrack were very difficult. Camera movement was restricted by large housings holding both the camera and the cameraman so that noise did not intrude into the set. Optical soundtracks were recorded simultaneous with the picture on a separate sound camera and could not be played back until the film was processed and printed and the print processed. Microphones were insensitive, so actors had to speak loudly and clearly. Silent movie actor's careers were on the line, as it was discovered by audiences that many of them had foreign accents or high, squeaky voices.

Today, the technical impediments of early sound recording have been removed. Acting styles are much more natural, with it more likely that an actor will "underplay" a scene because of the intimacy of the camera than "overplay" it. Yet the quality achieved in production sound is still subject to such issues as whether the set has been made quiet and whether the actor enunciates or mumbles his or her lines. Many directors pass all problems in speech intelligibility to the sound "technician," who, after all, is supposed to be able to make a high-quality recording even if the director can't hear the actor on the set!

A CONFUSION WITH DIRECTING ACTORS

One confusion for actors is that the frame of reference for what is left and what is right changes between theater and film. Actors have had the notion of left and right beaten into them, that it is from their vantage point facing the audience, called *stage left* and *stage right*. However, film and television employ the opposite convention. Called *camera left* and *camera right,* the point of view is that of the camera. This confusion has slowed down more than one production over the course of time, when the director yells "Go left," and the actors move camera right.

THE DIMENSIONS OF A SOUNDTRACK

The "dimensions" of a soundtrack may be broken down for discussion into frequency range, dynamic range, the spatial dimension, and the temporal dimension. A major factor in the history of sound accompanying pictures is the growth in the capabilities associated with these dimensions as time has gone by, and the profound influence this growth has had on the aesthetics of motion-picture soundtracks. Whereas early sound films had a frequency range capability (bandwidth) only about that of a telephone, steady growth in this area has produced modern soundtrack capabilities well matched to the frequency range of human hearing. Dynamic range capability improvements have meant that both louder and softer sounds are capable of being reproduced and heard without audible distortion or masking. Stereophonic sound added literally new dimensions to film soundtracks, first rather tentatively in the 1950s with magnetic sound release prints and then firmly with optical stereo prints in the 1970s, which have had continued improvement ever since. Still, even the monophonic movies of the 1930s benefited from one spatial dimension: adding reverberation to soundtracks helped place the actors in a scene and to differentiate among narration, on-screen dialog, off-screen sound effects, and music.

This page intentionally left blank

Objective Sound

AN OLD STORY

A tree falls over in a wood. Does it make a sound? From one point of view, the answer is that it must make a sound, because the physical requirements for sound to exist have been met. An alternate view is that without any consciousness to "hear" the sound, there in fact is no sound. This dichotomy demonstrates the difference between this chapter and the next. A physicist has a ready answer—of course, there is a great crashing noise. On the other hand, a humanist philosopher thinks consciousness may well be required for there to be a sound.

The dichotomy observed is that between objective physical sound and subjective psychoacoustics. Any sound undergoes two principal processes for us to "hear" it. First, it is generated and the objective world acts on it, and then that world is represented inside our minds by the processes of hearing and perception. This chapter concentrates on the first part of this process—the generation and propagation of physical sound—whereas Chapter 2 discusses how the physical sound is represented inside our heads.

The reason the distinction between the objective and the subjective parts of sound perception is so important is that in finding cause and effect in sound, it is very important to know the likely source of a problem: Is there a real physical problem to be solved with a physical solution, or does the problem require an adjustment to accommodate human listening? Any given problem can have its roots in either domain and is often best solved in its own dominion. On the other hand, there are times when we can provide a psychoacoustical solution to what is actually an acoustical problem.

An example of this is that often people think there is an equipment solution to what is, in fact, an acoustical problem. A high background noise level of a location cannot be solved with digital recording, for instance, although some people give so much credit to digital recording that they wonder whether this might not be true. "It's digital, so we won't need to do any postproduction, right?" has been asked naively of more than one sound mixer.

PROPERTIES OF PHYSICAL SOUND

There are several distinguishing characteristics of sound, partly arising from the nature of the source of the sound (is it big or small? does it radiate sound equally in all directions, or does it have a directional preference?) and partly from the prevailing conditions between the point of origin and the point of observation (is there any barrier or direct line of sight?). Sound propagates through a medium such as air at a specific speed and is acted on by the physical environment.

Propagation

Sound travels from one observation point to another by a means that is analogous to the way that waves ripple outward from where a stone has been dropped into a pond. Each molecule of the water interacts with the other molecules around it in a particular orderly way. A given volume of water receives energy from the direction of the source of the disturbance and passes it on to other
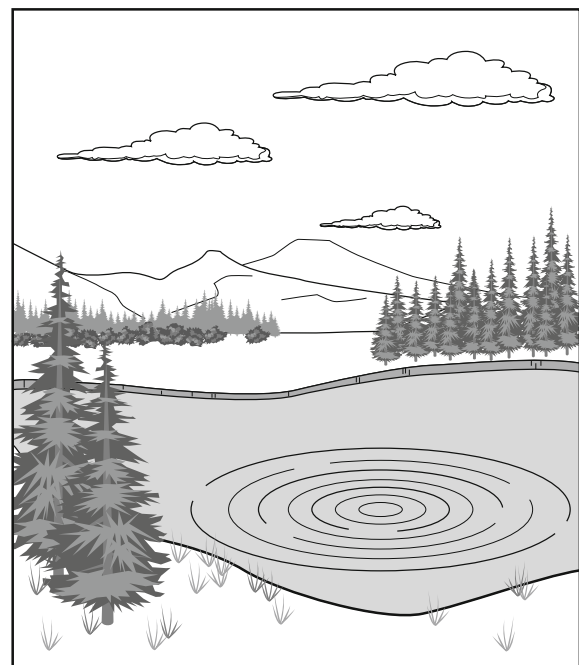
FIGURE 1.1 The waves resulting from a stone dropping into a pond radiate outward, as do sound waves from a point source in air, only in three dimensions, not two.
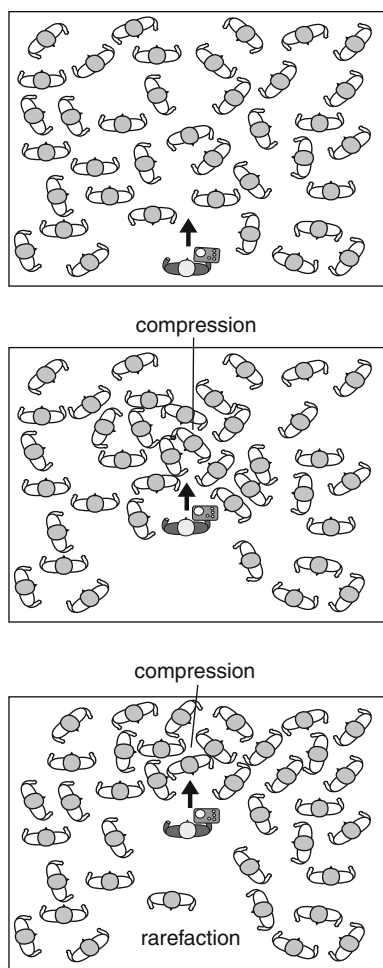
FIGURE 1.2   A waiter on the dance floor compresses dancers in front of him and leaves a rarefied space behind him.

of his way, and when they then start to bump into their neighbors, the neighbors move away, etc. The disturbance may be very small by the time it reaches the other side of the dance floor, but the action of the waiter has disturbed the whole crowd, more or less. If the waiter were to step in, then out, of the crowd, people would first be compressed together and then be spread apart, perhaps farther than they ever had been while dancing. The waiter in effect leaves a vacuum behind, which people rush in to fill. The two components of the disturbance are called *compression,* when the crowd is forced together more closely than normal, and *rarefaction,* when the spacing between the people is more than it is normally.

The tines of a tuning fork work like the waiter on the dance floor, only the dancers are replaced by the air molecules around the tuning fork. As the tines move away from the center of the fork, they compress the outside air molecules, and as they reverse direction and move toward one another, the air becomes rarefied (Fig. 1.3). Continuous, cyclical compression and rarefaction form the steady tone that is the recognized sound of a tuning fork.

Our analogy to water ripples can be carried further. In a large, flat pond, the height of the waves gets smaller as we go farther from the origin, because the same amount of energy is spread out over a larger area. Sound is like this, too, only the process is three-dimensional, so that by spreading out over an expanding surface, like blowing up a balloon, the energy farther from the source is even less. The "law" or rule that describes the amplitude of the sound waves falling off with distance is called the

water that is more distant from the source, causing a circular spreading of the wave. Unless the stone is large and the water splashes, the water molecules are disturbed only about their nominal positions, but eventually they occupy about the same position they had before the disturbance.

Consider sound in air for a moment. It differs from other air movement, like wind or drafts, by the fact that, on the whole, the molecules in motion will return to practically the same position they had before the disturbance. Although sound is molecules in motion, there is no *net* motion of the air molecules, just a passing disturbance.

Another way to look at how sound propagates from point to point is to visualize it as a disturbance at a dance. Let us say that we are looking down on a crowded dance floor. With contemporary dancing, there isn't much organization to the picture from above—the motion is random. A waiter, carrying a large tray, enters the dance floor. The dancers closest to the waiter have to move a lot to get out

FIGURE 1.3   The tines of a tuning fork oscillate back and forth, causing the nearby air to be alternately rarefied and compressed.

inverse square law. This law states that when the distance to a sound source doubles, the size of the disturbance diminishes to one-quarter of its original size:

$$\text{Strength of sound at a distant point} = \text{original strength}/\text{distance}^2.$$

Track 4 of the DVD that accompanies this book illustrates the inverse square law effect of level versus distance.

The inverse square law describes the fall-off of sound energy from a *point source,* radiating into free space. A point source is a source that is infinitesimal and shows no directional preference.

In actual cases, most sources occupy some area and may show a directional preference, and most environments contain one or more reflecting surfaces, which conspire to make the real world more complex than this simple model.

> One of the main deviations from this model comes when a source is large, say, a newspaper press. Then it is difficult to get far enough away to consider this to be a point source, so the falloff with distance will typically be less than expected. This causes problems for narrative filmmakers trying to work in a press room, because not only is the press noisy, but the falloff of the noise with distance is small.
>
> Another example is an explosion occurring in a mine shaft. Within the shaft, sound will not fall off according to the inverse square law because the walls prevent the sound pressure from spreading. Therefore, even if the sound of the explosion is a great distance away, it can be nearly as loud as it is near its source, and quite dangerous to the documentary film crew members who think that they are sufficiently far away from the blast to avoid danger.

The water analogy we used earlier falls apart when we get more specific. Ripples in water are perpendicular to the direction of propagation—that is, ripples in a pond are up and down. These are called transversal waves. Sound waves, on the other hand, are longitudinal; that is, they are in the direction of travel of the wave. Visualize a balloon
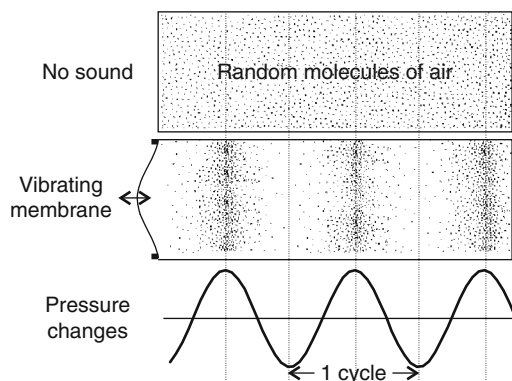
blowing up with a constant rate of inflation equal to the speed of sound, while its surface is oscillating in and out, and you have a good view of sound propagation.

A Medium Is Required

Sound requires a medium. There is no medium in the vacuum of outer space, as Boyle discovered in 1660 by putting an alarm clock suspended by a string inside a well-sealed glass jar. When the air was pumped out of the jar to cause a vacuum and the alarm clock went off, there was no sound; but when air was let back in, sound was heard. This makes sense in light of our earlier discussion of propagation: If the waiter doesn't have anything to disturb on the dance floor, he can hardly propagate a disturbance.

For physicists, the famous opening scene of *Star Wars* makes no sense, with the rumble from its spaceships arriving over one's head, first of the little ship and then the massive Star Destroyer. No doubt the rumble is effective, but it is certain that somewhere a physicist groaned about how little filmmakers understand about nature. Here is an example of where the limitations of physics must succumb to storytelling. Note that although radio signals and light also use wave principles for propagation, no medium is required: These electromagnetic waves travel through a vacuum unimpeded.

Speed of Sound

The speed of sound propagation is quite finite. Although it is far faster than the speed of waves in water caused by a stone dropping, it is still incredibly slower than the speed of light. You can easily observe the difference between light and sound speed in many daily activities. Watch someone hammer a nail or kick a soccer ball at a distance, and you can easily see and hear that the sound comes later than the light—reality is out of sync! Filmmakers deal with this problem all the time, often forcing into sync sounds that in reality would be late in time. This is another example of film reality being different from actual reality. Perhaps because of all of the training that we have received subliminally by watching thousands of hours of film and television, reality for viewers has been modified: *Sound should be in hard sync with the picture,* deliberately neglecting the effects of the speed of sound, unless a story point sets up in advance the disparity between the arrival times for light and sound.

The speed of sound is dependent on the medium in which the sound occurs. Sound travels faster in denser media, and so it is faster in water than in air and faster in steel than in water. The black-hatted cowboy puts his ear to the rail to

FIGURE 1.4 Sound is the organized pressure changes above and below ambient pressure caused by a stimulating vibration, such as a loudspeaker.

hear the train coming sooner than he can through the air, partly because of the faster speed of sound in the material and partly because the rail "contains" it, with only a little sound escaping to the air per unit of length.

Sound travels 1130 ft/sec in air at room temperature. This is equal to about 47 ft of travel per frame of film at 24 frames per second. Unfortunately, viewers are very good at seeing whether the sound is in sync with the picture. Practically everyone is able to tell if the sync is off by two frames, and many viewers are able to notice when the sound is one frame out of sync!

> Because sound is so slow relative to light, it is conventional lab practice to pull up the sound on motion picture analog release prints one extra frame, printing the sound "early," and thus producing exact picture–sound sync 47 ft from the screen. (Picture and sound are also displaced on prints for other reasons; the one frame is added to the other requirements.) In very large houses, such as Radio City Music Hall or the Hollywood Bowl, it is common practice to pull up the sound even more, putting it into sync at the center of the space. Still, the sound is quite noticeably out of sync in many seats, being too early for the close seats and too late for the distant ones. Luckily, this problem is mostly noticeable today only in those few cases in which the auditoriums are much larger than the average theater. Because of the one-frame pull-up built into all prints, for a listener to be two frames out of sync, the listener would have to be three frames away from the screen, or about 150 ft. The Hollywood Bowl measures 400 ft from the stage to the back row, so the sync problems there are quite large and are made tolerable only by the small size of lips when seen from such a large distance (see Figure 1.5). The speed of sound is fairly strongly influenced by temperature (see *speed of sound* in the Glossary), so calculations of it yield different results in the morning compared to a warm afternoon, when the speed of sound is faster.

Amplitude

The "size" of a sound wave is known by many names: volume, level, amplitude, loudness, sound pressure, velocity, or intensity. In professional sound circles, this effect is usually given the name *level.* A director says to a mixer "turn up the level," not "turn up the volume," if he or she wants to be taken seriously.
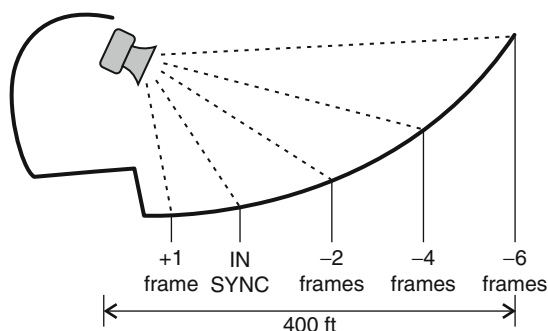
FIGURE 1.5 The Hollywood Bowl is over 400 ft long, and sound from the front of the house is significantly out of sync by the time it reaches the back.

The size of a sound disturbance can be measured in a number of ways. In the case of water ripples, we could put a ruler into the pond, perpendicular to the surface, and note how large the waves are, from their peak to their trough, as first one and then the other passes by the ruler. This measurement is one of amplitude.

> When reading the amplitude of a wave, it is customary to call the measurement just defined "peak-to-peak amplitude," although what is meant is the vertical distance from the peak to the trough. Confusion occurs when trying to decide which dimension to measure. If asked to measure the peak-to-peak amplitude of a wave, you might think that you should measure from one peak to the next peak occurring along the length of the wave, but that would give you the wavelength measurement, which is discussed in the next section, not the peak-to-peak amplitude.

In sound, because it is more easily measured than amplitude directly, what is actually measured is *sound pressure level,* often abbreviated SPL. *Sound pressure* is the relative change above and below atmospheric pressure caused by the presence of the sound.

Atmospheric pressure is the same as barometric pressure as read on a barometer. It is a measure of the force exerted on an object in a room by the "weight" of the atmosphere above it, about 15 lb/inch^2. The atmosphere exerts a steady force measured in pounds per square inch on everything. Sound pressure adds to (compression) and subtracts from (rarefaction) the static atmospheric pressure, moment by moment (see Figure 1.6). The changes caused by sound pressure during compression and rarefaction are usually quite small compared with barometric pressure, but they can nonetheless be readily measured with microphones.

> Although measuring sound pressure is by far the most common method of measurement, alternative techniques that may yield additional information are available. For our purposes, we can say that all measures of size of the waveform—including amplitude, sound pressure level, sound velocity, and sound intensity—are members of the same family, and so we will henceforth use sound pressure level as the measure because it is the most commonly used.
>
> Sound intensity, in particular, provides more information than sound pressure because it is a more complex measure, containing information about both the amplitude of the wave and its direction of propagation. Thus, sound intensity measurements are very useful for finding the source of a noise. Sound intensity measures are rarely used in the film and television industries, though, because of the complexity and cost of instrumentation.

Wavelength and Frequency

Wavelength

Another measure of water waves or sound waves we have yet to discuss is the distance from one peak, past one trough, to the next peak along the length, called the *wavelength.* Note that wavelength is perpendicular to the
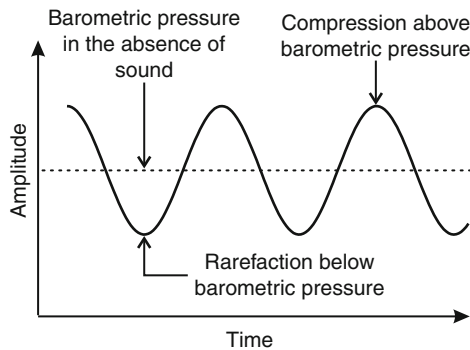
FIGURE 1.6   Sound pressure adds to (compression) and subtracts from (rarefaction) the static atmospheric pressure.

amplitude dimension, so the two have little or nothing to do with each other. One can have a small, long wave or a large, short one (a tsunami!). The range of wavelengths of audible sound is extremely large, spanning from 56 ft (17 m) to $^3/_4$ inch (1.9 cm) in air.

> Notice how our purist discussion of objective sound has already been circumscribed by psychoacoustics. We just mentioned the *audible* frequency range, but what about the inaudible parts? Wavelengths longer than about 56 ft or shorter than about $^3/_4$ inch still result in sound, but they are inaudible and will be covered later. The wavelength range for visible light, another wave phenomenon, covers less than a 2:1 ratio of wavelengths from the shortest to the longest visible wavelength, representing the spectrum from blue through red. Compared to this, the audible sound range of 1000:1 is truly impressive.

Track 5 of the DVD contains tones at 100 Hz, 1 kHz, and 10 kHz, having a wavelength in air of 11 ft 3 inches (3.4 m), $13^1/_2$ inches (34.4 cm), and $^3/_8$ inches (34.4 mm), respectively.

Frequency

Wavelength is directly related to one of the most important concepts in sound, *frequency*. Frequency is the number of occurrences of a wave per second. The unit for frequency is hertz (abbreviated Hz, also used with the "k" operator for "kilo" to indicate thousands of Hz: 20 kHz is shorthand for 20,000 Hz).

Wavelength and frequency are related reciprocally to the speed of sound such that as the wavelength gets shorter, the frequency gets higher. The frequency is equal to the speed of sound divided by the wavelength:

$$f = c/\lambda$$

where f is the frequency in Hz (cycles per second), c is the speed of sound in the medium, and λ is the wavelength. Note that the units for speed of sound and wavelength may be metric or English, but must match each other.

Thus the frequency range that corresponds to the wavelength range given earlier is from 20 Hz to 20 kHz,

which is generally considered to be the audible frequency range. Within this range the complete expressive capability that we know as sound exists. The frequency range in which a sound primarily lies has a strong storytelling impact. For example, in the natural world low frequencies are associated with storms (distant thunder), earthquakes, and other natural catastrophes.

When used in film, low-frequency rumble often denotes a threat is present. This idea extends from sound effects to music. An example is the theme music for the shark in *Jaws*. Those four low notes that begin the shark theme indicate that danger lurks on an otherwise pleasant day on the ocean. Alternatively, the quiet, high-frequency sound of a corn field rustling in *Field of Dreams* lets us know that we can be at peace, that there is no threat, despite its connection to another world.

Infrasonic Frequencies

The frequency region below about 20 Hz is called the *infrasonic* (or, more old fashioned, the *subsonic*) range, although the lowest note on the largest pipe organs corresponds to a frequency of 16 Hz, and this is still usually
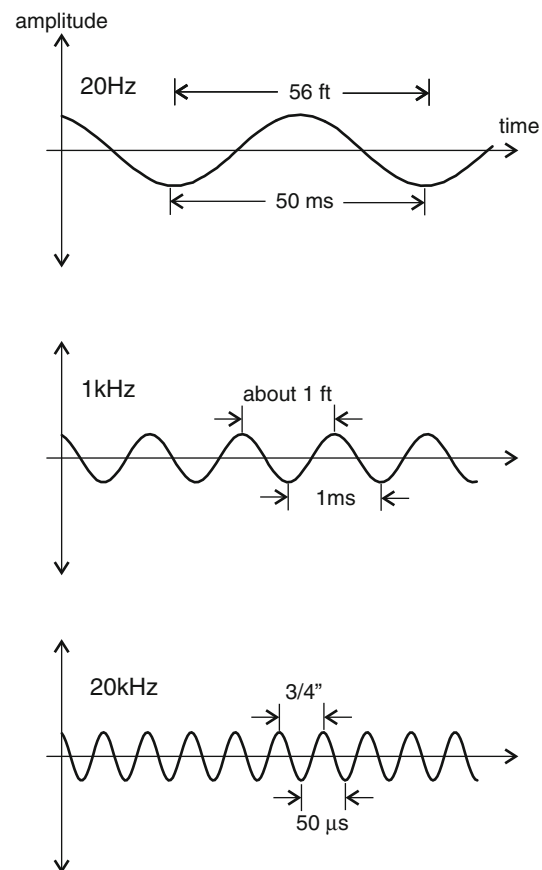
FIGURE 1.7   Wavelength and frequency of sound in air over the audible frequency range.

considered audible sound, not infrasonics.[1] This region is little exploited deliberately in film and television production, because problems can arise on sets when there is a very large amount of infrasonic noise, a not uncommon finding in industrial settings. The infrasonic sound level can be so high that it modulates voices—that is, it "roughens" their sound, making them sound as though spoken with a gurgle.

The frequency range around 12 Hz, when present at very high amplitude, may cause another problem—nausea. It has been found that people working in buildings with large amounts of structure-borne noise at 12 Hz may become ill from the high level of vibration and the consequent infrasonic sound. Fortunately, in most cases this effect has yet to be exploited by filmmakers.

There was a sound enhancement system used on several pictures called Sensurround™, which employed large amounts of low-frequency energy to cause a "rumble" effect, useful to simulate ground movement during the film *Earthquake* (1974) and for the aircraft carrier takeoffs and landings in the film *Midway* (1976). The frequency range of the loudspeakers was from 15 to 100 Hz, so the system probably did not have enough energy as low as 12 Hz to stimulate this effect. On the other hand, Sensurround pointed the way to the expressive capability of very low frequency sound, which was followed up over the years by the addition of subwoofers in theaters and separate soundtracks prepared for them.

Filmmakers exploit the lack of audibility of infrasonic sound on sets sometimes. "Thumpers," which put out very low frequency sound, are used to cue actors in dance numbers, for instance. Subsequent postproduction practice can filter[2] out such low-frequency sound, retaining the actor's vocal performance.

Ultrasonic Frequencies

The frequency region beyond 20 kHz is called *ultrasonic*. (The word supersonic is generally relegated to speed of aircraft, not frequencies of sound.) Although some people can hear high-level sounds out to as much as 24 kHz, most sound devices use 20 kHz as the limit of their range, focusing on the huge importance of sound below this frequency compared with the very minimal effects above it. There are several types of devices that employ frequencies above the audible range, but they are of little common interest to filmmakers.

They include acoustical burglar alarms (although many that claim to be ultrasonic actually operate around 17 kHz and can

be heard, and perhaps worse recorded, on location at times), some television remote controls, and specially built miniature loudspeakers made for acoustical testing of models.

One common problem in recording *stages* (the film term for recording spaces, different from the music term *studio*) is the use of conventional television monitors. Older NTSC color television uses a horizontal sweep rate of 15,734 Hz, well within the audible range. Many old video monitors emit strong acoustic energy at this frequency and must be acoustically shielded from microphones or else a filter must be employed to avoid recording this audible sound.

Importance of Sine Waves

So far what we have been considering are waves that have two dimensions, amplitude and wavelength. We have not yet considered the shape that the wave takes, the waveform.

For simple sources, like a tuning fork, in which the motion of the tines oscillates back and forth like the swing of a pendulum, the waveform traced out over one cycle is a *sine wave*. Sources more complex than a simple tuning fork emit more complex waveforms than a sine wave because the motion that produces their sound is more complicated. A violin string playing a note is a good example. The string exhibits complex motion, with parts of it moving in one direction and adjacent parts moving in the opposite direction at the same instant, in a complex manner. In 1801, a French mathematician, Jean Baptiste Fourier, made a very important theoretical breakthrough. He found that complex systems, such as a moving violin string, could be broken down into a number of basic ingredients, which, when added together, summed to describe the whole complex motion of the string. Fourier found that all complex motion, including sound, could be described as a summation of multiple sine waves.

A waveform may change rapidly in time, as it does when a violinist goes from note to note, and even within
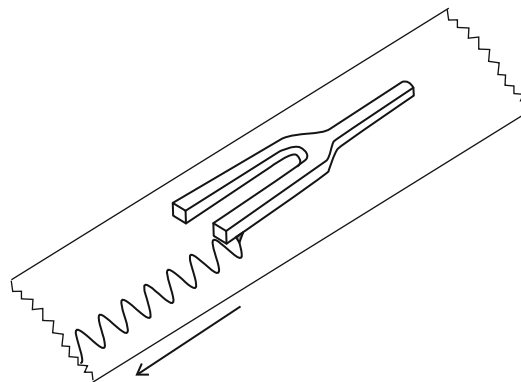
FIGURE 1.8 One tine of a tuning fork moving traces out a sine wave as time goes by. The shape of the wave versus time is called the waveform.

---

[1]You can see that the 20-Hz low-frequency limit is a little "fuzzy."
[2]See Chapter 12.
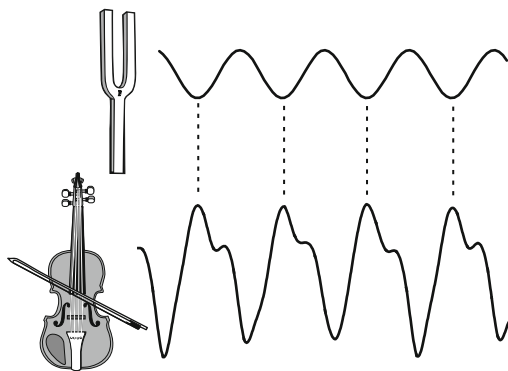
FIGURE 1.9 Whereas a tuning fork produces a sine wave, more complex motion, such as of a violin string, results in a more complex waveform, because of the addition of harmonics to the fundamental.

one note when vibrato[3] is applied, but for each point in time a *spectrum analysis* (also called *frequency analysis*) can be performed to tell us what underlying sine waves are being added together to produce the final complex composite waveform. Fourier found that not only were the constituent, fundamental parts of a complex waveform multiple sine waves, but also the sine waves were, for many sounds, related to one another in a specific way—harmonically. This means that a waveform can be broken down into a *fundamental frequency* and a series of *harmonics*. The harmonics lie at two times, three times, etc., the fundamental frequency. For example, a violin playing middle "A" has a fundamental frequency of 440 Hz, and it also has harmonics at 880, 1320, 1760 Hz, etc.

You can better understand how such complex motion can arise by thinking about the motion of a guitar string, tied at the two ends and plucked in the center. The fundamental frequency corresponds to the whole length of the string involved in one motion, up and down, with the amplitude varying from greatest at the center to nonexistent at the clamped ends.

The string can also vibrate simultaneously at harmonic frequencies. In one instance, one-half of the string moves up, while the other half moves down, relative to the fundamental. The string acts for this harmonic as though it is clamped at the two ends, just like the fundamental, but also as if it is clamped in the middle. This harmonic radiates sound at twice the fundamental frequency, because each half of the string vibrates separately, at twice the rate of the fundamental. This is called the second harmonic.

> Do not be confused by the fact that this frequency is actually the first harmonic found. It is still called the second harmonic because it is at twice the fundamental frequency.

This process also occurs at three, four, and more times the fundamental frequency, leading to multiple harmonics. A string can vibrate in more than one mode at one time, moving at both the fundamental frequency and the harmonic frequencies, leading to a complex motion in which the constituent parts may not be readily apparent. For any given point along the length, the shape of the curve traced out over time results from adding together the effects of the fundamental frequency and the harmonics.

One of the most important techniques to identify various sources of sound involves using their patterns of harmonics. The relative strength of harmonics plays a large role in determining the different sounds of various instruments playing the same note. A violin has a structure of harmonics that typically is not very extended and in which the harmonics are not as strong as those of a trumpet. A trumpet is thus called *brighter* than a violin. Spectrum analysis of a trumpet versus a violin shows a more extended and stronger set of harmonics for the trumpet than for the violin.

> Alternative names for harmonics are *overtones* and *partials*. Despite this relatively straightforward definition of harmonics, real-world instruments are more complex than this simplified discussion. They may radiate sound at other frequencies as well as at the fundamental and harmonics.

Fundamental sine wave
+
1/3 amplitude 3rd harmonic
+
1/5 amplitude 5th harmonic
+
1/7 amplitude 7th harmonic
+
⋮
+
all higher odd harmonics
with amplitudes
in inverse proportion
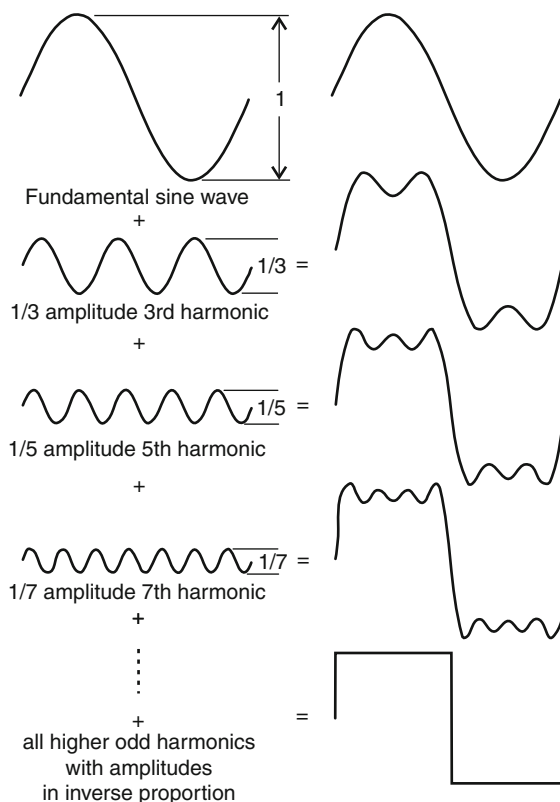
FIGURE 1.10 A complex waveform such as a square wave is built up from the summation of harmonics—the square wave includes the third harmonic at $\frac{1}{3}$ the amplitude of the fundamental, the fifth harmonic at $\frac{1}{5}$ amplitude, and so forth.

---

[3]A moment-by-moment frequency variation that adds interest to the sound of many instruments.
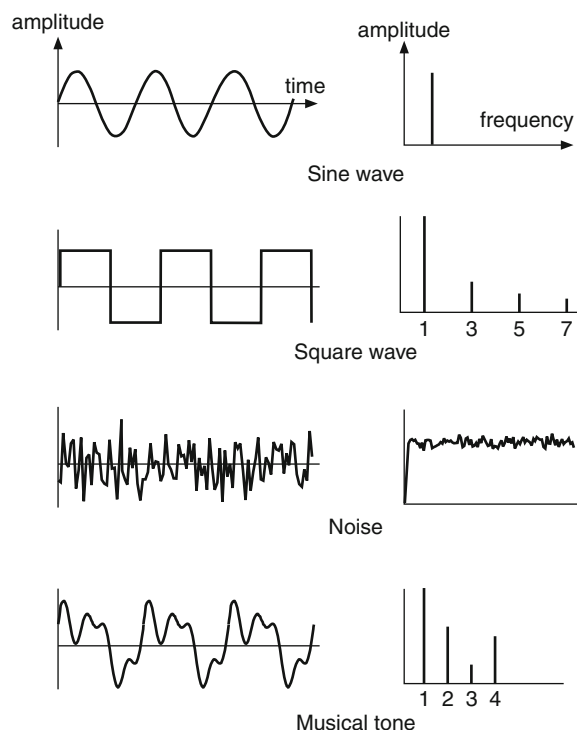
FIGURE 1.11 The amplitude versus time curves of some common signals and the corresponding spectrum analysis for each.

Track 7 of the DVD contains various waveform signals at the same frequency and amplitude, illustrating some of the differences that waveshape makes.

Whereas the generation of harmonics for tonal sounds such as those produced by notes played on an instrument should now be clear, what about sounds that have no explicit pitch, such as speech or waves heard at a beach? Do they have harmonics? Although less evident, perhaps, speech too consists of fundamental frequencies with harmonics, although both change rapidly in time. It is more difficult to assign a fundamental frequency to speech than to singing (which, if the singer is in tune, is the note written), but a fundamental frequency is nevertheless present.

Waveforms that have a clear fundamental frequency are generally called *tonal* by acousticians, whereas those for which the fundamental frequency is less clear may be called noise-like. *Noise* has several definitions. In the most common popular usage, noise means unwanted sound. An acoustician, on the other hand, would call the sound of surf a noise-like signal, because it is impossible to extract a fundamental frequency from it, despite the fact that the surf "noise" may be a desirable or undesirable sound, depending on your point of view at the time. Noisy sounds, like surf, consist of a great many simultaneous sinusoidal frequencies. The difficulty with separating the frequencies into fundamental and harmonics is that there are so many frequencies present at the same time that no particular order can be determined, using either an instrument or the human ear.

The discussion of fundamental frequency and harmonics can be extended to *subharmonics*. Some devices and instruments can radiate sound at one-half, one-quarter, etc., the typical fundamental frequency, especially at high levels. In an instrument, this can add a desirable feeling of "weight."

> In these cases, determining which is the fundamental and which are subharmonics is usually done by spectrum analysis; the strongest of the multiple frequencies is usually the fundamental.

Subharmonics can be synthesized by a device in film production, and their addition adds body to the recorded sound. The voice of Jabba the Hutt in *Return of the Jedi,* for instance, was processed by having subharmonics added to it. Although the technical addition of subharmonics was important, the primary consideration was casting.

Sympathetic Vibration and Resonance

If one tuning fork that has been struck is brought near to a second one that is at rest and tuned to the same frequency, the second will receive enough stimulus that it too will begin to vibrate. This *sympathetic vibration* can occur, not just in deliberately tuned devices such as tuning forks, but also in structural parts of buildings and can cause problems with room acoustics. An example is a room in which all the surfaces are covered in 4×8-ft sheets of $\frac{1}{4}$-inch plywood nailed to studs only at their perimeter. This room acts like a set of drum heads, all resonant at the same frequency. The frequency at which they resonate is determined by the surface density of the sheets of plywood and the air space behind the plywood. Any sound at a particular frequency radiating in the space will cause the surfaces to respond with sympathetic vibration; the surfaces are said to be *in resonance*. There will be an abrupt change in the room acoustics at that frequency, producing a great potential for audible rattles.

> Sympathetic vibration can create a major problem in the room acoustics of motion picture theaters. This is because the sound pressure levels achieved by soundtracks are quite high at low frequencies in movie theaters, more so than in other public spaces. In some cases, the sound system may be capable of producing sufficient output at the resonant frequency of building elements that very audible rattles occur, distracting the audience at the very least.

Phase

We have already described two main properties of a waveform, amplitude and wavelength. These two properties together are enough to completely specify one sine wave, but they are inadequate to describe everything going on in a complex wave, which includes a fundamental and its harmonics. To do this we need one more concept before we have a complete description of any wave—phase.
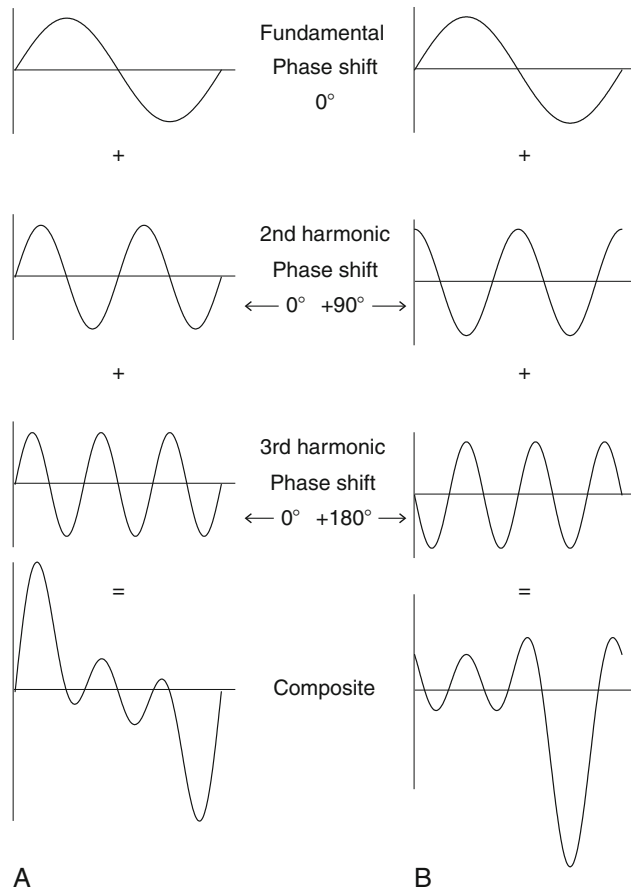
FIGURE 1.12   The addition of harmonics in various phases changes the overall waveform. In this example, harmonics of the same number and amplitude have been added together to make up two different composite waveforms. The only difference between examples (a) and (b) is the phase shift applied to each of the harmonics relative to the fundamental. Although the resulting waveform is quite different, it may sound only slightly different, if at all, because of the relative difficulty of hearing phase effects.

In our example of a string vibrating in different modes, there was one thing in common among all of the modes, that is, all motion ceased at the two ends of the string, by definition, where the string was attached. But not all generators of sound have such fixed end points; for instance, an organ pipe has one open end, and the harmonics may not have zero amplitude at the open end.

Phase is a way to describe the differences in the starting points on the waveform of the various harmonics relative to the fundamental. Because one sine wave cycle is described mathematically as occupying $360°$ around a circle, phase is given in degrees of shift, comparing the phase of each of the harmonics to that of the fundamental.

The reference for phase is not usually the peak of compression or the trough of rarefaction, but rather the point at which the waveform goes through zero, heading positive. If the second harmonic is at the crest of its compression at the moment that the fundamental is heading through zero on its way positive, we say that the second harmonic is shifted by $90°$ phase leading. Conversely, if it is in its trough at the same point, we say it is shifted by $90°$ phase lagging. In 1877, Helmholtz, based on some

experiments he conducted, thought that phase shift was inaudible, and his finding dominated thinking for a long time. Later research has shown situations in which it is audible. Added phase shift between a fundamental and its harmonics, which is caused by microphones, electronics, recorders, and loudspeakers, may be audible in certain circumstances. Another way to look at phase shift is called *group delay*. Group delay expresses the fact that leading or lagging phase can be converted to time differences between the fundamental and its harmonics. The time difference is only relative—there is no way to "beat the clock." Sound does not arrive earlier than it went into a system in the case of a leading phase shift—what happens is that the fundamental is more delayed than the higher frequency in this case.

Let us say that a recording system is so bad that it delays the 10th harmonic by 1 sec relative to the fundamental. This amount of delay distortion is obviously audible because the fundamental may stop and the 10th harmonic will play on for a second. So it is not the audibility of such effects that is in question, but rather the amount of delay that it takes to make the phase shift audible.

The concept of phase shift has been used incorrectly to describe all that is done well or poorly by a recording system but that remains a mystery. Phase shift is audible in large amounts, which result in time delays in one part of the frequency range compared with another above an audible threshold, but not so much as to make it a kind of magic ingredient that can be hypothetically adjusted to make everything perfect.

Influences on Sound Propagation

Up until now we've been discussing pretty abstract cases of sound propagation, with point sources and strings radiating into a "free field," that is, without encountering any objects. In the real world, sources are more complicated than this, and a number of influences affect the sound before it may be received by a listener or a microphone. These include absorption, reflection, diffraction, refraction, constructive and destructive interference, and Doppler shift. Most of these play some part in the overall sound of a film or television show.

Source Radiation Pattern

The first of these is due to the complexity of the source itself and is called its *radiation intensity*. Most sources do not radiate sound equally in all directions but instead have a preferred direction that often changes in a complex way with frequency. The fundamental may be radiated most strongly in one direction, but one harmonic may radiate mostly in a different direction, and another harmonic in yet a third direction. This is critically important to understand because if sources have priority for certain frequency ranges in certain directions, where is one to place a microphone to "capture" the sound of the source—which position is "right"? Inevitably one shows a preference for one frequency range or another by the forced choice of microphone position.

So let us say we don't have to be practical and instead of one microphone we can use a whole array of microphones equally spaced all over a sphere surrounding the source. We connect these to a multitrack recorder and record each microphone signal on a separate track. Then we arrange a whole array of loudspeakers, connected to the tracks, in accordance with the microphone positions, radiating sound outward from the recorded source. With such a system, we have a means to capture all of the relevant details of the sound field produced by the source, especially the complex way in which it radiates sound directionally into the world. Let us say that we have just described how to capture the sound of a clarinet completely. We've got one instrument of an imaginary orchestra finished, and now let us start on instrument No. 2, a flute .... You can see how quickly such a system, meant to be absolute in its ability to actually reproduce an audible event, falls apart. Thousands of microphones, recorder tracks, and loudspeakers later, we can reproduce the complete sound of an orchestra with great fidelity, but that is so impractical that no one has ever tried it.

So there is a fundamental theoretical problem with recording sound—no practical system can be said to "capture" completely the sound of most real sources in all their spatial complexity. What production sound recordists, boom operators, and recording engineers become highly proficient at is choosing one, or a few, microphone positions that instead *represent* the sound of the source, without making a valid claim to actually *reproduce* the source completely. This idea is commonly used on motion-picture and television sets, although it is probably expressed here in a manner different from that used by practitioners. What has been developed over time is microphone technique that permits adequate capture of sound for representation purposes. Microphone technique is presented early in this book because of its importance. Here, what is important to understand is that the choice of microphone technique is highly dictated by the requirements of the source, especially its radiation intensity.

Of course, in movies and television, perhaps the most important source much of the time is a human speaking voice. Talkers radiate various frequency ranges preferentially in different directions, similar to orchestra instruments. Thus exactly the same microphone, at the same distance, will "sound" different when recording a voice as it is moved around the talker. The practical direction that is preferred for the largest number of cases is overhead, in front of the talker, about 45° above the horizon, in the "boom mic" position, which was named for the device that holds it up, not for the microphone itself. In many cases, this positioning stimulates debate on the set over the relative merits of various microphone positions, with the camera department holding out for placement of hidden microphones on set pieces and underneath the frame line, all in an effort to avoid the dreaded boom shadow. The experience of the sound department, on the other hand, shows that the overhead microphone position "sounds" more like the person talking naturally than other positions. For example, with a position that is located below the frame line, the microphone is pointed at the mouth but is closer to the chest of the talker than when in the overhead position, and the recording is often found to be "boomy" or "chesty" in this position compared with the overhead mic. This difficulty is caused by the radiation pattern of the voice.

Another difficulty occurs when the actor or subject is capable of moving. If we use a microphone to one side of the frame, this may sound all right in a static situation, but when the actor turns his or her head to face the other side of the frame from the microphone, he or she goes noticeably "off mic." Thus the more neutral overhead mic position is better in the case in which the actor's head may turn.

Track 8 of the DVD illustrates recording a voice from various angles, demonstrating the radiation pattern of the voice.

Absorption

Sound may be absorbed by its interaction with boundaries of spaces, by absorptive devices such as curtains, or even by propagation through air. Absorption is caused by sound interacting with materials through which it passes in such a way that the sound energy is turned into heat.

The atmosphere absorbs sound preferentially, absorbing short wavelengths more than long ones (and thus absorbing high frequencies more than low ones). Thus, at greater distances from the source, the sound will be increasingly "bassy." While this effect is not usually noticeable when listening to sound in rooms, it is prominent out of doors. It is atmospheric absorption that causes distant gunfire sound to have no treble content but a lot of bass. This effect is used very well in the jungle scene in *Apocalypse Now*. Among the sounds that we hear in the jungle is a very low frequency rumble, which we have come to associate through exposition earlier in the film with B52 strikes occurring at a distance:

EXT. BOAT — DAY
A strong low-frequency rumble is heard.

CHEF
Hey, what's that?

CAPT. WILLARD (in a normal voice)
Arc Light.

ANOTHER MAN
What's up?

CAPT. WILLARD
B52 strike.

SOMEONE ELSE

Man.

BOAT CAPTAIN

What's that?

CAPT. WILLARD (louder)

Arc Light.

FIRST MAN

I hate that—every time I hear that somethin' terrible happens.

The later jungle scene has a sense of foreboding that is heightened by hearing, even more distant than in the exposition scene, another B52 strike.

The effect that the atmosphere produces on sound is also frequently simulated by sound effects editors and/or mixers. In a long shot showing the Imperial Snow Walkers in *The Empire Strikes Back*, the footfalls of the walkers are very bassy. As the scene shifts to a closer shot, with a foot falling in the foreground, the sound takes on a more immediate quality—it is not only louder, but also brighter; that is, it contains more treble. In this case, the increased brightness was created by literally adding an extra sound at the same time as the footfall, the sound of a bicycle chain dropping onto concrete. Mixed together with the bassy part of the footfall, the treble part supplied by the bicycle chain makes it seem as though we have gotten closer to the object, because our whole experience with sound out of doors tells us that distantly originating sound sources sound duller (less bright) than closely originating sources. Thus, in this case the filmmaker has indicated the distance to the object subliminally, when, in fact, no such object ever existed. The sound is animated, in a sense, like the corresponding picture that uses models. A principle of physics has been mimicked to make an unreal event seem real.

The surfaces of room boundaries (walls, ceiling, and floor), as well as other objects in a room, also absorb sound and contribute to room acoustics that are discussed later. Many materials are deliberately made to absorb sound, whereas other materials, not specifically designed for this purpose, nevertheless do absorb sound more or less. Some rules of thumb regarding absorption are as follows:

- Thick, fuzzy materials absorb more lower frequency sound than similar thin materials.
- Adding a few inches of spacing off the surfaces of the room to the absorption greatly improves its low-frequency absorption.
- Placement of absorption in a room is often important. Placing it all in one plane, such as on the walls only, will cause a problem when floor to ceiling reflections are considered.
- Absorption is rated on a scale of 0 to 1, with 0 being no absorption, a perfectly reflective surface, and 1 being 100 percent absorption, equal to an open window through which sound leaks, never to return.
- Absorption changes with frequency, yielding the 0 to 1 rating that is given in tables as a function of frequency.

Quite commonly on motion-picture location sets "sound blankets" are hung around the space, to reduce reverberation. The area available for absorption is made most effective by using thick absorption mounted a few inches from the wall or ceiling—a thinly stretched sound blanket doesn't do as much good.

Reflection

Sound interacts with nonabsorbent surfaces in a manner that depends on the shape of the surface. Large, flat, hard surfaces reflect sound in much the same way that a pool ball is reflected off the edges of a pool table. Such "specular" reflection works like light bouncing off a flat mirror—that is, the angle of incidence equals the angle of reflection, or the angle incoming is equal to the angle outgoing, bisected by a line perpendicular to the reflector.

Parabolic reflectors concentrate incoming sound from along the axis of the parabola to its focal point. Parabolic reflection is used for specialized microphones that are intended to pick up sound preferentially in one direction. By placing a microphone at the focal point of a parabola, the whole "dish" concentrates sound waves that are parallel to the axis of the parabola on the microphone. The problem with parabolic reflectors is that because of the wide wavelength range of sound, they are effective only at high frequencies unless the reflector is very large.

There are two prominent uses for parabolic microphones in film and video production. One is by nature filmmakers, who must capture bird song at a distance without distracting the birds; because the bird chirps are high frequencies, the dish size does not cause a limitation. The second is in sports broadcasts, in which the super directionality of a parabolic mic can pick up football huddle and scrimmage high-frequency sounds and add immediacy to the experience. (The bass is supplied by other microphones mixed with the parabolic mic.) The parabolic mic supplies a close-up of the action, whereas other microphones supply wide-frequency-range crowd effects, etc. Notice that this corresponds to the finding that greater amounts of high-frequency sound make the scene seem closer, as in *The Empire Strikes Back*.

Elliptical reflectors are used in "whispering galleries," usually located at museums. Here a person whispering at one focal point of an elliptical reflector can be heard clearly by a listener positioned at the second focal point of the ellipse, despite the remote spacing of the two foci, such that others standing away from either focal point cannot hear the conversation. This results from the large area available to gather sound from one focal point of the ellipse and deliver
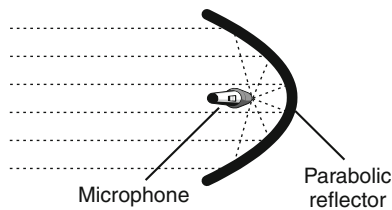
FIGURE 1.13 A parabolic reflector concentrates incoming waves on a microphone. The combination is called a parabolic microphone.

it to the second focal point of the ellipse, yet outsiders do not hear the effect of the concentration of sound energy.

An architectural feature found in some auditoriums is a spherical- or elliptical-shaped dome. A great difficulty with these shapes is that they tend to gather sound energy and concentrate it on parts of the audience. A whispering-gallery effect is quite common in domed spaces—you can hear other members of the audience under the dome, even if they are whispering. The only solution to this problem is to make the domed surface quite absorptive or to make the focus of the dome well away from listening areas.

An inside view of a parabola or an ellipse shows sound waves converging on focal points. What about the outside of such surfaces? Sound impinging on such "bumpy" surfaces is scattered, spreading out more rapidly than if reflected from a flat surface.

Such surfaces are called *diffusers* and play an important role in room acoustics, which are discussed in the next section.

Diffraction

One of the most profound differences between seeing and hearing is that sound is heard around corners, whereas sight stops at an opaque object. Diffraction occurs when waves interact with objects; the sound "flows" around corners in much the same way as incoming parallel water waves interact with an opening in a breakwater. Going past the gap the waves are seen spreading out in circles. A second generation of waves also occurs when a set of sound waves encounters the edge of a barrier (see Figure 1.14). Although the sound in the *acoustic shadow* may not be as distinct as sound with a direct, undiffracted path, it is nevertheless very audible.

Sound diffraction, and especially the everyday experience we have with it, has strong effects on how sound is used in movies. Whereas the frame line of the picture is practically always an absolute hard line, there is no such boundary for sound. We expect to hear sound from around corners, that is, from off-screen as well as on. Screenwriters often refer to off-screen action in scripts, with the attention of actors being drawn outside the frame line, and the easiest way to indicate this to an audience is with sound. With picture, the way to accomplish this is with a cut away, a shot that literally shows us the object that is to draw the attention of the actors. This may often seem clumsy because it is too literal, whereas off-screen sound can seem quite natural because it is part of our everyday experience.

The situation in media is parallel to what occurs in life: Seeing is limited to the front-hemisphere view, but sound can be heard from all around. To make use of this fact of life, which is based on the diffraction of sound waves about the head to reach the ears, the film industry has been developing *surround sound* for more than 25 years, a great expressive medium that is underutilized, but growing, in television. Surround sound is based on the differences between viewing and hearing caused by diffraction. Not surprisingly, perhaps, some of the best surround sound on television is heard in certain commercials.

Refraction

In addition to complete reflection, sound waves can also be bent by changes in the density of the atmosphere. Following the same principle by which lenses bend light waves, sound is refracted when stratification occurs in the atmosphere because of differing temperatures (and thus densities) in different layers. Be careful here to distinguish the terms *refraction* (bending) and *rarefaction* (opposite of compression), as they mean quite different things.

Constructive and Destructive Interference and Beating

Sound waves interact not only with objects in the room but also with one another. For example, one wave having a sinusoidal waveform with a specific wavelength and amplitude is flowing from left to right and is joined by a second wave, having the same wavelength and amplitude, but flowing from right to left. The result at any observation point will depend on the moment-by-moment addition of the two. Even in this simple case, there is a wide range of possible outcomes. This is because although we specified the wavelengths and amplitudes as equal, we failed to specify the difference in phase between the two waves. If the two waves have identical phase, that is, if their compression and rarefaction cycles occur at the same time, then the two waves are said to be in phase, and the result is addition of the two waves, resulting in doubling of the amplitude.

In the opposite case, if the two waves are completely *out of phase* with each other, with the peak of one wave's compression cycle occurring at the same time as the trough of the other wave, then the result is subtraction of the two waves. Because the two are now equal but opposite, the outcome is zero, which is called a *null*. So the range over which two simple waves can interact is from double the amplitude of one of the waves to complete cancellation! The former case is called *constructive interference,* whereas the latter is called *destructive interference.* Between these two extremes the results change smoothly. This represents
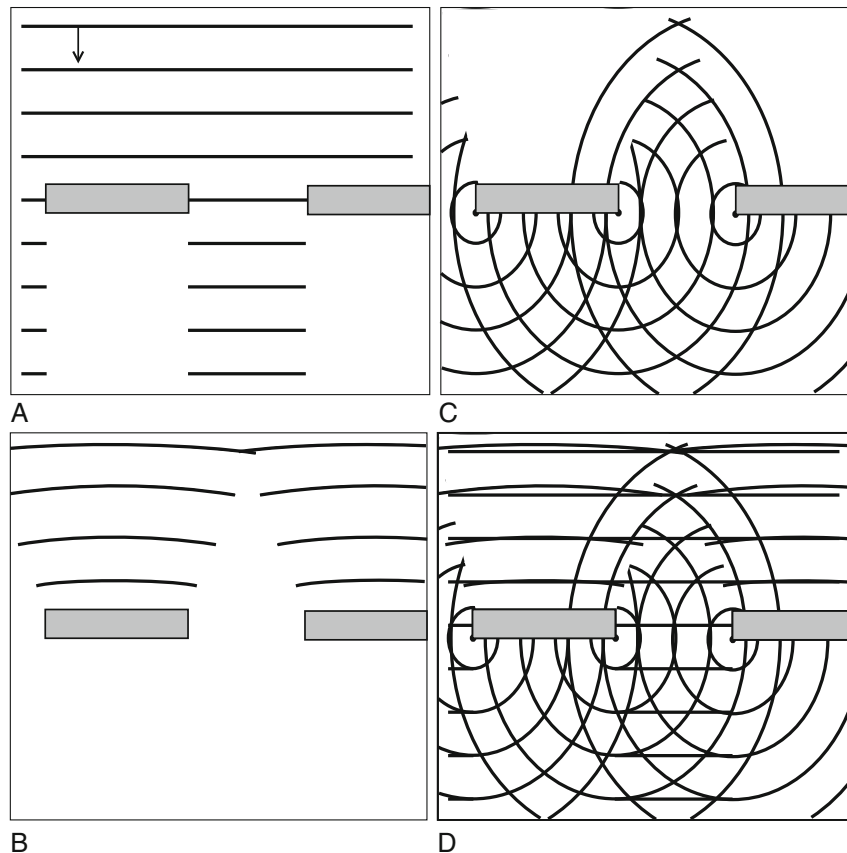
FIGURE 1.14 The four frames show various effects of waves interacting with barriers. In (a) the incoming wave is shown, and it passes through the barrier unimpeded, corresponding to the direct sound. In (b), the waves reflected from the barriers are shown. In (c), the diffraction off the edges of the barrier is shown. In (d), the combination of the three sets of waves is shown, those due to direct waves, reflected waves, and diffracted waves.

the condition, more common in actuality, in which the waves are not quite in nor out of phase, but rather have some phase difference between them. We will see how this affects recording in the section on room acoustics.

If, instead of the two waves having an identical wavelength, they have slightly different wavelengths, then the result is that at some points in time the waves add, and at a slightly later time they cancel, with the difference in the times being equal to the difference in the frequencies of the waves. For example, if waves at 1000 and 1001 Hz are mixed together, what one observes is a variation in the level of the sum at a rate of 1 Hz. This effect is called *beating*. It is what is often heard in out-of-tune pianos when playing only one note and is caused by the piano using multiple strings for one key, which may not be tuned to exactly the same frequency.

Film sound designers make use of beating in making certain sound effects. By manipulating various soundtracks, for example, slowing them down slightly and mixing the slowed version with itself, the kind of "waa-aaah, waaaaah, waaaaah" effect heard when opening the ark in *Raiders of the Lost Ark* is produced. This is far more interesting than the simpler low-frequency rumble that might have been used in the same film had it been made 10 years earlier.

Doppler Shift

In 1842 an Austrian physicist, Christian Johann Doppler, published a paper describing an effect he thought must exist but had yet to be demonstrated, Doppler shift. Given that the speed of sound is fairly slow, Doppler questioned what would happen to sound generators in motion. As they approach a point of observation, the sound waves should be "crowded together" by the velocity of the source and, likewise, should spread apart while receding into the distance. This crowding results in a shorter wavelength on the approach, and thus a higher frequency, and a corresponding lowering of the observed frequency as the object recedes.

Three years after Doppler's paper was published, his theory was put to the test. A train was loaded with 15 trumpeters who played the same note continuously while the train approached, passed, and then pulled away from a train station. Sure enough, while listeners on the train heard a constant pitch, those in the train station heard the pitch being raised and then lowered as the train passed the station.

Filmmakers fake this effect to simulate objects in motion and sometimes exaggerate for emphasis. In the Philip Kaufman version of *Invasion of the Body Snatchers* (1978), when two characters are riding in a car just before
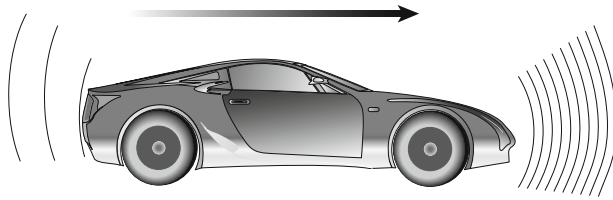
FIGURE 1.15   The sound waves emitted by a race car are affected by its speed, being crowded together in front of the car and spreading apart behind it, leading to the characteristic change in pitch as a moving object passes a listener.

a man is struck and killed by another car (the man being the star of the earlier version of the movie, a kind of tongue-in-cheek homage), we hear other cars pass us on a busy San Francisco street. The exaggeration is in the amount of pitch shift used: those cars would have to be traveling in excess of 90 mph to produce that effect!

A more complex example is heard in *Indiana Jones and the Temple of Doom,* when Indy has made it across to the far side of a chasm but is being chased by the bad guys, with arrows being shot at him. The challenge was to simulate an arrow sound throughout its flight, as it is exceedingly difficult to attach a microphone and recorder to an in-flight arrow. The steps that the sound designer took were as follows:

- A recording was made of an arrow passing by a microphone. The first attempts were futile, because good arrows make very little sound. So the arrow's tail feathers were ruffled and a new, more useful "take" was made.
- Of course, this recording was far too short to simulate a full flight, being just a "zip" past the microphone. So the recording was transferred to a digital audio sampler, and "looped" to extend it to the length of the flight. The usable sound was so brief that there was not enough time to hear a Doppler shift in this original recording.
- The recording was processed by a digital pitch shifter, which increased the frequency at the beginning of the flight and decreased it toward the end, producing a Doppler shift.
- During mixing, the arrows were panned so that their flight moved across the screen to match the action.

Doppler shift is also very commonly used for simulating transportation devices in motion. There are several types of sound effects that an effects recordist can make to "cover the bases" so that all uses of a device in a film can be covered by the smallest number of effects. For an automobile, such as the titlemaker in the movie *Tucker,* a recordist will record at least:

- Engine idling
- Acceleration from stop
- Braking to a stop

- A *pass by,* in which the microphone is at a fixed location and the car moves by it
- A *steady,* in which the microphone moves with the car using an exterior perspective
- An interior steady

With these in hand, a sound effects editor can make up a trip of nearly any length by methods to be described later, blending them together to make a functionally single-sounding whole. The greatest challenge in this list is simulating a realistic pass by, because there could be a great many of them possible in a film, at various speeds. It is possible to turn the steady into a pass by using a faked Doppler shift as a primary processing tool to make it sound realistic. Because the speed of the pass by is controlled by a knob in postproduction, one steady can be adequate to produce a large number of pass bys.

Doppler shift is demonstrated on Track 9 of the DVD.

ROOM ACOUSTICS

The acoustics of interior spaces are extremely important to filmmakers because so many of them are involved sequentially in the final perception of the sound of a film or video. There are the acoustics of the live-action shooting environment; those of various postproduction studios, including recording and monitoring stages; and those of the theater or viewing room.

The classic art forms of theater and music obviously depend on the room acoustics of the spaces in which they are performed, as is well known. Cinema and video depend on room acoustics as well, but, as we shall see, film and video producers seek to control what is heard through recording and reproducing under very specific conditions so that the sound of their productions translates from one environment to another. This is a fundamental concept behind recorded media: The performance is captured in time, for future display in many venues and at different times. This concept depends on the ability of the display venues to deliver the performance, ideally free of alteration, to the listener.

The modern story of room acoustics starts in 1885 with a junior assistant professor of physics at Harvard, Wallace Clement Sabine. Sabine was given a job that no one else wanted, to solve the problem of speech intelligibility in the newly built Fogg Art Museum lecture theater in Cambridge. Students had complained that they couldn't understand the lecturers in the theater. Sabine conducted an experiment to answer the question, How long does it take sound to die away to inaudibility after the source is abruptly stopped? To avoid contaminating the measurement by his presence, Sabine built a wooden box with a hole in the top of it for his neck, so that his body was essentially not present in the room, acoustically speaking.

Then he used organ pipes with valves so that their own sound stopped very quickly when cut off to make sound in the room. What he heard after the organ pipe stopped was the reverberation of the room. He used a stop watch to determine how long it took the sound to die out audibly and called this the *reverberation time*.[4]

The time was rather long, on the order of seconds. Realizing that when speech is heard in the room each syllable stimulates its own *reverberant tail,* Sabine found that these reverberant tails overlapped with new syllables uttered, causing the intelligibility problem. To test his hypothesis, Sabine added units of absorption to the room that would soak up the reverberation. For Sabine, one unit of absorption was one Sanders Theater seat cushion, borrowed at night from the venue across the street. With the added absorption, in various amounts, Sabine found a curve of reverberation time versus absorption in the room, which is still used today to predict the reverberation time of rooms before they are built and furnished.

In addition to the mathematical prediction capability of Sabine's reverberation time, it was also his understanding of the damage that reverberation does to speech intelligibility that is used widely in the film industry today. We make motion picture theaters "dead" acoustically, for the same reasons that Sabine found absorption useful. By the way, using the knowledge he gained starting with the experiment just described, Sabine went on to design Symphony Hall in Boston, widely considered to be one of the best-sounding concert halls in the world.

> Sabine's equation, although still used today, was modified to cover a wider range of conditions by later acousticians. In particular, in the 1930s, Carl Eyring found that Sabine's work did not predict the reverberation times of rather dead spaces well, and he found a better equation to describe such spaces. It is interesting to note that Eyring was working on the problem of spaces coupled by recording, in particular, motion-picture shooting stages and theaters, when he added new information to the entire field of acoustics.

Sound Fields in Rooms

Direct Sound

In all practical rooms, there are three sound fields to consider. The first is direct sound, which is sound radiated in a straight line between the source and the receiver and that arrives at the receiver first, before any reflected sound. If the observer is far enough away that the source can be considered a point, the direct sound follows the inverse square law falloff in level versus distance.

---

[4]This was later codified to be the length of time required for the sound to decay by 60 dB in level, a factor of 1000 in sound pressure level. See the end of this chapter for more information.

> Note that in certain circumstances, such as when sitting behind a column in a theater, there may in fact be no direct sound. This is one of the reasons (in addition to visual impairment) that such seats are undesirable.

Discrete Reflections

The second sound field consists of discrete reflections. Reflections are first produced from one "hit" off surfaces or objects in the room, such as off the floor in front of the listener at home or off the side wall of a theater. Over time, later and later first reflections from more distant surfaces are produced and are joined by second-order reflections, which involve two hits, and so forth. The first, second, and higher order reflections are all considered discrete reflections, until there are so many reflections that they merge into reverberation, which is discussed in the next section.

The audible significance of discrete reflections varies, depending on their time of arrival and direction, their relative strength compared with direct sound, and their relative strength compared with the third sound field to be discussed, the reverberant field. Early discrete reflections may cause audible changes to direct sound, but not usually in such a manner that they can be distinguished from the direct sound. Instead, they change the sound by adding *spaciousness,* one sense of being in a room; they may affect localization; and they may *color* the direct sound by changing its timbre, to be defined in the next chapter. The direction of the reflections matters as well, with reflections having nearly the same angle as the source being less audible than reflections at great angles to the source.

Adding artificial "reflections" by use of digital delay in postproduction can change the character of a sound. Darth Vader's voice is processed by adding an apparent reflection 10 ms after the direct sound (about $\frac{1}{4}$ of a frame) about equal in amplitude to the direct sound. This gives the voice the "sound in a barrel" quality that makes it sound mechanical.

Later-arriving discrete reflections may potentially form echoes if they arrive late enough and are strong enough relative to the direct sound. Even without hearing obvious echoes, they may also color the sound field.

Track 10 of the DVD shows the effect of adding one strong early reflection (at 10 ms, about $\frac{1}{4}$ frame of film) to the voice.

Reverberation

The third sound field in a room is the reverberant field. Reverberation occurs after there have been enough discrete reflections for no one reflection to be distinguishable from the others. Instead a "cloud" of energy, generally having no apparent direction, fills the room. What is usually desired for good acoustics is a smooth, diffuse decay of sound, having no particular signature or pattern. There

are a variety of deviations from this ideal, with potentially damaging results.

> For example, flat, hard, parallel walls lead to *flutter echoes,* a particular pattern of reflections in which a "ta-ta-ta-ta-ta" effect is heard in larger spaces and a kind of "sizzle" is heard in smaller spaces.

Reverberation time in a room was defined by Sabine as the time that it takes an abruptly terminated sound to die away to inaudibility. Sabine found that there were only two factors involved in computing the reverberation time: the volume of the room and the absorption of the surface area in the room, including the walls, ceiling, floor, and objects in the room. The *room volume* is the height times length times width, in units such as cubic feet. The *absorption* is expressed as the surface area times the average absorption, on a scale from 0 to 1, of each of the surfaces in the room.

> Note that the placement of absorptive materials for Sabine's equation does not matter; however, it does matter in lower reverberation time rooms where placement certainly affects discrete reflections, at the least, and probably the reverberation time as well.

Reverberation time must be evaluated across a range of frequencies. This is because most room surface treatments do not absorb sound equally well at all frequencies and often have less absorption at low frequencies than at high ones.

In addition, air absorption enters the evaluation at high frequencies in larger rooms.

This usually leads to a characteristic reverberation time curve versus frequency that has longer times at low frequencies, and shorter times at high frequencies, than in the middle frequency range. This can be important to filmmakers, because it can change the perception of certain effects dramatically. In the film *2010,* a loud, low-frequency rumble was used for the spaceship interior, and silence was used for the vacuum of space (physicists take note: a filmmaker followed physical law this time). Abrupt cuts between an interior shot and an exterior shot thus call for a loud, low-frequency sound to cease abruptly. At the Coronet theater in San Francisco, where I heard the film, the approximately 5-sec (!) reverberation time at very low frequencies caused the cut to seem, well, silly—the rumble just floated away slowly, negating the desired effect. Whereas the mid-frequency reverb time of this theater was all right, so that voices were not affected, it was the huge increase in reverb time at low frequencies that caused this problem. Room acoustics conspired against the filmmakers in this case, even though they had followed the requirements of physics closely.

A headline in the *San Francisco Chronicle* some years ago highlighted how little is known among people, even newspaper editors, about acoustics. The headline read "Six second echo in largest church rivals Grand Canyon." The article went on to describe how the church of St. John

A

B

C

D
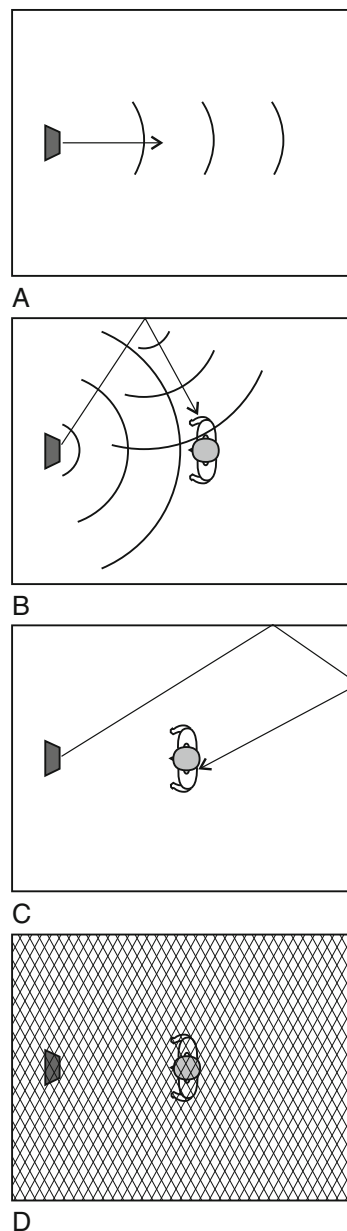
FIGURE 1.16 Sound fields in a room. (a) Direct sound, (b) first-order reflected sound, (c) higher-order reflected sound, and (d) a reverberant field in which there are so many reflections per unit of time that each becomes indistinguishable from its peers.

the Divine in New York has a 6-sec "echo," and that is how long it takes sound to cross the Grand Canyon and return. The problem is that St. John the Divine does not have a 6-sec echo; it has a 6-sec reverberation time. A 6-sec echo (that is, organized, intelligible sound arriving 6 sec after the direct sound) does not sound at all similar to a reverberant decay that requires 6 sec to die away to inaudibility.

> Most books treat room acoustics as having three sound fields: the direct sound, the *early reflections,* and the reverberant field. Unfortunately, in the room acoustics of motion-picture theaters,

which are relatively dead, that is, they have both a low reverberation time and a low level of reverberation, discrete reflections may occur at any time, early or late, and still be detectable by ear or by instrument. In many large rooms, like concert halls, discrete reflections become covered up with reverberation fairly quickly over time. This does not happen in cinemas, where even the effect of the port glass for the projector in an otherwise absorptive back wall of a theater may cause a delayed reflection to the front seats in the auditorium that is large enough to be audible! The way to fix this problem is to tilt the port glass so that the first reflections are sent to the ceiling of the theater rather than the seats.

I did something related in the Hollywood Bowl for the film presentation of *The King and I* years ago. A booth erected near the center of the seating reflected sound off its face, causing an echo for patrons in expensive seats. I bought a large piece of Plexiglas and tilted it in front of the booth, sending the reflection skyward, away from those expensive seats.

Track 54 of the DVD illustrates reverberation effects.

Sum of Effects

So the actual sound heard in a room is a sum of the effects we have been discussing. First the source radiates sound, probably with a preference for one or more directions, and the straight-line path to the observer contains the direct sound. Discrete reflections start to arrive and then reverberation builds up, as time goes by. Although discrete reflections may arrive at any time after the direct sound, they may be blended into reverberation or not, depending on how late they arrive and how strong they are compared to the prevailing reverberation. The sum of these effects is extremely complicated, but there are some interesting examples of methods that have evolved, especially in the motion-picture industry, to achieve reasonable overall results. They also illustrate an adaptive industry that has changed over time.

In the late 1920s the introduction of sync sound to movies was mostly done in converted vaudeville theaters, because specifically designed and built cinemas had yet to come into being. Many theaters had a live show and then a film. Over time, live stage show venues had reached a compromise on reverberation time between desirably low times for good intelligibility and higher times, which were basically cheaper.

Live stage actors changed their performance to accommodate the acoustics of the space, slowing down and enunciating as necessary to be heard. Because absorptive materials were fewer in type and quality at that time than today, most auditoriums were fairly live, that is, reverberant, and treatments were minimal, using only drapery to make speech intelligible, if not optimum. This resulted in fairly long reverberation times in the venues where movies were first shown with sync sound. It was known in Hollywood at the time that if reverberation was also recorded on the soundtrack, the combination of recorded plus live reverberation put intelligibility over the edge.

For this reason, shooting stages were virtually *anechoic,* that is, practically 100% absorptive, with thick mineral wool covering the walls and ceiling, held in place with acoustically transparent cloth or wire. Thus, one heard only one reverberation time in the final auditorium screening, because the recorded reverberation was so small as to be negligible.

This empirically derived method of working with reverberation from the 1930s has been completely reversed in recent years. Instead of live theaters and dead shooting stages, we now have dead theaters and permit relatively more lively shooting conditions. The reasons for this quantum shift are several:

- If the listening space always adds its own reverberation to everything that is heard in it, it is difficult to "transport" an audience from the acoustic sound of the back seat of a car to a gymnasium on a picture cut between the two. It is better for the listening room to be relatively dead and then use recorded reverberation to move the audience from scene to scene. This idea was greatly enhanced in the stereophonic era because reverberation is, by definition, a spatial event, and stereophonic systems can reproduce spatial effects, whereas a monophonic system cannot.

- Today we have far better means to add controlled reverberation in postproduction than in earlier times. This means that recording in a dead studio, and playing back in a dead auditorium, can be enhanced by deliberately added high-quality and flexible reverberation, which is one of the primary processes done during postproduction to provide good continuity from shot to shot and to "place" the sound, that is, make a space for it to live in.

- The cost of production on sets in Hollywood studios has become prohibitively high, so the trend has been to shoot on location. Locations are almost never as dead as Hollywood shooting stages, so we live with increased reverberation in recordings, even if the microphone technique is identical in the studio and on location.

- Sometimes the tolerance for reverberation recorded on location is taken too far. In *Midnight in the Garden of Good and Evil,* for instance, the large, live rooms portrayed in Georgia mansions look good, but sound too lively.

Another effect of summing the direct, reflected, and reverberant sound fields is heard when listening at the source and then moving away from it, along the direction of preferred radiation. When the observation point is near the source, we say it is *in the direct field,* or is *direct-field dominated.* For small sources measured nearby, the drop-off in sound level as we move away from the source follows the inverse square law, as long as we remain direct-field dominated. Far away from the source

in a very reverberant room, we are *in the reverberant field,* or *diffuse-field dominated.* Despite moving the point of observation around in the reverberant field, the level does not change. This is because one part of the definition of a reverberant sound field is "everywhere the same," so the rules are obeyed and we find no difference in level.

In between the region where the level falls off as an inverse square with distance and the region where the level is constant with distance, there is an elbow in the curve of level versus distance. The midpoint of this elbow is the place where the direct sound field and the reverberant sound field levels are equal, called the *critical distance* from the source. All other things being equal, the critical distance varies in the following ways:

- As a room becomes more reverberant, for example by removing absorption, the critical distance gets shorter (because there will be more reverberation).
- As a source becomes more directional in the direction we are measuring, the critical distance increases (because this tends to increase the direct field compared with the reverberant field at our measurement point).

For rooms that have more ordinary reverberation time than the example of Fig. 1.17, such as say a living room, the falloff of sound level versus distance is likely to be closer to a straight line on the graph.

This discussion may seem fairly abstract in light of the day-to-day concerns of filmmakers, but it has a profound effect on the perception of film and video sound, because of the differences in the venues in which the programs are

heard. Today, with films destined for multiple lives, first as theatrical releases and then as home video, the problems associated with transfer of the experience from theaters to homes has received greater attention.

> In motion-picture dubbing stages and first-rate theaters, the primary listening locations are direct-field dominated, because the reverberation times are low for the room volume and the speakers are directional. In homes, however, listening rooms are typically more reverberant for their room volume than theaters, and speakers are less directional. Therefore, listening at home is reverberant-field dominated, and the viewer frequently asks, "What did he say?" A frequent comment from video viewers of film material is that the dialog is not as intelligible as it should be, and this is one of the reasons—the acoustical conditions are very different.

Another feature of this discussion that is important to filmmakers is the relatively low reverberation times present in theaters today, creating a strong propensity for uneven sound coverage. The sound level near the screen can be very high compared with the back of the room; after all, the inverse square law is working in the direct field, and we are direct-field dominated in a low-reverb-time space. Although this is true, theater sound system design has found a way out of this dilemma. Film screen loudspeakers are located and aimed in such a way that they deliberately send less sound in the direction of the front rows and more toward the back rows, overcoming, to a large degree, the falloff from front to back that occurs along the axis of radiation. In other words, the sound system is aimed over the heads of the audience to promote uniformity from front to back.

Standing Waves

So far what we have been discussing involves waves moving from point to point and the effects on them. These moving waves are called *progressive waves,* but there is another kind of wave, called a *standing wave.* The lowest frequency standing waves occur when a wave is propagated within a room and the wavelength being radiated "fits" precisely into the dimensions of the room. The longest wavelength of sound that fits this requirement is twice the distance between the two walls (because there needs to be a full wavelength, including a compression and a rarefaction, to complete the wave). What occurs is that the reflection of the wave reinforces the incoming wave, and the pattern generated seems to "stand still." This process goes on and on, with reflection after reflection contributing to the standing wave pattern. It also occurs at higher harmonics of this fundamental such that more than one wavelength "just fits" into the room dimensions. Standing waves, also called *room modes,* occur between two parallel walls, as described, and there are higher orders of standing waves that involve two and three pairs of surfaces as well. Simple, parallel-wall standing waves are called *axial,* whereas those involving two pairs of surfaces are called *tangential,* and those involving three
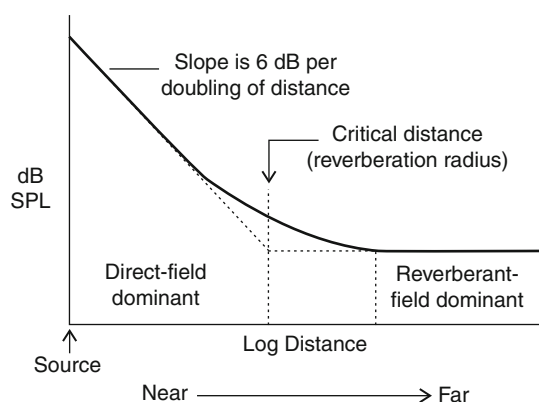
FIGURE 1.17 The falloff in sound level versus distance from a point sound source in a very reverberant room. Near the source, the sound level drops off at 6 dB per doubling of distance (see the end of this chapter for a definition of dB), and far away from the source the level is constant. The distance from the source along the line of principal radiation where the level of the direct sound and reverberant sound is equal is called the critical distance. The critical distance is made longer by a more directional source, or a less reverberant room (one having a lower reverberation time), and vice versa.

pairs are called *oblique*. It is usually the axial modes that show the strongest effects.

The consequence of standing waves is exaggeration or diminution of certain low frequencies at one particular point in the space and strong variation in these perturbations as you move around the space. This problem is worse in smaller rooms and at lower frequencies. When sitting in one position in a small room one hears one thing, but when sitting in another position there may be a quite different impression, caused by the pattern of standing waves in the room.

For production sound, the consequences can be a boomy quality to a sound recording, such that male voices sound like they have an exaggerated "uh-uh" quality to everything said, usually made increasingly worse with increasingly smaller and more live (less absorptive) rooms. In addition to reverberation, this is the reason that men like to sing in the shower—it gives bass boost to their voices. However, the same phenomenon is very troubling for production sound and would call for a great deal of treatment in postproduction to make a recording from a lively shower useful (as well as to reduce the noise).

One principle that applies is that all of the sound pressure "piles up" at the frequencies of the standing waves at the boundaries and intersections of the room (see Figure 1.18). The bass is thus stronger at the walls, stronger still in the corner of two walls, and strongest at the joint of two walls and the floor or ceiling. To filmmakers, this has the following two consequences:

- Most producers sit at the back wall in dubbing stages, where they can talk on the phone while a mix is in progress. Unfortunately, this is just where the bass is exaggerated in the room, so they receive a bass-heavy impression of the mix.
- The effectiveness of low-frequency absorption varies depending on its placement.

Noise

We've used one technical definition of noise already—noise as a sound without tonality, such as a waterfall. The more popular definition of noise is unwanted sound, which of course varies from time to time and from person to person. Thus, no strict definition of noise is possible.

Unwanted sound is a problem for filmmakers at many times and places. In production sound, it can have a strong effect on budgets, because shooting in a noisy location may well mean the actors need to return for an ADR[5] session (also called looping) so that their voices can be rerecorded. In theaters, background noise may be so loud that it obscures the quieter passages of a soundtrack.

[5]Automated dialog replacement.
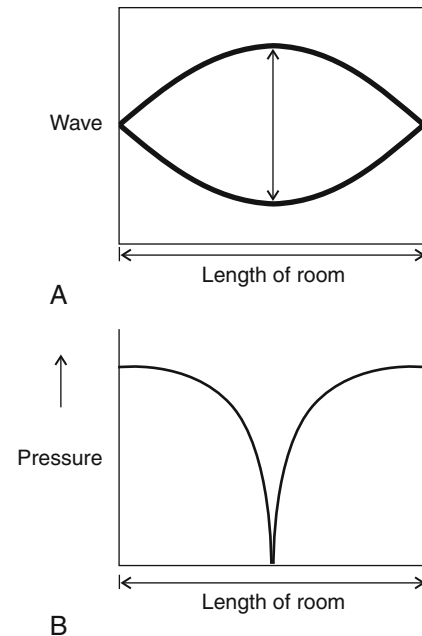
FIGURE 1.18 At the frequency at which one-half of a wave fits the room dimensions precisely, a standing wave is formed. At this frequency, the pressure is high at the two walls and low in the center of the room.

Noise reaches listeners and microphones by two possible paths. Either it is airborne or it is structure-borne, which then radiates from the room surfaces to become airborne. In more serious cases, it may even generate sufficient vibration such that the noise becomes noticeable, not because of its audibility, but because it directly shakes the listener. Airborne noise leaking directly into a set is prevented by having an air seal all around the space. This especially means providing a method to get cables in and out of the space through some kind of air lock; otherwise, the shooting space is in direct contact with the noisy outside world.

Airborne noise sources include transportation noise, air handlers, and many other parts of everyday life. These sources are an extra problem for modern-day shooting, in which the full expressive range of acting is expected to be captured by the production sound crew, with no technical limitation. In the 1930s actors were expected to speak loudly and clearly, and for several years after the introduction of sound there was a sound director in charge of the movie's sound, including the voices of the actors. The sound director was even known to fire actors for their inability to speak clearly. Some careers of silent film stars were ruined by the transition to the talkies because the actors did not sound like what the audience expected. Today, if an actor can't be understood it is routinely thought to be a technical fault, no matter how much of a mumbler the actor is. These problems are made only worse by intrusive noise on location.

One of the first applications of the new science of room acoustics, and particularly noise control, was the building of MGM studios by Louis B. Mayer. During planning it was realized that on one stage there could be a quiet love scene being shot, while on the next stage over a battle was to be raging. A proposition was made to MGM that a wire grid be strung up between the stages to ensure noise control. Luckily, someone at the studio saw through this poppycock and went looking for a scientist willing to tackle the problem (which was by no means small). They found a physicist at the University of California at Los Angeles, Vern O. Knudsen, to help them out, and he proposed massive walls and huge, solid doors for noise control. This led to the problem of heat. Because the shooting stage was a sealed box, combined with the extensive lighting needed for the slow film speeds of the day, actors were strongly affected by the temperature. The first air-conditioning noise control system was invented by Knudsen on a cross-country train trip to see the vendor so that he could meet the simultaneous needs of keeping the actors cool and keeping the noise low enough that production sound could be recorded at MGM.[6]

Structure-borne noise is often even more insidious than airborne noise, and harder to solve, because the whole room may be moving from the vibration. Let us say that there is a subway close to a cinema. The subway in operation puts vibration into the ground, and, in turn, the walls of the theater move in response to the vibration, creating sound emitted in the theater. This problem affected the old Astor Plaza theater in New York city, for example. The only possible fix is to break the connection with the ground and suspend the theater on vibration-isolating mounts, which can be done, but the expense is very large. It was done for cinemas in Tokyo, where subways and buildings are often in close proximity and where the vibration mounts can also be part of a system to reduce the effects of earthquakes.

Filmmakers working on location can usually do nothing about structure-borne noise, except to be aware of its effects and to use special microphone mounting devices, called *vibration isolators* or *suspensions,* to avoid conducting vibration directly to the microphone. Airborne noise, on the other hand, can be treated to some extent by sealing the environment away from the noisy world. Producers should be aware that shooting in pretty but noisy locations will increase postproduction costs, because more treatment of the sound will be required, perhaps including the need for actors to loop their performances.
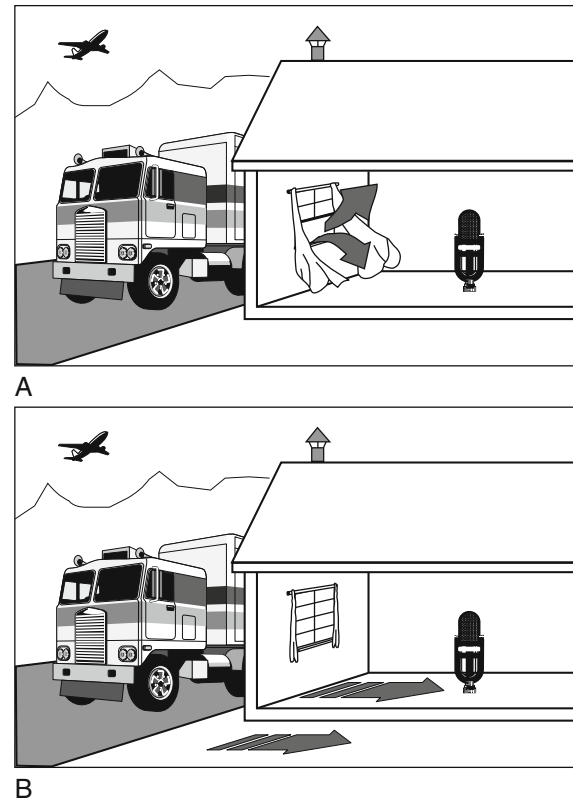
FIGURE 1.19 The two principal noise sources are (a) airborne and (b) structure-borne.

Scaling the Dimensions

Frequency Range

We already know that the frequency range of audible sound is large. To describe the range, we could use a linear scale from 20 Hz to 20 kHz, but we would then devote one-half of the scale to the range from 10 and 20 kHz. In this high-frequency region, only the high harmonics of some instruments and sounds are found, and the region is less important than what lies below it. For this reason, it is more common to think of the frequency scale as being composed of octaves, that is, of having equal increments along a scale of 2:1 units of frequency. One such scale is given in Table 1.1. Audible sound thus covers 10 octaves of frequency (whereas visible light ranges from 400 to 700 nm, a range of less than 1 octave).

Track 6 of the DVD illustrates the audible frequency range, with a "boink" in each octave from 31.5 Hz to 16 kHz, covering a wide part, but not all, of the audible frequency range.

Amplitude Range

We have left the scaling of amplitude range until last in this chapter because it forms a bridge to psychoacoustics.

[6]Knudsen, V. O. (1970). Reminiscences: part 1, noise; part 2, sound stages. *Journal of the Audio Engineering Society* 18, 436–439.

TABLE 1.1 Octave Band Center Frequencies for the Audible Frequency Range

Band number	1	2	3	4	5	6	7	8	9	10
Octave Band	31 Hz	63 Hz	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz	16 kHz

We've already seen how human hearing has influenced the way in which we view the frequency range of sound, despite the fact that there may well be sound outside the audible frequency range. What about an amplitude scale?

Normal sounds encountered in the world cover an enormous range of amplitudes, with a range of about 1:1,000,000 in the sound pressure level difference from the softest sound we can hear to the loudest we can normally stand. To represent this world conveniently, it is commonplace to use a logarithmic scale, in which equal increments along the scale are given in powers of 10 as 1, 10, 100, ... This log scale is used principally for convenience in representing the extreme range of amplitude encountered in real life. With some mathematical manipulation, the log scale is represented in the unit decibel (dB) and this is the most widely used unit for noting the amplitude of physical and recorded sound.

The term decibel always refers only to a ratio of amplitudes. Although very often lost, a reference is always at least implied if not stated. In the case of physical sound, the decibel scale for sound pressure level (SPL) has its reference at 0 dB SPL, roughly equal to the threshold of hearing, the softest sound one can hear. Quiet background sound such as that found standing in a still field at night is about 30 dB, face-to-face speech averages 65 dB, typical film dialog around 75 dB, and the loudest sounds in theatrical films are up to 115 dB, all in dB SPL.

The decibel scale for sound pressure level for audible and tolerable sound runs from 0 to 120 dB SPL (Table 1.2). Of course, there are sounds louder than we can tolerate as listeners, and sounds softer than we can hear, so this scale can be carried to both lower and higher levels. The reference is approximately the threshold of hearing, 0 dB SPL, which is a pressure of 20 μN/m^2.

Along a decibel scale, 3 dB is twice the power and 6 dB is twice the voltage, but it takes about 10 dB to sound twice as loud. Thus stereo salesmen who claim that a 60-W receiver plays much louder than a 50-W receiver are wrong—the difference is only 0.8 dB, and the loudness difference is very small!

See also Table 6.1 in Chapter 6 for further sound pressure levels.

Track 11 of the DVD illustrates 80 decibels of the audible dynamic range, with a special noise signal recorded in 10-dB steps.

TABLE 1.2 Typical Sound Pressure Levels Relative to 0 dB SPL

120 dB	Threshold of sensation; "tickle" in the ear
118 dB	Loudest sound peaks in digital theatrical films[a]
90–95 dB	Loudest sound peaks in 35 mm analog theatrical films[b]
80–90 dB	Typical loudest sounds from television
75 dB	Average level of dialog in films
65 dB	Average level of face-to-face speech
50 dB	City background noise (but varies greatly)
30 dB	Quiet countryside
20 dB	Panavision camera running with film measured at 1 m
0 dB	Threshold of hearing in silence

[a]In digitally equipped theaters playing at the standardized volume control ("fader") setting, measured with a wideband, peak-reading, sound pressure level meter. This combination of ingredients leads to the highest possible reading, but one that does *not* represent the noise exposure of an audience to sound level. This is because a movie has constantly changing levels, with rare peak moments at such a high level. See Chapter 2 for more information on the loudness level of movies.
[b]Measured as in footnote a.

This page intentionally left blank

Psychoacoustics

INTRODUCTION

Many issues in human perception of sound are a direct result of the physical acoustics of the head, outer ears, ear canal, and so forth, interacting with sound fields. Although not strictly in the realm of psychoacoustics, this kind of physical acoustics occurs only because a person is present in a sound field, so we will take it up here. The head is a rather large object, acoustically speaking, so there are many interactions between the head and a sound field. A sound wave arriving from the front must "spread around" the head from the front to the sides through diffraction, interact with the outer ear structure principally through reflection, progress down the resonant ear canal to the ear drum, and so forth. This interaction between the sound field and the object observing the sound field is fundamentally different from an interaction with, say, a tiny measurement microphone. The small size of such microphones is specified by the need to measure the sound field with only minimal disturbance, almost as if the microphone were not there.

The placement of the head in a sound field also matters. As children, we first hear reflections off the ground shortly after direct sound, because we are close to the ground. As we grow taller, the reflection is heard later than the direct sound. In neither case is the sound late enough to be separately perceived from the direct sound but rather affects the direct sound. The difference between the two conditions is a result of the physical acoustic differences of the conditions; nevertheless, we incorporate them into "perception." We are, after all, continuously training our perception because we use it to function in the world all the time. We learn that a certain pattern of reflections represents our bodies standing in the real world. Let us look at those parts of perception that are most influenced by physical acoustics first, and then examine the more psychological aspects.

THE PHYSICAL EAR

The first part to consider in talking about "the ear" is actually the human body, especially the head. As stated earlier, incoming sound waves interact with the head as a physical object, with sound waves "flowing" around the head via

diffraction. (A relatively minor effect even results from sound reflecting off the shoulder.) The interaction differs for various incoming angles of incidence, making at least the levels and times of arrival of the sound wave at the two ears different.

After reaching the outer ear structure, called the *pinna*, various reflections and resonances occur because of the structure of the outer ear convolutions. Older treatments on how this worked concentrated on the pinna's hornlike quality, its ability to gather sound from many directions and deliver it to the ear canal. Since the 1970s, the detailed role of the pattern of reflections and resonances caused by the pinna has come to be better understood. The details of the pinna structure play an important role in localizing sound because the pattern differs for sound arriving from different directions. We come to learn these patterns and rely on them for finding direction.

After interaction with the head and pinna, sound waves enter the ear canal. The length and diameter of the ear canal tube cause it to be tuned, like a whistle, to the mid-high-frequency range. This ear canal resonance increases the sensitivity of hearing around 2.5 kHz by a factor of 20 dB. An increase in sensitivity in these middle-high frequencies proves highly useful in survival, because it is the range in which many threatening forces might make noise, such as the sound of a snapping twig behind us or the sound of lions moving through tall grass.
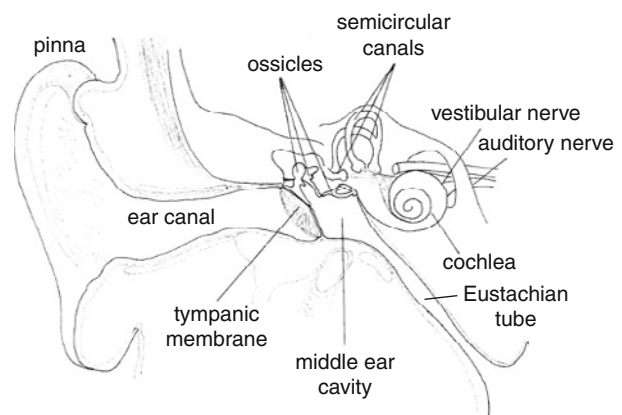

FIGURE 2.1 A cutaway drawing of the ear showing the primary parts described in the text.

At the end of the ear canal lies the eardrum, which stretches across the end of the canal and seals it. This *tympanic membrane* vibrates in response to incident sound. Sound is airborne up until it strikes the eardrum; thenceforth sound is represented as vibration of the structures further inside the ear, although for convenience it is often still called *sound* (to distinguish it from *vibration,* which is felt by the body rather than heard). The inside cavity beyond the eardrum is called the *middle ear* and is supplied with air by way of the *Eustachian tube,* which exits into the throat. The purpose of the Eustachian tube is to keep the eardrum from being pushed to the limits of its travel by pressure changes in the atmosphere. The Eustachian tube equalizes the ambient pressure on the two sides of the eardrum, thus letting the eardrum come to rest at the center of its possible range of displacement, rather than becoming "stuck" to one side. It is these tubes that become clogged when you have a cold. When they are unable to equalize pressure on both sides of the eardrum, it is pushed all the way to its limit of travel, and hearing suffers greatly. It may be dangerous to our hearing, or at least very painful, to fly with a cold and go through the air pressure changes involved in air travel, first to lower pressure while ascending and then to a higher pressure while descending. During such airplane flights, or even while going up or down in an elevator, it is a good idea to try to get your eardrums to "pop" to relieve the pressure before it reaches an extreme by moving your jaw around, chewing gum, or holding your nose shut while attempting to blow air out it gently, thus forcing air up through the tubes into your middle ear.

Inside the middle ear, the eardrum is connected to three tiny moving bones, called *ossicles,* organized in sequence. These three bones form a mechanical lever that magnifies the motion of the eardrum by another factor of 20 dB, albeit peaking in a lower frequency range (250–500 Hz) than the ear canal, before delivering it to the inner ear. Attached to these bones are two small muscles, whose contraction and relaxation are affected by the sound level. A loud sound causes the muscles to tighten, producing the *aural reflex.* Tightening the muscles reduces the transmission between the eardrum and the inner ear. The aural reflex is most active with intermittent, intense impulses of sound, but it does take time to act. For these reasons, the first burst of gunfire in an otherwise quiet sequence sounds loud, and later bursts may sound softer as the aural reflex begins to act. Filmmakers often ignore the fact that continuous loud sound is no longer perceived as loud by the audience because the aural reflex "turns down the volume," making the scene less effective than expected. Many contemporary action–adventure films clearly suffer from this problem, providing no time for the audience to relax the aural reflex before another loud sound is presented.

The inner ear, or *cochlea,* has several roles. It is the organ that converts the vibrations received from the middle ear bones into nerve impulses destined for perception by the brain, and it also has an important role in maintaining physical balance. The cochlea is a small, snail-shell-shaped organ with a "window" at one end to which the last bone in the series in the middle ear is attached. The length of the cochlea is bisected by a stretched *basilar membrane,* on either side of which are fluids. About 30,000 hair cells are present along this basilar membrane and are attached to nerve cells. It is the motions of the hair cells that get converted to nerve impulses that register with the brain as sound.

Sound waves are magnified by ear canal resonance, which is converted into mechanical vibration by the motion of the eardrum, and are magnified again by the lever action of the middle ear, producing a combined nine times gain in amplitude. The inner ear selectively converts this mechanical motion into electrical signals via the aural nerves that correspond to the hair cells.

Specific sounds are destined to stimulate specific nerves because the basilar membrane is stretched stiffly suspended at the entrance for sound and is stretched loosely at the far end. Like tuning a drum, the tighter the drum is stretched, the higher the pitch is when the drum head is struck. The membrane in the cochlea acts like a continuous series of drum heads, each tuned to a progressively lower frequency as the distance from the input increases. Thus if the incoming vibrations are relatively fast, corresponding to a high frequency, they cause the greatest stimulus where the membrane is stiffly suspended, near the entrance. A lower note causes more movement where the membrane is more loosely suspended, farther from the entrance. The hair cells that move the most put out the largest number of nerve firings and thus indicate to the brain that a particular frequency range is strong in the sound signal. The cochlea is thus basically a frequency analyzer that breaks down the audio spectrum and represents it as level versus frequency, rather like Fourier frequency analysis described in the last chapter[1] but with important differences that we will consider later.

Hearing Conservation

Evolution recessed the eardrum to protect it from damage and increase the sensitivity to important sounds. The cochlea is in one of the most protected regions of the body, inside the skull. These are examples of the importance nature placed on hearing. Two main factors cause hearing to deteriorate: aging, and exposure to high sound levels.

Hearing deteriorates naturally with age, which has been found even in quiet, primitive societies. Hearing loss

---

[1]See page 6.

with age primarily affects high-frequency sensitivity, and there are statistically significant differences between men and women, with women faring better with age. It is not known whether this is an innate biological difference or is caused by more men than women working in noisier industrial environments and having noisier hobbies.

The fine hairs of the inner ear can be permanently damaged by one very loud sound; even the U.S. Army does not allow recruits to be exposed to sound pressure levels over 140 dB, even briefly. Gunfire close at hand is the most likely source of sound that may cause a permanent loss with as little as one exposure. A European study of 17,000 men and boys found significant hearing losses in those that had been exposed to as little as one traumatic hearing event, such as having a toy horn blown in their ear as a child.

Luckily, film and television sound systems typically cannot play loudly enough to cause this kind of near-instantaneous damage. On the other hand, lower levels of sound, when accumulated over time, may lead to hearing loss. An engineer working for a chain of theaters throughout England found in the 1960s that the volume control was set higher, on average, in movie theaters in the north of England than in the south. The audience in the north consisted largely of textile mill workers, still working with noisy 19th century machinery. The average hearing loss of the population was greater in the north than the south of England, because of occupational noise exposure.

Noise-induced hearing loss begins to occur when the sound level is above a certain threshold, set by standards to typically 80 dB,[2] called the *threshold level*. Exposure to sound below this level is ignored as not playing a significant role in sound-induced hearing loss. Above the threshold there is a trade-off: *louder* sound is permitted, but only for a *shorter* time. A *criteria level* of 85 dB may be allowed for 8 hours, 88 dB for 4 hours, and so forth. This is called a 3 dB (level versus time) trading rule. Thus continuous noise sources of many industrial machines can be ranked for how much exposure is created by them, and a program of abatement of the noise source or wearing required hearing protection can be put into place.

Listening to movies and television shows makes life more complicated than simple continuous industrial noise because they have constantly varying level. A method of averaging the level over time is required, and a widely used one is called L_{eq}. Using the concepts of threshold and criteria levels, the trading rule, and L_{eq} together results in a percentage value called a *daily noise dose*. Going to

the movies in theaters that play them at their original post-production levels (note: many theaters turn the sound down from the original level), results in noise doses of between 2 percent, on a dialog-driven picture with one loudish scene (*Tea with Mussolini*), and 55 percent, for a space adventure (*Episode I: The Phantom Menace*); thus going to the movies, even daily, does not expose one to enough sound to cause any statistically significant greater hearing loss than normal aging.

Conversely, working on films and television shows for long hours at a time, especially dubbing loud action features, may well produce noise doses that are above the amount that will cause sound-induced hearing loss greater than aging. Professionals who work on loud films may find themselves confronted with the dilemma of working now versus working over a long career. Some mixers have resorted to wearing hearing protection on the dubbing stage, because their exposure to the loud passages of a movie are so much greater than that of their ultimate audience, given the hundreds of times they may listen to a loud passage. Manufacturers make hearing protection devices that lower all levels across frequency equally, so the high-frequency rolloff of most hearing protection systems is not a problem for these special types.[3]

There are undoubtedly sound professionals who have suffered various degrees of hearing loss in the pursuit of their professional duties, although the evidence for this comes mainly from music recording, not film and television work. In rock music recording, an average monitor sound level may be 10 dB higher than it is for film dubbing and, without controls over that level, can go much higher.[4] Television mixing is, in turn, done at a level typically 5 to 7 dB lower than film dubbing, so the potential for hearing loss is certainly less for film and television mixers than for rock music mixers.

Avoid attending live amplified concerts without wearing hearing protection. The sound pressure level of these concerts is routinely more than 10 dB louder than the loudest sound possible in a film, and this level is sustained for much longer times.

> Each person varies in his or her susceptibility to hearing damage. One person at a specific rock concert may lose a significant amount of hearing permanently, whereas another person sitting next to him may be affected only short term. There is no way to know in advance whether you are a particularly susceptible person.

One indication that you are exposed to too much sound level over time is if you experience tinnitus after the event. Tinnitus may variably be described as an internal

---

[2]This number is measured with an *A weighting* curve that emphasizes middle-high frequency range sound over low- and high-frequency sound, to match better the human sensitivity to sound levels.

[3]E.A.R. makes ER-15 for 15 dB of attenuation and ER-25 for 25 dB. ER-15 is suitable for film dubbing.

[4]Film dubbing is done at a standardized "volume control" setting, but music mixing generally uses no such standard.

whistling, ringing, buzzing, hissing, or humming heard when you are in a quiet environment.

Certain drugs, such as some of the most powerful antibiotics, may have negative effects on hearing, and their use is restricted to life-threatening cases and is accompanied by monitoring of hearing. Even aspirin may affect tinnitus, with the onset of hearing the effect indicating too large a dose. High-impact aerobics, professional volleyball, and high-mileage running have been implicated in high-frequency hearing loss and balance problems for some fraction of the participants in these sports.

AUDITORY SENSITIVITY VERSUS FREQUENCY

Threshold Value—the Minimum Audible Field

Human hearing does not respond equally well to all frequencies in the audible range. This is because evolution has caused our sensitivity to be increased in frequency ranges at which threats and voices might be most easily perceived, while being relatively less sensitive at other frequencies. As a broad statement, humans are most sensitive around the middle to high frequency ranges, with the worst sensitivity at the low and more extreme high frequencies. Scientists can measure the threshold of perception versus frequency with relative ease because it is largely a matter of finding a space quiet enough to be able to hear the softest sounds, and then presenting calibrated sound-pressure-level sine waves at various frequencies to large numbers of listeners (usually headphones are used to eliminate the effects of room acoustics, which otherwise must be accounted for). By averaging a large number of observations, individual variations disappear, leaving a curve of auditory sensitivity versus frequency, the *minimum audible field*. The reason that this is a straightforward psychoacoustic experiment is that we ask the subjects whether or not they hear something.

An experiment involving scaling loudness is much more difficult, because it is a lot harder to get common agreement on what is twice as loud, comparing one sound to another, than to determine whether a subject hears any sound at all. *Hearing loss* of an individual is rated in terms of the increase in hearing threshold for various frequencies compared with the standard minimum audible field.

Of course, such average curves do not characterize any one individual, but they are accepted in international standards as being representative of the population as a whole. The standard deviation on these experiments is on the order of ±5 dB, that is, 67 percent of the normal population will lie within ±5 dB, and 90 percent within ±10 dB, of the average curves. The greatest sensitivity corresponds to the bottom of the curve in Fig. 2.2 because
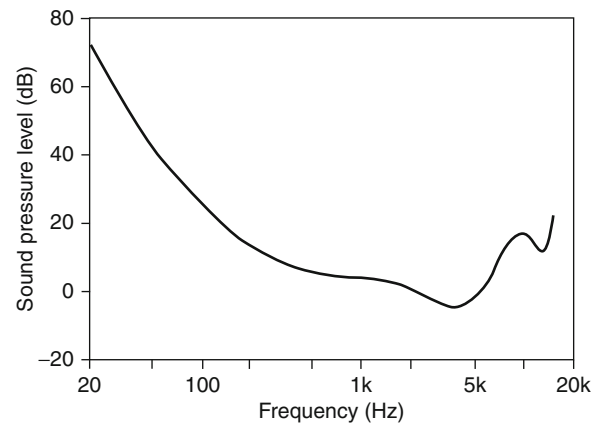
FIGURE 2.2 A curve of the minimum audible field versus frequency, showing the greatest sensitivity for human hearing in the 2–4 kHz range and greatly diminished sensitivity at higher, and especially lower, frequencies.

that is where the smallest possible sound level can be detected. Note that 0 dB SPL is roughly the bottom of the curve (constructed such that the numbers will almost always come out positive), but the minimum audible field does go below zero in the most sensitive range, at around 3 kHz.

In light, if we are properly dark adapted, we can see one photon, the minimum divisible packet of light energy. In sound, we can hear displacements of molecules in air that are about one-third the size of the orbit of the sole electron spinning about the nucleus of the hydrogen atom. Perception is so finely tuned to the environment that any further development would not be very useful. For example, the minimum audible field has a higher threshold at lower frequencies than at midrange frequencies, as shown in Fig. 2.2. This is thought to provide the evolutionary advantage of helping humans sleep soundly. If we had greater sensitivity at low frequencies, we would spend excessive time listening to internal bodily functions such as breathing.

Equal-Loudness Curves

At levels higher than threshold, perceptual scientists ask another question of large numbers of people, and the scientists get good results. The question is, How strong must each frequency be to sound as loud as the same sound pressure level at, say, 1 kHz? That is, how does loudness vary with frequency? The outcome of such an experiment produces curves of equal loudness.

You can think of them in this way. The background "grid" on the graph is the physical world, whereas the lines on the graph are the psychoacoustic representations of that world (see Figure 2.3). For instance, take a sine wave tone at 1 kHz and 20 dB SPL. This is a clear, although soft, tone. Now change the frequency to 100 Hz. What we see is that at 100 Hz we have now dropped below the threshold of hearing and no sound is perceived at all,

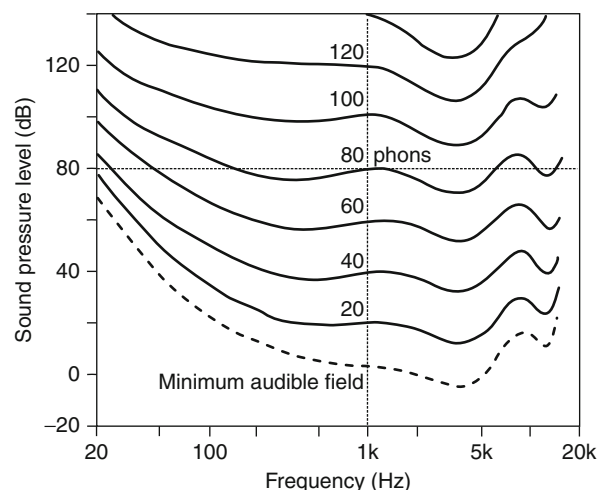**FIGURE 2.3** The curves of equal loudness versus frequency. The background grid represents the objective world, and the curves represent the response of an average human listener to that world in loudness. The contours are labeled in *phons,* one unit of subjective loudness, which is equal to the sound pressure level at 1 kHz. Taking the point 80 dB at 1 kHz as a reference, and following the 80 phon contour down to 20 Hz, shows that about 110-dB sound pressure level is needed to sound equally as loud as an 80-dB SPL 1-kHz tone. Human response to sound pressure level does not stop abruptly at high frequencies as shown, but this is the limit of experimental data.

simply by a change in frequency. So the frequency of a tone affects loudness, as does its sound pressure level.

Another finding has to do with the shape of the curves. Notice that at no level is hearing flat with frequency. Because we perceive the whole world in this manner, we do not find this unnatural, but it does have some serious consequences for film sound. For example, all of the curves go up at low frequencies.

This means that more energy is required at low frequencies to sound equally as loud as midrange sounds. Recently designed film sound systems include an extra low-frequency only channel called Low Frequency Effects, with greater level playing capacity in the low bass than in the midrange to account for this psychoacoustic effect.

> Yet there are plenty of places in audio recording and reproduction where this lesson has not been applied, such as analog optical soundtracks. For example, *Return of the Jedi's* analog optical track reveals that Jabba the Hut's basso profundo voice uses up practically all of the available area of the soundtrack[5] yet is not very loud. The reason for this is that the technique used by the sound designer made Jabba's voice very bassy and thus bigger and more threatening sounding (remember *Jaws*). In order to sound even normally loud, the level of Jabba's voice had to be adjusted upward, which used up practically all of the available area of the soundtrack. Conventional optical soundtracks have a maximum capability that is flat with frequency, whereas hearing requires more low frequencies than high

frequencies to sound equally loud. A director who demands "more bass" is, of course, right, because that is what he or she thinks necessary to make the point, but getting what a director wants may be limited by the capability of the medium in which he or she is working.

Another effect occurs because the equal-loudness curves of perception converge at low frequencies, which is called the *loudness effect.* This occurs when sounds are reproduced at a level higher or lower than the original sound. At a lower level, sound seems to lack bass compared with playing it at its original level. This is often a problem in film production, in which music, faded underneath dialog, becomes thin sounding, that is, lacking in bass. A postproduction mixer can make up for this effect by adding bass as the level is turned down. Track 12 of the DVD illustrates the loudness effect.

WHAT'S WRONG WITH THE DECIBEL— MAGNITUDE SCALING

Although a logarithmic scale serves well to "compress" the range of the real world into manageable numbers, the relationship between a strict log scale such as decibels and hearing perception is not simple. Research studies have not settled on one single amplitude scale because experimental results are affected by the method used to obtain them. After all, we are dealing with human perception; we cannot hook up a loudness meter inside a brain to measure the sensation level but must rely on listener reports, such as "This sound is twice as loud as that sound." Depending on the experiment, "twice as loud" falls along the decibel scale between 6 and 10 dB, tending toward the higher value more often. So for everyday purposes, we say that 10 dB is twice as loud.

> Once again, note that twice as much power is 3 dB, twice as much voltage is 6 dB, and twice as loud is 10 dB. This means that to play music twice as loud without distortion as, say, a 50 W receiver requires a 500 W receiver!

The minimum difference in level that is detectable on program material is usually said to be 1 dB; however, this is a number for untrained laypeople. Trained sound mixers can set sound levels so they match to within ½ dB.

LOUDNESS VERSUS TIME

The foregoing discussion regarding loudness concentrated on more or less continuous sounds, such as that of a kitchen fan. Although ranking constant sources for perceived loudness is certainly important, it is also useful to know how loudness changes over time for more rapidly changing sounds.

This factor is called the *integration time* of hearing, because it takes a finite amount of time for loudness to

---

[5]The recorded width of the optical soundtrack corresponds to the amplitude of the waveform.

grow to its full value. A high-level sound presented only briefly, say, for less than one frame, will not achieve the full loudness as a sound having the same level that is sustained for about one-third of a second. There are many variations in the types of experiments that have been done in this area, including center frequency of the phenomena, number of occurrences within a period of time, etc., and consequently a large range of results can be obtained for the "time constant" of hearing. The range of time for a high-level but brief sound to reach nearly its full loudness is between 35 and 300 ms (one to eight frames at 24 fps). Track 13 of the DVD illustrates loudness versus time.

SPECTRUM OF A SOUND

Critical Bands of Hearing

The audible frequency range can be divided into 24 frequency regions, called *critical bands,* each representing about 0.05 inch in length along the basilar membrane and encompassing about 1300 nerve receptors. The frequency width of the critical bands varies, being wider in the bass than the midrange and treble, but typically they are about one-third of an octave wide.

When two tones of equal amplitude that lie within one critical band are added together, the result is an increase in loudness. If the same amplitude tones are farther than one critical band apart, the result is an even greater increase in loudness. A typical meter displaying the amplitude will have the same reading in either case, but in the second case, we perceive a louder sound. So the loudness of sound depends not only on the level measured, but also on the *spectrum* of the sound. The spectrum is a plot of level versus frequency, for example, showing all the harmonics of a sound.

> Note that the critical bands are not fixed at rigid frequency boundaries but rather "slide" to fit the stimulus. The critical band concept thus involves the selectivity of the ear, that is, how narrow a frequency band the effect covers, not the absolute frequency limits.

In general, the more spectrum a sound takes up, the louder the sound will be. For a low-frequency rumble to sound louder, it is useful to add other higher frequency components to the sound, such as adding the sound of gravel being poured from a dump truck to the sound of a bassier rumble. So, to maximize loudness, frequency components spanning a wide range are used, usually by the addition of added sound layers editorially.

FREQUENCY MASKING

Louder sounds cover up softer ones, especially those that are nearby in frequency, which is called *frequency masking*. The idea of masking is a major plot point in Alfred Hitchcock's *The Man Who Knew Too Much.* Assassins plan to shoot an ambassador during a concert in London's Royal Albert Hall. The chief plotter uses a phonograph record of the music to be played at the concert to train the assassin when he can shoot and not be heard by firing the gun simultaneous with a big cymbal crash. The record is played several times to get the point across. Whether masking would in fact work in this context is a matter of some conjecture and depends on many factors, but it certainly works in this movie, which has a 12-min section with no dialogue surrounding the shooting. In the end, Doris Day screams in the silence just before the cymbal is to crash and distracts the assassin just enough that his aim is spoiled and the ambassador is only wounded instead of killed.

To make the maximum number of different sounds audible at one time, it is useful for them to be spread out across frequency to minimize the masking of one sound by another. Editors and composers employ a strategy of choosing different frequency ranges for a variety of effects and music, just to minimize the masking of one sound by another.

At low levels, frequency masking effects are contained in a relatively narrow frequency range about the masker (the louder sound that does the masking). As the level increases, the masker becomes more effective at masking than just a level change would predict, and this effect is greater toward higher frequencies than toward lower ones, as shown in Fig. 2.4. Thus, there is an upward spread of
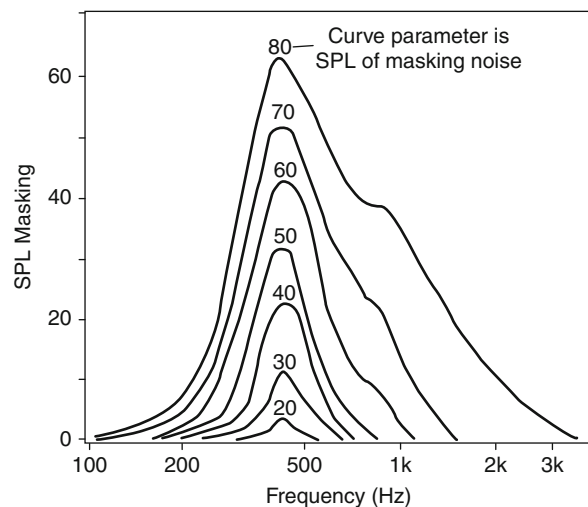
FIGURE 2.4 Frequency masking. A narrow frequency band of noise centered near 400 Hz masks sound at nearby frequencies, and the effect grows to cover a greater frequency range, especially toward the higher frequencies, as the level increases. The curves are numbered with the sound pressure level of the masking noise and show the level versus frequency that is masked. In other words, any frequency components lying below the curves are masked when a band of noise at 400 Hz and the stated SPL is present. Adapted from Egan, J. P. and Hake, H. W., "On the masking pattern of a simple auditory stimulus," *J. Acoust. Soc. Am.* 22, 622–630.

masking with increases in level and more of a masking effect at higher levels.

Frequency masking is routinely used in another way by sound editors. Let us say that we have a production sound-track recording of a scene that is reasonably good, containing fine performances by the actors, etc., but there is a flaw: the background noise changes from shot to shot within the scene in a noticeable manner. This could easily arise if the scene was shot out of order over the course of a day and the noise level at the location changed over time. The background noise difference might not have been noticeable at the time of shooting, but when the shot made in the morning is cut together with that made in the afternoon, the change is heard at the edits. The abrupt changes in room tone, as the background noise is called, distract from the feeling that the scene is one continuous whole. Various processes are used in postproduction to clean up this background noise, but there may still be a residual change at the edits.

To prevent the audience from hearing small background level changes at edits a second track is provided, running in parallel with the first and containing an ambience appropriate to the setting, and it is mixed together with the production sound. The continuous nature of the ambience and its use in masking the actual recorded background noise produce smooth-sounding edits and good continuity.

Track 14 of the DVD demonstrates frequency masking.

TEMPORAL MASKING

A loud sound can mask a soft sound that does not occur at precisely the same moment. This seems clear from the example of, say, a gunshot covering up a soft sound for a time after the shot, and is called post or forward masking. A higher level masker (such as our gunshot) extends the time of masking further than a lower level one, as might be expected.

One of the most astonishing findings of psychoacousticians that, at first glance, confounds a scientific mind is that temporal masking works in the other direction as well. What is amazing is that the same gunshot covers up sound momentarily even before the shot! This occurs because the louder sound is perceived more quickly by the brain than the softer one, and this is called premasking or backward masking. The extent of this effect is not very large, with most of its utility occurring within 10 msec, or one perforation of 35 mm film, $^1/_4$ frame, at 24 fps. Nevertheless, sound editors use backward masking frequently to cover up discontinuities at edits. A sound editor will edit music "on the beat" by cutting just before a loud cymbal crash is to occur. The cymbal crash then covers up the fact that there may be a momentary discontinuity at the actual edit point, using premasking.

Track 15 of the DVD illustrates temporal masking.

Knowledge of the details of frequency and temporal masking has been fundamental to the development of a relatively new branch of audio, low-bit-rate digital coding, which is discussed on pages 186–187. The basic idea behind this technology is that there are sounds that are inaudible to human listeners because they would be masked, and thus there is no need to store and transmit such sounds. This topic is controversial, as are many in which psychoacoustics are used, because agreement among a great many listeners must be reached to say what is truly inaudible.

PITCH

The subjective sensation of pitch is often thought of as being interchangeable with the fundamental frequency of a note, but there are several factors on which pitch depends other than frequency. No instrument can measure whether a musician is in fact playing in tune, because pitch depends not only on fundamental frequency but also, less strongly, on level and possibly other factors, so expert listeners must determine whether a musician is playing in tune.

Another example of the difference between the objective measure of frequency and the subjective measure of pitch occurs if a recording is made and the fundamental is removed. Perception will recognize a "missing" fundamental in the pattern of harmonics and will supply it. Subjectively one hears the missing fundamental, even if it is not present, and associates it with a pitch (of a non-existent frequency in the sound). Naturally the idea of pitch applies to musical notes, but what about to voice or other sounds? The speaking voice is also said to have pitch, despite the fact that it is not singing. Raising the pitch will make voices seem more feminine, and it is likely that films that use men in drag also alter their spoken voice by deliberate upward pitch shifting in postproduction.

Another example of pitch being associated with nonmusical sounds is *repetition pitch*. In 1693, the French astronomer Christiann Huygens, standing near a staircase, was struck by how the sound of a fountain nearby seemed to contain a pitch sensation. Repetition pitch occurs when direct sound is heard along with a pattern of reflections. The fact that the reflections are evenly spaced in time reinforces certain frequencies, while partially canceling others, because of constructive and destructive interference. Because the steps are a regular structure, so is the pattern of reflections, and their effect on the direct sound is also patterned in such a way that we hear the effect.

An example of pitch shift causing changes in a filmmaker's work occurs routinely with transfers of films to the PAL system of video used in Europe and elsewhere. PAL video uses 25 fps, not 24, so theatrical film is sped up by 25/24, or about 4%, for transfer to European videos. Increasing the speed of playback raises the frequency of sound, and thus the pitch. For example, Darth Vader in *Star Wars* takes on a less menacing quality as James Earl

Jones' voice is misrepresented by the pitch being raised. Some films are transferred through a compensating pitch-shift device to remove this error, but this compensation regrettably is used infrequently.

SPATIAL PERCEPTION

Spatial perception is one of the areas in which technology and art are most interrelated. Perceiving the sonic world in three dimensions is something we are accustomed to doing every day, and film and television sound systems are designed to provide the means to reproduce the most salient directional characteristics of the real world, with an ability to place sound around you. The aesthetic difficulty comes in choosing what sounds to represent where, because the picture has boundaries, but the sound is unbounded in the same sense. We take this idea up later, but first let us examine the underlying psychoacoustics to see how it has an impact on film.

Transients and the Precedence Effect

Localization is best for transient sounds. A *transient* is a brief sound such as fingers snapping or a drum hit. It is worst for relatively steady-state sounds such as pipe organ notes. Striking a piano note first produces an onset transient, which then tends toward steady-state sound over time. Nevertheless, the brief transient attack at the beginning of the note gives us enough information to localize it, even in the presence of a lot of reflections and reverberation. This attention to the first-arriving sound is called the *precedence effect*. The precedence effect says that we will locate a sound in the direction of the first-arriving wavefront, unless later-arriving sound is even higher in level than the direct sound. The effect is quite strong and has obvious utility in survival, because locating the tiger early is critical in avoiding it.

Influence of Sight on Sound Localization

Vision is obviously also important for localization and can overwhelm aural impression. Sight dominates sound for localization. Nevertheless, mismatches between the position of a sound source visually and aurally do cause cognitive dissonance, which tends toward limiting the suspension of disbelief usually sought. Here, professionals and laypeople differ on their level of perception and the annoyance experienced from mismatches. For professionals, just a 4° mismatch in the horizontal plane between the position of visual and aural images is noticeable, whereas it takes a 15° mismatch to annoy average laypeople.

An example of vision dominating sound in film is the "Exit Sign" effect. In *Top Gun,* when jets fly left to right across the screen and then exit screen right, what may be perceived aurally is the jet flying off-screen as well, right into the exit sign. In fact, the multichannel film sound system does not have the capability to accomplish this effect technically, but it is nonetheless perceived because of vision overwhelming auditory spatialization.

Localization in Three Dimensions: Horizontal, Vertical, and Depth

Human localization is best in the horizontal plane because our ears are located on the sides of our heads, not the top and bottom, and, crudely speaking, triangulation works. The reason for this evolutionary adaptation is obvious: threats in our primeval environment most often came in the horizontal plane and so needed to be localized well, and quickly! The triangulation idea of using both ears for direction finding is a crude one because what actually occurs is considerably more complex. Conventional triangulation would rely on two spaced pickup points "seeing" the whole sound field, unobstructed. Instead, there is a large object in the way—the head. Let us think about the head as a simple hard sphere, with pickup points for sound located where ear holes would be.

Even in this simple model, sound waves impinging on the head from a direction of, say, right-front in the horizontal plane reach the right ear hole first. Slightly delayed, sound reaches the left ear hole by diffraction about the head, with the delay being caused by the extra time it takes for sound from the right to wrap around the head, plus the "as the crow flies"[6] time.

> Human perceptual capability is so good that JNDs (just noticeable differences) of the time difference between ears in the most tightly controlled experiments are on the order of 10 µsec! Four thousand of these time intervals pass by in one frame of time at 24 fps. That is, moving a source in the horizontal plane so that it arrives 1/100,000 of a second earlier at one ear than the other, compared with the reverse, can be apparent under the most sensitive experimental conditions. This is an astonishingly fast speed, much faster than brain transmission processes, but note that it is not absolute time that is being measured but the match between two paths from the two ears, which must match within the brain remarkably well.

For this sound field coming from right-front, we say the left ear hole is *in the acoustic shadow* of the head. The shadow effect varies with frequency and reduces the level at the left ear relative to the right. At low frequencies, sound diffracts around the head easily, so the level is nearly the same at the two ears, but for high frequencies the head appears to be a larger object because of the shorter wavelengths of high frequencies, so the level is substantially reduced at the left ear.

---

[6]An English idiom that means "by the most direct path."

So there is both a time difference arising out of the geometry and a level difference due to the shadowing effect. The time difference is used primarily as the localization strategy up to about 1 kHz, and the level difference is used from about 4 kHz up. This leaves a hole between the two strategies, which might suggest poor localization performance in this range, and that is the experimental finding.

For the model just described, we used a sphere of average head dimensions with ear holes as pickup points for illustration, and its use demonstrated amplitude and time differences well. In fact, such a sphere with embedded microphones in the positions of two ear holes is on the market as a stereophonic microphone, which is claimed to be capable of reproducing the level and time differences of hearing. The pinna were left out of our model but play a part in localization as well. As stated earlier, the convolutions of the pinna cause various interactions with sound waves, depending on their direction of incidence. Although there are a variety of effects, the main one is the difference caused by reflections off the *concha,* the principal cavity leading to the ear canal, combining with the direct (or diffracted) sound. The combination of the direct sound with a reflection causes constructive and destructive interference at the entrance to the ear canal.

Pinna reflections occur, of course, for sound in every plane, including the horizontal, but because the horizontal plane uses two-eared listening time and amplitude differences effectively, pinna effects are more important for localizing sound in the vertical dimension than in the horizontal. Because the pinna are fairly small acoustically, the effects occur mostly at high frequencies, at which the wavelengths are short, greater than 6 kHz. The perceptual outcome is that vertical resolution is quite a bit worse than horizontal, so film sound systems are designed to produce localization effects largely in the horizontal plane, because that is the most effective plane.

The third dimension of localization is the depth dimension. In this dimension, perception is the most rough, because there is less information to distinguish sound distance than in the other dimensions. Still there are several mechanisms for hearing to obtain depth estimates. These include:

- Amplitude and brightness of the source, compared with experience: a closer source is louder and brighter than a distant one, as occurs acoustically (see Chapter 1);
- Audibility of the ground reflection and how it changes with time out of doors;
- Doppler shift for moving objects;
- In rooms, the pattern of early reflections tells us information about the size of a space and helps us locate the surfaces (something that the blind do remarkably well; sighted persons rely much more heavily on vision for this);
- In rooms, reverberation; longer reverberation times usually mean larger spaces.

Even though the depth dimension is the "worst" perceptually, it is nevertheless very useful and is manipulated continuously by filmmakers. Differences in the depth dimension can be used even in the simplest, monaural (one track or channel) productions to "place" sound in space, at least in one dimension, and thus have been used since shortly after the introduction of film sound.

Examples of the use of the depth dimension are many. A few follow:

- Making a voice-over narration much less reverberant than on-screen action, thus separating the narrator to a "voice inside the head," is used widely in documentaries and also in narrative films, such as the voice-overs in *Apocalypse Now*. There, closeness to the narrative voice was accomplished by recording Martin Sheen in a small, dead room, which added no acoustics of its own to the recording, close mic'd on a very bright microphone, with all of the technique aimed at achieving intimacy. The result is a very "in your face" style of recording that sounds quite different from his on-screen appearances, which have the reverberation of the set and a less intimate ("looser") microphone technique.
- Making a voice-over narration much more reverberant than the on-screen action is a method for indicating that we are hearing the inside thoughts of a character. Dating back to radio plays, this method seems rather crude today but, nevertheless, when we see a cut to a contemplative character and hear his voice reverberated without his lips moving, we know what to think—these are the character's inner thoughts. A primary user of this effect is soap operas.
- There is a process used deliberately to add the room-like character of a venue to a recording. Let us say that we want to be present at a high-school dance, circa 1960. We obtain records from that era and dub them into the film, but the sound is too direct; it lacks any sort of room sound except that of the original recording. What is done in such a case is to *worldize* the sound by rerecording it over a deliberately less than great sound system, which thus has all the "flaws" of typical reproduction in the new recording. An addition to this effect is to move the loudspeaker and microphone continuously while recording, making the sound "swirl" by the constantly changing the acoustical path. This was done for the music in the gym scenes in *American Graffiti* by sound designer Walter Murch and filmmaker George Lucas, who picked up the loudspeaker and microphone and moved them while making the new recording*.

---

*See http://filmsound.org/terminology/worldizing.htm

- In a complex scene representing several layers, the various layers are likely to employ different methods of recording and rerecording to make them more or less reverberant. Starting from "in front" of the screen, dryly recorded narration stands out. The next layer back is often the foreground production sound, complete with the reverberation present on the set. Within that context would be source music, for example, music playing on a radio in a scene. Further back may be off-screen effects, and the deepest part of the depth dimension is often scored music, especially if the score is orchestral.

So even relatively simple monaural production includes a strong potential for the use of "spatial" hearing because the depth dimension can be used. In more elaborate production, stereophony is used. *Stereophony* is the use of two, or preferably more, sound channels from production through distribution to the end-user environment, delivered by more than one loudspeaker, spaced apart. Stereo offers two vital perceptual features that make it important to film and television sound: the ability to localize sound in various directions and the ability to create enveloping, spacious sound having no particular direction but reproducing recorded reverberation more correctly spatially than any monaural system can do.

> These two factors, localization, on the one hand, and envelopment, on the other, are the two limits of a continuum. Sounds may be pinpoint localized, or a little vague or spacious, depending on your point of view, or they may be completely directionless, like the diffuse sound field we expect from high-quality reverberation. A stereo sound system will reproduce these two effects within limits imposed by the number of channels.
>
> Stereo film sound systems routinely employ five or more channels, whereas home stereo systems before the introduction of home theater used only two channels; the increase in the number of channels is so that fewer compromises are made in the dimensions of localization and envelopment.

The Cocktail Party Effect (Binaural Discrimination)

Standing at a party in a reverberant space, with background music and many conversations going on, we are able to understand the one conversation in which we are participating with a friend. If we replace ourselves with a microphone, make a recording, and listen to it, we find that the recording is usually completely unintelligible. The fact that we can understand the conversation only when present is apparently caused by a number of factors:

- Spatial hearing allows us to concentrate on sound coming from one direction.
- Visual cues—we lip read to some extent.
- Both participants in the conversation are likely to share a huge amount of background, restricting the range of possible topics, messages, etc.

- When all else fails, we fake it, smiling knowingly and filling in gaps in the conversation from our shared background and experience.

Sound recording is at an enormous disadvantage to actually being there when the cocktail party effect is considered, because it is much less likely that the factors that make the effect work in person can be made to function in a recorded medium. The technical name for this effect is *binaural discrimination,* in other words, the ability to discriminate sounds better through the use of two ears rather than through recording; but every psychoacoustician knows this by its slang name.

The cocktail party effect has a strong impact on the method of recording on sets. In a bar scene, with a master shot, several close-ups, cutaways to the crowd, etc., the most rigorous way to proceed is as follows:

- Record the master shot by having only the principal performers speak their lines. Everyone else mimics conversations.
- For close-ups, record the actors normally, in silence. If some other people show up in the background, they can mimic conversation, as in the master.
- For cutaways, record the extra (nonprincipal performer) saying his or her lines so that there is sync sound to cover obvious lip movement (or we would be left with *lip flap,* a defect in which we see but do not hear a person speaking), but watch out, for once a performer has a "speaking role" his or her pay goes up.
- Record "room tone" to serve as a continuous background presence under the scene, easing the way across cuts as described earlier.
- Record a "babble" or "walla" track, either in production or in postproduction, with the right number of people, genders, and level of activity.
- Direct the speaking actors to keep their energy level correct for the finished scene, not what they are encountering in shooting. That is, if the scene is a noisy bar, they need to speak up over the background. Keeping the energy level correct from take to take, and not slipping, marks good direction and acting.

Using all of these methods, a sound editor can prepare a scene, cutting multiple soundtracks to represent principal dialog, background dialog, walla, and presence, and build a complete structure in which the sensation of being there is invoked.

AUDITORY PATTERN AND OBJECT PERCEPTION

Up to this point we have discussed traditional psychoacoustics. Although the principles in this field are useful to film sound, with concepts such as loudness compensation, some decades ago it was realized that further work

along this path, although interesting, was not coming any closer to answering some fundamental questions. The main such question was, How do listeners separate auditory objects from one another and from the background? An *auditory object* is a sound that can be distinguished from other sounds: It can be thought of as a sound "molecule," composed of component parts, like atoms, but perceptually indivisible and distinct. For example, we can distinguish two actors speaking; each is an auditory object to us. In rerecording, we often combine many sounds, such as dialog, music, and sound effects, into one channel. The waveforms of all of the sources are added together and essentially cannot be taken apart by technical means once combined. It is the human perceptual capability that resolves the various elements into separate auditory objects, at the very end of the chain.

This is a little more difficult to think about than the same problem in vision, because anyone looking at a scene can tell the difference between a table and a chair. People do not have the same facility with the variety of sound objects presented in films without training, and this is a core idea. It explains why sound is so valuable to filmmakers. Its relative subtlety compared with picture elements allows filmmakers to manipulate an audience's emotions in a way that is not obvious. The flip side of this effect is that it produces a frustration in those who specialize in sound, because if their work is good, it will never be "understood" by a wider public than specialists. (About the only time one ever hears a comment from the general public is when the dialog is hard to understand!)

So here we examine those aspects of perceptual sound that have received less treatment by classical psychoacoustics and that may help illuminate the processes by which listeners are able to separate the various sounds presented to them into a sensible internal representation of the world.

Information Used to Separate Auditory Objects

Timbre

We have described frequency and its primary correlate, pitch, as well as amplitude and its correlate, loudness, extensively. What is left after differences in loudness, pitch, and duration are made equal is called *timbre*. Timbre is multidimensional; unlike loudness and pitch, which can be placed as a point along a line, timbre is more complex. It depends on:

- Spectrum, that is, the relative amplitudes of the fundamental and its harmonics. For example, turning up the treble control on a stereo makes the sound brighter by increasing the level of higher frequencies in the sound. This is a change in timbre. Track 16 of the DVD illustrates various timbres from different instruments.

- Onset transients. A piano played backward in time does not sound like a piano anymore, but more like a kind of reedy organ, because the attack transient of hitting the strings comes at the end of a note instead of the beginning. Track 17 of the DVD illustrates the importance of onset transients.

The way that the spectrum changes over the duration of the event is important to timbre. A starting transient may have a wide frequency range, but this range may narrow and become more tonal as time goes by. The spectrum changes during the course of the sound event, and we come to associate these changes in time with particular sounds, particularly musical instruments.

Reproducing timbre well is a primary goal of high-fidelity reproduction systems, although reproducing timbre may in practical situations yield to even more important goals, such as dialog intelligibility. There are many films in which the dialog sounds honky or nasal because the midfrequency range has been overemphasized[7] to promote intelligibility, given the competition from music and sound effects. Listening to these films sometimes sounds as though we are hearing two different movies simultaneously, because the dialog seems to have a narrow frequency range and an emphasis within the narrow range of midrange frequencies, and yet the music and sound effects sound wide range and well balanced. The compromise that has been reached in these cases comes from the observation that understanding speech is little interfered with by such tactics, and in fact this emphasis may promote intelligibility when the film is heard under suboptimal conditions.

Sound editors and composers know intuitively about this effect and avoid putting content in the frequency range of the primary components of speech in important dialog regions. Although speech occupies a moderately wide region, the range of the greatest intelligibility is from about 500 Hz to 3 kHz, so this is the region that sound editors and composers keep a little clear in a dialog scene.

Reaching the goal of accurate timbral reproduction generally requires that all the equipment in the chain from the source to the listener covers the full audible frequency range, without emphasizing one frequency band over another. All notes of a piano should be reproduced with the same strength with which they were played during the original performance, so that the reproduced sound accurately represents the instrument.

Fundamental Frequency

The fundamental frequency of various sounds is used to separate them. Although we do not generally listen to a

---

[7]Either through the choice of microphone and technique of its use or by postproduction manipulation, called *equalization*.

fundamental and each of its harmonics individually (although we can be trained to do so), a better representation of what we typically hear is the fundamental plus the timbre (in music, the pitch of the note being played and the instrument playing it).

If we wish to increase the richness of a sound effect, one way would be to layer two sounds together. To get them to merge into one auditory object, one trick can be to match the fundamental frequency of one of the tracks to the other. This can be accomplished by simply pitch shifting with a plug in to a Digital Audio Workstation, or by playing an analog tape deliberately off speed until the pitches match.

Correlated Changes in Amplitude or Frequency

If the component parts of the above-mentioned example change together, such as all the harmonics fading out at the same rate, this will also promote the formation of a single auditory object composed of all the parts that are changing together.

Location

The localization mechanism helps greatly to separate auditory objects. Spatial separation for sound is akin to spatial separation for vision: If we hear a set of sounds from a single location, we are likely to combine them into an auditory object. For instance, all the various sounds constituting a given dinosaur in *Jurassic Park* are placed together through a process of "panning" the sound effect elements all to the same place. An auditory object can be formed using location, even in a monaural soundtrack, by matching the apparent depth of the various elements. This is accomplished by adding more or less reverberation to each of the components so that they match one another in depth.

Contrast with Previous Sound

A negative example of this principle is what occurs when background sound is mismatched on picture cuts within a scene: We perceive a change in the "space" that we do not wish to perceive. The mismatched cut adds contrast where there should be none, and a separate auditory object (new background sound) is formed, drawing the audience's attention away from the content.

A second negative example comes from documentary production. In such shooting, the filmmaker usually cannot exercise as much control as in narrative production, and some sounds are recorded that are inevitably annoying. Adding filters for these background sounds also affects the foreground, desired sound. It would be possible to "switch in" such filters, by editing or in mixing, only during the occurrence of the noise, but the effect this has on the foreground sound has to be considered. It is better

to leave the filter process in for the entire scene than to switch it in just as needed. This is because of the contrast that would be caused from moment to moment in the foreground sound, something to be avoided because we wish to obtain good continuity, implying a lack of contrast.

A unique example of the use of contrast by film and television program makers is using *prelap* edits. If the sound changes ambience to that of a new scene before the picture changes, we draw the attention of the viewer/listener in an interesting way to the edit. Although this technique is subject to overuse, it may also be an extremely good one for certain cases. It is well used extensively in *The English Patient,* for instance. Picture and sound editor Walter Murch gives credit to director Anthony Minghella for suggesting this method, which involves complex storytelling with interwoven flashbacks associated with particular sounds.

Time-Varying Pattern

The rate at which the elements of a soundtrack change gives us a perception of the size of the sound object: A large object probably moves more slowly than a small one. The distant thumps of the footsteps of the dinosaur in *Jurassic Park* are well separated from one another, illustrating this principle. The footsteps are also, of course, bassy, following the principle of bass representing threats aesthetically.

In sound effects as well as music, rhythm causes the formation of an auditory object, because a rhythm forms an expectation in time that something is going to happen in the future at just a certain moment—beats, as in musical beats, will occur. Another rhythmic factor is that musical rhythms have a preferred higher order pattern of beats, the notion of the downbeat. In a waltz rhythm, for instance, we come to expect not only the beat, but the downbeat as well, emphasizing every third beat.

Gestalt Principles

Starting in the 1920s, a group of psychologists in Germany examining classical psychoacoustics came to the conclusion that there were other, more holistic methods to apply to how humans perceive sound. One of their ideas was the necessity of separating auditory objects into *figure* and *ground,* terms borrowed from painting.

Similarity

Sounds are grouped together into one perceptual "stream," in a process called *auditory streaming,* if they are similar in pitch, loudness, timbre, and location. An application of this principle in filmmaking is combining together various sounds to make one composite that we think of as the "sound" of a particular auditory object. The dinosaurs in

Jurassic Park were formed from multiple recordings, mostly of existing animals, modified and summed together. The summing process had to consider the similarity of the elements being summed, so as to produce one sound from the whole.

Good Continuation

Smooth changes, with all the constituent parts correlated with each other, probably arise from one sound object, whereas abrupt changes usually indicate that the source must have changed. The example of background noise from shot to shot is valuable here as well: If two shots within a scene have mismatched background noises, it is more acceptable to cross-fade the sound between the shots around the picture cut than it is to cut abruptly, in order to promote good continuation.

Common Fate

If two components of sound undergo the same changes in time, they will be grouped together and heard as one. This is essentially the same factor as *correlated changes in amplitude or frequency* mentioned earlier, but was given the interesting name *common fate* by the Gestalt psychologists.

Belongingness

A single element can form a part of only one stream at a time. Hearing will try to impose such an order upon hearing multiple sounds.

Closure

A sound that is intermittently masked is perceived as continuous, provided there is no direct evidence to the contrary. An example of the use of this principle is a scene with music underneath that is the wrong length and needs to be edited to fit. Music imposes a lot of restrictions on where edits can be made, because to make sense, the melody, rhythm, tempo, orchestration, etc., must match. This may make it impossible to perform a cut, but if there is a loud sound effect in the scene, an edit can be made in the music "underneath" the loud effect, which is concealed by masking. If the discontinuity in the music is not too great, and the sound effect is loud enough and long enough to cover the edit, closure will work and the edit will appear seamless.

Attention to Streams

Usually listeners pay attention to one auditory stream at a time, with that stream or object standing out from the others for that listener. This does not mean, however, that there needs to be only one sound at a time. Interestingly, different listeners will latch onto different streams, at least some of the time, and in listening multiple times listeners may find themselves attending to different streams upon various hearings of the program. Walter Murch, sound designer for *Apocalypse Now,* explains this idea in different terms. For him, hearing can attend to only one of three sounds presented at a time, so he would argue for up to a maximum of three foreground audio streams occurring simultaneously. He says "there's a reason why circuses use three rings; more is confusing." If there are more principal sounds at one time, the listener is unlikely to be able to separate them out into individual streams.

Forming streams places constraints on attention. Probably the most relevant example from film and television sound is the necessity that the audience feels to hear the dialog. Dialog may well suffer from speech intelligibility problems because of the competition from sound effects and music, and the filmmaker may even want it that way, but, nonetheless, the people in the audience demand to hear the words, and they spend all of their concentration on forming the dialog stream. When this process is frustrated by masking from sound effects and music, they become highly annoyed, and poor-performing sound systems and room acoustics exacerbate the problem.

Auditory stream formation is strongly impacted by visual information. Probably the simplest definition of sound in a Hollywood movie is "See a car, hear a car." That is, everything that we see on the screen that we expect to make a sound does make a sound. What we don't know upon first hearing the film is that it is a conscious decision on the part of the sound editors to record, edit, and mix sound effects to cover all of the objects on the screen.

A related topic is that auditory streams sometimes need visual explanation. For instance, in a documentary film, if there is an aquarium out of the picture during an interview, but bubbling, the audience will spend its time wondering what the heck that sound is rather than paying attention to the interview. A cutaway to the aquarium, or its inclusion in a master shot, provides an explanation to the audience, who can then concentrate on the content.

Multisense Effects

Sound for film and television is by definition presented along with a picture, and sound and picture have interactions. As we have seen earlier, picture has an impact on sound localization, and sound has impacts on picture as well. Most people who see a complex scene with the soundtracks broken down are surprised by how much faster it seems to run when all of the sound is present than when it is heard a layer at a time. Items like the title crawl in *Star Wars* seem to take a long, languid time rolling up the screen when shown silently, but seem to march along when heard with the theme music.

The size of presentation of the picture also has an impact on the apparent speed of action. Larger picture displays, probably because of their greater stimulation of

more faculties of the brain, seem to cause action to move more quickly than smaller editorial displays. For this reason, it is a bad idea to edit film silently on a small editing system because when sound is added and the display size is large, the pace will seem faster, almost frenetic, compared with the action in an editing room.

Recognition Gestalt

In speech perception at least, there can be a "eureka" moment, a recognition that something is understood that up until that time has been unclear. For me, the most salient recognition Gestalt I ever had was listening to the title song from the television show *Friends* a great many times and, despite trying to, being unable to understand one line, which turned out to be "Your love life's D.O.A." The use of the letters to stand in for the term "dead on arrival" was simply too much for me to understand in the context of the song, with the masking provided by the music. Once I got it, then every time I hear that song, I hear it clearly.

SPEECH PERCEPTION

Speech seems to be a special case for perception. There are several pieces of evidence for this, and the idea has a strong impact on the way that film sound is practiced. In 1967 Philip Lieberman showed that the basic blocks of speech, *phonemes,* could be understood even when sped up to a rate of 30 per second, whereas the time boundary for other sounds to cross into confusing order was around 10 per second. This finding was made use of in devices that sped up playback of recordings that, when combined with a corresponding correcting pitch shift, allow listeners to perceive the spoken word at greatly sped up rates. At 10 notes per second, musicians could not tell the order of notes, but subjects could tell what the content of speech was at the three-times-higher rate. Brain scans taken while subjects listen to speech versus music show different brain activity centers stimulated by speech compared to other sounds.

The overall pattern of sound in speech is what is important: frequency, amplitude, timbre, and how they change in time—a multidimensional effect. This is known by adding deliberate distortions to any of these factors and finding that speech is still intelligible despite very dramatic distortions. For example, for World War II bomber crews, the earphone sound was restricted to a very narrow high-frequency range because the aircraft noise was so loud, with lows predominating. Despite this reduction in bandwidth (frequency range), speech was still intelligible. The multidimensional nature of speech is equivalent to high redundancy, making it difficult to corrupt to the point of complete unintelligibility (although high quality is a different matter).

Speech for Film and Television

A special issue facing speech in the movies is the interaction between two perceptual factors. If a localization error of 4° is noticeable to professionals and 15° is annoying to laypeople, what about placing speech sounds to be coincident with the picture of the actor speaking?

When stereo sound was introduced into films in the 1950s, dialog was routinely "panned" from left to right on the screen to match the action. This process has dropped out of use, except in certain circumstances. One possible explanation for the demise of stereo dialog could be the fact that producers won't pay for the time necessary to get it right in postproduction nowadays, and there might be some truth to that, but it isn't the right answer. In fact, there is a problem of perceptual mechanisms competing to determine what is going to cause the greater disturbance, localization error or something else of importance. Let us say that a scene, such as the interior of General Allenby's office in *Lawrence of Arabia,* is shot with a wide master shot encompassing a large office and that close-ups are shot of each of the characters. When the scene begins, picture editors almost always start with the master shot, because the very width of the shot sets the stage for the action. It is really an expositional shot, showing us the scenery, in effect, although of course it may also contain content. Then, as the scene progresses, more and more use is made of close-ups, helping to build the tension and intimacy of the scene. This is a standard progression we know from seeing it many times, and we are never confused about where we are and what we are looking at. Now consider the sound. In the master, sound positioned to match the performers is all right, coming from the left or right as the characters face off across the desk. At the cut to the first close-up, with the actor centered in the frame, there is a jump in the sound: It must cut from the side where the character was to the center. This cut seems unlike the picture edit; it is not smooth but rather jarring. The principle of good continuation has been violated by the jump[8] cut. The reason for this jarring quality may well lie in the fact that so much training has gone into the way we look at pictures and their cuts that we "know" the film grammar of picture cutting so well that it seems completely natural, having grown up seeing so much film and television. This kind of sound edit, on the other hand, is something with which we have practically no experience. The question might be resolved by showing edited films to primitive tribespeople who have had no exposure to film or television and seeing what they make of it, but we have to leave that to some visual anthropologist who is not afraid of disturbing his or her own object of research.

---

[8]This is a new use for the term that is applied to picture edits that violate continuity rules, such as "crossing the line," or that are otherwise abrupt.

Certainly, a factor in how jarring this kind of edit seems is how well matched the sound of the two loudspeaker systems is. If the side and center loudspeakers don't match perfectly for timbre, then we are distracted by the timbral change from position to position that we would not notice if the character stayed in one place. This problem was addressed industry-wide in the 1970s, and although not quite a thing of the past, it has been greatly reduced, leaving the first reason as the most prominent.

So stereo dialog positioned line by line into the correct position has dropped out of use due to these competing perceptual factors. The standard is now to place most dialog in the center channel, despite its not matching the picture in each shot.

Panned dialog does have its uses, however:

- Off-screen lines are routinely panned to far left or right as makes the most sense to distinguish them from on-screen dialog.
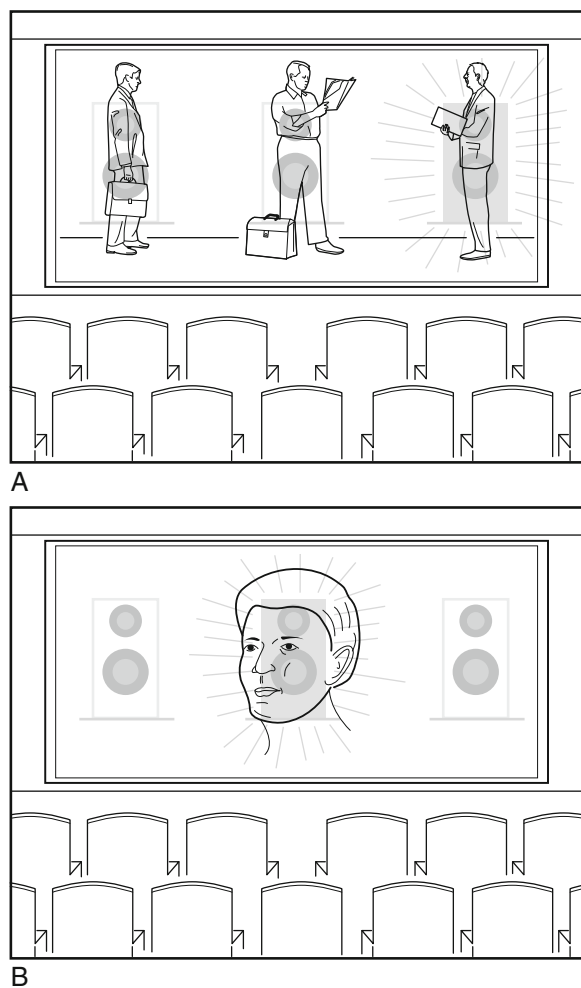
- Dialog lines that are well separated in time from others and for which the character speaking is well off-center may be panned.
- Slight panning off-center for two characters speaking to one another is occasionally used, because it is less jarring in cuts to jump by a smaller amount than the large amount from the extremes.

Influence of Sight on Speech Intelligibility

It is remarkable the extent to which speech can be understood despite interfering noise. Human listeners use a variety of tactics to hear speech, even buried underneath noise, which in film sound we call music and effects. As the subject of an experiment at the House Ear Institute in Los Angeles, I was able to understand speech some 16 dB underneath noise when I was shown a video of the talker, and this was about 2 dB better than when I could not see the talker. This was despite the fact that I was a very poor lip reader.

The Edge of Intelligibility

Professional facilities have controlled room acoustics and fine sound systems. However, when films go out into the world, and television programs even more so, the conditions of playback are not so pristine and can vary a lot. The movie *Backdraft* is a case in point. This film about firemen pushed the limits of the envelope on intelligibility during fires, and it was understandable in the Stag Theater at Skywalker Ranch where it was auditioned, but when turned down and played over worse sound systems in worse auditorium acoustics, the intelligibility suffered so that the audience missed things.

Also, the needs of directors and producers to hear a "new" experience in postproduction may lead them to deemphasize dialog intelligibility, as they have lived so long with the project that one can shut off the dialog and still hear and understand the speech. One of the reasons that the most competent directors and producers hire the best mixers time after time is that they provide a new set of ears that will make certain that an audience can understand the movie the first time out. Experience gives good mixers the ability to know how far they can push the envelope and still be intelligible in less than pristine conditions.

FIGURE 2.5 The master shot (a) calls for dialog to be panned right, whereas the close-up (b) calls for it to be panned center. The cut between the two shots thus causes the sound to jump, an aural discontinuity, which leads to most dialog being panned to the center in most films.

CONCLUSION

We have observed in this chapter issues in hearing and perception that affect filmmakers daily. In sound production, many of these factors are given consideration, either

explicitly or, more often, through the training and experience of practitioners. Sound editors may not call it temporal masking, but they have come to know where the best places to make edits are, such as just before a "plosive"[9] in speech or the downbeat in music.

Rerecording mixers may not have the equal-loudness contours in front of them, but they know that music, faded under, becomes thin sounding, worthy of correction. The lesson of perception is that film and television editors and mixers tailor soundtracks to fit the human psyche, just as much as they fit it into any technical requirement.

---

[9]Speech phonemes having a hard edge, such as p's and t's.

Audio Fundamentals

AUDIO DEFINED

Audio is the representation of sound, electrically or by various methods on media, not sound itself. We usually say *audio tape,* not *sound tape,* for instance.[1] The advertising tag line that was tried some time ago, "It's audio that surrounds you," is also wrong, because you would be wrapped up in electrical signals or in tape, not in sound, by this image. *Sound* is the term used for acoustical energy, whereas *audio* applies to electrical signals and magnetic and optical recordings.

Sound is the input for audio processes by way of microphones, which are *transducers,* turning sound energy into electrical energy, at that point called generically a *signal. Recording* is the process of converting the electrical signal into a form that is stored on a medium, from which it may be played back and converted to an electrical signal. *Mixing* involves a variety of processes to manipulate audio and, in this way, winds up indirectly manipulating sound. Finally, for conversion from electrical signals back to sound, loudspeaker transducers are used.

We have already seen one primary consequence of the overall process of soundtrack preparation—mixing together sounds from various audio sources. For all intents and purposes, this means that they cannot be separated again by technical means, but rather the parts are separated by the final listener into a variety of auditory objects or streams, using perceptual processes. This causes us a lot of trouble and makes for a specific method of working in both production and postproduction mixing, which we discuss later.

TRACKS AND CHANNELS

Technically, the word *track* refers to the space on the medium for the audio representation of sound. Thus, we say that we use 24-track tape recorders because there are 24 parallel stripes of area on a 2-inch-wide piece of tape on which we can record separate signals. Physically speaking, the soundtrack on a piece of film is the area devoted to sound and the recording made on that area. In nontechnical usage, the term applies more broadly to everything recorded and its overall effect, as in "That was a dramatic track."

The word *channel,* on the other hand, does not apply to the representation on the medium but is a more abstract term describing a signal path. It may describe a pathway for signals inside a piece of equipment, as in, "Assign the first input channel of the console to the output channel," or in a broader sense, as in "The transfer channel is not working at this time" (meaning that transfers from one medium to another cannot be made now). This second usage is really a Hollywood one, little heard outside postproduction, it must be said.

SIGNALS: ANALOG AND DIGITAL

There are two primary ways to represent sound as audio, by analog and by digital means. Digital techniques dominate most areas of sound for picture today, although some parts of the chain remain principally analog. Those are the microphones and their connection to equipment, some consoles, and most power amplifiers and loudspeakers. So the two very ends of the chain are still dominated by analog, and analog appears in some of the "middle." Of the areas that remain analog, digital microphones have distinct advantages described in Chapter 6, and it seems as though the handwriting is on the wall for analog consoles. The longest holdout will probably be power amplifiers, as the advantage of digital amplification brings with it attendant complication that can lower reliability, so there is no persuasive case for it yet in cinemas.
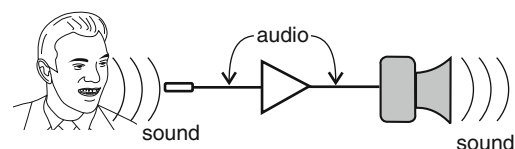

FIGURE 3.1 Audio is distinguished from sound by its being the electrical representation of sound, a physical phenomenon.

---

[1]Although of course we do say "soundtrack." Are we talking about the physical track on the film representing sound or what we hear? This is usually ambiguous because a mixer will say, "That's a great soundtrack," and an engineer will say, "The density of the soundtrack shows that it is underdeveloped." They are both right, thus it is the term itself that is ambiguous.

Both analog and digital methods may be practiced well or poorly and may result in good or bad sound at the end of the day, although there are large differences in what the potential problems are. Probably the popular notion is that digital audio is good and analog is passé, because of the large success of the compact disc as a music medium and its obvious improvements over the phonograph record. When it comes to film and television, however, things are more complicated. It is the great complexity of post-production for large-scale films for which digital techniques are for all practical purposes today required, as their productivity is so much higher.

Years ago I had a summer intern count all the splices in the analog magnetic film units for *The Ewok Adventure,* a made-for-television movie with some, but not a great deal of, complexity. I then looked up the time cards for each of the sound editors and calculated their productivity in terms of the number of sounds found, cut, and documented per working hour. The number hovered around 4 for the various disciplines, dialog, music, and effects editors. So a single sound in the analog days took about 15 min to deliver to the mixing stage. If you were to tell this story to any sound editor working on a digital audio workstation today (or their bosses), they would be amazed that anyone ever worked that slowly. With centralized library sound effects and production sound servers delivering audio to workstations over a network, etc., today the productivity increase associated with digital techniques is so great that no one who has the equipment available will do things the old-fashioned way.

You can see this by comparing the credits of two action films made 12 years apart. Take *Con Air,* a 1997 action film cut on mag film and budgeted at $75 million. Thirty-seven sound editors or designers worked on the picture. Compare it to a similarly budgeted 2009 movie (counting inflation to $100 million) cut on digital, *Public Enemies.* Sixteen sound editors/designers are listed on www.imdb.com, a 56 percent decline.[2]

One of the strongest suits for digital technology, beyond productivity, is replication. The nature of digital recording and distribution is such that they are *potentially* far more impervious to outside influences than are analog recording and distribution chains. The word "potentially" is emphasized because an underlying medium impervious to outside influences is also needed to realize the most important benefit of digital—its permanence. This is what makes digital superior at distribution technologies, such as digital soundtracks for films and the compact disc, compared to competing analog technologies.
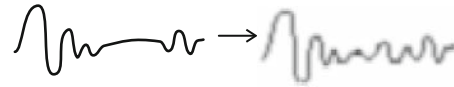

FIGURE 3.2 The difference between digital copying and analog copying is the difference between copying numbers and accurately representing a waveform in a medium. This gives the advantage to digital copying, for if the copy can be read at all, it can be restored to its original values.

The fundamental difference between analog and digital representations of sound occurs in the amplitude domain. In analog, the amplitude of the audio waveform is represented as an *analogy* to the original waveform, whether electrically in wires or recorded on a medium. That is, there is strict proportionality between the original sound waveform, its electrical representation, and its amplitude[3] in the medium of choice. These analogies can take many forms, such as the displacement of the diaphragm of a microphone, the consequent strength of a magnetic field on audiotape, the electrical voltage in a console, or the motion of a loudspeaker cone turning audio back into sound, all looked at moment by moment.

Note that for both analog and digital methods there are two dimensions that describe a signal. The instantaneous amplitude is one dimension, and time is the other. Amplitude varying over time describes a waveform.
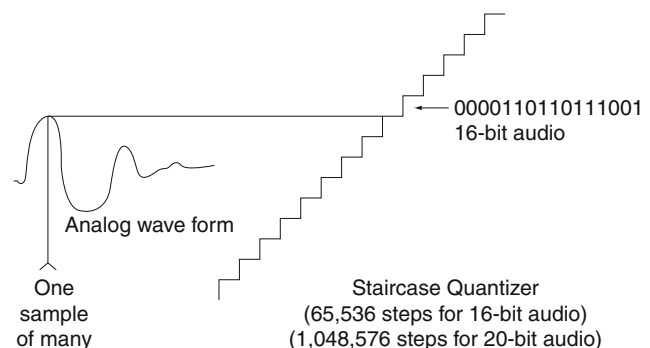

FIGURE 3.3 The heart of digital audio is quantizing, comparing the amplitude of each sample in turn to the steps of a staircase quantizer, and assigning the correct number for that step.

---

[2]Although citing only two pictures does not capture the entire scene, nonetheless the conclusion does seem to be true. This could be fertile ground for further research.

[3]In the strictest sense, the amplitude of the audio waveform may be represented by analog means other than by amplitude in a medium. Schemes such as FM (frequency modulation), used in radio microphones, convert amplitude variations to frequency variations for transmission and then back to amplitude variations for the output and are still considered analog methods.

The fundamental difference between analog and digital signals is that the amplitude domain in digital is *quantized*. Imagine a set of bins stacked on a staircase. Quantization is the process of measuring the practically instantaneous amplitude of a waveform and "binning" it, assigning a particular numerical value by using the number of the bin closest to it in height. The number is what is stored instead of a signal strictly proportional to the waveform itself. The number is no longer strictly analogous to the waveform, because analog-to-digital conversion has been performed.[4]

The beauty of storing the amplitude as numbers is the "ruggedness" with which numbers can be stored; that is, despite many corrupting influences, the numbers can still be read. By adding powerful *error codes* to the quantized numbers, especially the Reed–Solomon code,[5] the numbers are made highly robust. Because most errors in their transmission can be corrected there is no change in the sound under these conditions. When the errors become so numerous or dense that they cannot be corrected, that fact is known within the reproduction equipment. Then a variety of tactics can be brought into play, such as guessing what the original waveform was, ultimately giving way, in the worst case, to no sound. Here is the Achilles' heel of digital audio: It can easily lull one into thinking that all is well, when what is actually happening is that we are operating closer and closer to the edge of not working at all, and there is no way to tell that from the sound quality. Manufacturers do not typically provide any means to tell you how close your reproduction really is to disaster, despite knowing it within the equipment, for if you did, you might actually think something was wrong! Luckily, these problems have been lessened through time and the use of more rugged media. It must be said that DAT (digital audiotape) is not much lamented, for instance, as it could fail with no prior warning.

> The first strategy that digital equipment uses on finding an error is *error correction.* In error correction, the error decoder completely restores the original numbers, and there is no change in the sound whatsoever. On the other hand, if there are many occurrences of error correction it means that the medium is potentially damaged, which is likely to lead to more serious problems later.
>
> The next strategy used, after the error correction mechanism is overwhelmed, is *error concealment.* Here the playback decoding circuitry makes an educated guess about what the numbers were, based on what came before and what comes after the missing data. The sound cannot be said to be identical to that produced by the original, but may be good enough for all but

mastering purposes. If error concealment occurs on a digital audio master, the master is usually remade. Finally, when both error correction and error concealment can no longer be used because the data are so corrupt, most equipment "mutes," that is, switches to silence.

Analog systems, on the other hand, tend to have a gentler curve of failure—that is, they often sound increasingly bad before outright failure, giving some time to take corrective action before no sound is heard. Another advantage of analog recording is the "soft overload" characteristics of analog tape and film media. Film saturation is a gradual process, whereas digital overload is a "hard" one, with a rapid onset of really very distorted sound. So one can more often use an overrecorded analog tape more easily than a digital one.

The analog method of representing sound as analogies to the amplitude of the original waveform has the problem that the analogy is only a representation of the original, not the waveform itself. The difficulty that this causes is that with multiple-generation copying, something is inevitably lost during each generation, ultimately resulting in audible quality problems. Distortion and noise, to be discussed later in this chapter, increase from generation to generation, sometimes by tolerable amounts, but they nonetheless do inevitably increase. The popular term for this is *generation loss*.

> I spent a great deal of time on *Return of the Jedi* (1983) trying to sort out why the 70 mm prints sounded more distorted than the master in brief passages. Naturally the first place we worked was on the printing process itself, because that seemed to be indicated. In the end, it was found that this stage was not at fault because it was essentially equal in quality to all of the other generations, but rather it was simply that we had accumulated distortion from generation to generation to the point of audibility. (This affected only a few moments of the movie.) On the other hand, by improving the foregoing generations before the printing process, more difficult scenes had inaudible distortion by the next year in *Indiana Jones and the Temple of Doom* (1984).

Digital copying and transmission do not rely on making an analogy each generation, but rather the correct copying of numbers, a far easier task because even if the numbers are "blurry," they can still be read. So all other things being equal, and with certain assumptions, a 10th generation digital copy is indistinguishable from the 1st generation, whereas a 10th generation analog copy will certainly show audible defects.

PARADIGMS: LINEAR VERSUS NONLINEAR

In film and television production the terms *linear* and *nonlinear* are used to describe the means of access to the portions of a program to be edited or mixed. Linear in this usage means that the material is recorded along the length of a medium, which could be an analog or a digital tape.

---

[4]Although we could say that the numbers are still proportional to the waveform, the proportionality is no longer strict, because in the act of binning the amplitude, there is a range of possible waveform amplitudes that still fit within one bin before the quantizing device "snaps" to the next bin.

[5]This was a key technology that made the compact disc, among other digital audio media, possible and was developed by Professor Irving Reed of the University of Southern California.

Examples include 24-track analog and digital tape machines. To get from one part of a program to another, winding the tape or other medium is required over the intervening portions of the tape, and this can take a considerable amount of time. Nonlinear means that access to the material is available by jumping over all of the intervening material. An example is a phonograph record, where you can lift the tone arm and jump from cut 1 to cut 10 with reasonable ease. More importantly, computers store their files in a way that permits nonlinear access. Digital audio workstations generally operate in a nonlinear way, able to jump from one part of the program to another, saving time.

> Some picture and sound editors, notably Walter Murch and Randy Thom, have pointed out that there are drawbacks to the enormous speed advantage of nonlinear editing, because viewing the material at high speed, such as on a flat bed editing table, can yield ideas for editors; but nonetheless the sheer speed gains of nonlinear systems make them the method of choice, with at least one notable exception, Michael Kahn, Steven Spielberg's long-time editor. Pictures that Kahn cuts bear the end title "Edited on the Moviola."

In digital editing systems, linear storage devices are rapidly giving way to nonlinear ones, because if a decision over the length of a shot is changed on a linear editing system, all the subsequent work in that particular reel will have to be redone. This problem for linear systems, and the rapid access to material of the nonlinear systems, means that for all practical purposes today beyond starter video editing systems, nonlinear ones are dominant. The terms linear and nonlinear mean something different when describing the audio quality of a product or system, which we will take up later.

LEVEL

The amplitude dimension of a waveform may be represented in a variety of ways, such as:

- An electrical voltage, for example, at the output of a microphone;
- The strength of a stored magnetic field on an analog audiotape;
- Numerically, as in digital audio.

Because program[6] waveforms constantly change amplitude with time, the value of each of these representations is also constantly changing. This constant motion makes thinking about the relative level in the various parts of the system difficult, so the idea of level is usually simplified to that which corresponds to that of a simple sine wave. For sound in air, 0 dB SPL was set as a reference at about the threshold of hearing. Practically all acoustical measurements are referenced to this level, which is a pressure of 20 μN/m^2, giving the scale to acoustical measurements shown in Table 1.2.

Microphone Level

To characterize the output of a microphone a reference at 0 dB SPL is inconvenient because it is difficult to obtain spaces quiet enough to make a sound of 0 dB SPL without masking by room noise, and it is nowhere near the SPL typically seen by the microphone. Thus, for most microphone measurements it is commonplace to choose as a reference sound pressure level 94 dB SPL, which is incidentally called 1 pascal.[7] The microphone is rated for *sensitivity,* delivering a specific voltage at 94 dB SPL. Conventional microphones may deliver anywhere between 2 and 60 mV[8] under these conditions, depending on their type. This is a possible range of 30 dB from one microphone type to another, a very large difference.

Although 94 dB SPL is a relatively high sound pressure level, the electrical voltage is still quite small, so microphones are routinely connected to *microphone preamplifiers,* which amplify the output of the microphone to an electrical level that is more useful, with the amount of amplification depending on what is being recorded and the system requirements of the following equipment. The wide range of output levels from various microphones means that preamplifiers must be matched to the microphone type. We say that the low output level of microphones is at *mic level* and that the output of the microphone preamplifier is at *line level.* A typical microphone level taken as a snapshot on speech is 2 mV, and a microphone preamplifier may boost this to a 1.2-V line level.

Today with the introduction of digital microphones, issues about the low levels of microphone output and thus the needed microphone preamplifiers has been overcome with the preamplification and digitization built into the microphones. This is taken up in Chapter 6.

Line Level

Line level signals and analog audio signals are those that are routinely interchanged within a studio environment, used for connecting various pieces of equipment together for signal processing, for instance. The method of routing such line level signals may be by cables connected to the input and output connectors on pieces of equipment or by way of patch bays (which look like the old-fashioned

---

[6]*Program* is derived from broadcasting practice, in which *program* means the desired material to be heard by the listener, as opposed to test tones or leaders that may also be on the same medium.

[7]And which is equal to 1 N/m^2, a measure of pressure: force per unit of area. Thus 0 dB SPL = 20 μN/m^2.

[8]A millivolt (mV) is one one-thousandth of a volt.

TABLE 3.1 Line Levels

Application	Level (dB re: stated reference)[a]	Voltage (rms volts)
Consumer and semi-pro equipment	−10 dBV	316 mV
Pro 1	+4 dBu	1.228 V
Pro 2	+8 dBu	1.946 V

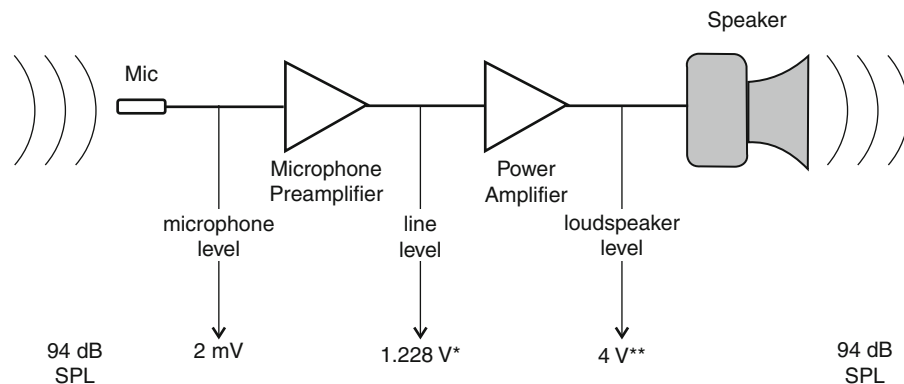[a]*dBV reference is decibels relative to 1 Volt. dBu reference is decibels relative to 0.775 Volts.*

equipment. Typical consequences of patching in equipment without compensation for the reference level include excessive distortion or noise. A proper method for using consumer equipment in a professional studio is to attenuate the signal at the input to the consumer equipment, to reduce the higher studio level down to the consumer equipment level, and to use an amplifier on its output to restore the level. Boxes are available to make both these changes for the input and output of the consumer equipment to install them in pro studios. They are called *match boxes*.

Speaker Level

The third level used in a recording chain is speaker level. It is higher than line level and is provided by power amplifiers capable of delivering up to hundreds of watts to loudspeakers. A typical speaker voltage level is 4 V to produce 85 dB SPL in a theater, a typical reference sound pressure level in large-room mixing.

Level Comparison

There are thus three principal levels used in analog systems—mic, line, and speaker—which correspond to a few millivolts, around 1 V, and a few volts (and with high-power capability), respectively. Although speaker connections are rarely confused with line and mic connections, being of different connector types, that still leaves mic level and line level connections to be confused with each other. Perhaps no problem is so prevalent in audio as mixing up these two levels, with gross consequent distortion or noise. Often, the two levels may be presented even on the exact same connector, such as at the output of a mixer and on the input to a professional

telephone operator switchboards on which they were based) or electronic switches. The electrical voltage of line level varies depending on the studio and whether the equipment is professional or consumer grade (Table 3.1). The most common professional reference line level is +4 dBu, which is 1.23 V, called here Pro 1. Pro 2 is an older reference level still found in a very few broadcast applications. The most common consumer equipment line level is −10 dBV, which is 0.316 V. However, CD, DVD, and Blu-ray players are more likely to have 0.2 V for a reference level, some 4 dB less, so there is confusion even among consumer products. CD mastering makes up for this by recording the program material "hotter" than the reference level, but DVD and Blu-ray, being associated with a picture, normally does not.

Many problems in professional audio relate to interfacing equipment intended for different line levels. Patching consumer equipment into a professional studio, for instance, is difficult because of the large level difference in the reference levels of the various pieces of

* with insignificant power capability
** with significant power capability

FIGURE 3.4   An audio system diagram showing typical voltage levels at various points in the system. Note that although line level and loudspeaker level are similar in voltage, the loudspeaker level comes with far higher power capacity than is available at line level.

video camera, and the nominal level is set by means of switches.

> If a microphone is inadvertently connected to a line input, it will not see the correct power (see Chapter 5), and even if not requiring power from the connected equipment, the level will probably be so low it won't be heard, or if the gain is advanced enough to be heard, the system will be very noisy.
>
> Likewise, if the output of a mixer is connected to a camera input, and the mixer is set to mic level, and the camera input is set to line level, the result will be excessive noise.
>
> Conversely, if the output of a mixer is set to line level and the camera input is set to mic level, gross distortion at least on signal peaks will almost certainly be the result.

A common problem in audio is connecting a microphone to an input jack of a portable recorder, such as a Betacam SP unit, and setting the level, forgetting that there is a switch that sets the input sensitivity of that jack to mic level or line level. Connecting a microphone to a line level input will result in excessive noise if the microphone can be heard at all. Connecting a line level source to a microphone level input will usually result in gross distortion.

A major problem in professional audio today is that the same connectors are used for both analog and digital connections, and for mic and line level analog signals, apparently simply because of the popularity of one type of connector. Analog and digital signals are NOT interchangeable and never should be interconnected. There could be equipment damage as a result, or even hearing loss.

ANALOG INTERCONNECTIONS

Signals are usually routed from microphones to preamplifiers, from consoles to tape machines, and from power amplifiers to loudspeakers over conductive wiring, usually copper. There are two principal ways to do this, by balanced lines, used mostly in professional applications, and by unbalanced lines, used mostly in consumer applications. Because of the differences in balanced versus unbalanced systems, different connectors are used, generally distinguishing the types.

Conventional home high-fidelity system wiring is unbalanced. That is, there is one signal conductor contained inside a shield, and then the whole cable is wrapped in an outer insulator. The outer shield serves as an electrical ground, as well as providing shielding to prevent electrostatically induced hum.

The difference between unbalanced and balanced wiring schemes occurs when there is interference from external *magnetic* sources, such as the magnetic field set up around lighting cables on a set by virtue of their carrying large amounts of current. In an unbalanced connection, the magnetic field induces a voltage in the principal conductor, which the receiving equipment sees as the same as the desired signal; thus, hum may be heard at the end of the chain.
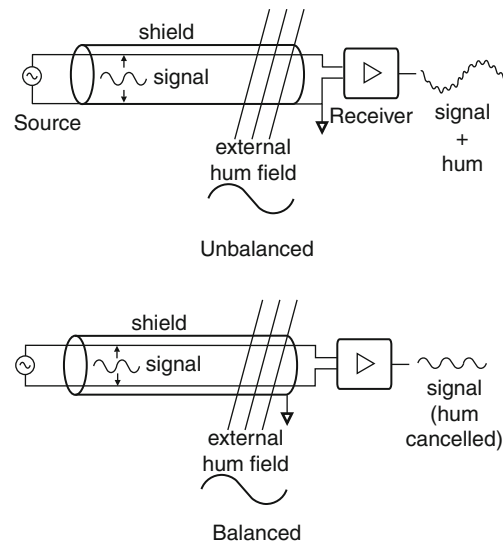
FIGURE 3.5 An unbalanced system is susceptible to hum pickup by magnetic fields being converted into voltages. The balanced system is much less susceptible because a balanced input is sensitive to the differences between the two conductors, both of which see more or less identical magnetic fields.

Balanced wiring provides a means to reject hum caused by stray magnetic fields. It uses two signal conductors contained within a common conductive shield. Thus, there are three conductors altogether. At the instant when one of the two conductors has a positive-going signal voltage on it, the other one will have an equal and opposite negative-going signal voltage on it. External magnetic fields induce a voltage in the conductors, but it will be equal in magnitude and polarity in both conductors. The receiving equipment is deliberately made sensitive only to the *difference* between the two conductors, and because the voltage induced by the magnetic field is the same in both conductors, it will be rejected by the receiving equipment.

> These two modes of signals in balanced wiring are called the *differential mode,* for the desired signal that is in opposite polarity in the two wires, and the *common mode,* for the induced hum that is in phase in the two wires. The measure used to quantify this effect is called the *common mode rejection ratio,* a metric that compares a deliberate differential mode signal to a common mode signal and expresses the difference in decibels. A good common mode rejection ratio is 80 dB, providing a reduction in hum of a factor of 10,000:1.

Balanced wiring is considered essential for microphone wiring, because the signal voltages are low at mic level and the cables are often in hostile environments, leading to audible induced noise. Balanced wiring is also used in professional studios in between pieces of equipment, although the added expense is sometimes unneeded when the signals stay relatively local, such as within one rack of equipment, and unbalanced wiring is sometimes used in these instances.

There is one problem possible with balanced wiring compared with unbalanced: If the two signal leads inadvertently become interchanged, through wrong cable wiring, for example, then the signal will be inverted and yet still work. This condition is called *polarity reversal,* or more commonly, *phase reversal.*

If the same inversion occurs in unbalanced wiring, the signal is shorted out, because the outer shield is grounded at its ends. The inadvertent "polarity reversal" of a miswired balanced line may be inaudible for some purposes, but if two microphones cover a scene and the performer is equidistant from the two, then the subsequent addition of the signals in mixing will cause the two signals to at least partially cancel. Thus the polarity of all cables must be observed in balanced wiring to prevent such cancellation.

Impedance Bridging versus Matching

Today, most audio wiring proceeds in a manner familiar to anyone who has ever plugged in multiple lamps on one electrical circuit: no matter how many are plugged in, the same voltage is delivered to the lamps (they don't get dimmer as more are plugged in), up to the capacity of the circuit breaker. So audio signals can be routed freely, and a single source can feed multiple devices, up to a reasonable limit. Such a system is called *bridging* because each of the devices connected to a source is said to bridge across the output of the source.

> The technical description of this condition involves a concept not yet presented—the idea of impedance. In the case of bridging systems, we say the source impedance is low and the load impedances are high, which is the same condition as exists with electrical generators supplying house wiring. This means that the source will, practically speaking, maintain the same voltage despite the number of loads bridged across it.

The alternative system of matching impedance is a system in which each source is *terminated* in a specific design load impedance. It is principally useful today in very long lines, such as transmitting audio over telephone lines, but has little utility in studio environments. Still, there are some holdover applications in broadcasting in which matching is employed, because it was commonplace in earlier eras when a premium was placed on everything in the studio being able to drive very long lines.

Connectors

Unfortunately, there are a great many connectors used for audio (Table 3.2). It is useful to know the names and area of application of audio connectors because they so frequently must be interconnected that even for the simplest jobs one must often specify an adapter for connecting

TABLE 3.2 Some Connectors Typically Used for Audio[a]

Name(s)	Conductors	Photo	Usage
XLR Canon	3 – 5		Most widely used connector in professional audio, for microphone and line level analog signals and digital AES 3 (two channels on one cable). Pin 2 positive signal, pin 3 negative signal, and pin 1 shield ground.
¼″ mono phone	2		Mono headphones, other monaural uses such as microphones and line level signals. Tip positive signal, sleeve shield ground.
¼″ stereo phone ¼″ balanced phone	3		Stereo headphones with the tip conductor the left channel, the ring the right channel, and the sleeve ground; balanced inputs on some equipment with the tip positive signal, ring negative signal, sleeve ground.
¼″ TRS patch bay	3		TRS=tip, ring, sleeve. For patching balanced lines in patch bays (note the tip diameter is smaller than on a conventional ¼″plug: the two are interchangeable only in some jacks). Tip positive signal, ring negative signal, sleeve ground.
Tiny-T TT	3		A smaller version of a balanced patch bay connector

continued

TABLE 3.2 Some Connectors Typically Used for Audio[a] *(continued)*

Name(s)	Conductors	Photo	Usage
3.5 mm mini mono plug	2		Mono consumer headphones, microphones (may be called ⅛″).[b]
3.5 mm mini stereo plug	3		Stereo consumer headphones, microphones (may be called ⅛″).[b]
2.5 mm micro mono plug	2		Miniature recorder input plug.[b]
2.5 mm micro stereo plug	3		Miniature recorder input plug.[b]
Phono Pin Plug RCA Cinch	2		Common unbalanced hi-fi system interconnects at consumer levels including for both analog and digital signals.
BNC	2		Professional audio test equipment, video, some professional digital audio use especially in video facilities. Digital audio on BNC connector is to standard AES 3id. The BNC connector is also used for word clock to synchronize digital audio devices to a master clock.
Banana	1 per lead		Nagra tape recorder outputs, test equipment.
Tuchel	2–8		Nagra portable recorders.

[a]*Many other types are used for specific circumstances, such as multi-pin connectors for multi-channel use, special connectors for radio microphones, etc.*
[b]*The variation in body style among these four is typical of varaitions found among connectors in the field, and some large diameter connector bodies prevent full insertion into corresponding jacks due to obstructions. Also, variations in the precise dimensions, particularly the diameter of mating surfaces, vary from jack to jack. Thus if a large diameter plug has been plugged into a jack, that jack may subsequently not make good contact with a smaller diameter plug, within the tolerances of parts found in the field.*

two types, and this must be done by naming the types correctly. In professional use, connectors coming from a source, such as the output of a microphone, are usually equipped with male plugs, and connectors accepting signals are usually female, thus giving an immediate indication of the direction of the signal, although there are exceptions to this rule. Male connectors are those equipped with pins for the signal conductors and are called *plugs,* whereas female connectors are equipped with receptacles to accept the pins and are called *jacks,* so the "sex" of a connector is determined by the conductors, not the outer shell.

In consumer use, it is common to use hi-fi-style interconnects, in which both ends of the cables use male connectors and chassis connectors are female. Thus the "direction" of the signal is not indicated by the sex of the connectors. A common problem that arises as a result

is finding jacks labeled "Tape Out" on the back of a receiver. What is meant in this instance is actually "Out to Tape," that is, a signal destined for the *input* of an external tape recorder. One could easily think the converse, that Tape Out should be connected to the output of the tape recorder, but that would be wrong.

QUALITY ISSUES

Audio equipment is assessed by means of measurements and listening tests. Measurements can quickly tell if certain factors are optimum in a given piece of equipment and whether problems found are likely to be above or below audibility. Nevertheless, it is also necessary in the final analysis to conduct proper listening tests to check for factors that may escape conventional measurements. Generally, when something is heard in such listening tests, a measurement can be devised to quantify the effect.

Dynamic Range: Headroom and Noise

One difficulty facing film and television production equipment is the very wide volume range, from the softest sound to the loudest, present on a typical set. The background noise may be at 20 dB SPL, and an actor shouting can easily reach over 120 dB SPL. The consequent more than 100 dB volume range is a challenge to record. There are two limitations to the ability of any item in the audio chain to reproduce the volume range from soft to loud. At the bottom end of the range, noise is the limiting factor, and at the top end distortion is the limit.

Noise is inevitable in microphones, amplifiers, tape, analog-to-digital and digital-to-analog conversion, and ancillary electronic equipment. It arises from the underlying randomness associated with the electrons comprising the signal-carrying mechanism. Even the simplest dynamic microphone (see pg. 82), containing no electronics and sitting in a vacuum, produces noise. This noise is caused by the Brownian motion of the electrons in the conductor comprising the voice coil and wiring of the microphone at room temperature. The only way to eliminate this noise is to cool the microphone to absolute zero temperature, at which all motion ceases. Thus, at practical temperatures a noise floor is established right at the microphone, below which desired signals are masked. Ultimately, then, the way to achieve minimum noise is to capture lots of signal by using high-sensitivity microphones close to the source.

Unfortunately, using high-output microphones close to the source leads to a potential problem at the other end of the dynamic range: So much signal will be picked up when the source is loud that it may easily overload or clip the electronics and cause severe distortion. The term *clipping* means that the signal peaks are literally truncated

by the electronics, grossly changing the waveform and producing distortion that is quite likely to be audible.

Choosing a reference level for the electrical level in a studio, or the recorded level on tape, is like using a gray card in photography: Making an exposure based on a gray card puts midrange brightness tones in an image in the middle of the exposure range of the film. Likewise, reference levels in audio are used to represent an average "exposure" or recorded level of a program. From this reference level, the two limits to the dynamic range are measured. The dynamic range from the reference level to the noise (such as tape hiss) is called the *signal-to-noise ratio,* and from the reference level to the maximum undistorted level is called the *headroom.* Adding the headroom in decibels to the signal-to-noise ratio, also in decibels, results in a single number for the dynamic range of the device or medium, which is certainly one of the most important aspects of performance.

The signal-to-noise ratio is so named because it is an expression in decibels relating the reference level (the signal) to the noise. The signal-to-noise ratio may be *unweighted,* that is, the measurement instrument may respond to all frequencies equally, or it may be *weighted,* that is, the meter is made to respond more like a human listener by emphasizing those frequency regions in which human hearing is most sensitive, with decreasing sensitivity in regions in which the ear is less sensitive.

Headroom is the amount above the reference level that a system can produce without severe distortion. Headroom may be flat with frequency, with frequency components anywhere in the audio spectrum overloading at the same level, or it may not be flat with frequency, often overloading or clipping first at the frequency extremes.

An example is given by the performance of open reel analog tape recorders operating at the studio speed of 15 inches/sec compared with cassette decks operating at 1⅞ inches/sec. Although a cassette recorder can do a credible job most of the time, one primary difference lies in the ability to record high levels of high frequencies; we say that the open-reel machine has much more *high-frequency headroom,* and thus a cymbal crash, consisting mostly of high frequencies, recorded on the open-reel
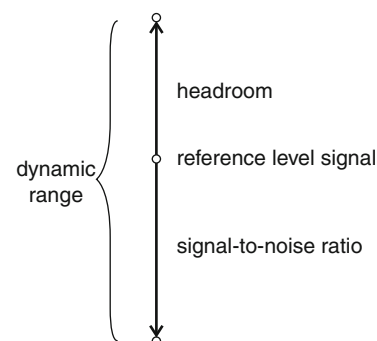
FIGURE 3.6 The realationship among dynamic range, headroom, signal-to-noise ratio, and reference level.

machine is undistorted, whereas recorded at the same relative level on the cassette machine it is distorted.

Audible distortion is more often a problem in film and television production than is audible noise caused by a source such as tape hiss. That is because the relatively high acoustic noise levels present on most sets thoroughly mask the noise from technical sources. (However, the noise from technical sources may become important in very quiet recording situations such as Foley.) That leaves distortion as the most obvious manifestation of dynamic range problems in most situations.

There are many points in an audio chain at which the signal may get to be so large that clipping or distortion occurs:

- In the microphone, especially in those equipped with their own electronics.
- In the microphone preamplifier.
- In the analog-to-digital conversion for digital recording.
- In subsequent signal processing to "improve" the sound quality.
- When multiple sources are summed together in a mixing console: One might not be too great a signal, but many added together may be too great.
- On any intermediate recording stage.
- At the final recording stage to the release medium.

Mixers call optimizing each of these stages for the best compromise between distortion and signal-to-noise ratio *gain staging*. Suffice it to say at this point that the signal in each stage in the chain should be optimized for level so that the widest dynamic range is preserved. This is like optimizing the exposure of film, not only on the negative, but on subsequent interpositive, internegative, and release-print stages so that the "signal," the desired picture, does not become under- or overexposed at any point in the chain.

Distortion added at any point in the chain can for all practical purposes not be undone at a later stage of the chain, and thus it is important to keep it low in every stage for the final result to sound undistorted. Likewise, noise accumulates from stage to stage, so underrecording is not the solution to distortion.

Linear and Nonlinear Distortion

Linear Distortion: Frequency Response, Amplitude, and Phase

Any change in a waveform constitutes distortion in the broadest sense, but some distortions under this broad definition are benign or even beneficial, whereas others are quite detrimental. The first class of distortion is called *linear distortion*. These distortions change the waveform, but the effect of the change can be "undone" by equal and opposite signal processing. If, for instance, the treble is boosted a few decibels by a piece of equipment, an equal and opposite treble cut introduced subsequently in the audio chain will restore the waveform to its original spectrum precisely, and essentially nothing is lost.

A linear distortion then is a change in what is called the frequency response of the system. A misnamed term, it would probably better be called "amplitude response with respect to frequency," but that being too long, the term has been shortened to the more familiar one, *frequency response*. It means how much one part of the audio frequency range or spectrum is accentuated or attenuated. For instance, a frequency response rating may read ± 1 dB, 20 Hz to 20 kHz, meaning that there is no more than a 2-dB variation from the minimum to the maximum of the response across the range.

> Frequency response actually has two parts, the amplitude response and the phase shift, both with respect to frequency. The two together completely describe the linear distortion that the waveform undergoes.

Many sound qualities are attributed to frequency-response variations, some of them very far away from the expected definition. For instance, midrange sounds around 2 kHz have a larger effect on the perception of distance than do other frequencies. Boosting this frequency range makes sound sources seem to be closer, whereas cutting it makes them seem farther away. Some console manufacturers, knowing this, have gone so far as to label equalization knobs for this frequency range (tone controls affecting a narrow frequency range only) in the boost condition presence and in the cut condition absence. There are many other examples of frequency response variations being ascribed subjective effects perhaps well beyond the expected range of a "tone" control.

In the most general sense, what we most often seek is a flat frequency response (sometimes called a *linear frequency response*) from most items in the audio chain. For instance, a tape recorder that discriminated against bass frequencies and boosted treble frequencies would not be desirable because all sounds going through the recorder would be affected. Although certain sounds might "sound better" with such a nonflat response, the overall average of sounds would not be improved. Thus, flat response is generally desired in most parts of the audio chain, and this is surely one of the most important specifications of any piece of equipment. In fact, in careful experiments, a deviation of as little as $^1/_2$ dB over several octaves of frequency range is an audible change. For equipment that is not supposed to change the sound quality, specifications on this order of magnitude ought to be considered necessary, especially when considering the multigenerational nature of film and television production, in which each sound is recorded and played an average of six times before it reaches the listener, and thus errors accumulate.

Exceptions to the requirement for flat response include the deliberate response changes made by equalizers and

filters to improve timbre and reduce noise, which are covered in Chapter 12. Microphones are most often distinguished audibly by two factors: their frequency response and how this varies with the angle to the microphone. There are reasons to make microphones nonflat, such as a high-frequency rise in shotgun microphones to make their typically distant placement seem closer.

Nonlinear Distortion

Nonlinear distortions are a class in which the waveform cannot be restored to its original shape by equal and opposite compensating equalization. In nonlinear distortions, new tonal components are added to the original ones, and no ordinary process can remove these added components. One of the most egregious examples of nonlinear distortion is *clipping distortion*. In clipping distortion, an amplifier or other device is driven beyond its capacity, with the result being literal clipping off of the peaks of the waveform. Clipping a sine wave, for example, which is by definition only a single frequency tone, results in a great many overtones being generated because the waveform is changed dramatically. Clipping distortion is sometimes heard in production sound recordings because it is difficult to exercise adequate control over all the gain-staging factors in a production sound mixer and recorder. It can usually be completely avoided by correct settings on the various pieces of equipment at hand, including the microphone itself, the mixer, and the recorder. Clipping distortion results in added frequency components at many harmonics of the original tone. If the clipping is perfectly symmetrical for positive and negative excursions, the resulting harmonics are odd order (the third, fifth, seventh, ninth, etc.,) and go out a long ways.

Distorting analog tape also creates harmonics, although the brick-wall clipping effect is not so pronounced. Analog tape tends to distort more and more as the level is driven higher and higher above the reference level, with the ultimate limit being complete *saturation* of the magnetic oxide. For this reason, it is conventional practice to modulate tape so that the peaks of the program reach only a certain distortion, without going all the way to saturation, except on nontonal sounds such as gunshots, for which the added distortion is generally inaudible.

The measure usually used for the distortion described is *total harmonic distortion* (THD). THD is the sum of the energy in all of the harmonics compared with that in the fundamental, expressed as a percentage. A typical maximum level for THD in an analog tape recording system is 3 percent on the maximum peaks of program material. However, THD is not an adequate measure for all types of audible distortion, for several reasons. Just 1 percent THD of clipping is more audible than 3 percent of tape saturation, because the greater number of higher harmonics of clipping is more audible than the simple lower harmonics (principally third) of the tape recorder. Because

digital recording is subject to clipping distortion, it is far worse to overrecord a digital medium than an analog one, with its more benign distortion harmonics.

To measure various distortion mechanisms beyond simple clipping or tape saturation, *intermodulation* (IM) distortion measures are used. IM distortion comes in a variety of types, but all are distinguished from harmonic distortion by the fact that the test signal contains more than one frequency, and it is the mutual effects of one frequency tone on another that are examined.

> For measuring SMPTE (Society of Motion Picture and Television Engineers) intermodulation distortion, for example, two tones at a low and a high frequency are used to drive the system being tested, with the low-frequency tone being 12 dB greater in amplitude than the high-frequency tone. What is looked for is a change in the high-frequency tone as a result of the larger low-frequency tone being present in the system. In a perfect system, the high-frequency tone would be unaltered by the presence of the low-frequency tone, but in a practical system, intermodulation distortion makes the high-frequency tone change level over the cycle of the low-frequency tone. Changes in level of the high-frequency tone at the low-frequency rate are heard as a "roughening" of the high-frequency tone, a kind of gurgle effect.
>
> High-frequency difference tone distortion tests send two relatively closely spaced high-frequency tones into a system and measure the resulting *difference tone intermodulation* at the difference frequency between the two tones. For example, 19- and 20-kHz tones mixed in a level ratio of 1:1 are sent into a system,
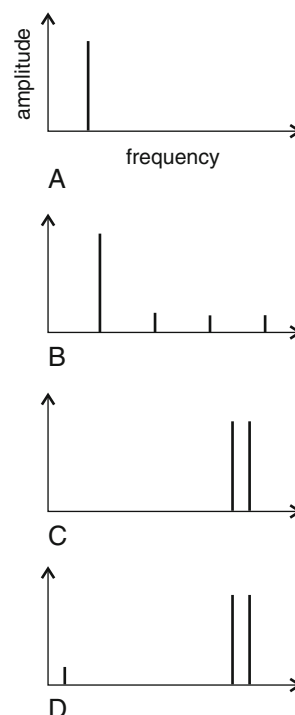

FIGURE 3.7   (a) The sine wave input to a device under test; (b) the distorted output showing the fundamental plus harmonic distortion; (c) the two sine wave inputs for a difference-tone intermodulation distortion test; and (d) the distorted output showing the original sine waves plus distortion at the difference frequency.

and the amount of 1 kHz resulting from distortion coming out of the system is measured.

Generally speaking, the most audible of these distortions is clipping, which must be avoided for all but the briefest instants to remain inaudible; the next most audible is harmonic distortion on overrecorded analog tape or film; and the least significant typically is intermodulation distortion. Still, there are special cases in which each one of these distortions can come to prominence in film and television production. For example, early in the history of tape recording, an ornithologist found that bird song recorded on a tape recorder contained lots of audible distortion in the form of low-frequency thumps accompanying the high-frequency bird song. What was found was a design problem with difference-tone intermodulation distortion in the recorder model used that had been overlooked by the designers because conventional test signals did not stimulate the effect.

Wow and Flutter

Analog mechanical tape and film transports, and phonograph records, are subject to pitch variations as the speed of the mechanism varies slightly around the normal playback speed. After all, off-speed transfers are often made of sound effects to make them seem something other than what they are. For instance running a tape machine at half-speed produces frequencies that are one-half the original on playback. It should thus not be surprising that speed variations result in pitch variations. Human hearing is particularly attuned to pitch variations and is able to distinguish a very small fraction of 1 percent variation under optimum conditions.

Analog tape machines, film transports, and analog optical playback from projectors in theaters and home videotape machines are all subject to wow and flutter, which are pitch variations arising from speed variations. Wow and flutter were originally distinguished as *wow* having to do with once-around variations in the once-around speed of phonograph players and *flutter* having to do with higher speed variations. Today, the two phenomena are lumped together and are thought of as one, wow and flutter.

Wow and flutter measurements are standardized differently in different parts of the world, so the numbers derived in Europe, the United States, and Japan are not necessarily comparable. Reliable measures of this statistical phenomenon are also hard to make. Unfortunately, like noise, wow and flutter is something that accumulates over generations, and performance that is fine for one generation may well be audible when accumulated over multiple generations. Wow and flutter is typically most audible on music, including solo instruments such as oboes and piano.

Humans are most sensitive to wow and flutter at a rate of around four variations in frequency per second and are less sensitive both below and above that frequency. For this reason, most wow and flutter measurements are made with a weighting curve emphasizing this frequency range.

Another form of speed modulation occurs at even higher frequencies than conventional flutter, *scrape flutter.* This is what happens when the stretched tape acts like a violin string and vibrates quite quickly. This causes *modulation noise,* which is a form of noise that is not present in quiet, but occurs only when a signal is present. You can hear this easily on a Nagra recorder by recording the reference-level oscillator while listening to the playback in headphones. Stopping the roller closest to the record head with your finger will raise the scrape flutter to audible levels, whereas letting it move freely will essentially eliminate the audible noise.

Digital Audio-Specific Problems

All of the dynamic range and distortion measures outlined earlier apply to digital audio systems. Wow and flutter, however, can be made vanishingly small, because of the nature of digital recording. Any speed variations in tape transports can be eliminated in playback by storing the digits coming from the unevenly played tape in an electronic buffer and withdrawing them at an even rate. Digital systems also provide the potential for no generation loss. These properties constitute some of the best features of digital audio systems, at which they are unequivocally better than analog systems when equipment and processes reach the full potential of digital audio.

On the downside, digital systems also come with their own peculiar distortion problems. Unlike analog systems, digital systems quantize the amplitude dimension. This function may cause difficulties, especially when compromises are deliberately introduced to save space on a medium. Particularly in fitting audio to the capacity of some low-end computer uses, such as CD-ROM or game boards for PCs, these compromises are likely to be so great that they become audible to the casual listener. Thus, the informed user should know what problems will arise when such measures are invoked.

The digital audio-specific problems discussed here apply to the most common digital audio method of representation, called linear *pulse code modulation* (PCM). In PCM digital audio, quantization occurs by comparing the amplitude of the waveform to the height of a series of stair steps. Each of the stair steps is of equal height. The quantizer (the heart of the analog-to-digital converter) compares the amplitude of the waveform to the height of the stair steps and assigns a number corresponding to the number of the nearest step.

There are other problems that occur in practical equipment; the ones outlined here are problems inherent in the basic PCM digital method. For example, an analog-to-digital converter in which all of the steps are not ascending in order (with, e.g., one step missing) is clearly defective.

Resolution

Resolution is the number of bits being used to represent the amplitude dimension, the number of steps in the stairs. For the compact disc, this number is 16 bits of binary (0 or 1) information, or 65,536 steps. Sixteen-bit representation yields a dynamic range of nearly 96 dB,[9] because each "bit" of resolution buys 6 dB of dynamic range ($16 \times 6 = 96$). On the other hand, some low-end computer boards, programs, and CD-ROM recordings are

---

[9]*Nearly* is an important word here. Nearly, for two reasons: no practical device reaches the theoretical and because of the need for dither, which is explained later.

made at only 8 bits, for a dynamic range of 48 dB. This produces audible noise accompanying almost any program material, because there is practically no program material that will mask noise only 48 dB below the maximum signal level. If we were to assign a reference level just 8 dB below the maximum, that is, 8 dB of headroom, then there is a signal-to-noise ratio of only 40 dB. A 40-dB signal-to-noise ratio means that the noise will be $^1\!/_{16}$ as loud as a signal at reference level and, thus, clearly audible. Eight decibels of headroom is also very little to accommodate louder sounds.

In addition, there is an inherent problem with complex productions using digital audio. For the result to be of a certain resolution, for example, 16 bits, if only one track is sent to one output, a 16-bit source will do. As the number of source tracks that are delivered to one output grows, however, the noise from each of the sources will add, and the result will be a decrease in resolution. So a 16-bit multitrack recorder has an inherent design problem: With any degree of mixing down to fewer output channels it is impossible to deliver an output that has a 16-bit dynamic range. Although a formula can be given for determining the number of additional bits necessary to produce the desired resolution, the formula does not take into account varying levels among the channels, equalization, etc. One emerging trend in digital audio is toward greater resolution in professional equipment, to 20 and 24 bit, which is a useful improvement in large-scale production.

Sampling and Aliasing Distortion

Although quantizing is at the heart of digital audio, another process must occur beforehand, sampling. The procedure is to measure the audio signal so many times per second that all of the nuances of the signal that are in the audio frequency band are captured. The sample rate required is a little more than twice the highest frequency in the desired bandwidth. With 20 kHz usually considered the highest audible frequency, the sample rate for virtually all sound accompanying a picture is 48 kHz, and the compact disc is 44.1 kHz. (A lower rate of 32 kHz is used in some broadcasting applications, and some available equipment uses 96 kHz sampling, coming down on the side of extending the bandwidth into what is generally considered to be the ultrasonic domain.)

To save space on computer discs and CD-ROMs so that more audio can be stored, it is common to sample at lower rates, usually submultiples of 44.1 kHz, such as 22.05, 11.025, etc. The problem encountered with sampling at these lower rates is that frequency components at more than one-half of the sampling frequency are likely to be in the signal. The sampling process "confuses" these with lower frequency signals and produces an output tone that is different in frequency from the input. This is called *aliasing distortion*.

> One-half of the sample rate is called the *folding frequency* because aliasing distortion "folds" the signal frequencies around one-half the sample rate. So any input frequency above one-half the sample rate will alias and appear as a new frequency in the output. For instance, if the sample rate is 11 kHz and the signal frequency is 7 kHz (speech recorded for a CD-ROM will contain such a frequency in an "s" sound), one-half of the sample rate is 5.5 kHz and the 7-kHz signal will be seen as a 4-kHz signal (7 kHz is 1.5 kHz above the folding frequency, and 4 kHz is 1.5 kHz below the folding frequency).

An example of aliasing distortion in picture images is a shot of wagon wheels appearing to run backward when photographed, or sampled, at 24 fps. Up to a certain speed of the wagon, photography renders an accurate representation of the speed of the wheels. But at the point at which the camera shutter is open, then closed, then reopened just as the spokes have moved, for example, $^1\!/_6$ of a rotation for a six-spoke wheel, the film "sees" a stationary image of a moving object: they appear frozen. This is picture aliasing distortion. At other relative rates between the photography and the speed of the wheels, they may even appear to run backward, which is clearly an artifact.

In audio, the sound of aliasing distortion is distinctive. On speech as a signal, it sounds like a chirp that accompanies "ess" and other high-frequency sounds in speech. It may be fairly benign, or very nasty, depending on the strength of "esses" in the speech and the sample rate (lower ones are worse). The way to avoid aliasing distortion is to use a filter that removes the frequencies higher than one-half the sample rate before sampling. This filter is called an *anti-aliasing filter*. Unfortunately, many cheap low-sample-rate analog-to-digital converters are not equipped with anti-aliasing filters, and plainly audible aliases are the result on much program material.

At the output end of a digital audio chain, another filter is needed to strip off the ultrasonic components of the
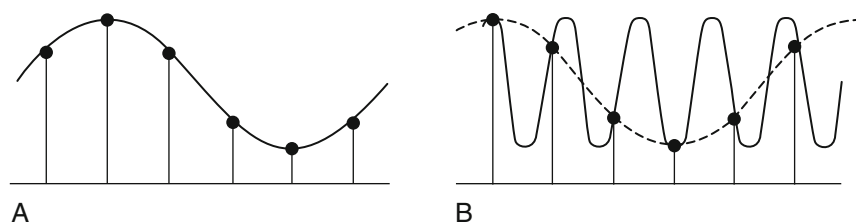

FIGURE 3.8 (a) shows a signal properly sampled; (b) shows a signal sampled too infrequently and thus producing an alias, the dotted line, from an input signal that is above the folding frequency.

A

B

audio that are artifacts of sampling. Sampling will cause "images," repeated audio spectra at ultrasonic frequencies. A *reconstruction filter,* sometimes called an *anti-imaging filter,* is used to filter out these frequency components, leaving an artifact-free audio band output.

Jitter

An issue in sampling and in digital-to-digital interfaces is called *jitter*. This is the variation in time from sample to sample from the calculated time due to imperfections and can result in pure tones becoming "rougher" sounding, much like scrape flutter, only generally jitter is smaller in effect than scrape flutter. Jitter can even be present in digital-to-digital interfaces. The input of properly designed digital audio equipment "reclocks" incoming jittered signals and restores them.

Quantizing Distortion

The process of quantizing the amplitude dimension of the signal also carries with it the potential for a particular kind of distortion, called *quantizing distortion*. Quantizing distortion arises even in a perfect linear PCM system when the amplitudes of the signals are very small, unless measures are taken to prevent it (see *dither* later).

Let us say that a signal is just slightly larger than one step of the staircase and is a pure sine-wave tone. It will be converted as it first crosses above and then below the level of the first tread of the staircase. The digital representation of the sine wave will simply alternate between the higher and the lower bit levels. The result, upon conversion back to analog, will be a square wave, not a sine wave. The reason is that the converter is coarse at these low levels and cannot discriminate the waveform, only the fact that the one tread has been alternately crossed. Thus, what came into the analog-to-digital conversion process as a sine wave is actually converted as a square wave, producing a huge distortion. Furthermore, a just slightly lower level signal is not converted at all, because it never crosses the tread of the staircase. Thus any lower level signals than the smallest step (called the *least significant bit*) are discarded, another important distortion. There is a way around these distortions, and it is called *dither*.

Dither

The effects of quantizing distortion can be fully eliminated by the addition of some deliberately added random noise, which may sound like hiss. Although noise is usually considered to be a detriment in any system, adding dither noise in a digital audio conversion process randomly "agitates" the signal so that even the smallest signals cross the treads of the staircase. The noise pushes the signal up and

down randomly, causing it to cross the threshold often. By averaging the signal plus the dither over time, like the hearing mechanism does, signals even far below the threshold of the smallest step can be perceived. The noise effectively smears out the steps in the staircase, turning the stair steps into a linear ramp when averaged over time. The amount of dither is about equal to one step in the staircase, so the added noise to produce these benefits is quite small, although it may become noticeable when many channels are added together.

Another important role for dither is in reducing the number of bits of resolution when a source has more resolution than a copy. If 16-bit recordings are made in the studio, and then transferred to 8 bit for release on CD-ROM, very significant audible distortion results. These effects can be minimized by adding the proper amount and type of dither. With 20-bit storage capability on digital videotape, and 20-bit converters growing in use, conversion to the final consumer format of 16 bits, for example, must add dither so that the coarser steps of 16 bits compared with the 20-bit original do not become a source of quantizing distortion.

Dither may be added in a number of forms. The noise needed may be "hidden" in the less audible parts of the spectrum, by shaping the frequency response of the noise to be like the equal-loudness contours, with little noise in the 2- to 3-kHz most-sensitive region of hearing, and increasing it above 10 kHz, where sound is less audible. Called *noise-shaped dither,* and also by trade names such as Super Bit Mapping, this process yields an audible improvement in dynamic range over the restrictions of the output resolution. Perceptibly, a 16-bit system can be made to sound as though it has the dynamic range of a 19-bit system in this way.

A Digital Audio System

There are a number of processing steps in making a digital audio system, caused by the needs outlined above. In addition, it should be pointed out that we are dealing with only one kind of digital audio in this exposition, and that is linear pulse-code-modulated digital.

A PCM system has the following parts to its block diagram, in the order that a signal encounters them:

- Anti-aliasing filter;
- Dither noise generator;
- Summer for the signal and the noise;
- Analog-to-digital converter;
- Digital circuitry to add error-protecting codes and to condition the signal for recording or transmission;
- Medium for storage or transmission of the digital bits;
- Digital circuits to decode the error codes and correct for errors or to interpolate between known-good samples if correction cannot be performed;

- Digital-to-analog converter;
- Reconstruction filter, the equivalent on the output of the anti-aliasing filter on the input, which constrains the output spectrum to the audible one.

Specialized application areas, such as digital release prints, employ other techniques, because there is not enough space on prints for conventional digital audio. Linear PCM is the most common method of recording and storing audio today, but it consumes a lot of digital audio storage for a given portion of track time. For instance, sampling at 16 bits with a 48 kHz clock results in 720,000 bits per second per channel. Because of this high number, lower sample rates and resolution are used for CD-ROM and other such purposes to save space. Another solution is perceptual coding, which works by "throwing" away signals that would be masked by human hearing mechanisms, achieving a more than 10 times bit-rate reduction with potentially few audible artifacts. See Low-Bit-Rate Audio below.

Oversampling

Two dimensions of an audio signal are amplitude and frequency. It is possible to trade one of these off against the other, because the information-carrying capacity of a channel is related to the product of these two factors. Radio microphones, for instance, convert the audio bandwidth and dynamic range from their microphone to a wider bandwidth and a smaller dynamic range for transmission in the radio-frequency channel. The signal is represented by a different means (FM radio), but the information content is still the same. The two dimensions are represented in digital audio by quantizing and sampling, respectively. Oversampling can be compared with the process used in radio microphones, trading off bandwidth and dynamic range. By sampling at a higher than normal sample rate, less dynamic range in the A-D and D-A converters is needed to produce results associated with a wider dynamic range in the audio channel.

Oversampling is a process one frequently hears about in terms of the number of times of oversampling that a particular circuit employs, from "4 times oversampling" to "128 times oversampling." The number of times refers to the sample rate; for example, a 4-times oversampled professional audio system samples at 192 kHz (4 × 48 kHz). The basic idea behind oversampling is that sampling at higher rates spreads the noise of the analog-to-digital conversion process out over a wider than audible frequency range. The portion of the noise that is ultrasonic is inaudible, so the more spreading the better (the higher the sample rate or number of times of oversampling). When converted at the other end of the process back to audio, just the audible frequency range noise counts, so it is possible for an oversampled system to come closer to the ideal than one using conventional sample rates. Oversampling also simplifies the requirements of the anti-aliasing and reconstruction filters, because they need not filter so steeply because the sample frequency is raised so much. On the other hand, there are practical problems involved in oversampling, such as the fact that *jitter,* small variations in the time that the samples are taken from one to the next because of imperfections in the process, becomes relatively more important than in a less-sampled system. So it is by no means clear that the system with the highest oversampling rate is necessarily the best.

Low-Bit-Rate Audio

It is easy to calculate how much space linear PCM content takes up. Multiply the sample rate times the number of bits per sample (also called resolution, word length, or bit depth) times the number of channels times the number of seconds of content, and you have the number of bits that must be stored. For computer files, this is a close estimate, with only a little overhead. For media like the compact disc, it is only a fraction of what must be stored because of the needs of error and channel coding.
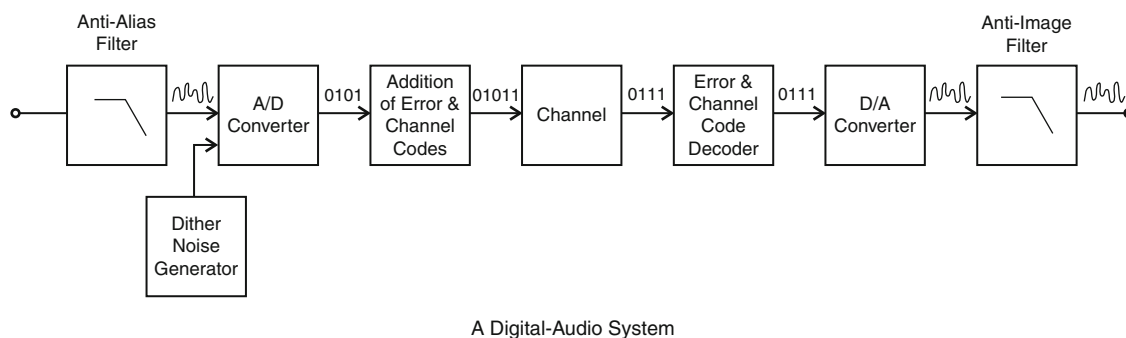

A Digital-Audio System

FIGURE 3.9 The parts of a linear PCM digital-audio system from analog input to output.

For 1 hour of mono content that might come from an ADR studio, for example:

48,000 samples/sec $\times$ 24 bits/sample $\times$ 60 sec/min $\times$ 60 min/hour $=$ 4,147,200,000 bits/hour. Divide the number of bits by 8 to get 518,400,000 bytes (8 bits $=$ 1 byte).

For media like digital optical sound on film, the bit rate is so high that it is impractical to record on the space outside the picture area on film. Take a 5.1-channel soundtrack (actually 5.005 mathematically, described in Chapter 13) to be stored digitally between the 4 perforations on one edge of the print for each frame on 35 mm motion picture prints. There are 4 perforations per frame, and 24 frames per second, so there are 96 little spaces in between the perforations where we might record bits photographically. But how practical is this? We also want a greater than 16-bit (CD performance) resolution, to get more dynamic range, so let's go for 20 bits, or 117 dB dynamic range. So here is the calculation:

- 96 spaces/sec, where each little space between the perforations to photograph pixels is about 0.1-inch square. The area per second of 96 spaces that are 0.1-inch square is 96 spaces $\times$ (0.1 inch/space)^2 = 0.96 square inches, or roughly speaking, about 1 square inch.
- 48,000 samples/sec $\times$ 20 bits/sample $\times$ 5.005 channels $=$ 4,804,800 bits/sec.

- The nearly 5 million pixels must fit in less than 1 square inch of film, about 2200 pixels per inch linearly. (The square root of 5 million.) That number is better than the resolving power of film, which is about 2000 pixels per inch.
- So it won't fit. Furthermore, these are only audio "payload" bits, because there need to be error correction ones, synchronization pattern ones for recovery off the film, and others.

For this reason, low-bit-rate coding was developed. By utilizing the known frequency and temporal masking characteristics of human hearing, it is possible to reduce the number of bits required by a significant factor, such as greater than 12:1. Such perceptual coders are well known today, because MP3 players are ubiquitous, and they use this type of coding.

For the purposes of professional film and television work, perceptual coding is applied only at the last stage, putting the sound on film digitally or broadcasting it, for example. Recording and using it for originals, or compressing the audio in postproduction to save space on a hard drive, is a bad idea because the variety of perceptual coders available have unknown properties when they are used sequentially, and artifacts that might be inaudible in one or another coder may become annoying when two are used in series.

Capturing Sound

INTRODUCTION

In the earliest days of sound films, cameras had to be housed in small padded booths on the large shooting stages in use at that time to prevent their very audible mechanical noise from being picked up by the microphone(s). They sounded approximately like a sewing machine, because the pulldown in the camera is something like a machine stitch. The "blimped" camera, with noise-reducing padding around the camera mechanism, had yet to be invented.[1] The resulting static pictures seemed like photographed stage plays, not cinema as it had developed over the previous 30 years.

> The artistry of silent film that empowered the camera with incarnating not only the world and people in it as moveable, but subjectivity in terms of perspective (image size, angle, movement), just had to stop being expressive until they could figure out how to capture sound without camera noise while getting back to all camera could and should do.[2]

Able Gance's 6-hour epic silent film *Napoléon* was released in April 1927 at the Paris Opera, and it featured scenes with very fluid camera movement, including a camera swinging in and out of a scene. Commercially viable sound came in October of the same year, with *The Jazz Singer*,[3] but this then had to be made with a static camera. Sound movies of 1927–1930 including, for example, *The Royal Family of Broadway,* were composed of very static camera shots. Despite the drawbacks caused by the static camera, the new-fangled sound film overwhelmed the silent film at the box office in an amazingly short time, with millions of dollars in silent film inventory orphaned. In fact, Hollywood, knowing a good story when it saw it, later made a movie about tensions that occurred in the silent to sound movie transition, *Singin' in the Rain*.

One preeminent silent filmmaker, though, saw through the early sound problems, and said this:

> Talkies, squeakies, moanies, songies, and squawkies . . . whatever you call them, I'm absolutely serious in what I have to say about them. Just give them 10 years to develop and you're going to see the greatest artistic medium the world has known.[4]

That filmmaker was D. W. Griffith.

Today we have come to the point at which sound people are expected to get clean dialog in every circumstance, even when you as a director cannot hear the actors when sitting next to the camera! Although this might make sense in some scenes (a very long shot), in general the ability to hear the actors on the set without interfering noise and reverberation is important. The more naturalist style of film acting leads some actors to be so afraid of "overacting" in a theatrical way that they play at a low energy level that can lead to recording difficulties. When this happens, the sound recordist should say something to the director, although the euphemism "energy level" will probably go further than just telling the actor to talk louder!

There are other concerns about the interaction of the sound crew with the rest of the crew, and those will be taken up later.

MICROPHONES IN GENERAL

Microphones are the lenses for sound. Their position and the direction they are pointed affect sound profoundly,[5] as do the acoustic conditions of noise and reverberation present in the recording environment. The analogy to lenses extends to lens focal length: a long focal length lens has a narrower field of view—it is said to be a telephoto lens. Likewise, microphones have an angle of acceptance, which may range from all around, called omnidirectional, to the narrow beam of a "shotgun" mic.

---

[1] The Mitchell BNC camera, with the "B" standing for blimped, was introduced in 1934, 6 years after commercial film sound.

[2] University of Southern California Professor Drew Casper, private communication, October 2008.

[3] Although there were a great many earlier experiments with systems of both sound on film and sound on a separate medium.

[4] From an exhibition at the Hollywood Roosevelt Hotel that, until recently, occupied the balcony there for many years.

[5] The comment about the direction they are pointed applies to most microphones used on booms, podiums, and the like. Most lavaliere microphones are omnidirectional, and there are reasons to use omnidirectional microphones in other circumstances described in the text.

However, the analogy to lenses breaks down when the separate mechanisms by which light and sound propagate are considered. For practical purposes light does not make it around corners; sound does. This means that a picture has a hard frameline, whereas a microphone operating on the same scene can offer only the tendency to pick up what is in the frame better than what is out of the frame. This fact causes problems on locations, in that a camera can pan off an offending sign while making a Western in the desert, but a microphone cannot completely ignore the airplanes flying overhead.

Microphones are, of course, fundamental to film and video sound for use in capturing the sounds present on a set, in a sound-effects gathering expedition, on a scoring stage, or elsewhere, and in converting the sound into an electrical voltage proportional to the amplitude of the acoustic waveform,[6] varying moment by moment. This voltage is generally conducted over wires to a microphone preamplifier. Alternatively, it may be the input to a wireless transmitter that conveys the audio content by radio frequency transmission to a companion receiver. A microphone is called more generally a *transducer,* because it converts energy from one form (acoustical) into another form (electrical). Likewise, a loudspeaker is also a transducer, working in the opposite direction.

> There are other transducers as well as microphones used very occasionally in filmmaking. These include hydrophones (underwater microphones) and contact microphones or accelerometers (vibration-sensitive pickup devices that are directly attached to the source, thus obviating air transmission).

PRODUCTION SOUND FOR FICTION FILMS

A production sound crew for a feature film usually consists of a boom operator, a production sound mixer, and a utility person. The boom operator runs the boom or fishpole with the microphone at the business end of it. The production sound mixer operates the technical equipment consisting of a microphone mixer, a recorder,[7] wireless radio receivers, and a variety of auxiliary equipment, often housed on a sound cart. The utility person helps by running cable, making minor repairs to equipment, and working as a second boom operator in complicated scenes.

Principal photography occurs with a first unit, a set of personnel and equipment that includes the director, principal actors, and crew. The production sound crew generally works on the first unit, but depending on the schedule, might be asked to record sound during the shooting of a second unit. Second units usually do not involve the principal actors and may or may not record sound, depending on the nature of what is being photographed. If second units are deployed at the same time as first units, depending on what is being shot, a second sound crew may be hired for second-unit work.

Sound crews, often being among the most technical people on the set, often also look after such ancillary items as walkie–talkies; headphones for the director, script supervisor, and others; setup of sound systems for monitoring, public address, and the like; sound systems for stars' trailers; repair of equipment; and many other sorts of technical jobs. The utility person may be assigned the job of keeping track of the time that all the batteries have been operating, especially radio mic batteries, and changing them or charging them before their discharge would cause problems like distortion due to low battery voltage.

An added crew position of playback operator may be needed for instances in which music is prerecorded and the actors perform to the track. In some instances full performances are prerecorded, and the actor "lip syncs" to the track, played to him or her over loudspeakers or by way of wireless earphones. In others, part of the actor's live performance may be prerecorded, and part recorded live. By cueing the actor with a prerecorded track part of the time, that part of the performance is kept in sync. One of various technical means may be used, such as click-track or cue-track receivers and earphones for the actors or "thumpers" playing only the low bass beat of a track over subwoofer loudspeakers so that the actors can maintain sync while performing. Later in postproduction the low frequencies of a thumper can be filtered out. This leaves a "clean" track of the live performance, but in sync with a preexisting track, so that it can be easily combined with a musical score.

PREPRODUCTION—LOCATION SCOUTING

On feature films preliminary location scouting is usually performed by a location manager. He or she provides numerous photos (but never sound recordings!) from which likely sites are selected, and then the production designer and director of photography, and sometimes the director, visit the sites with the location manager for final selection. Throughout this process it is not common for the sound personnel to accompany them, so it is important for those visiting the sites to be aware of background noise and other problems such as excessive reverberation in the environment.

- For *Brokeback Mountain,* location manager Darryl Solly explains in an interview[8] that one site selected

---

[6]These concepts are discussed in Chapters 5 and 6.

[7]Sometimes the functions of a mixer and recorder are combined in one unit.

[8]Online at http://www.findingbrokeback.com/Downloads/Solly_Interview .pdf. See also http://www.findingbrokeback.com/Downloads/Benz_Interview .pdf for an interview with production manager Tom Benz. Accessed 15 October 2008.

was very quiet, but trains came by the site, for which they would have to stop production to let the noise pass.

- For *To Wong Foo, Thanks for Everything! Julie Newmar* the crew members scouting the locations told production sound mixer Michael Barosky about the unpaved gravel main street of the small town in which much of the action is set, so that some action could be taken to minimize the sound of the crew moving on the roadbed.

- In a funny but sad scene in the documentary *Lost in La Mancha* (2002) about a failed film production, director Terry Gilliam takes on with much swearing the excessive reverberation of a so-called "shooting stage"—it seems those who designed this Spanish stage expected film production to work in the Italian style, in which everything is postdubbed, with the director talking to the actors continuously, and thus to throw away everything recorded there!

MICROPHONE TECHNIQUE—MONO

Monaural recording is defined as recording with one or more microphones, but if more than one, the result is not intended for stereophonic (more than one primary channel) presentation. In production sound recording of motion pictures and television shows most of the primary sound will wind up in the center loudspeaker channel, for reasons discussed below and in the chapters on editing and mixing. There is basically single-point mic'ing for each important sound source, with only unintentional overlap among the microphones. The recording may take place on one or more tracks of a recorder, with "multitrack mono" not being a contradiction in terms. A variety of microphone techniques to be discussed—boom mics, planted mics, lavalieres— may be recorded on separate tracks, but they are not meant to be a part of a stereophonic scheme such as left, center, and right.

For a further explanation, see the definition of stereo recording under Microphone Technique—Stereo. Monaural production sound recording, especially on multiple tracks of a recorder, is the norm today, even if the final show will be in multichannel. The reasons for this are several:

- Although multichannel stereo may sound great within a single shot (and thus at dailies), especially at better representing the space of the scene, it brings about major problems in editing. Cutting from a master shot in which a character is on the right side of the frame to a close-up, in which he is in the center, sounds discontinuous. We have come to expect the grammar of picture editing, with smooth editing (without jump cuts) as a typical goal. Most sound edits that change perspective dramatically sound abrupt, as we do not have as much experience with sound perspective changes as with picture cuts, and thus they create an aural jump cut.

- Keeping separate mics, albeit not ones intended for stereo coverage, on separate tracks permits cutting out noises that arise in one kind of coverage, such as lavaliere clothing rustle. With stereo microphony, if one mic's signal has to be cut out, a collapse of the stereo sound field will be in evidence.

- Matching the background noise (room tone, presence) from shot to shot is greatly complicated for the dialog editor if not only the level and timbre have to match, but also the sense of perspective including spaciousness. Having room tone cut for every single perspective is impractical and can cause loss of continuity, although later I discuss using a mid-side (MS) shotgun on the boom for capturing dialog and ambience simultaneously.

Distance Effect

Decreasing the distance between a microphone and a sound source decreases reverberation and increases direct sound. In production sound recording, and for sound effects recording both on Foley stages and in the field, the general practice is to minimize reverberation as much as possible. This is because it is always possible to add reverberation in postproduction and practically impossible to reduce it significantly. Thus, all other things being equal, there is a desire to record a source relatively close up. Of course, in narrative filmmaking this must be tempered by the need to keep the mic out of the frame.

Decreasing the distance between a microphone and a sound source also may decrease the recording of extraneous acoustical noise, depending on the direction and nature of the noise. It is often useful to aim not only the "hot" side of the microphone at the desired source, but also to aim the "null" or dead side of the microphone at the source of noise.[9] Although all noise will surely not be eliminated, because the noise as well as the source has its own direct sound, discrete reflections, and reverberation[10] and no mic directional pattern can cope with all of these, nonetheless a significant noise improvement can be gained by this technique.

Decreasing the distance increases the direct sound, and because the noise level of the microphone and microphone preamplifier that is heard as hiss is fixed, decreasing the distance improves the ratio of the desired sound to the electronic noise. Working at a great distance from a weak source will cause the recordist to "turn up the gain" and that will pull up the fixed electronic noise, so close working generally improves the signal-to-noise ratio.

---

[9]This is explained further on pages 80–81.
[10]Explained in detail in Chapter 1.

If a directional microphone is being used, and with the exception of lavalieres, one almost always is, moving the microphone in really close will generally need to be tempered by the proximity effect[11] of the particular microphone in use. If it is too close the bass will be significantly boosted, causing an audible timbre problem. Although this could be ameliorated in postproduction by turning the bass down with an equalizer, varying the distance from the microphone changes the amount of the effect, so it can be very difficult to vary the equalization from moment to moment to compensate for this effect as the actor moves to and fro. In some cases it may even be better to change to an omnidirectional microphone moved in closer to the source instead of using a directional microphone farther away. This consideration does not usually apply to boom mic'ing in production sound because of the distances involved, but it certainly is applicable to voice-over or narration recording, to a "public-address" microphone into which the actor is speaking, and to many kinds of effects recording: too close can be a problem.

Microphone Directionality

Using a more directional microphone generally leads to recording with a higher ratio of direct sound to reverberation (and sometimes, echoes), with the intended source emphasized and reverberation deemphasized. This is why practically all recordings made with boom mics use directional microphones, usually those with the highest directivity, hyper- or supercardioid or interference tube club shape, described under *Polar Patterns* in Chapter 5 and shown in Fig. 5.8.

On the other hand, it may be impractical in a given shot to mic from the overhead boom position, and body mics may be necessary. In this case, the required small size of the microphone favors an omnidirectional type because the space to make the microphone directional across a wide frequency range just isn't available. In this case, we rely on the fact that the microphone is much closer than would otherwise be possible to increase the ratio of direct sound to reverberation and noise.

Microphone Perspective

Perspective is the matching of the sound recording to the characteristics of the picture, in particular, the match of the sound reflection and reverberation properties to what is seen. In early sound recording there was an attempt to match camera perspective, shot by shot, to what was seen. A wide master shot was thus more reverberant than the associated close-ups. Then, when the scene was cut

together, there was a very noticeable change in the amount of reverberation, as the microphone perspective was "loosened" for the master and "tightened" for the close-up. This tended to draw attention to the soundtrack because of the unnatural changes at the cuts, which are not duplicated in life. Modern practice thus most often uses relatively small changes in the microphone perspective to correspond to large picture changes and the reverberation changes are consequently kept fairly subtle—audible if you listen for them, but probably not noticeable to the average viewer who is not "tuned in" to them.

For a microphone perspective that matches the perspective of the camera, you might think that a mic over the camera would do, moving from a wide shot to a close-up just as camera lenses do. This does not work; a microphone recording just does not have the same perspective as being there. In fact, the best way to cover a scene that is obviously shot in a reverberant space, such as a large marble-covered interior of a museum, is to record all of it with a constant and tight perspective and then decide in postproduction whether the judicious use of added reverberation for the wide shots makes sense.

The Boom—Why, Isn't That Old Fashioned?

Although it may be old fashioned, mics on booms usually work best for dialog recording. The best position is on a boom or fishpole above the frameline and in the center of the frame. Among other things, this is because the perspective matches that of the camera. If an actor should happen to turn and face out of the frame we hear the change of going from "on mic" to "off mic" as natural. The on-mic position sounds clearer and less reverberant, whereas off mic sounds a little duller and more reverberant, and this matches our common experience.

Capturing the sound of the voice well, its timbre, is the second reason an overhead boom is desirable. Talkers sound clearer to the front and above compared to the side and below the axis of their mouth. Because mics in front of talkers are on a line with the camera, this position is only possible occasionally when the mic is allowed to show in the shot, such as on a podium from which a speech is being made by an actor, making the overhead position the best available for most situations.

A microphone centered underneath the frameline can be considered to be a fallback position, but it is not as desirable, because this position often sounds "chesty," emphasizing midbass, and less clear than above. Also, if the microphone thus has to be near the floor, it will receive interfering reflections off the floor that can color the sound.

If the microphone has to be placed to one side of the frame, in the case in which there is a very wide and low

---

[11]Explained further on page 79.

shot that does not permit either an overhead or an under-hand position, then perspective problems arise. Whether this is successful depends on the blocking of the scene. If an actor should turn toward and then away from the microphone while speaking, he or she will sound on mic when facing one direction and off mic facing the other, even though both angles may be the same to the camera. This then sounds artificial as we come to understand at least subconsciously where the microphone is.

A large potential problem exists with microphones that must be close to surfaces but cannot be made nearly integral to those surfaces. The problem lies with the strong potential reflection off the surface, which arrives slightly later at the microphone than the direct sound. This gives rise to constructive and destructive interference, which results in a regular series of peaks and dips in the frequency response that can be quite audible. All in all, microphones prefer a lot of "air" around them, so that nearby reflections are minimized, or they should be made an integral part of the surface, as in the boundary-layer method to be discussed.[12] If a boom microphone needs to be placed near a ceiling because of low ceiling height on a location set, the reflection off the ceiling may be ameliorated by taping an area of absorbing material to the region of the ceiling above where the microphone has to work. Thicker material covering more area will be more effective than thin material covering a small area.

For boom mics in their ordinary location above the frame, there is a large difference between film and video cameras to be aware of. Film camera viewfinders show a larger area than do most video cameras. It is common for film cameras to have etched lines showing the frameline for the format in use in the viewfinder and for the operator to be able to see a fair amount outside the frameline. This gives a film camera operator an advantage as they can tell when the boom mic is about to intrude into a shot. Video cameras do not have this feature: their video viewfinder shows just what is scanned. So a videographer is perhaps naturally more insistent that the boom mic be placed

higher than we might consider necessary, because he or she has no "early warning" system to know when the microphone might intrude into the shot. A few high-end digital cinema cameras, which are upgraded high-definition video cameras, have optical viewfinders to overcome this limitation of most video cameras and work in the same way as film cameras with respect to the viewfinder.

Booms and Fishpoles

Appropriate microphones may be put on a boom or a fishpole. The choice of one of these depends on many factors, with, among other considerations, the boom being less tiring to the operator and the fishpole more mobile. This leads to uses on fixed sitcom sets that prefer the boom, whereas shoot-and-run documentaries prefer the fishpole. Feature film production falls in between these two types of productions, so either choice may be made, and the choice may even change fluidly throughout a production.

The industry standard J. L. Fisher booms may be operated mounted on floor stands on wheels or from catwalks overhead in studios. See Figs. 4.1 and 4.2. They provide a means to rotate and tilt the microphone silently at the end of the boom, as well as rotating and tilting the entire boom arm and extending the boom. Operation of these is a specialty a little like being able to rub your stomach while patting your head, so it takes some training, but a trained operator can perform remarkable feats of positioning and aiming the microphone to capture best the performance of the actors.

Fishpoles are generally extensible and are sometimes made of carbon fiber for low weight. They may contain coiled microphone cables inside, so one connects the microphone at the far end and cables to the mixer or to a radio mic transmitter at the operator's end. Some models are available with side-mount XLR output connectors, which means that the fishpole can be rested on one's foot between takes, an important consideration. K-Tek makes a wide range of models, one of which is illustrated in Fig. 4.9 They also manufacture brass weights, which, when added to the end of a fishpole, permit the operator to work at the fulcrum of a beam instead of at one end of a long arm with its accompanying torque moment,

---

[12]On page 62.

FIGURE 4.1   A J. L. Fisher boom in use. Photo courtesy J. L. Fisher Inc.

FIGURE 4.2 The business end of a Fisher boom, showing the mechanism that rotates and tilts the microphone under the control of the operator at the other end of the boom. Photo courtesy J. L. Fisher, Inc.

producing less fatigue. This may be counterintuitive, but adding weight may make the job easier in this case.

Boom and Fishpole Operation

"Booming" is such an important job that a boom operator can make or break a recording, even though he or she is subordinate to the production sound mixer in the staffing hierarchy. For fiction filmmaking, the boom operator learns the scene and positions the microphone from moment to moment to best effect. The operator learns the script extremely well, as well as anyone on the set. Rehearsals are essential. Cinematographer Haskell Wexler says "I can't light a set; let me see the rehearsal," because it is the actors he is lighting, not the set. Likewise the boom operator needs rehearsal to optimize the mic location through the course of a scene. Having stand-ins for the stars to do the lighting usually does not help the sound department, because stand-ins are not trained actors who speak lines. It is often said that actors must just "hit their marks and say their lines," because it is essential for the camera focus puller to obtain sharp focus by having the actors "hit their marks," often mapped out on the floor. Although it would be a good idea to be able to hear, in advance of a camera rehearsal with the actual actors, how they are going to sound with a particular microphone setup, this happens infrequently. The rehearsal with the stand-ins can reveal problems with boom shadows though, and can give a rough idea of booming, so it is nonetheless useful.

Boom operators have a much more interesting job than one might think at first glance. Although sometimes it is thought of as a job just anybody can do—"just a big kid out of high school with strong arms," as producer Gene Corman explained it to me (while I was doing the boom operator's job on a show!)—boom operators have perhaps the most input to microphone perspective and thus coverage; they act to favor weaker actors over stronger ones, and they often may have an important interaction with the actors.

Good boom operators, after a long day on the set, read the script for the next day's shooting, and memorize it. During complicated master shots, for instance, the microphone is likely to be constantly in motion, getting the actor who is speaking on mic, and missing a cue would cause an obvious change in perspective, so must be avoided. This is why the boom operator must know the script.

On many sets today, the director is huddled over a video monitor during takes, shielded from the sun. It reminds one of early photographers under their black cloth so they could see the ground glass of the camera. The camera operator is looking through the viewfinder. The script supervisor is sitting underneath the camera, and the eyeline of the actor would look wrong pointed in his or her direction. There is lots of lighting from multiple directions, effectively blinding the actor from looking in those directions. All these being considered, that leaves the boom operator as a point of human contact. There are stories of actors coming at the end of the day to the boom operator, who has shared an emotional moment by shedding a tear with the actor, and thanking the boom operator for the connection.

A major issue in the operation of either a boom or a fishpole is the potential for boom shadows. Stanley Kubrick was a still photographer before becoming a director. He decided to light a loft interior in New York himself for his second film, *Killer's Kiss*. He threw everybody out of the room, and proceeded to light this large, white space evenly over the course of several hours. He then called in the crew and actors. The moment the boom was put up, multiple shadows were obvious. And when the actors moved and the boom followed, those pesky shadows moved too, making them all the more obvious. Kubrick asked the sound mixer if the boom was necessary. Told that it was, Kubrick fired the sound crew and recorded the dialog on a bad, nonsync recorder just to know what was said (important because actors don't always follow the script exactly). He then spent the next 4 months himself recording and postsyncing the dialog, adding footsteps, ambience, etc. On his next show, he hired a professional gaffer![13]

The best position for a boom or fishpole operator is usually to the left side of the camera facing the scene and a little in front of the camera. This is because the camera operator is virtually always on the left side of the camera, and communicating with the camera person is far easier when one is on the same side of the camera as its operator. With the boom operator's body to the left side of the boom or pole, he or she can follow the scene by turning his or her head left or see the operating side of the camera to the right. The camera person can give an

---

[13]This story was told by the soundman, Nat Boxer, and is from Vincent LoBrutto's book *Stanley Kubrick: A Biography* (Da Capo Press, 1997).

index finger up in the air to tell the boom operator that the boom is too low, saying "move it up" nonverbally, or an index finger down to say it can come in lower (now that's a good camera person who is thinking holistically of picture *and* sound quality). A "slash" in the air with the fingers thrown horizontally means that the boom is at the right height.

A second reason for the orientation of the operator to the left side of the camera is that the boom operator can see the marks on the lens, particularly on those of a zoom lens, and know how wide the camera shot is. If the marks are not prominent enough, white camera tape on the zoom ring of the lens marked with arrows can give the boom operator information about how wide the lens is set, moment by moment. A good boom operator will also be cognizant of the discussion about what focal length is in use on a shot when fixed focal length (prime) lenses are in use, such as 21mm for a wide shot or 150mm for a very tight one. In some instances, a videotape output of the camera may be fed to a lightweight monitor that can be mounted on the boom or fishpole, but note that these rarely show the full viewfinder image for film cameras, only showing what it is in the video-recorded area.

The overhead fishpole generally requires for good performance that the operator be strong and able to hold the fishpole overhead for extended scenes. The handgrip at the end of the fishpole may be grasped firmly in the right hand, whereas a more open, "Y-shaped" left hand can permit the right hand to rotate the fishpole and thus effectively pan the microphone left and right. For this to work, the microphone will typically be tilted to an angle of about 45° from the end of the fishpole. It is less desirable to hold the pole at an angle so that it is tilted up, although physically easier to do, because then the pole may intrude into the corner of the frame, particularly when anamorphic photography with its wider aspect ratio is in use. For extended scenes, in this position it is possible to gently lower the fishpole onto the top of your head to relieve some of the load on your arms.

The degrees of freedom of movement are, then:

- To rotate the body and thus swing the fishpole around a vertical line lying through the operator's body;
- To tilt the fishpole, swinging it up and down, to get the microphone in the closest that it can safely be, considering camera framing and tilts during a shot;
- To pan the microphone left and right by rotating the fishpole with the right hand to turn the microphone left and right;
- To move one's arms up together to change the height of the mic;
- To move on the ground or floor to follow the action; you may find this is helped by using a stance with the knees slightly bent. Needless to say, to do this, quiet shoes are a necessity.

All movement of the fishpole, however, must be done in such a manner as to prevent even the smallest noise, because direct conduction of the motion to the microphone may cause a strong acoustic output. For this reason, shock mounts are necessary, and special limp mic cable is used between the connector at the end of the boom and the microphone.

All in all, good boom operators have an incredibly athletic and intellectual job: they know the scene as well as anyone in it by learning the script pages the night before the shoot so they can anticipate the action, and they collaborate as much as anyone on the set for a good result.

Sometimes scene coverage calls for the use of two or, rarely, even more boom mics. In this case, the third person on the crew will be pressed into a boom operator role, or a set production assistant might be asked to do the same job. This would be the case, for instance, when there are two separate areas of action that are impractical to cover with a single boom.

The photographs in Figs. 4.3 through 4.7 illustrate good and bad uses of the fishpole.

Checklist for Boom/Fishpole Operation

It is essential for all parts of the boom or fishpole to be appropriately tight[14] or loose, so that no small rattles or scraping sounds occur in operation. Because the parts are sometimes very close to the microphone, and direct mechanical conduction as well as air conduction of even small noises may be possible, attention to this factor is important.

All cables must be properly "dressed," arranged so that they have the degree of freedom needed, but constrained from making any sound. In some of the best cases, special cable will be used between the fixed part of the microphone boom/fishpole and the microphone body itself for flexibility. In an earlier age, special cable with a knitted cloth exterior was available for this function, but because each boom needs only about 1 ft of this cable, I have not seen this product available for some time. Some microphone cables of more ordinary construction claim to be super flexible, though.[15]

Planted Microphones

Many times in complex scenes, especially long master shots, it is impractical to get a boom in over the frameline and directed at the actors at all times. Shots may involve the camera going backward through doorways, for instance, and the boom if used would drop into the shot. In these instances, it is common to use planted microphones, fixed in place and hidden by set pieces that can be arranged with

---

[14]For screw-together parts the memory aide is "righty tighty, lefty loosey."

[15]For example, Mogami W2582.

Sound for Film and Television

FIGURE 4.3 Proper use of a microphone on a fishpole, overhead at about 45° above and in front of the actor.

FIGURE 4.4 Much less desirable use of a fishpole below the frameline, and not quite pointed at the mouth. This position will require equalization to get it to sound as good as overhead.

the set dresser. The classic planted microphone was used in the earliest days of the talkies. The microphone was literally housed in a plant on a table between the actors, who could be seen leaning into the plant and speaking loudly and clearly, with some accompanying hilarity.

An example of the use of a boom and a planted mic occurs in *Field of Dreams.* Near the beginning of the movie, Kevin Costner comes into the house. At first we hear him off mic and reverberant, but as he approaches a doorway and wipes his hands on a towel off-screen, we hear him come on mic. Clearly there was a planted microphone behind the doorway, probably above it. Then he enters the kitchen, coming closer to camera, but gets more reverberant as he comes onto the kitchen boom microphone, a reversal

of the normal order in which coming closer should sound less reverberant. Although this is not a good example of the utility of planted microphones, it does show the effect clearly and also some of the pitfalls: the reverb should have been matched in postproduction, but this did not get done, probably because of a lack of time.

The best microphones to use for planting usually are boundary layer microphones, ones that are relatively flat plates with the diaphragm essentially flush with the plate, especially designed to be set on large, flat surfaces. These are based on the pressure-receiver principle (see page 77), which yields a hemispherical pickup. The diaphragms are stretched to a greater tension than in more directional types, just as omnis are, so these microphones are less

FIGURE 4.6 Boom operators must be sensitive to the comings and goings of actors as they run the risk of hitting them in the head with the boom.

susceptible to structure-borne noise and, incidentally, to wind. This makes them a good candidate for use on, say, a desktop that an actor might move objects around on—this type will be less susceptible to direct structure-borne noise than others. They are also well suited to automobile interiors, where they can be mounted on the roof between the actor and the camera and provide the correct perspective. A good example of this type is the Neumann GFM 132. It has a deliberate high-frequency rise in its frequency response, which is good for making distant recordings sound more intimate. In the case in which the rise is noticeable as emphasized "esses" on dialog, simple treble equalization can improve the response. Any boundary layer mic is better the larger the flat surface on which it is mounted.

Directional microphones may also be used up against a barrier, where they will exhibit a sort of "folding over" of their polar pattern, with a directional preference that might be an advantage over a hemispheric type in certain circumstances.

A "poor man's" boundary layer microphone is to use an omnidirectional lavaliere microphone and mount it essentially flush with a surface. This certainly works in a pinch.

Lavaliere Microphones

Body microphones, or lavalieres, are small microphones worn by the actors. They have the advantage of working at a close distance, thus reducing reverberation and acoustic noise, but the angle to the mouth, and the mounting on the chest, lead to disadvantages in the recording of the timbre of the voice. This can be ameliorated with sophisticated equalization, described in Chapter 12 on mixing. However, another problem is that the microphone has a fixed relationship to the actor, not to the camera. This means that if the actor turns upstage, away from the camera, the sound perspective will not change naturally.

In some instances, we accept as a movie convention that we can hear people who sound intimately recorded despite the fact that they are far away in a wide, exterior shot. These recordings could be contemporaneous with the shooting or rather simply dubbed in later. If done while shooting, then the use of a lavaliere is dictated.

For fiction films in which the willing suspension of disbelief means we are not supposed to see the deus ex machina[16] —that is, the machinery of the production— the body microphone has to be concealed. This is usually under one or more layers of clothing that typically absorb

---

[16]In the original sense of the term, the machine that held up the gods in Greek theater.

FIGURE 4.7 Boom operator Gerard Loupias captures the action by being quick on his feet. He became known as the ballet master of boom operators for his dance with the actors. Note his hand positions that allow for boom rotation as well as changes in other dimensions.

high frequencies, making the sound "dull." If it is not too great a high-frequency attenuation, this problem can also be overcome by equalization in postproduction.

Another factor is that the clothing around the microphone may rub against it and cause direct, unnatural, noise. Noise generated by rubbing can be ameliorated through taping together the clothing near the microphone. In this way it cannot rub against either the mic or the nearby skin. Performer's movements could go so far as to hit the chest too, causing tremendous contact noise around the microphone (this happens more often than planned, sometimes during a "take" that was not rehearsed with the breast beating).

Some mounting advice is given in Figs. 4.8–4.12.

FIGURE 4.9 Taping a loop to the chest prevents cable-induced handling noise.

FIGURE 4.8 Wrapping a Band Aid® most of the way around a lavaliere microphone will hold the lavaliere in contact with the skin and not allow it to move.

The microphone output goes to a body pack transmitter, with FM or digital transmission. The radio frequency part of radio microphones is covered in Chapter 6.

As for all recording, the microphone output level depends on the vocal effort of the actor. The range of vocal effort from one actor to the next and from scene to scene with one actor may vary greatly. Because radio transmitters containing mic preamplifiers are usually used with lavaliere microphones, an adjustment may be necessary to accommodate

FIGURE 4.10 The cabling, even under a T-shirt, is not too noticeable.

FIGURE 4.11 However, under cross lighting, the microphone cable becomes obvious. The only way to fix this is to change the wardrobe.
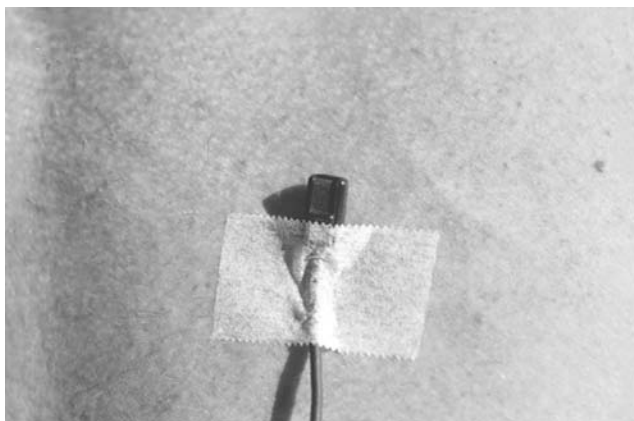
FIGURE 4.12 The side-address type of lavaliere microphone produces a thinner overall package than cylindrical models, and thus it is more easily concealed.

the smaller dynamic range (soft to loud) capability of the transmitter/receiver system compared to the larger dynamic range of the actor. Most models show the appropriate level by way of some kind of metering, which may be just two LEDs, with one lighting for normal level and the other for overly loud speech. The operator adjusts the input level control on the transmitter to light the "signal" LED much of the time, and the "overload" light rarely, with the actor speaking his or her lines at the level of the performance.

Lavalieres worn for documentary and reality programming may be visible or made invisible by the methods used for fiction material. In all cases, it is best to "dress" the wire lead between the microphone and the transmitter inside clothing, because a wire running down the front of a subject is distracting. In the case of television interview shows, note that good practice places the lavaliere microphone on the lapel of a suit on the side facing the moderator, so that when the subject turns toward the moderator they come on mic, rather than going off mic. You can see this by watching a high-quality interview show such as NBC's venerable *Meet the Press:* each of the subjects sitting around the table is wearing a microphone mounted toward the side facing the moderator, and the moderator's microphone is centered on his tie—all good practice.

There are mounting gadgets, such as "vampire clips," that have prongs that stick in clothes, and various other kinds of clips, usually supplied with lavaliere microphones. Other considerations include working with the costume and makeup departments to supply methods for mounting microphones, such as pouches within costumes and shaving chest hair for taping directly to skin. Figure 4.8 shows the use of a BandAid®, for instance.

Using Multiple Microphones

Many practical situations require the use of more than one microphone. An actor may walk out of range of a boom mic to another part of the set, there may be many actors in the scene who must be covered equally, or there may be many sites of activity such that if a single microphone were used at a distance (e.g., matching the camera) it would result in such a loose perspective that no one area of the scene could be well heard.

In fiction filmmaking, doing rehearsals makes practical the use of multiple microphones, combined with judicious mixing, for recording to multiple tracks to cover these situations. When the sound from the various microphones is combined, however, whether in the production sound mixer before the original recording or even when separate edited tracks from the various microphones are ultimately combined in postproduction mixing, several effects may occur that could be unexpected and cause difficulty:

- If one microphone should happen to have been wired *out of phase* (with the two balanced leads reversed) and the pickup of the two microphones overlaps, there will be partial cancellation, particularly in the bass, of the sound.
- Even in cases in which the microphones are wired properly, if the source is not precisely centered between two microphones with overlapping pickup, then the sound will arrive at one microphone before the other. In this case, when the electrical output of the microphones is summed there will be constructive and destructive interference effects, leading alternately to peaks and dips in the frequency response of the sum. This may sound as though the source is being recorded in a barrel, or like Darth Vader, the principal signal processing of which involves a 10-msec delay (the sound is repeated about ¼ frame apart to make his voice sound mechanical).
- An "out-of-time" summation is particularly true of recording with the belt and suspenders approach of capturing a voice on both a boom and a lav. One might want the flexibility in postproduction to use a track from one of them, but adding the two together can result in *some comb filtering* (ripples in the frequency response of the sum) because the arrival time of the sound at the two mics is probably different.

In the case of two redundant podium microphones, these effects are reduced by placing the mics in use very close to one another, thereby reducing the time of arrival differences between the mics. The worst situation possible would be a pair of spaced podium mics wired out of phase, with the performer precisely centered between them. The effect will be nearly complete cancellation in the bass and thin and barrel-like sound above the bass frequencies.

Following are guidelines to minimize the effects of multiple microphones when the outcome is to be reduced to a single channel, as is usually the case:

- Check that absolutely every microphone and cable in the recording system is wired in phase (just putting one cable in line wired backward, with pins 2 and 3 of an XLR cable interchanged, will change the polarity, commonly called *phase*). Polarity checker boxes are available to do this, and it is the cable person's responsibility to see that all the cables have the correct polarity.
- Choose directional microphones and minimize overlap in their coverage.

When multiple microphones are in use, it is important to do everything possible to match the frequency response of each of the mics, so that the timbre remains relatively constant throughout a shot and scene. Using mixed microphone types may lead to identifying a different timbre from an actor as he or she moves from mic to mic and thus give away the secret that multiple mics are in use. Even with perfectly matched microphones there may be large differences in response in a room due to different angles toward the voice and standing waves and their positions relative to the microphones.

If both boom and lav coverage is to be provided, then the use of at least a two-track recorder is dictated so the tracks can be kept separate throughout postproduction. Usually only one will get used at a time, but it may be worthwhile nonetheless to provide both types of coverage so that good sound can be produced throughout a complicated scene. Having the lav on a separate track permits its signal to be equalized so that it better matches the boom, for instance.

Multiple microphones are usually used simply for coverage reasons. In one film production, though, their use had a profound impact on the show. *Nashville,* directed by Robert Altman and with production sound mixer Jim Webb, used multiple radio microphones recorded to seven of eight tracks of a portable multitrack recorder to allow spontaneity and ad libbing on the part of the actors, who walked around wide master shots and interacted with each other in groups. *Nashville* was a fascinating experiment and movie, but the exigencies of cutting not one but seven tracks in the days of mag film made this a one-shot effort.

To summarize, the boom mic is preferred in all instances in which it can cover the scene properly. It has the perspective that matches the picture (lavs do not change sound when the actor turns and faces away from the camera; booms do). Boom mics have better representation of the timbre of the actor's voice, because they are not hampered by a body-mounted position. Still, in many instances lavs solve production problems and can be subtler in documentary situations in which their use is dictated.

Typical Monaural Recording Situations

Production Sound for Feature Films

As we have seen, there is a preference for boom[17] mic recordings over other types, because the radiation pattern of the voice is such that "clearer" sound is achieved for recordings made from overhead and in front of the actors, at about 45° above them. On the boom the preference is for quite directional mics, with the hypercardioid[18] or shotgun types preferred by most practitioners. Some older shotgun types develop their high directivity only at high frequencies, so off-axis sound is dull. In these cases, aiming

---

[17]Equivalently, a fishpole may substitute for a boom.

[18]And supercardioid; in practice, there is very little difference between a hypercardioid and a supercardioid.

FIGURE 4.13 A Schoeps CMIT 5 U shotgun microphone. Photo courtesy of Schalltechnik Dr.-Ing. Schoeps GmbH.

the microphone is especially important. A relatively new design that overcomes this problem is the Schoeps CMIT 5 U shown in Fig. 4.13. Its off-axis frequency response is similar to its on-axis response, albeit attenuated.

Production Shooting and Microphone Technique

Typically a scene will be shot from several angles to offer the postproduction editor a range of choices in developing the scene. In conventional film-style production, this is accomplished with multiple setups, for example, a master shot, a two-shot, and close-ups, all shot separately by having the actors repeat the scene over and over. The set-ups may be shot over a period of hours, as the lighting is adjusted for each new camera angle. The temptation in sound is to match the camera perspective for each new shot. Although this may even make great-sounding dailies (on which the quality of production sound work is judged), later on in postproduction it may become clear that all is not well, because intercutting may not be as smooth as desired.

Let us say that you are the boom operator for a scene from the latest LA cop show. The duty officers are resting on their motorcycles overlooking a southern California freeway, alert for cars making too-rapid lane changes. The scene will be covered by one master and two close-ups, observing the correct screen direction and not "crossing the line."[19] If the scene is recorded simple-mindedly, the microphone will face the two actors in the master shot and each of them in turn in their close-ups. To be able to cut between them, they need to be looking at each other, rather than having both actors look at the freeway. So conventional wisdom is to have one looking toward the freeway and the other toward the first actor, both in the master and, again, in the close-ups. Also, the camera must move to different angles with respect to the actors for the three shots, because that is a part of the "grammar" of filmmaking, the way that we expect cuts to be made.

In the master, the microphone and camera "see" the two men on motorcycles with the freeway in the background. It is important that the freeway be seen in the wide shot

because that establishes where they are, what they are doing, and, from our point of view, why there is so much background noise. A certain ratio is established between the sound levels of the men and of the freeway. In the close-ups, we have to shoot first looking one way, say, away from the freeway, and then looking the other, that is looking toward it. *If the microphone literally follows what the camera sees, there will be lots of freeway noise in the close-up pointing toward the freeway and much less in the shot pointing away.* Although the production sound crew might get away with this at dailies, when the editor puts the scene together the failing will be obvious: Even if the voices of the actors match between the master and the close-up, the varying background noise of the freeway from one shot to another will cause huge continuity problems, probably requiring a looping session, and the unhappy producer will wind up paying for the actors and director to come into an ADR (automated dialog replacement) stage to replace their dialog or else live with an obvious problem.

So a primary rule for production sound recording is: Get the big picture. The cleanest dialog for each of the three shots, combined with more or less constant sound effects of the freeway, will probably make for the most convincing scene.

The problem with recording this scene is that there are two problems being faced simultaneously: microphone perspective and matching the background sound from shot to shot. The microphone perspective is most affected by the distance from the microphone to the actors and is secondarily affected by the angle of the microphone to the performer. The background sound, in this instance, is most affected by the angle of the microphone to the freeway. Because we must aim the microphone at the actors, the best way to mic this scene is to use the overhead boom microphone, keeping the same angle to the freeway in each shot, panning the microphone boom arm to cover each actor rather than rotating the mic at the end of the boom, because that would change the angle with respect to the freeway and consequently the background noise.

In another case, let us say that the camera is in an ante-room off a concert hall where an orchestra is rehearsing. Important action is to take place within the hall, and then the action moves to the foreground space. What is called for is an *audio zoom,* a perspective shift that matches the camera shot, and probably a focus pull of the camera. We first want to hear distinctly what the performers are saying when they are in the background, but we also want

---

[19]The line is a line perpendicular to the camera in the master shot. Using this idea means that when we cut to a close-up of one person speaking to another, the screen direction they face, left to right or right to left, is dictated by the camera not going more than $\pm 90°$ from its starting position, thus the close-ups match the direction established by the master shot.

to know they are in a hall; conversely, we wish to hear the dialog covertly as they move closer to us. There are at least two ways to proceed.

One would be to use either a boom or a planted microphone for the activity in the hall and a separate boom for the foreground in the anteroom, judiciously (i.e., sneakily so that it is unnoticeable) cross-fading from one microphone setup to the other. The mic in the hall can be placed to record appropriate reverberation, or a second reverb mic can be used such that by adjusting the relative levels of the "direct" mic and the reverb mic a proper balance can be struck. It is best if all three mics (direct hall mic, reverb hall mic, direct anteroom boom mic) be recorded on three separate tracks for best editorial flexibility. Some setups allow recording each mic separately, and a contemporaneous mix be made for use at dailies and throughout picture editorial, but making available the individual tracks when it comes to sound editing.

> One way to mic the reverberant-field component of the sound field, while picking up practically no direct sound, is to use a directional microphone for the reverb mic, with its null pointed toward the source of the direct sound. Thus the back side of a cardioid, or the side of a figure-8 pattern mic, can be used. One problem with this technique is that the cancellation of the direct sound is often highly variable with respect to frequency, so you may wind up hearing low bass and high treble of the direct sound in the reverberant-field mic. Experimentation with different microphones is definitely needed for this to work, but it is potentially a fine technique.

Another way to approach this same scene is to just use a radio mic on the performer who must move and leave the changes in perspective and reverberation to postproduction mixing. One difficulty with this is that the dailies will sound a long way from the final product, yet by proceeding this way it is known that high-quality direct sound is available for the scene. The biggest problem with this approach is the lack of natural change when the actor turns his or her head; for example, when turning around and facing away from the camera the microphone perspective stays the same, but the picture perspective changes.

Other Sources for Dialog

There are three sources available to sound editors for dialog besides the "circled takes" shown at dailies. The circled takes are the ones deemed by the director, after the on-set ritual of checking the camera gate for hair or dirt, as takes to be seen at dailies. These takes are used for picture editing and consequent sound, but once heard over good monitoring may not be of high enough quality for use in a finished soundtrack. Note that "It sounded all right on the Avid" is one of the most common frustrations heard about sound nowadays, because "the Avid" usually means it sounded OK over a 4-inch speaker in an acoustically bad space with a lot of computer noise, and when heard properly, mismatches of the actor's timbre

and the background presence from shot to shot are obvious.

These ideas have an effect all the way back to production sound, because after principal photography the actors get busy on other projects or even disappear and are often not available to do an ADR session. So there are a total of four types of dialog recording for use by sound editors:

- Principal photography takes chosen for dailies.
- Outtakes. These are an important source for sound editors, who will troll all the recorded material to fix problems with actor's performances, background noise, etc.
- Wild lines. These are typically recorded on the set, perhaps at the end of the day of shooting or sometimes along with the scenes being shot. They are recorded in the same acoustic setting, but are likely to have the microphone closer and involve more noise control. For instance, if wind machines are in use on a shot, the sound is likely to be unusable. Killing the wind machines and having the actor speak (or shout!) the lines the same way as in the camera shot may prevent the need for a subsequent ADR session.
- ADR with the actor coming in for a special recording session may be necessary if none of the above work. See the section on ADR recording on page 68 about acting issues, which also applies to wild lines.

Production Sound for Documentary

Although many of the foregoing considerations apply to documentary filmmaking, there is one startling difference. You almost always get only one crack at a scene. Boom operation becomes even more difficult than for fiction films, because rehearsal is forbidden by circumstance. In this case, a sixth sense is necessary to anticipate what may happen among subjects in a scene. The boom operator must be very alert to what the camera operator is doing, too, and anticipate where the next camera move may be. A primary factor in the ability to do these tasks is listening carefully to what the subjects are saying, while still managing to get all the technical details right. This includes, usually above all, keeping the microphone out of the shot and from casting shadows visible to the camera, just as in fiction filmmaking, while getting it in close enough to get good sound.

That having been said, production sound for documentaries also involves usually doing two jobs at once: boom operation and production sound mixing. The mixing parts of the job are covered in Chapter 7. A special type of fishpole meant for documentary and electronic news gathering applications allows one person to both operate the fishpole and a portable production sound mixer. Its bottom end is a mushroom-shaped elastic piece that can go up against your belt, in a break from the method used in fiction filmmaking. It is shown in Fig. 4.14.

FIGURE 4.14   An articulated fishpole helps in being able to both boom and mix at the same time. Photo courtesy of K-Tek® USA, M. Klemme Technology Corp.

It is a trend in documentary recording to employ radio mics on the principal characters, and some scenes might get coverage only on radio mics. A problem with this is that the radio mic output contains little location sound, so can sound overly dry. For most cameras in which the recording occurs (single-system recording), only two tracks are available. In this case, it is often best to put primary tracks, say two lavaliere microphones on an interviewer and an interviewee, on track 1 and a boom mic, generally on the interviewee, on track 2. Track 1 is the one usually employed in editing, but track 2 can be used as a backup and may possibly be valuable in mixing to add some room sound. The reason for this separation is that the lavs are likely to need similar equalization in postproduction, separate from the boom.

Some formats of cameras record four audio channels. In this case, more separation between mics and channels can occur, with up to three radio mics and a boom on the four channels. This allows greater control in postproduction.

Reality Recording

Reality television employs all of the techniques described for feature filmmaking. What distinguishes it is the large number of cameras, recorders, and microphones in use. Shooting ratios, the ratio of original camera hours of videotape to the finished show, are usually enormous. Planted microphones are used as almost "spy" mics on the subjects. Often the original sound is of poor quality because the number of recordings being made simultaneously is too

great to monitor; thus a great deal of work has to be done in postproduction to make the sound acceptable.

Although it happens in other types of show making, there is one thing that comes up quite a lot in reality television: the hug, performed with radio mics on the subjects. If talking while hugging, the sound gets muffled and noisy. With automated equalization in postproduction it is sometimes possible to turn up the treble just at the moment of the hug and ameliorate that problem. Still, we live with the clothing rustle and even mic hits that the hug causes. For this reason, a "belt and suspenders" approach of recording the radio mics and a boom, all to separate recorder tracks, is very useful. The boom mic sound can be substituted through mixing at the moment of the hug and prevent large noises.

Narration or Voice-over

Narration or voice-over recording usually occurs in a purpose-built studio offering low reverberation time and background noise and relative freedom from room standing-wave effects. Direct-conducted vibration can be eliminated with a shock mount and by isolating the microphone mounting from anything the performers might touch. Breath noise can be eliminated with a windscreen and/or a silk disc.

One common source of a problem in narration recording is the script stand. If it is extensive and massive, it will reflect sound well and the microphone will receive first the direct sound and then a strong reflection off the surface, causing constructive and destructive interference. It is better to use a fold-up music stand and loose script pages to minimize the reflection off these surfaces.

The choice of the best microphone will often depend on the precise interaction between an individual voice and the range of microphones at hand. I have tried different types with one voice and have found the best mic for that voice, only to find it not as good with another voice, for which a different choice is needed. Thus, the absolute quality of the microphones is not in question, but rather the "match" between the timbre of the voice and the response details of the microphone that matters. If this were not the case, one microphone type would have come to dominate all recording, probably the flattest one. To be able to make such judgments, though, one first must know with great certainty what the conditions of monitoring are because, for instance, a dull monitor will lead to the choice of a bright microphone.

Once a voice and a microphone are chosen, then the details of the working distance can be managed. Given the quiet environment already specified, there is little trade-off if the microphone is used at 1 or 2 ft, except for the proximity effect varying the response.[20] Another

---

[20]See page 79.

factor is whether the performer is seated or standing, with some attendant changes in voice due to the shape of the chest cavity in these states. Standing usually sounds better than sitting.

Microphone placement for narration recording is usually best straight in front and a little above the horizontal plane, rather than below a horizontal plane that includes the head. This is because of the radiation pattern of the voice, which is generally "chesty" sounding below the horizontal plane. Placing the microphone somewhat up in the air helps to reduce the likelihood of "pops" caused by direct breath rattling the microphone diaphragm.

> You can experiment with pops by baring your arm and placing it vertically, palm toward you, and with your open hand about 8 inches in front of your mouth. While standing, pronounce "P" and "T," in the sentence "John Kennedy piloted a P.T. boat" and you will feel the puffs of wind on your hand (for the P) and a little lower (for the T). Thus, straight in front and below the axis of your mouth are both problematic positions for microphones.

ADR Stage Recording

Automated dialog replacement, or *looping,* stages have many of the same considerations as narration recording studios, with several added requirements. The first is that the reverberation time, which may be almost arbitrarily short in a narration setting, cannot be so short in an ADR stage. Rooms that are near anechoic are difficult for actors to work in, because they do not hear enough energy coming back from the room and tend, as a result, to force or stress their voices, which shows up as timbre change. In narration and some ADR recording, it is commonplace for the performers to wear headphones so that they can get back all the level desirable by simply turning up the headphones (but watch out: loud headphones can leak sound, which gets to the microphone, changing the source timbre by effectively adding a reflection at the time of the spacing of the headphones and microphone; in the worst possible case, this could potentially even lead to acoustic feedback, for which the British offer the more colorful term, *howl round*).

The upper limit on reverberation time[21] is set by the consideration that we do not wish to impose audible reverberation on the recording at all and would prefer a perfectly dead recording, which we could then liven up to taste in postproduction. A typical compromise reverberation time is 0.4 sec for reasonable room sizes, flat with frequency, and containing little in the way of discrete reflections, especially those aimed at the microphone. This, combined with a low background noise level, will make for adequate conditions for ADR stages.

Otherwise, the conditions are similar to those for narration recording, except that ADR recording may deliberately use greater microphone distances from the actor to try to match perspective to that of the production sound recording. Microphones on boom stands are usually employed—sometimes more than one—at varying distances from the actor recording to different channels of a medium to provide several perspective choices later on in postproduction.

Another factor that may differ from narration recording is the dynamic range of the actor. An actor may scream or whisper, and the recordist should be prepared with pads and low-noise preamplification to accommodate the full range of a performance. Capturing the full dynamic range is discussed in the next chapter.

A performance disadvantage of ADR recording is that the actor faces a dead and quiet room with little acoustic response compared to what is happening on screen. In these circumstances it is commonplace for actors to underplay their performance, only to be found later as too low in energy to "read" through the rest of the sound. Hearing Harrison Ford's ADR only in the rolling boulder scene from *Raiders of the Lost Ark* illustrates this. His performance seems "over the top," but he and Steven Spielberg, the director, know that in the final mix there are going to be many more sound elements present that would bury a subtle performance. By the way, Harrison performs all his own vocal "efforts," those grunts and groans that accompany activities on the screen.

Failure to match microphone perspective between production sound and recordings made on a looping or ADR stage is one of the most telltale differences that may prevent the two from sounding alike. This means that the sound mixer for an ADR session should have access to the studio to move the microphone to get the right perspective, or multiple mics may be used, spaced closer to and farther from the actor to accomplish the same thing. In this case, only one mic's output would be used at a time, to match the perspective of the shot.

There are basically no postproduction sound techniques that can change the stress in an actor's voice. Postproduction can enhance intelligibility and smooth out the level variations, and so forth, but no technical process available can change the underlying performance, so it must be delivered by the actor.

Excessive "lip smack" in ADR and narration recordings can sometimes be ameliorated by the actor taking a few bites of a green apple or spraying water throughout the mouth to moisten it.

The ADR stage equipment shuttles the picture according to instructions given it, usually by an assistant picture editor, prepares for recording a segment (automatically goes into and out of record), and alerts the actor to the start of the line with three beeps, with the start of the performance expected where the fourth beep would go— beep, beep, beep, "now. . . ."

---

[21]The time for an abruptly stopped sound to fade away to inaudibility.

A Hollywood story is told of experienced sound people leading on a young person. A neophyte sound editor was given the task of working on an ADR stage with Henry Fonda, logging the takes and so forth. In those days, physical loops of film had to be cut, and the name "looping" was really true, and it was a slow, tedious process. The projectionist would put up a loop, and Mr. Fonda would look at it over and over and over. The neophyte began to think that the crew would be at this for weeks. Older, more experienced editors only smiled. For when Henry Fonda did "Take 1," not only was it perfectly acted, but it was also perfectly in sync! The older guys would watch for the younger one to heave a sigh of relief.

Another useful Hollywood story is that of soundman James G. Stewart at work on ADR for Orson Welles' *The Magnificent Ambersons*. An exterior day winter scene was actually shot in a cold meat packing warehouse in downtown Los Angeles, to get the cold necessary to see the actors' breath in the air. With the compressors running to keep the cold, there was no getting usable production sound. The actors rode along in an early open motorcar, and upon ADR-ing them, and adding car noise, the sound seemed flat—the actors didn't seem in the scene—to Wells. So Stewart, not to be confused with the actor Jimmy Stewart, got a carpenter's horse, a big board, and some hefty grips to form a teeter–totter, with the actors on one end and the grips on the other to shake them in sync with the motion in the picture. Each of the six actors was recorded separately with the shakes. It worked.

A variation of ADR stage recording is *Walla* recording. Performed by usually anonymous actors, walla is a substitute for the background crowd sound of a scene. For instance, the background crowd in a restaurant would mimic speech while the principal actors play out in the foreground of the shot. To substitute controllable crowd sound, finely tuned to the scene and not just a library effect, walla will be recorded.

Foley Stage Recording

Recording on a Foley stage bears a lot of resemblance to ADR stage recording, but with some added considerations. Foley sound effects, by their nature, may include everything from a quiet clothes rustle to gunshots. Thus, there is a premium placed on all factors that affect the dynamic range, including the background noise of the room, microphone and preamplifier noise, and microphone and preamplifier headroom.

Second, the reverberation time considerations of the conventional ADR stage are relaxed because there is no need to support actors speaking, and any added reverberation is generally undesirable, as it is preferable to add reverberation in mixing. Thus, Foley stages are usually very low reverberation time spaces, although at least one studio prefers some reverberation in its recordings so that none needs

to be added in mixing. The problem with this is that all the Foley recordings then take on the sense that they are in the same space, when in fact it is better to match the scene.

The lowest noise level microphones can be those with the largest diaphragms, all other things being equal. Although large-diaphragm microphones are not generally associated with the flattest response on axis, and certainly with less uniform control over directionality with frequency than smaller types, Foley stages are a place where the trade-off in favor of low noise will often be made for two reasons:

- Capturing the exact timbre of the source is not as important in sound-effects recording as capturing the much more familiar sound of a voice. Voice and music timbre are of supreme importance because of our everyday familiarity with them, with sound-effects timbre running well behind these in importance.
- There is just not much off-axis sound for many Foley stage sources. Remember how very dead the stages are, making the off-axis response of the microphone unimportant.

Given these considerations, a microphone such as the Neumann TLM-103 becomes valuable. A large-diaphragm microphone, it has a very large dynamic range, from a noise floor corresponding to 7 dB SPL A weighted to an undistorted sound pressure level of 138 dB, a remarkable range. Decibels and sound pressure levels are discussed in Chapter 1 and Appendix I. The 131 dB range of the microphone corresponds to a ratio of the largest sound pressure to the noise floor of 3.5 million to 1!

Typical Problems in Original Recordings

There are some problems that are frequently heard and that can be avoided with some care. These are often audible on television news shows, because of their lack of postproduction time; somewhat less noticeable on made-for-television productions, as the worst problems are probably eliminated in postproduction; and hardly audible at all in the best theatrical films, for which, if the sound is not good, there has been the time and money available to loop the scene. In any of these productions, though, these situations do cause problems; if not audible problems, then ones associated with spending time and money to fix them. Following are some typical problems:

- The use of lavaliere microphones on an agitated subject who is moving around. The worst example seen was the morning news show that put a lavaliere microphone on a new mother holding her baby, who of course turned to her mother, spotted the shiny microphone, and took it for a toy to play with, causing loud noises.
- News stories reported by on-the-scene reporters in which the "stand-up" part, the on-camera narration, is

mic'd with a lavaliere mic, but the voice-over is mic'd with a hand-held microphone. As soon as the editor cuts to the voice-over, the sound is clearer because of the microphone quality and location compared with the lavaliere.

- Teamed news reporters sitting side by side, but with their microphones wired out of phase. When the two mics are live, the bass is attenuated so that the sound is thin and low in level, but as soon as one microphone is faded out, the remaining mic sounds much better. This occurs because at low frequencies the microphones are close together compared with the wavelength of sound, so wiring them out of phase and adding the two together in the mix will cause bass cancellation.

- Low-frequency noise from moving the microphone or from wind, which can be limited or eliminated by the choice of microphone (omnis are better than directional mics, all other things being equal), shock mounting, wind screening, and filtering, defined in Chapter 5.

- Simply too much noise and/or reverberation in the environment for good recording. Location scouting should include a sound person, but because of budget constraints it is rare if there is travel involved. However, it is useful to bring along a sound pressure level meter to show people just how loud the sound is. People use hearing tactics such as binaural discrimination[22] when standing under the freeway listening to you, but the microphone lacks this ability. Sometimes pointing at a sound level meter that shows levels around those of speech can set the question to rest. "But it looks just right" does not make it right for shooting. There are far too many examples of shooting in which the background looks great, but you have to remind directors and location scouts that sound cannot pan off a noise source like the camera can pan off an unsuitable object.

Microphone Technique of Singers

Watch a good vocalist who has a lot of experience with microphones and you will notice a few techniques that he or she employs with hand-held microphones to improve the natural pickup of sound. Ella Fitzgerald was brilliant at this, although younger performers may be more used to standing in a studio with a microphone on a fixed boom in front of them.

- Sing across the end of the mic, not directly into it, to avoid popping the diaphragm. The effectiveness of this technique depends on the exact microphone and its

external and internal windscreens. Some microphones may benefit, whereas others do not, from this technique.
- Adjust the working distance to account for the proximity effect and loudness.
- Dynamically move the microphone closer and farther away to help crescendos and diminuendos.

MICROPHONE TECHNIQUE—STEREO

Background

Up to this point, although multiple microphones might be in use, their purpose has been to provide coverage of the action, and the reproduction is usually from a single channel centered on the picture. Recording could be on multiple channels of a storage medium, yet we don't say that the channels are stereophonic because the relationship among the channels has not been constructed in that way. The word "stereo" is derived from the Greek for solid. It means that the spatial characteristics are meant to be reproduced: the sense of location and potentially size of a source. For film sound, it is usually expressed as a 5.1-channel sound system, described below. For television sound, although 5.1 is in widespread use for dramatic programs, more ordinary programming is often two-channel left–right stereo. News programming is still most often monaural.

Using stereo produces two effects that cannot be obtained with monaural systems. The first of these is spatial localization at more than one place. Disbursed localization of auditory objects helps to separate objects psychoacoustically, as described in Chapter 2. Thus, all other things being equal, in stereo a more complex soundtrack can be separable into its component objects by listeners. The second effect is spaciousness, the capability to reproduce diffuse sound fields such as reverberation in a manner that is spatialized, that is, not occupying a single point in space.

A monaural system produces localization, but it is all at the one point where the loudspeaker is, centered on the picture in film applications. It can also produce one dimension associated with spaciousness, the depth dimension, principally by adjusting the direct-to-reverberant sound ratio. In this way, along with others, the voice-over in the beginning scenes of *Apocalypse Now* is distinguished as a separate object from the on-screen voice of the same actor—the amount of reverberation is essentially nonexistent for the voice-over, but we hear the room acoustics of the hotel stairs and room in the production soundtrack as the officers arrive to roust the protagonist out of bed.

Multichannel stereo offers the potential to localize sound on the screen coincident with the picture of the object that is expected to be making the sound, left to

---

[22]Explained in Chapter 2.

right, heightening reality. Also, surround sound helps to engage the listener by enveloping him or her in the action. In film sound, the release format standard today is typically 5.1 channels,[23] not the two channels that most people think of when they think of stereo. The 5.1 channels are:

- Three front channels coincident with the left, center, and right of the picture;
- Two surround channels, left and right, arranged around the audience seating area, sometimes supplemented with a separation into left, rear, and right components;
- A low-frequency, and thus low-bandwidth, enhancement channel (the $\frac{1}{10}$ of a channel).

At its broad introduction to film sound in the 1950s,[24] stereo had three screen channels and one auditorium "effects" channel, but when introduced into the home, the number was reduced to two, left and right, because that is all that the geometry of the phonograph record groove could store. Thus we came to know "stereo" as two channels in widespread parlance, but in film sound we always mean a minimum of four, left, center, right, and surround, with 5.1 the current most widely practiced standard.

Techniques

There are four principal methods of multichannel stereophonic recording:

- *Spaced microphones,* usually omnidirectional, spread across the source, with spacing from left to right depending on the size of the source, but never so large that a discrete echo can be formed wherein a source to, say, the left of the scene is heard first from the left and then from the right as an echo. The limit on the size depends on the size of the source and reflections within the environment, but as a rule of thumb I have used up to 60 ft across a five-mic array from left to right without problems in the Los Angeles Coliseum recording the University of Southern California marching band. This method depends on time of arrival and level differences among the microphones. In scoring stages, added room microphones fill out the array for 5.1 recording of orchestras. Another application for spaced omnis is for the crowd noise at sporting events.
- *Coincident or near-coincident directional microphones,* which use the polar patterns of closely spaced microphones to distinguish among the channels. This method

depends mostly on level differences between the microphones because the time differences are minimized by their close spacing. There are a number of such methods.[25]

- *Dummy head recording,* also called *binaural recording,* using microphones placed at the position of the ear canals in an anatomically correct head. This method, when heard over headphones, may be quite spectacular in spatial reproduction of sound from many directions. However, sounds that originated from center front often suffer from "in head" localization. This is the effect you hear from headphones when a voice in a recording seems centered between your two ears. Transforming dummy head recordings into recordings suitable for loudspeakers has proved to be an intractable problem, and no method for doing so has yet made dummy head recordings mainstream.
- *Close mic'ing,* usually done with directional mics, to provide good isolation among the microphone channels. This method is also popular for scoring, because it minimizes recording time and allows for a great deal of control in postproduction. The difference between this type of mic'ing and what one would do for multichannel mono is that these mics are intended to be panned into a position in the stereophonic sound field in postproduction mixing.

The spaced microphone approach was used in production sound recording experimentally at the beginning of the stereo era but dropped out of use because of editorial problems, as well as the clumsiness of the multi-mic boom. A film that was recorded in this manner is *Oklahoma!,* winner of the sound Oscar in its time.

Spaced microphones have some disadvantages when it comes to the accuracy of timbre and sound imaging because the spacing brings about large amounts of time delay among the microphone channels. There are two problems:

- If the outputs of the channels, spread out in time, are ever combined, such as being mixed together to produce a monaural mix, then the time delay gives rise to constructive and destructive interference, frequency by frequency, which is audible and is called comb filtering. Thus the *mono compatibility* of stereophonic spaced microphone recordings is not very good.
- Even without combining the channels electrically, there is a question regarding constructive and destructive interference because the "triangulation" of the source through the channels to the ears creates time-of-arrival differences.

This still remains a popular technique for stereophonic recording, although it is not often applied in mainstream

---

[23]Digital optical releases are 5.1-channel or a quasi-6.1-channel format, and one system offers 7.1 channels. The analog system is used as a backup for the digital and in secondary theaters as primary. It has four channels, LCRS.

[24]An earlier one-film experiment for *Fantasia* used 5 channels.

[25]These are covered in the author's Focal Press book: *Surround Sound Up and Running.*

film and video production sound because of its complexity. However, it is the most popular in film scoring applications.

For film sound, spaced microphones are used mostly in music recording, with some use for stereophonic sound effects. More often perhaps, coincident microphone techniques are used for sound effect recordings, for their convenience and their special properties. Convenience is obvious; there is only one principal pickup location, although there are several microphones present at that location, so it is easy to pick up and move the whole stereophonic microphone array, which is much harder to do with spaced microphones.

One technique in particular that is useful is called MS for "mid-side." This technique uses two coincident microphones, with a forward-facing cardioid or hypercardioid and a side-facing figure-8 pattern mic. The combined microphone is positioned facing the source, such that it is in the middle of the pickup pattern of the cardioid (or hypercardioid). The source is also centered on the null of the figure-8 pattern mic, such that it picks up practically no direct sound. Knowing that the two halves of a figure-8 pattern mic are different in polarity, described in Chapter 5, allows one to understand the following operation: Summing the output of the two elements together results in a left-facing channel (if the figure-8 pattern mic in-phase lobe is left facing), and taking the difference results in a right-facing channel. This is a method that is compatible with mono mixdown, because if you mix the two resulting signals together in a 1:1 proportion, you get back to just the forward-facing microphone's output. Second, it does not suffer from comb filtering in reproduction because there is no time-of-arrival difference between left and right.

> This makes it very useful in systems that employ an amplitude-phase matrix for release (Dolby stereo, Ultra stereo) and thus is a preferred method for sound-effects recording.

Although conventional MS works well for two-channel reproduction, it cannot resolve all the directions of multi-channel sound. Double MS, with one system forward facing and the other rearward facing, resolves this problem. A setup for it employing a short shotgun for the middle mic, a bidirectional for the side-facing mic, and a cardioid for the rear-facing mic, is shown in Fig. 4.15.

MS stereo recording was tried on the principal photography of a picture some years ago and wound up not being used. The sound recordist reported that the dailies sounded great but the editing problems of stereo soundtracks discussed earlier caused the production soundtrack not to be usable in stereo. So the M channel was used, just as in conventional work.

Today we might think of an MS microphone as recording direct sound and ambience with its two elements to two recorded tracks, with no "decoding" into left–right space.
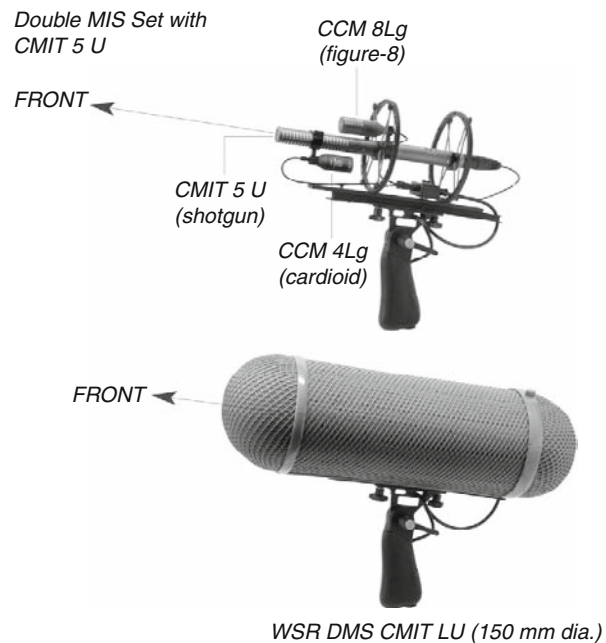
FIGURE 4.15 A Schoeps setup for double MS. Photo courtesy of Schalltechnik. Dr.-Ing. Schoeps GmbH.

Thought of in this way it may well be suited to production sound recordings. In fact, a classic short shotgun mic from Sennheiser is now available as an updated model with an added side-firing bidirectional mic, as shown in Fig. 4.16.

So, although the field of stereo recording is large—offering many techniques—certain methods dominate in specific areas of filmmaking: spaced omnis in music recording and coincident mic'ing for sound effects. It should be pointed out that many monaural sound recordings are combined into complex stereo recordings in postproduction, with the spatialization created by stereophonic reverberation of the original monaural source.

FIGURE 4.16 This Sennheiser MKH418 combines both an interference tube shotgun receiver and a bidirectional receiver in one mic body. Photo courtesy of Sennheiser electronic GmbH & Co. KG.

Although this does not satisfy a purist as to how stereo should be done, it is the dominant method used today. The monaural recorded parts of a soundtrack are usually the dialog, both production and ADR, Foley sound-effects recordings, and some "spot" effects. Other effects, especially ambience, and music are most often original stereo (including more than two channel) recordings.

Scoring

Scoring sessions occur on specially equipped scoring stages, or sometimes in the venue in which an orchestra routinely plays. Multitrack recordings, with stereo techniques that tend toward the spaced omni approach with added spot microphones, are typical. Spot microphones are usually directional ones, such as cardioids, located near the sources, so that they can be separately addressed in postmixing. The differences from large-scale orchestral recording include the ability for the conductor to follow a projected picture or at least a video monitor. The video has inserted in it a "streamer," a left-to-right line across the picture that hits the right side of the screen when a particular cue, or point in a cue, is to happen. Also distinguishing scoring stages from ordinary orchestral recording are various headphone feeds, needed to maintain strict synchronization.

Event Sound

Sports and live television variety and awards shows are special categories unto themselves, but they generally employ a mixture of the methods described above. Some of these are among the most complex productions that ever occur, because of the live nature and complexity of the event. Producing audio for the Academy Awards for instance is a very sophisticated operation, involving four people mixing in four different environments (one for house sound located in the auditorium, one for feedback to the stage and cueing located above and to one side of the orchestra pit, one for orchestra mixing in its own truck, and one for the final show mixing in another truck[26]) and a host of support people, including one that manages the radio microphone frequencies to minimize the likelihood of interference when so many radio mics are in use.

Although the production of programs such as the Academy Awards is extremely complicated, the microphone techniques employed may be broken down into sets of the methods described above, layered over one another for the final presentation:

- The main podium has two microphones, located one above the other to minimize visual impact, one a cardioid and the other a hypercardioid. The cardioid is used for groups

at the podium, and the hypercardioid is used for single presenters.
- The orchestra pickup uses a very large number of close mics, practically one for each player, as well as some that pick up a little wider perspective. Most of these microphones are cardioids. The pickup by a great many microphones permits balancing in mixing rather than acoustically, which is harder to control.
- There are house microphones dead hung over the audience area used both for front and for surround channels.
- The on-stage production numbers involve some studio prerecording such as for a chorus that would otherwise be hard to pick up well in a show in which the scenes are constantly changing. However, principal singers who are subject to close-ups on camera employ hand-held live radio mics.

RECOMMENDATIONS

Some general recommendations for microphones can be given for various recording situations. The table titled "Microphone Applications" at http://booksite.focalpress.com/Holman/SoundFilmTV/ gives these recommendations. Note, however, that your own experience may vary, and owing to my own limited experience the list is doubtlessly incomplete. For those who find these solutions expensive, I sympathize. I have tried some inexpensive alternatives and have usually been disappointed. I would like to point out, however, that the microphone supplied with the Canon XL-1 camcorder was surprisingly good, beyond expectations, so it is not always true that inexpensive microphones cannot work well.

For the best performance in wind, try the following in order:

- For directional mics employ a basket- or zeppelin-style windscreen, trapping a volume of air between the windscreen and the microphone, the larger the better.
- Place a fuzzy fur-like device over the windscreen.
- Use an omnidirectional mic if possible, which will have to be used closer for the same ratio of direct sound to reverberation and noise.
- Use a large foam-type windscreen on the omnidirectional mic.
- Use a pad between the microphone capsule and the microphone electronics, either by activating a built-in switch or by screwing in a pad between the capsule and the electronics. This applies to both directional and omnidirectional mics, although it is of potentially more use for the directional microphone, which is inherently more affected by wind noise. The pad reduces the level into the following preamplifier and prevents its overload. This is described fully in Chapter 5.

---

[26]I am indebted to mixer Ed Greene for inviting me to several events to observe at close hand all the complexity of live awards shows.

- Use a high-pass (low-cut) filter between the capsule and the microphone electronics, possibly in addition to a pad. This applies to both directional and omnidirectional mics.

MICROPHONE DAMAGE

Microphones must respond to extraordinarily small movements of air. The threshold of hearing corresponds to vibrations that are the size of one-third of the orbit of a single electron in the smallest atom, hydrogen, and microphones approach this sensitivity. On the other hand, they might be asked to respond without significant distortion, such as "bottoming" the diaphragm, to a level 165 dB greater in a gunshot mic'd at about 3 ft. That's a ratio of pressure of 178 million times from the smallest to the largest level, although no one microphone can handle this full range. In fact, some microphones, including some expensive electrostatic ones, can be damaged by such high levels, whereas some cheaper electrodynamic ones may work well. Some specialized types are rated for such high levels.[27] It is worth noting that rental houses in Los Angeles receive expensive conventional microphones damaged by gunfire with some regularity. In cinematography there is a concept called a "sacrifice camera," an inexpensive one that may be run over by a thundering herd of buffalo, for instance. Sound people might do well to adopt the same approach and have available cheap mics for situations that include very high levels.

WORLDIZED AND FUTZED RECORDING

Rerecording sound through a deliberately poor "channel," consisting of a loudspeaker and microphone, is called *worldizing*. It is not high quality that one seeks in such recordings, but rather, quite often, the opposite. Experimentation is the order of the day for worldizing. For the high-school dance scenes in *American Graffiti,* sound designer Walter Murch and director George Lucas rerecorded 45-rpm dance records of the 1950s through a loudspeaker and a microphone in a real space—a suburban back yard—and picked up and moved the two around during the recording so one got the swirling effect of constantly changing sound quality, enhancing the feeling of being there. During rerecording, both the "clean" track and the worldized one were edited so they could be mixed together or cross-faded between.

The use of a public address system almost always employs a film sound icon, which is that all sound systems *must* feed back just before use, telling the audience a sound system is in use. Feedback is very hard to control in production, so the squeal may well be added in postproduction, although a moment must be left for it by the actor. This icon is so embedded in the consciousness of patrons that what is remarkable about the opening lecture in *Jurassic Park III* is that the sound system does *not* feed back.

Futzed recordings are those made through a deliberately bad system so that an effect is added. Although this may be done electronically, a simple open-frame loudspeaker and low-quality mic located in a box, usually lined with absorbing material, and fed from an amplifier and rerecorded through the path, are often more convincing. A possible use for such recordings is in telephone conversations. Years ago, simply restricting the original sound to the midrange frequencies by means of an electrical filter was enough to cause the thin nasal quality called the *telephone effect*. With today's better transmission channel, simply filtering the sound is not enough, and futzed sound is commonly used to add nonlinear distortion as well. There is a correct "grammatical" method for portraying a telephone conversation in a film.

- When we see the person speaking, we hear him or her naturally, but if we hear the other end of the conversation at all, we hear it highly modified, by being filtered to a narrow frequency region and rerecorded through a deliberately poor set of transducers.
- When we cut from one end of a conversation to the other, the roles reverse with regard to sound.
- If we see first one person, then the other, we follow these rules at the picture edits, but, if we then see both of them (by way of a split-screen optical), we hear both of them "direct."

Some films deliberately play with the sensibility of whether the person at the far end of a phone conversation is heard at all, heard through the intermediary telephone, or heard directly, and they may change the perspective over the course of the film. In *The Deep End* (2001) a blackmailer played by Goran Visnjic makes phone calls to a mother played by Tilda Swinton about her son's involvement in a murder. Early in the film, telephone filtering is used, but as the film progresses and the threat becomes more real, less and less filtration is used, until the blackmailer is projected into the home of the mother despite speaking over the phone, as reported by rerecording mixer Mark Berger.

OTHER TELEPHONE RECORDINGS

For *All the President's Men,* production sound mixer Jim Webb developed a system for phone calls that encourages

---

[27]Some may be found with the advanced search engine on www.microphone-data.com. For instance, the Josephson Engineering C550F is rated to 170 dB SPL, but note that it also has a high noise floor of 82 dBA, as is doubtlessly necessitated by the design necessary to reach such high levels.

better performances from the actors and allows for dialog overlaps. The usual method has the script supervisor read the lines that will be heard from the far end of the phone line, to provide pacing for the actor. No overlapping of dialog is permitted, because editing it out must eliminate the script supervisor's voice, and then the far-end actor's performance is substituted. This artificial condition can lead to stiff performances and limits the far-end actor, who must fill in the precise time allowed in an ADR session after the fact.

The steps in the method used for *All the President's Men* are:

- Record the actor shown on camera speaking into his telephone with the boom mic on one track of a two-track (or more) recorder, but defeat the microphone transmitter of the on-camera telephone so that its signal is not present on the telephone line. On the other hand, make the telephone receiver earpiece functional, allowing the on-screen actor to hear the far-end actor.
- Place the far-end actor in an acoustically dead, quiet space isolated from the set.
- Have the far-end actor speak into a telephone.
- Arrange a headphone system for the far-end actor, with the output of the boom mic available on it so that he or she can hear the on-camera actor. Mix into that feed the far-end actor's own voice from the telephone system so that it sounds natural to him or her.
- Connect the telephone line to the second track of the recorder.

Using this method improves the acting of the scene, as now the actor has the actual performance from the opposite side of the phone to respond to. It has an effect principally on the dynamics of the scene, even allowing for dialog overlaps, which are otherwise something best constructed in sound editing.

This page intentionally left blank

Microphone Technicalities

PRESSURE MICROPHONES

If asked to make a microphone, one way to proceed would be to copy the way the ear works: Stretch a diaphragm over a sealed chamber (providing a leak, like the Eustachian tube, for pressure equalization with barometric pressure change) and measure the displacement of the diaphragm caused by the sound, converting its vibration into voltage by various means. This is the way the simplest microphones work. Compared with the ear, though, there need to be several "improvements." One is that the ear canal provides a resonant tube in front of the eardrum, increasing the sensitivity in just one frequency range. Because we usually want our microphone to keep timbre constant, we typically want a flat frequency response, so the tube will have to go: The diaphragm should be exposed as much as possible to the sound field. (Some compromise is usually made here, and the diaphragm is slightly recessed to prevent damage.) A microphone constructed this way is called a *pressure microphone,* because the motion of the diaphragm, and the consequent voltage, results from the pressure variations caused by the sound.

Notice that for wavelengths of sound that are large compared with the diaphragm, it does not matter from what direction the sound pressure comes; any compression part of a sound wave pushes the diaphragm into the cavity behind it and produces a positive voltage at the output terminals. This is because the sound waves from behind the mic diffract around the microphone body, and the local increase in air pressure caused by the compression pushes the diaphragm into the body of the mic. Pressure microphones are thus inherently *omnidirectional,* accepting sound from all directions.

> At high frequencies, sound makes its way around the body of the microphone and through any acoustically transparent covers, through processes of diffraction and reflection. The positive-pressure part of the pressure wave still presses in on the diaphragm, no matter what its direction of origin is, which is what makes the microphone omnidirectional, often called just an omni.
>
> Where the wavelength of the sound waves becomes comparable to the dimensions of the microphone, the omnidirectionality breaks down because, for sound coming from the direction behind the face of the diaphragm, the body of the microphone casts an acoustic shadow. Nevertheless, sound pressure from the back makes it to the diaphragm through diffraction, but is

somewhat attenuated in doing so. Thus, practical-sized pressure microphones "roll off" sound coming from the rear at high frequencies.

Something else happens to high-frequency sounds coming from the front. If the microphone were not present, the sound wave would flow freely. Placing the microphone in the sound field interrupts the field, but this is hardly noticeable at wavelengths at which the microphone appears small, so across most of the audible frequency range there is little effect. At high frequencies, however, pressure "congestion" in front of the diaphragm raises the level; this is the same effect as seen at the walls of rooms, in which sound pressure is raised locally because of the pressure increase at a barrier. So without corrective action, the frequency response of a pressure microphone will rise on axis, increasing its output as frequency goes up. This is not a small effect, amounting to 9 dB at 20 kHz for a microphone equipped with a ½-inch diameter diaphragm.

The high-frequency rise on axis in an uncorrected pressure microphone was initially used to add "clarity" to recordings because almost inevitably in older recording equipment there would be a high-frequency loss somewhere that the microphone was useful in overcoming. Then, as the equipment and methods grew better over the years, a market arose for mics with a flatter frequency response, and the design was modified to make them flat on axis.

Today, both types, those with rising response on axis and those without, are offered as recording microphones. This is because mics with rising response are seen as especially useful when placed at long distances, for which there are air absorption losses to overcome, and flat on-axis mics are used for close work or in smaller rooms.

In precision-measurement microphone terminology, an uncorrected mic (with rising response on axis) is called a pressure microphone, whereas the models corrected for flat response on axis are called *free-field microphones.* Free-field precision measurement mics are used where there is a single direction for the sound field, such as in anechoic chambers. Pressure microphones are used in real rooms with mixed-sound fields, with the diaphragm oriented perpendicular to the direct sound, thus eliminating the effect of the congestion on the direct sound. In such an arrangement, the frequency response for the direct sound and for diffuse-field reverberation match reasonably well for ¼-inch-diameter and smaller microphones.[1] This is why you will see technicians aiming measurement microphones at the ceiling when tuning dubbing stages and theaters; they are using pressure mics.

---

[1]The difference is about 3 dB at 20 kHz for a ½ inch-diameter microphone and 1 dB at 20 kHz for a ¼ inch mic.

Most recording microphones are used as end-address, although a few are side-address. Measurement microphones for use in rooms (as opposed to anechoic chambers) are pressure types aimed at 90° to the direct sound.

> The high-frequency effects of shadowing by the microphone body and pressure congestion in front of the microphone are worse for larger diaphragm diameters and are better for smaller ones. On the other hand, a larger diaphragm captures more sound energy and is thus more sensitive, producing a better signal-to-noise ratio. Microphones designed for the best trade-off among these factors typically have diameters of about ½ inch. Smaller microphones are usable, as in lavaliere mics, with some noise compromise, and larger ones are even desirable in some music applications, in which the "defects" of larger size may be used to some benefit to make a particular sound quality, such as compensating for distance by the rise on axis.

For low frequencies, a pressure microphone may respond all the way down into the infrasonic region. This causes a problem, because just taking the microphone up in an elevator could push the diaphragm to one extreme and cause it to stick there, just like your eardrum in an elevator when you have a cold. For this reason, a small pinhole-size pressure equalizing vent is provided to relieve the pressure changes from the very lowest frequencies. This "leak" is typically the only limit on low-frequency response.

In sum, pressure microphones are naturally omnidirectional, generally accepting sound for all directions equally, except at high frequencies (short wavelengths), at which the microphone body and diaphragm are acoustically large objects, so there are differences with respect to direction: High-frequency sound from the rear is attenuated, and that from the front is boosted. The latter may be corrected, but to do so "tips" the whole response downward so that sound coming from the rear is even more attenuated once sound coming from the front has been made flat. This is why microphones used on sets most often have small-diameter diaphragms, whereas recording studio microphones often have larger ones. The difference is that on the set, one can never know for sure the direction of the sound source at all times, but in the studio, under more controlled conditions, the actor or musician can be placed in front of the microphone all the time.

Boundary-Layer Microphones

A special pressure microphone has been developed for placement on a large barrier, such as the walls or ceiling of a room. These microphones are specially constructed as practically flat plates with the microphone element flush with the plate.

> The advantage of this construction is that the microphone benefits from the pressure buildup at the surface across all frequencies; the "congestion" occurring in front of the diaphragm at high frequencies in conventional mics hung freely in space is made to work across all wavelengths because the diaphragm is essentially made a part of a large surface. Thus the
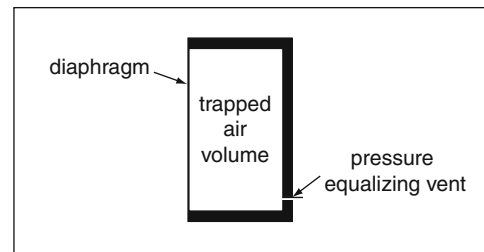
Pressure microphone cutaway drawing
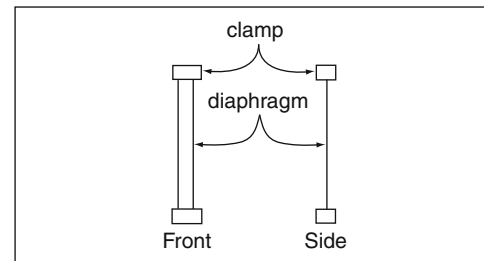
Pressure-gradient microphone

FIGURE 5.1 The basics of the pressure and pressure-gradient microphone types. In a pressure microphone, the diaphragm is stretched over a cavity, whereas in the pressure-gradient microphone, both sides of the diaphragm are exposed to the air.

> high-frequency rise seen in conventional types does not occur on axis. The directional response is hemispheric because the "back wave" is blocked by the surface.
>
> Such microphones work to full advantage only when they are placed on very large barriers. Placed on discs of plastic several feet across they suffer from the effects of diffraction around the edge, with poor low-frequency response.

Practically speaking, any pressure microphone placed up against an acoustic barrier that is not too absorptive will display the boundary layer effect of increased output and hemispheric response.

Wind Susceptibility

Pressure microphones are the type least susceptible to wind-induced noise, because their sealed cavity exposes only one surface to wind and because of the relatively high tension to which their diaphragms are stretched compared to the other types to be discussed.

PRESSURE-GRADIENT MICROPHONES

In a pure *pressure-gradient* microphone, both sides of the diaphragm are exposed to the sound field, unlike the pressure mic. What this arrangement measures is the *difference* in pressure between the two sides of the diaphragm. The diaphragm, in reacting to the difference between its two sides, works in a way that is fundamentally different from a pressure microphone, leading it to have very different directional properties that make it an

essential ingredient in microphones used for a lot of film and television work. There are also consequences to this directionality that can be managed if they are understood.

The first type of pressure-gradient microphone developed was the ribbon microphone. In it a diaphragm, consisting of a thin and light-weight (2 mg) conductive metal ribbon, is suspended in a strong magnetic field. A compression wave, perpendicular to the ribbon and coming from the front, pushes in on the ribbon, and its movement through the magnetic field induces a positive voltage across the ends of the ribbon, converting sound into electricity. A compression wave from the rear also pushes the diaphragm, but in the opposite direction to sound from the front, so compression waves from the rear induce a negative voltage, a difference to be discussed later.

We have seen that sounds coming from both the front and the rear are converted to electrical energy, but what about sound coming from the side of the ribbon? Here the sound field faces the ribbon on its edge, and there is virtually no motion of the diaphragm and thus no electrical output. A trace of the output characteristic versus angle, made by moving around the microphone, is called the *polar pattern* of the microphone. In this case, the polar pattern is a figure-8 shape, called *bidirectional* or *dipolar,* indicating that the front and back are sensitive, whereas sound arriving from either side is severely attenuated.

> Note that the figure-8 pattern is three dimensional, the type of surface we would get by spinning the symbol for the number 8 around its vertical axis.

Pressure-gradient microphones exhibit a behavior that is not found in pressure microphones because they are sensitive to the pressure gradient instead of pressure. As the distance between the source and the microphone is decreased, the low-frequency response is boosted; this is called the *proximity effect*. It is illustrated by Track 24 of the DVD.

> This effect occurs because of the curved wavefront close to the source and its measurement at two points in space, the front and back sides of a pressure-gradient mic, rather than one. At larger distances, the spreading wavefront, which is actually spherical, when sampled at the two points by a small microphone, appears essentially flat.

Old-time radio announcers used the proximity effect to good advantage for their careers. They spoke into ribbon microphones closely, thus boosting the bass and their reputation at the same time.

Whereas the proximity effect is prominent in the frequency range usually considered to be bass, from 50 to 200 Hz, there is another effect that pressure-gradient microphones exhibit that can seem contradictory. At the very lowest frequencies, below 50 Hz for typical microphones, the response is attenuated. This occurs because for the very lowest frequencies, with the longest wavelengths, the difference between two closely spaced points in air is very

small as seen by a microphone. So unlike pressure microphones, pressure-gradient mics do not respond to extreme low bass and in fact tend to virtually no output at infrasonic frequencies. Tracks 25 and 26 of the DVD show the difference in the very low frequency response of pressure and pressure-gradient microphones. (Note that the sound system you use to play these tracks must have extended low bass response to hear these differences.)

Comparing an omnidirectional pressure microphone with a figure-8 pressure-gradient microphone operating in a room reveals some important differences. The room has direct sound, reflections, and reverberation, and the direct sound has a known direction, whereas the reverberation is practically directionless. If an omni mic and a figure-8 mic are installed side by side and aimed at the direct sound, and have equal sensitivity, the two will pick up the direct sound equally, but the omni will pick up more reverberation than the figure-8 mic. This is because the omni is sensitive all around, whereas the part of the reverberant field that is coming from the side is attenuated in the figure-8 mic.

Keeping reverberation low in recording is generally desirable because of the fact that it is essentially impossible to reduce reverberation in postproduction, and it is simple and routine to add it. Thus dead recording rules, i.e., studios are usually very absorptive and directional microphones are used, both to reduce recorded reverberation.

> Another difference between pressure and pressure-gradient microphone types is due to room acoustics. Heretofore this discussion has generally assumed a free field for direct sound or a diffuse field for reverberant sound.[2] At low frequencies, at which standing waves dominate in rooms, we do not observe a free or a diffuse field, but rather standing waves, and the pressure and pressure-gradient types can be quite different at these frequencies. Just rotating a pressure-gradient microphone about its axis in a room with standing waves can also produce very dramatic changes in the bass because the standing waves have particular directions. Pure pressure (omni) and pressure-gradient (directional) types respond respectively to the pressure and the pressure difference, different components of the sound field, producing potentially very different outputs. This can easily be heard using highly directional microphones often used for film sound: their bass frequency response is affected by the standing waves in the recording room.

Wind Susceptibility

Figure-8 pressure-gradient microphones are the most susceptible type to wind-induced noise, because both sides of the diaphragm are exposed to wind and because of the relatively low tension to which their diaphragms are typically stretched.

---

[2]See Chapter 1.

COMBINATIONS OF PRESSURE AND PRESSURE-GRADIENT RESPONDING MICROPHONES

Thus far we have discussed two polar patterns, omnidirectional and dipolar. These two "primitive"[3] patterns are useful directly in many ways that we discuss later, but most microphones in widespread use for film and television work use a combination of the two types.

Certainly, one of the most popular polar patterns is the *cardioid,* or heart shape. A cardioid (and variations thereon) can be made in one form by placing both a pressure and a pressure-gradient responding microphone element together in close proximity and adding their outputs together. You will recall that we said the difference between the front and the back halves of the ribbon microphone was polarity, with one side producing positive voltage for positive energy and the other producing negative voltage for positive energy. The back half is out of phase with the front half. It is this characteristic that we can put to use. By adding the negative voltage from the back half of the dipole to the positive voltage received from the omni, cancellation is achieved for sound coming from the rear: $(+1) + (-1) = 0$. Sound coming from the front, on the other hand, sums in phase: $(+1) + (+1) = 2$. The sensitivity compared to one of the mics alone is doubled. Tracing this characteristic output versus angle around the microphone results in the cardioid shape. Track 28 of the DVD illustrates adding the two types together to form a cardioid.

For such cardioid mics, the suppression of reverberation is about the same as that of the figure-8 mic. Although reducing reverberation is useful, it is the forward preference for sound that makes it favored in film and television applications compared to figure-8 mics.

The method for producing a cardioid polar pattern that relies on summing two mic elements is rather clumsy, but it was the first way in which cardioid mics were made, and some still function this way. After the importance of the

Pressure   +   Pressure Gradient   =   Cardioid
Omni   +   Bidirectional   =   Cardioid

FIGURE 5.2   A pressure receiver and a pressure-gradient receiver added in the proportion of 1:1 produces a cardioid.

---

[3]Primitive in the sense of serving as the basis for derived forms.

pattern was established in the marketplace, other methods of construction that achieved a cardioid pattern were found.

> The methods used include adding chambers, holes, silk, or metal mesh screens and other methods of acoustical phase shifting to sound applied from the outside to the rear of the diaphragm, all of which are aimed at producing a cardioid pattern over a wide frequency region. In an alternative method, two diaphragms that are electrically added together are used to produce various polar patterns. Some types so equipped can even be made with electrically adjustable polar patterns.

The cardioid microphone is without a doubt the most commonly used type among reasonable quality microphones in the world. However, it is used infrequently in film and television production sound, for which one of the types discussed below is more likely to be used. Sound-effects recording, music recording, and Foley recording are more likely to make use of cardioids.

Super- and Hypercardioids

There is a need for microphones that are more directional than cardioids, especially in film and television production sound, in which the microphone must often work at a large distance. The cardioid or the figure-8 mic, when used at a distance factor of 1.7 times as great as an omni mic, produces equal reverberant pickup. The cardioid pattern is generated by adding the outputs of two receivers, for pressure and for pressure gradient, in equal proportions, causing the null toward the back. If the levels of the two receivers are adjusted differently, the pattern can be made with two important differences from a cardioid. The sides are "pulled in" compared with the cardioid, so for sound at 90° the output is attenuated. In addition, there is a "lobe" or region of sensitivity pointed toward the rear. This polar pattern is either supercardioid or hypercardioid depending on the ratio of the pressure to the pressure-gradient component of the sound field. These two types are so close to each other in performance that, all other things being equal, they are essentially indistinguishable in practice. They are among the most valuable patterns for filmmaking for several reasons:

- Because of their tighter directivity, the response to reverberation is less than in any mic type discussed up to this point, having a distance factor of 2.0 for pickup of reverberation compared to a pressure mic. Although this is not a large difference compared to the cardioid factor of 1.7, with 1.4 dB less reverberation at the same distance, it is just enough to hear the difference and exploit it.
- Instead of a line pointing straight back from the mic being the null in the cardioid type, the super- and hypercardioid have their nulls at angles of 120° and 135°, respectively, from the front. This null is rotated in space and thus forms a cone shape from which the output is
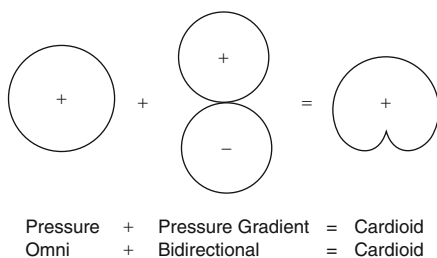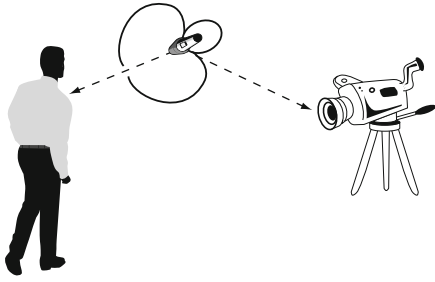
FIGURE 5.3   Super- and hypercardioid mics on overhead booms allow the boom operator to place the actor on the "hot" side of the microphone and the camera on the "cold" side, thus maximizing the desired performance and minimizing the camera noise simultaneously.

FIGURE 5.4   This series of capsules from Schoeps provides a wide variety of polar patterns and built-in frequency response variations to cover a wide range of applications. They include pure pressure (omnidirectional) and pressure-gradient (bidirectional) types, as well as cardioid, hypercardioids, subcardioids, and boundary layer types. Other variations include capsules that roll off low frequencies and ones that emphasize high frequencies, along with ones that demonstrate very flat frequency response.

minimal. In the overhead position on a boom, the front "hot" side of the mic can be aimed at the actor, and the null "cold" side can be aimed at the camera. Thus the actor can be recorded well and the camera noise rejected at one and the same time, which is important in motion-picture applications (movie cameras are noisy because of the intermittent movement of film through the gate; some video cameras have noisy fans).

FIGURE 5.5   Any of the capsules shown in Fig. 5.4 may be combined with one of a variety of electronics "bodies," such as this one, having various powering options to produce a complete microphone.

Subcardioid

If the summation between elements is made different, more in favor of the pressure type, then a subcardioid (also called a *limaçon*) pattern can be generated, with no complete null but with a forward preference. Used occasionally in music recording for an appropriate balance between direct sound and reverberation, it is rarely if ever used in production sound.

Variable-Directivity Microphones

Because these various patterns are generated by summing together the ingredients of a sound field, it is possible to make either fixed-directivity (fixed polar pattern) microphones or variable-directivity microphones, with the variation among patterns accomplished either mechanically or electrically. As a general statement, though, variable-pattern microphones are compromises in favor of flexibility over performance on any single pattern and therefore are used in applications in which there is an emphasis on flexibility. In film and television production sound, it is rare to find switchable directivity microphones in use, although music recording studios make more use of these types.

> Note that the degree of the proximity effect will vary depending on the pattern to which a variable-directivity microphone is set, so in practical situations, changing the pattern could result in hearing different amounts of bass. This is not a defect but an expected result arising from the various methods of operation.

Interference Tube (Shotgun or Rifle Microphone)

The final directivity pattern is one of the most important for film and television uses. It was developed originally to solve the problem of keeping the microphone out of the camera shot while still providing a high ratio of direct-to-reverberant sound. To achieve this, even more directionality was required than the sum of two receivers—pressure and pressure gradient—could provide, as in the supercardioid. The interference tube microphone developed in the United States for the introduction of television in 1939 has undergone several generations of development, with the current one described here.

If a tube is arranged with slots along its length, the slots are covered with acoustical resisting material such as silk, and the end of the tube terminates in a supercardioid microphone transducer, then sound waves progressing along the axis of the tube will be unimpeded. Sound incident on the tube from 90° will suffer interference effects within the tube and will not add together at the transducer. So, in sum, there is an increase in directivity compared with any other microphone type, with an accompanying reduction in reverberation and other off-axis sound.

In contemporary production sound, this is probably the most popular type used as a boom mic. It may suffer from the following problems, however:

- Off-axis sound is reduced through interference. In older models this causes peaks and dips in the off-axis frequency response, because the interference becomes constructive or destructive at various frequencies. Although attenuated, the off-axis sound is "colored," that is, the timbre is quite noticeably changed. This can have an effect on, say, footsteps, when the microphone is pointed at the mouth of the actor from 45° overhead. The feet are very far off the axis of the microphone, and they may sound as though they were recorded in a barrel.
- The longer the interference tube, the wider the frequency range over which the narrow directivity is achieved. Thus, the mics having the most uniform directivity with frequency use long tubes, which may become unwieldy.
- Wind susceptibility is relatively high. This means effective windscreens must be used, which are also necessarily large, to create a region around the mic that is relatively less turbulent.

MICROPHONE TYPES BY METHOD OF TRANSDUCTION

All microphones convert acoustical into electrical energy, but there are a number of ways to accomplish this, with varying areas of application.

Carbon

The first microphones were composed of a diaphragm stretched over a cavity that was loosely filled with granular carbon. Positive sound pressure presses in on the carbon, compacting it slightly. This reduces the electrical resistance of the mass of carbon, and when connected to a DC power source such as a battery, a voltage proportional to sound pressure can be generated. On 10 March 1876, Alexander Graham Bell said into such a microphone the famous first line ever sent by electrical means "Mr. Watson—Come here—I want to see you."

It is extremely difficult to make such microphones both sensitive and with a flat frequency response, and they suffer relatively high distortion due to the nonlinearity of the carbon mass. Still, such microphones formed the heart of the telephone industry for 100 years, so there are still some in daily use. Carbon microphones are rarely if ever

FIGURE 5.6 A Sennheiser shotgun microphone. Its greater length compared to shorter shotgun microphones allows it to maintain a narrow polar pattern over a wider frequency range. This type is often used out of doors, encased in a windscreen and, often, a Windjammer (see Windscreens in this chapter). Photo courtesy Sennheiser electronic GmbH & Co. KG.

used directly in filmmaking, but the output of telephone conversations may need to be recorded.[4] A magnetic-induction pickup coil, or direct wiring, can be used to record telephone conversations over phone lines, for which the source is the carbon microphone in the telephone. Such sound quality may well be more appropriate for a phone conversation than the use of a better microphone. Although postproduction techniques include a "telephone filter," sometimes it is best simply to record a telephone, with all its response variations and distortions.

Ceramic

Certain crystalline, glasslike materials, when struck by vibration, produce voltage directly by way of the piezo-electric effect. The vibration can be conducted from a diaphragm to a transducer element and thus form a ceramic microphone. Small blocks of such ceramic materials are very resonant, like striking a bell, so it is difficult to get a wide-frequency-range receiver. The primary place where such an element is in any use in film and video is in making hydrophones (underwater microphones), for which the very stability of the glasslike structure is highly useful. In this case, the microphone is supplied with it own set of electronics, which produce a line-level signal suitable for recording. On the other hand, underwater recordings can also be made with other microphone types enclosed in water-impermeable housings, but their sound may be muffled by the enclosure.

Electrodynamic (Commonly Called "Dynamic") Microphone

If a conductor of electricity such as copper is moved in a magnetic field, a voltage is induced across the ends of the conductor. The conductor can be insulated wire arranged in a coil and attached to a diaphragm. With the right shape of magnet, the diaphragm motion will cause the *voice coil* to produce a voltage at its ends as the magnetic lines of force cut across the coiled wire.

Dynamic microphones generate their own electricity, without an outside source of power, as is needed in the other types. They are also typically rugged compared with other types, withstanding both shock and variations in temperature and humidity better than other mics, and thus they are preferred as at least a backup in many kinds of filmmaking. They do contain strong magnets, and some models leak magnetic field, which means they should be stored at least a few inches away from audiotape. The simplest dynamic microphone to make is the omnidirectional

---

[4]Note that there are laws covering permissible recording of telephone conversations.

FIGURE 5.7 Cutaway drawing of a dynamic microphone.

pressure type, but other polar patterns are available, with the cardioid probably being the most popular among professional types.

Because of the particular combination of strengths in the dynamic microphone, it is typically used in film and video production in which requirements for ruggedness and reliability under adverse conditions prevail. The ultimate quality of a dynamic microphone is potentially limited by the requirement that the sound move the mass of the diaphragm and voice coil to produce an output, and the mass, although low, is higher than in the electrostatic microphone (discussed below). Although well-designed dynamic microphones can be very good, as a class they are not considered the ultimate transducer.

> Note that a ribbon microphone described in the section on pressure-gradient microphones employs the same principle of transduction as the conventional dynamic microphone, but the smaller amount of conductor in its magnetic field typically leads to much lower output levels, which must be stepped up to a more usable level by a transformer in the microphone. Even with the step up, the sensitivity of a ribbon microphone is usually too low for most film and television purposes but may find use in music studios, especially in front of louder instruments.

Electrostatic (Also Known as Condenser or Capacitor) Microphone

The electrostatic microphone has only one moving part, the diaphragm. The motion of the diaphragm is detected by measuring the electrical property capacitance between the diaphragm and a fixed back plate.[5] *Capacitance* is the ability of two conductors, separated by an insulator, to store charge. In the case of an electrostatic microphone, there is a fixed amount of capacitance caused by the spacing in silence, which varies up and down by the motion of the diaphragm due to sound. It is the changes

in capacitance with sound that are detected, using one of three methods.

> One method is to charge the capacitor formed by the diaphragm and back plate, using a polarizing voltage, generally 45 to 200 VDC. A set of electronics can be arranged to make the variations in charge resulting from the changes in spacing between the diaphragm and the back plate into an output voltage. This method is quite stable and is used in measurement and many recording microphones.
>
> The second method is to use the changes in capacitance caused by the sound to change the instantaneous frequency of an oscillator, typically operating around 10 MHz. The changes in frequency are detected by a frequency modulation detector, which works like the detector stage in an FM radio, and are converted to the audio output signal. This method avoids the use of a polarizing voltage, with a claimed attendant increase in reliability. This technique is labeled "RF microphone," which may be easily confused with radio microphones, for which the same term is sometimes applied, although one describes what is going on inside the body of the microphone and the other describes a wireless microphone.
>
> The third method is most commonly used in inexpensive electrostatic microphones, such as those found on answering machines, but there are a number of less expensive recording microphones that also make use of the technique. In this type, the polarizing voltage is applied permanently, by electrochemistry during manufacture, and thus does not require an external voltage source for polarization (although power is still needed for the electronics, which must follow the capsule). These prepolarized microphones are called *electret capacitor* (or condenser) microphones. Although the best of this type can be practically as good as air capacitor microphones, few of the highest quality microphones use this principle.

All three of these methods of converting capacitance variation to output voltage require electronics to do the conversion and thus require power. Generally the electronics must be placed in close proximity to the pickup capsule of the microphone, although for some types special active cables are available to space the pickup capsule away from the electronics. The power supply may consist of a battery in the microphone proper, in an external box, or in the unit into which the microphone is plugged or as a power supply provided as part of a microphone preamplifier that may be standalone or made a part of a mixer or console.

There is more than one method for supplying the power. In the simplest method, a battery is inserted into the microphone body itself or into a connector at the other end of a dedicated mic cable. These batteries are easy to forget about because they often last several hundred hours. They are often specialized types, which are not widely available, and thus it is essential to have spares. 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 lavaliere (Sony ECM-55B) *using the same battery*. Thus it is important to know the rated battery life of a particular microphone you use.

---

[5]This property gives this microphone type its common names: condenser and capacitor (condenser is an older word, now obsolete in other uses, meaning the same thing as capacitor). Most texts use one of these words, although the word "electrostatic" describes the underlying principle better.

There are some very inexpensive types of electret capacitor microphones available, which, if they have two leads, are powered by supplying the signal lead with DC. If they have three leads, one is the common ground, one is the output, and the third is for battery power. These are suitable for planting on a set or for nature shows, where they might be destroyed or eaten, for example.[6]

There are several methods for supplying power to microphones remotely over the required balanced lines. The most popular studio method is *phantom powering,* in which a positive voltage is applied to both balanced leads from the microphone and the negative voltage to the shield ground. Microphones using this type of powering usually have the letter "P" for phantom in their model number. Such microphones are suitable for connection to microphone preamplifiers that supply phantom power (which may often be switchable on and off). The most common voltage for phantom power is +48 VDC.

> Although one intent of phantom powering is to allow interchangeability between powered electrostatic microphones and electrodynamic ones, it is generally not good practice to supply unpowered microphones with power. What may happen is a connection accident in which one side of the balanced line becomes connected first, applying the DC voltage to a microphone not designed to handle it. Phantom power is thus often switchable, even on a per-channel basis, on devices so equipped.

A second, less common method of powering is to supply positive voltage to one of the balanced leads and negative voltage to the other balanced lead. Called *A–B* or *T powering,* this method is not as popular as phantom power, because of potential damage to other microphone types that are inadvertently connected. However, T powering came about in the film industry, so there are still many examples of these microphones in use for location sound purposes. Connecting a ribbon microphone to a microphone input supplied with A–B powering will probably damage the ribbon because it seeks to "transduce" the applied voltage in reverse and act like a loudspeaker; damage may also result to dynamic mics. Microphones using this method of powering usually contain the letter "T" in their model number. The most common voltage for T powering is 12 VDC.

> Another variation is that T-powered microphones may be supplied with "+" and "−" voltage models. This is a leftover from powering by early Nagra recorders that were Red Dot models, with the polarity of pin 2 negative compared to pin 3.

Electrostatic microphones using air as their insulator with nickel or titanium metal diaphragms and back plates and quartz insulation are extremely stable, and this construction is used for measurement microphones. Recording microphones more often use plastic diaphragms coated with a conductor such as gold. These too may be quite

stable. On the other hand, capacitor microphones depend vitally on the quality of insulation. Anything that reduces the insulation can lead to noise or failure. For instance, condensing humidity between the diaphragm and the back plate reduces the insulation greatly and causes the microphone to become noisy, increasing hiss, popping, or producing no output at all. Conditions leading to microphone failure are quite possible in the environment of production sound recording, in which microphones may be exposed to weather or fog machines on shooting stages. The advanced state of design of such microphones is indicated by the fact that more trouble does not occur. Capacitor microphones can sometimes be restored to their original condition by slow warming in a dry, 120°F environment if humidity was the cause of their malfunction.

> Electrostatic microphones have been used under arctic and jungle conditions. Here their reliability may be improved by several methods. First, they must be kept dry, and storing them when not in use in air-tight containers with desiccant can help. Desiccant is a salt that absorbs water from the surround atmosphere, and a small bag of it is often shipped in camera and lens cases. Second, they should not undergo thermal shock, that is, rapid temperature change, because that can lead to condensing humidity. In this way the microphones share a common problem with rotating-head tape machines, such as video recorders and DAT machines.

Overall, the principle of the capacitor microphone is applied to what are typically the highest grade microphones, because this principle works with the smallest moving mass and the highest efficiency of conversion from sound to electrical energy, but typically at a higher cost than dynamic types and with a need for power.

Electrostatic mics may be built as pressure or as pressure-gradient transducers explained on pages 77–79 in multiple polar patterns. Some of the pressure-gradient types work by using dual back-to-back diaphragms, whereas others employ acoustical phase-shift networks behind the diaphragm to accept and delay sound from the rear to produce cardioid, super- and hypercardioid, subcardioid, or interference tube designs. Some dual-diaphragm designs claim better signal-to-noise ratios than single-diaphragm designs.

MICROPHONE TYPES BY DIRECTIVITY (POLAR PATTERN)

A variety of microphone constructions lead to the following polar patterns, already discussed in terms of how they are made up of combinations of pressure and pressure-gradient receivers and, in one case, an added tube:

- Omnidirectional
- Bidirectional (figure 8)
- Subcardioid (limaçon)
- Cardioid
- Supercardioid

---

[6]One of these is Panasonic Part No. WM-61A102A, available from Digi-Key Corp. at www.digikey.com for $1.25.

● Hypercardioid
● Club-shaped (interference tube or shotgun)

This list is ordered from the most reverberant-field-sensitive to the least sensitive comparing the mics at equal distances from a sound source at which they are aimed. All other things being equal, the cost of a microphone having wide, flat, and smooth on- and off-axis responses and wide dynamic range generally increases in several steps as one goes down the list, because the conventional pressure microphone is the simplest and the interference tube the most complex of the constructions, with the others being about equal in complexity. Thus, at equal price an omnidirectional pressure mic may well be better than a superficially similar hypercardioid.

Tracks 22 and 23 of the DVD feature a "walkaround" demonstration of microphones having various polar patterns.

The way that users can typically distinguish a directional microphone from a nondirectional mic is that the directional mic will have more than one primary entrance for sound, whereas a nondirectional mic will have only one entrance (neglecting the tiny pressure-equalizing vent).

The axis of a microphone is usually along its length, so the design intent is for the barrel of the microphone to be aimed at the source. These are called *end-address microphones*. There are some microphones, however, that are addressed from the side, especially dual-diaphragm capacitor microphones, often with switchable directivity.
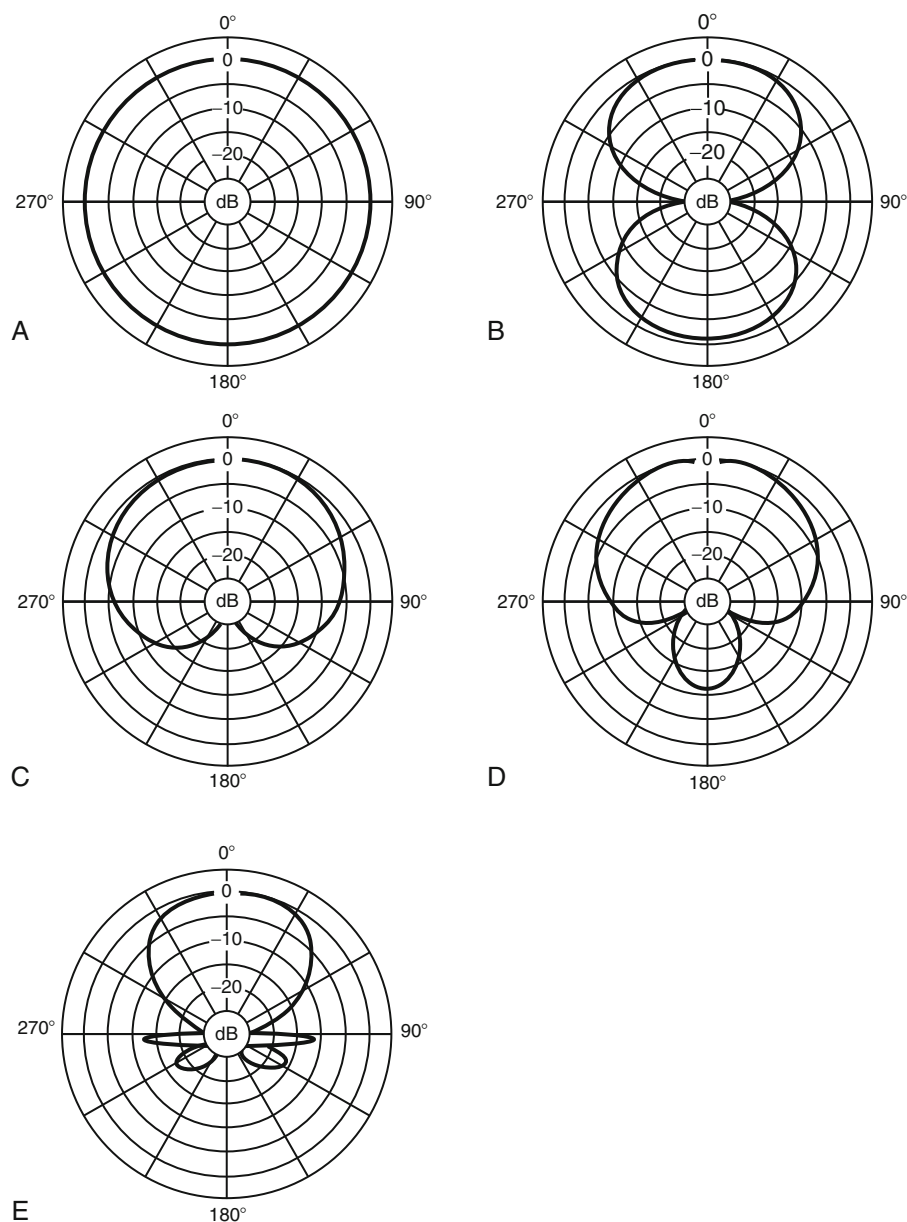
FIGURE 5.8 Polar diagrams of various microphone types. (a) Omnidirectional (pressure microphone), (b) bidirectional (pressure-gradient microphone), (c) cardioid, (d) hypercardioid, (e) interference tube. All are shown at only one, upper-midrange, frequency.

They may be distinguished by having a solid end cap or by the shape of the grille and are called *side-address microphones* or *vertical mics*. Because some of these mics are otherwise cylindrical, it is difficult to know which side is the front; thus, they usually have a red dot or a symbol of their directivity on the front side of the mic to indicate the side to be aimed at the source.

MICROPHONE SPECIFICATIONS

Today specifications for microphones are accessible via the web, with one valuable site being www.microphone-data.com. Although some of the response curves[7] on the site appear to be manufacturer's marketing data rather than engineering measurements, it is nonetheless highly useful for comparing other specifications.

Sensitivity

Sensitivity is the ratio of conversion of sound pressure level to electrical voltage or power. The most popular way is to rate sensitivity as the electrical voltage produced when the microphone is exposed to a reference sound pressure level of 94 dB at 1 kHz, a fairly loud sound level. The rating units are millivolts per pascal because the sound pressure level corresponding to 1 Pa is 94 dB (re: 20 µPa). On this scale, an insensitive ribbon microphone measures 0.5 mV/Pa and a high-output electrostatic type 60 mV/Pa.

> An older method of specifying microphone sensitivity applied to electrodynamic mics. It was called the *gm* method and it specified the output of microphones as the power they would deliver into a rated load impedance. The specifications on the web site cited above have been converted into mV/Pa from this older method.

When comparing two microphones by listening, it is important to adjust out sensitivity differences by trimming the level controls in the chain, because on A–B tests, the louder device, even if only ½ dB louder, will sound "better." This places a premium on high sensitivity, which may be misapplied, because an adjustment of level can remove the effect of mic-sensitivity differences. What is important is the dynamic-range capability of the microphone and preamplifier combination, compared with the volume range of the sound source: Can the volume range of the source be "fit" into the dynamic range of the microphone/preamplifier combination? More information on this important topic is presented later in this chapter and the next.

Frequency Response

Frequency response is an exceedingly important microphone specification, but it is difficult to tell microphones apart from their data sheet frequency response because the methods used are too variable, and the spec sheets too idealized, to make them comparable across manufacturers. Also, frequency response varies with angle of incidence for direct sound fields and also differs between direct and diffuse sound fields, making a single curve inadequate to tell the whole story. Figure 5.9 shows a typical good microphone frequency response. Note that the definition of frequency response, as described in Chapter 3, is the variation in decibels over a stated frequency range.

> For example, comparing an omnidirectional mic with a cardioid and a hypercardioid from the same microphone series, having practically the same axial response, results in hearing the sound become "harder" when switching from the omni through the cardioid to the hypercardioid. This is apparently caused by the differences in diffuse-field responses, which change the timbre of the reverberation.
>
> The microphones best characterized for frequency response are measurement microphones. These are instruments used by scientists and engineers to measure sound fields, not for recording. There is a great deal of reliable information available for these types, including on- and off-axis frequency response, changes due to accessories, etc.

Choice of Microphone Frequency Response

Many microphones are made with a deliberately nonflat frequency response for a variety of reasons, which have some very specific rationalizations:

- All directional mics suffer from the proximity effect. The microphone designer chooses the distance at which the microphone will be flattest. Generally they are rolled off in the bass when used at a distance, and the proximity effect boosts the bass as the microphone is brought closer to the source. Used at a distance, as on a boom, there is a net low-frequency rolloff. This rolloff may well be desirable because it reduces low-frequency room and boom noises.

Response is ± 2 dB from 100 Hz to 20 kHz
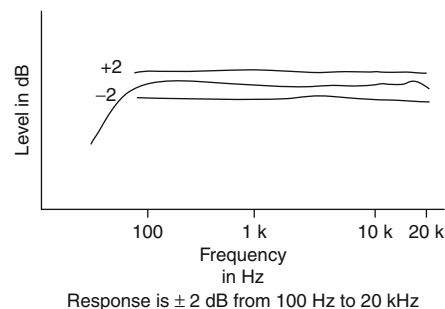
FIGURE 5.9   The frequency response of a microphone.

---

[7]Short for frequency response, defined in the Glossary and in Chapter 3.

- Also, in most rooms the reverberation time is longer at low frequencies than at middle and high frequencies. It is thought that longer reverberation time at low frequencies may lead to an impression of exaggerated bass in recording when pressure microphones are used, so the use of pressure-gradient types with their low-frequency rolloff at a distance helps to compensate for this impression.
- High-frequency boost is built into shotgun microphones to overcome anticipated high-frequency air absorption when working at a distance from the source and to deliver a greater sense of intimacy with a distant source than a flat microphone. A typical amount is +4 dB at 10 kHz.
- Midrange boost is built into many vocal microphones to increase the "presence" frequency region and to ensure that the outcome is intelligible, despite possible bandwidth and frequency-response limitations later in the chain. Recordings from such microphones tend to cut through other sonic clutter, but at the expense of deviating considerably from the timbre of the source.

Although we generally wish to preserve source timbre and thus desire a flat frequency response, the above-mentioned considerations can cause deviations from such thinking. In general, the microphones that are the most popular for production and sound-effects recording have a wide and smooth frequency response, although they are not necessarily flat (see Figure 5.10). They also have generally uniform polar patterns with frequency such that off-axis sound is only attenuated, not colored, by frequency-response anomalies.

In music recording, microphones are chosen for their frequency response more than any other single factor. This does not mean, however, that one would always choose the flattest mic as the one that necessarily best preserves the source timbre, because representing the source best from a single pickup point is a complex trade-off, as described in Chapter 1. Most practical sources have complex output directivity, and choosing and placing the microphone is an aesthetic choice designed to best represent that source in all its actual complexity.

> Music recording also suffers from a lack of standardization of the monitoring experience. By moving from studio to studio
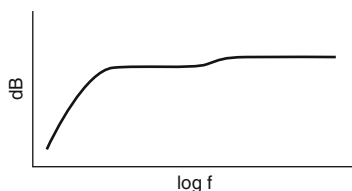
FIGURE 5.10 A useful frequency response may not be flat. This shape of response helps control boom noise and increasing noise and reverberation time as frequency goes down, and it has a high-frequency boost so that it can be used at a distance and still produce a high amount of presence.

and playing the same source, one finds very great differences in the octave-to-octave balance of the monitoring system. Thus, it is not surprising that competent users make different microphone decisions in different studios.

Polar Pattern and Its Uniformity with Frequency

The polar pattern is usually portrayed as a set of curves for various frequencies plotted on polar graph paper. An alternative method of display is to make a series of frequency-response curves at various angles. Generally it is desirable to maintain the same polar pattern versus frequency insofar as possible.

> On the other hand, the "defect" of narrowing directivity with frequency is sometimes useful in orchestral recording, with mics spaced at a distance from the players. In this case, the narrowing directivity and rising on-axis response with respect to frequency make for greater "clarity."

Equivalent Acoustic Noise Level and Signal-to-Noise Ratio

The equivalent acoustic noise level is often specified in dB SPL. This is the noise that the microphone itself makes in an extremely quiet space. When audible because a very small signal causes the person mixing to turn the gain up, it is usually heard as hiss, whereas most acoustic noise in rooms is heard as lower frequency rumble.

Modern large-diaphragm microphones can have a noise floor equal to 10 dB SPL, A weighted, while mics with ½-inch diaphragm diameters are somewhat noisier.

> Microphone noise measurements are usually weighted for audible effect by using a frequency-response curve that emphasizes the frequencies at which human hearing is most sensitive. One frequently used weighting curve is called *A weighting,* and another, producing worse numbers for the same noise, is *CCIR weighting.*

The signal-to-noise ratio measures the difference between the sensitivity of the microphone and the noise floor in decibels. It is almost always referenced to 94 dB SPL, so if only the signal-to-noise ratio is given as a spec you simply subtract the number from 94 dB to get the equivalent acoustic noise level.

It is a mistake to think that because electrodynamic mics contain no electronics they make no noise.

> A conductor at room temperature makes noise due to the Brownian motion of electrons, called Johnson noise, for its discoverer at Bell Labs. Johnson noise sets the noise floor of a dynamic mic, even if it is followed by a noiseless microphone preamplifier, which is only a theoretical condition but one that may be closely approached. Most professional microphones have an impedance of 200 ohms. This makes a Johnson noise at room temperature of 0.25 μV. Comparing it to the sensitivity of a typical electrodynamic microphone of 2 mV/Pa sensitivity places its equivalent noise floor at 13 dB SPL, a value that competes with electrostatic microphones.

Maximum Undistorted Sound Pressure Level

The maximum equivalent sound pressure level that a microphone can produce without exceeding a specified total harmonic distortion grows increasingly important as the source becomes closer and louder. It is essential for the microphone to remain undistorted, because any distortion occurring in the microphone or microphone preamplifier will generally not be able to be reduced in subsequent processing.

Microphone overload may occur typically in the range from 112 to 165 dB SPL, for electrostatic types and depending on the particular model.

> The maximum undistorted output level may vary with frequency, usually with less output capability at the highest frequencies and sometimes at low frequency. It is usually specified at only one frequency though, in the midrange.

What overloads or "clips" is the electronics made necessary by the electrostatic principle. Once a sound recording is clipped it is very difficult to recover undistorted sound. There is simply a limit beyond which the electrical output will not go, and this is the cause of an abrupt onset of audible distortion. In extreme cases, such as close-recorded gunfire, the diaphragm can move as far as it can and bottom out, or even become stretched and need replacement.

A pad that attenuates voltage and is installed between the pickup capsule of electrostatic type microphones may be part of a microphone design or may be available as a screw-in accessory. The maximum undistorted sound pressure level is increased by the value of the pad. Pads are covered under Microphone Accessories, later in the chapter.

Dynamic Range

The dynamic range of a microphone is the "distance" in decibels from the equivalent noise level to the maximum undistorted SPL. In high-quality microphones this amount may range to over 130 dB, a ratio of more than 3,000,000:1. This causes one of the largest problems in recording because there are few if any microphone preamplifiers that can handle both ends of this range simultaneously. The job of the recordist is to fit this "gallon" of material (the microphone output), first into a "quart" jar (the preamplifier input) and then into possibly an even smaller "pint" jar (the recording medium).

The factor that ameliorates problems in this area is that the volume range of most practical situations, from the loudest sound to be recorded to the background noise level of the space, rarely approaches 130 dB. There are several ways to bring the large microphone dynamic range to within the capability of the microphone preamplifier, described in Chapter 6.

Susceptibility to Wind Noise

This rarely specified feature of microphones can make or break a production sound recording, because the low-frequency noise made by wind "rattling" the diaphragm directly can pollute the recording so badly that it cannot be repaired in postproduction. This is not simply the sound of wind, but rather an added interaction between wind and the microphone. In general, directional microphones are more susceptible to wind than are omnidirectional mics and require a larger volume captured inside a turbulence-free zone, and thus a larger windscreen, for equal performance.

Susceptibility to Pop Noise

Associated with wind noise is the noise that many unprotected mics make when a performer is working near the microphone and his or her plosive "p"-like sounds pop the diaphragm. Again, this is an intolerable noise that is impossible to remove subsequently and must be prevented from being recorded in the first place. Some mic types are specifically designed for this use, with built-in *pop filters,* that is, mesh screens that reduce the air velocity while passing sound. External pop screens are also available as an accessory.

Susceptibility to Handling Noise

Microphones are exquisitely sensitive to sound pressure levels. Various models though may be more or less sensitive to direct handling noise because of their basic method of operation or their internal mounting details. Almost all applications of microphones require isolation of the diaphragm and other parts from the direct conduction of mechanical noise. This may take the form of internal isolating elements inside the microphone body, particularly likely with microphones intended to be hand held, to external shock mounts, always needed for high-quality microphones on a boom, for instance.

Susceptibility to Magnetic Hum Fields

Electrodynamic mics, including ribbon mics, are subject to the direct pickup of magnetic hum fields. Often associated with AC wiring such as for lights, these hum fields can be reduced by moving the microphone away from the source and by reorientation with respect to the source, because the magnetic fields are directional. Of course, the mic cabling may also pick up hum, as described in Chapter 3.

Impedance

> The impedance system employed today in a microphone-to-microphone-preamplifier interface is called the "bridging" system. The source impedance of the microphone is low,

typically 10 to 200 W, and the input impedance is relatively high, such as 3 kW. This means that virtually all of the voltage available from the microphone is available at the input to the preamplifier. Among other advantages, the bridging system—compared to the older "matching" impedance system—permits mic *splitters* so that one microphone can feed more than one microphone input in cases in which this is necessary.

Some years ago, a clear distinction was made between professional microphones, having a low output source impedance and using balanced lines, and consumer microphones, which used higher impedances and unbalanced lines. Today there are few microphones sold that are not low impedance.

Power Requirements

Electrodynamic microphones require no power source, because they are electric generators in and of themselves. This is a reason to prefer electrodynamic types in some instances. For example, having an electrodynamic microphone along on a recording expedition far from civilization would be a good idea when the batteries used to supply electrostatic microphones may be in short supply.

Electrostatic microphones are specified according to their type of powering, discussed under Electrostatic Microphone, above. Power requirements are given in terms of current necessary to supply the microphone at its rated voltage and may be used to determine battery life.

MICROPHONE ACCESSORIES

Pads

For electrostatic microphones, an attenuating pad may be built in between the capsule and the microphone's own electronics, raising the maximum undistorted sound pressure level in the range of 10–20 dB. These are represented as a switch on the microphone body, usually marked something like "−10 dB." For other microphones preelectronics pads are available as accessories, in which case they are screwed in between the capsule and the electronics, are usually available in values of 10 and 20 dB, and may be stacked (although the result will be less than 30 dB attenuation). For both of these types, the maximum undistorted sound pressure level is increased by the amount of the pad, but the equivalent noise level is also increased, so the dynamic range is not improved, but the capability of handling high sound pressure levels is. These pads are most often used for sound-effects recording, as well as in close mic'ing of musical instruments and shouting performers. I have measured an actor screaming at a boom mic location at 128 dB SPL, and this level overloads some electrostatic microphones. In particular, microphones intended for use at a distance, such as long shotguns, may overload at lower levels than this, when used fairly close to an actor.

There are also pads available for insertion "inline," that is, for insertion into the microphone cable after any microphone electronics, but before a microphone preamplifier. Although not protecting against electronic overload in the microphone electronics itself, they may protect from overload in the microphone preamplifier, which is discussed in Chapter 6. Some of these types will affect the power supplied to the microphone, so may not be used for phantom or T-powered mics. So special pads that pass power may be necessary in some cases.

High-Pass (Low-Cut) Filters

For some models a low-frequency filter is available that is inserted between the microphone pickup capsule and its electronics. Inserting the filter improves low-frequency headroom, attenuating wind and boom motion noise before it can overload the microphone's own electronics or the microphone preamplifiers. The Schoeps Cut 1 and Cut 2 filters provide a steep high-pass filter, 24 dB/octave, with a −3 dB frequency of 60 or 30 Hz, respectively. The 60-Hz Cut 1 will retain voice timbre in all but the most extreme cases, like James Earl Jones. The 30-Hz Cut 2 would normally be used on music or sound-effects sources with low-frequency bass that must be recorded outdoors. In addition, Cut 1 and Cut 2 provide an adjustable-frequency 6 dB/octave filter that can be used to overcome the proximity effect that boosts the bass and occurs when using a directional microphone close to a source.

Shock and Vibration Mounts

Protecting the microphone diaphragm and its suspension from direct exposure to shock and vibration is the job of special mounts that are either incorporated in the body of the microphone or offered as an external accessory, or both. Noise may be induced in microphones by shock or vibration through a direct mechanical path from the outside to the diaphragm. This noise typically takes the form of large amounts of low-frequency noise, such large amounts that subsequent treatment in postproduction is not likely to be able to make the sound usable. An example is using a lectern microphone without an internal or external shock mount and having a boisterous lecturer pound on the podium for emphasis. Not only do we hear the through-the-air sound of hitting the podium, but the sound is far worse because of the direct conduction of shock into the microphone pickup.

External shock mounts work by a mass-and-spring isolation system similar to a spring with a weight hung on it. If we wiggle the top of the spring quickly, the weight stands still; we are above the frequency of the mass–spring resonance, at which inputs to the system are filtered out by the time they reach the weight. This is the frequency region in which shock mounts work: they are the spring, and the microphone is the mass.

Conversely, if we move the spring slowly, the mass follows slowly too, but this condition is not a problem as it induces no noise in the microphone.

Because the mass–spring system of the microphone suspension involves tuning to a specific frequency, the suspension should be matched to the microphone. Too tight a suspension and little isolation is achieved. Too loose a suspension and the microphone will wiggle around so much when the boom or fishpole is moved that it can hit its maximum extension.

One suspension type introduced in 2007 provides a different resonant frequency in different planes, with the objective of further isolating the direction of the motion of the diaphragm from noise. This "lyre" shape was introduced by Rycote. An example is shown in Fig 5.11.

It is very important once microphones are suspended within the shock-mount cradle that the effect of the shock mounting not be circumvented by "short circuiting." Stretching the mic cable taut across the mount from the mic body to the attachment point will render the shock mounting ineffective. The cable must be adequately limp and formed into a loop to prevent shock from being conducted across the cable to the mic body. Figure 5.11 shows, along with the lyre suspension, an accessory cable device that makes a transition from a special limp, small-diameter microphone-to-suspension cable and a long microphone boom cable.

Mic Stands

There are a huge variety of mic stands on the market, with the principal differences being in the flexibility of the

FIGURE 5.11 A lyre-shaped shock mount having different resonant frequencies in several planes.

positioning of the microphone. Some larger "professional" mic stands may be a detriment to good sound, because the large size (of their "professional" parts) close to the mic body reflects sound from the source back into the microphone. Conversely, small, light-weight parts can rattle in a loud sound field and, being close to the microphones, be heard. So small but sturdy microphone stands are to be preferred.

Mic Booms and Fishpoles

Mic booms and fishpoles are described in Chapter 4.

Windscreens

Windscreens vary depending on whether they are for use on pressure or pressure-gradient microphones and whether they are to be used indoors as a pop filter, out of doors to tame a gale, etc. There are a variety of windscreen types, and the various ones are most suitable for particular mics:

- Tight-fitting foam type that completely encases the acoustic openings of the microphone body. These inexpensive windscreens, although often a supplied accessory with all kinds of microphones, are most suitable for omnidirectional microphones. Directional mics are better screened when there is a trapped volume of air inside the windscreen. However, this type may be used on podium mics for a simple pop screen (see below), even on a directional mic.
- Foam type with an enclosed trapped-air volume between the windscreen and the acoustic openings in the mic. This is better than the close-fitting ones for directional mics and may be adequate for mics on booms and fishpoles operated indoors. A windscreen is needed indoors for two reasons: panning the boom will generate noise, and room air may have drafts that could result in low-frequency noise unless windscreened.
- Basket type usually made of an open plastic grille or other such material covered with fairly tightly woven nylon or silk cloth. The mic itself should be placed as far from the outer covering as possible inside the basket windscreen. Usually this type is made in two pieces that attach to each other. These may be round, for most mic types, or cigar shaped for interference-tube microphones. In the latter case, they may be called zeppelins.
- Fur-like overcovering for a basket-style windscreen. These outer devices, which go by a number of trade names such as Windjammer, may provide up to 12 dB of added wind noise reduction. This is the reason you most often see these on microphones used for production sound out of doors.

Some considerations for windscreen usage are:

- Windscreening is always needed on podium and other microphones placed close to lecturers that would otherwise be subject to pop noise, unless such a feature is built into the mic and is effective (and there are not very many of those).
- When working out-of-doors under severe conditions, multiple layers of windscreening may be necessary, starting from the outside-in with fuzzy materials to slow down air but allow sound through to a woven silk windscreen. For directional mics, though, it is important to provide a region of low turbulence around the microphone, where the wind component is identical at the pickup entrances in the microphone body. For this reason, it is a bad idea to use a foam windscreen inside a stretched windscreen, which might otherwise be tempting

It is worth remembering that pressure microphones, although omnidirectional, are less wind susceptible, so if it is possible to get a microphone in close, it may produce better sound quality than a pressure-gradient mic at a distance, where there is wind.

Silk Discs

In studio work, a disc of woven silk, possibly having multiple layers, about 4–5 inches in diameter, can be placed between the talker or singer and the microphone to shield the mic from direct breath and prevent pop noises. One trade name for these is Popper Stopper.

Microphone Cables and Connectors

The short special limp cable that goes between the microphone body and its boom or fishpole was described on page 90.

Conventional microphone cable utilizes the balanced line approach described in Chapter 3 to reduce the susceptibility to hum. Several considerations apply especially to microphone cables, as opposed to most other professional cable uses:

- Microphone cables need to retain flexibility despite temperature variations because of their being used on location. One specialty type permits use in extreme conditions, to $-40\,°C$, which is incidentally also $-40\,°F$.[8]
- Microphone cables are subject to "microphonics." Moving the cable, or striking it, can cause extremely small output voltages from the cable alone, but given that microphone output is also small, this may be heard. Although all cables could be subject to this, it is only in the case of microphone cables that susceptibility to microphonics makes any practical difference. Some manufacturers have a test procedure for various cable types so that they can be compared.
- Balanced microphone cables help reject external magnetic fields created on sets mostly by cabling for lighting equipment. The very large currents involved in set lighting cause the cables to radiate significant magnetic fields, so microphone cables that best reject this noise are useful. This means that it is a good idea to space microphone cables away from lighting cables, and when they must be crossed, do it in a perpendicular manner. This problem caused a lost first day of production on the first English sound film, directed by Alfred Hitchcock.[9] No one understood the source of the hum that permeated the recordings until someone figured out that it was the lighting cables radiating fields and spaced the mic cables away from the lighting cables.
- One particular type of cable helps to reject the noise caused by dimmers and other items that cause the stray fields present on sets to be polluted with many frequencies. This type is called Star Quad and is made by a number of manufacturers. Two pairs of tightly twisted internal wires, wired in parallel (usually two white wires together and two blue wires together), reject magnetically induced noise up to 20 dB compared to ordinary microphone cable. This rejection occurs because of the tight twisting of the pairs yielding tight coupling that helps to reject noise.

---

[8]Mogami POLAR FLEX.

[9]Explained by sound man Edward Bernds in a talk to the Cinema Audio Society. He went on to direct many Three Stooges movies, proving sound men can make it into directing!

This page intentionally left blank

Handling the Output of Microphones

WHAT IS THE OUTPUT OF A MICROPHONE?

The output of a microphone may be analog or digital, and each will be applied to a corresponding analog or digital microphone input on a mixer or a camera. Alternatively a microphone output, usually analog, may be connected to a radio microphone transmitter for sending the signal to a corresponding receiver, which is connected in turn to a microphone or line-level input of a mixer or camera. All these cases are covered in this chapter.

Analog Microphones

Conventional analog microphones change or "transduce" acoustic energy into electrical energy represented as the voltage across the microphone's output. The output usually appears between pins 2 and 3 of a three-pin male XLR connector,[1] with pin 1 connected to the shield and body of the microphone, intended for grounding. The signal on pin 2 is said to be "in phase" with the acoustic energy, so positive-going sound pressure on the diaphragm results in a positive voltage on pin 2 relative to pin 3. Pin 3 is the inverted polarity signal; positive sound pressure results in a negative voltage at pin 3 relative to pin 2.[2]

In most cases, the output voltage of the microphone is delivered by means of balanced microphone cabling, described in Chapter 3, to a microphone preamplifier.

The potential voltage range coming from microphones is enormous from their noise floor to the maximum they can stand without distorting, easily exceeding 120 dB or 1,000,000:1. Luckily in most practical recording situations the whole range is not exercised, because room or outdoor noise is much greater than the microphone noise at the quiet end, and the loudest sound may not challenge the

maximum level that the microphone can handle. However, there are times encountered, especially in film and television production, in which a great deal of the practical range of the microphone may be used, and it may even need to be extended toward louder undistorted capability using a method described later. The range of most microphones used for film and television production is greater than that of the microphone preamplifiers to which they are connected: it is a little like putting a gallon of water into a quart jar—you cannot just do it simply. You can put in a portion from the "top," the "middle," or the "bottom," but not all of it simultaneously.

How can the dynamic range of a microphone that can range from microvolts to volts be contained to a range that a preamplifier can handle?[3] The first answer to this question is that it is far worse to have clipping distortion, because the microphone puts out so much level that it clips or overloads the input, than it is to have too little level and thus have to advance the gain ("turn up the volume") to pull the signal up and, thus, increase the noise. That is because, generally speaking, it is easier to reduce noise in postproduction than it is to reduce distortion, which may be impossible. Therefore, it is better to underrecord somewhat than to risk overload clipping, especially in digital systems, and "contain" the microphone output to a range that the input of the preamplifier can handle.

There are two basic methods to reduce the effect of large outputs from microphones: pads and preamplifier gain controls. Pads are attenuators that are placed either between the microphone capsule and its own electronics or between the microphone output and the microphone preamplifier input. A potential problem with the latter is that the pads may interfere with powering the microphone from the microphone preamplifier, so they must be assessed as to whether they will have an impact on powering. Some external battery-power boxes provide a pad function so that proper powering and the capability of padding is provided in one unit.

Microphone preamplifiers are also available with variable gain controls. These are rated and some are marked in

---

[1]Smaller connectors are used for microphones intended for connection to a radio mic transmitter, but the same issues apply.

[2]This is true for all microphones sold for audio pickup. However, because of a difference in convention, measurement microphones usually use the opposite polarity. However, their output is rarely on an XLR connector, and they are infrequently used for audio purposes, so little confusion exists in the field.

[3]An extensive article is at http://booksite.focalpress.com/Holman/SoundFilmTV/.

decibels of gain of the preamplifier, with typical gain from 20 to 60 dB or more. These are usually set by the experience of the operator for the sound levels to be encountered, the microphone type, and its spacing from the sound source. It is possible, however, to calculate the required gain, knowing the sensitivity of the microphone and the sound pressure level expected from the source.

Where to Put the Pad/Gain Function

There are three potential points in the system of an analog microphone connected by wire to a mic preamplifier at which the level can be affected, with consequences on headroom and noise. In the case of radio mics, this becomes even more complicated, as there are several more level controls with which to deal. So some rules need to be developed to describe where the best position is in the chain. Here are some rules of thumb:

- If an electrostatic microphone is being used and the sound pressure levels are quite high, exceeding the undistorted sound pressure level capacity of the unaided microphone,[4] then a preelectronics pad is necessary. This may take the form of a switch on the microphone body for attenuation, or it may take the form of an accessory screw-in pad. You can get an idea of the SPL by using a Radio Shack sound level meter, but note that even on its fast setting, it responds in $1/8$ sec, and we can hear distortion in as little as 2 msec, some 60 times faster. So it can give only a rough idea of the actual peak sound pressure level that can be handled without distortion. Almost all sounds are covered by adding 15 dB to the reading of a fast-responding sound level meter to get to the true peak value. So if the meter reads 120 dB SPL (C weighted, fast reading), assume the level is 135 dB SPL for headroom purposes.
- If an electrostatic microphone is of the RF type (not radio mic, but internally using radio frequency modulation and demodulation, such as Sennheiser MKH models), then the only solution for very high sound pressure levels is to move the mic away from the source.
- If very high sound pressure levels are encountered, one solution is to move to less sensitive electrodynamic mics with no electronics or to specialized high-level electrostatic ones (with a corresponding high noise floor). These would be useful for gunfire, for instance, for which the peak SPL may reach up to 170 dB SPL.[5] A 0.357-caliber Magnum pistol produces a measured 165 dB SPL for 2 msec. The database at http://www.microphone-data.com lists 3 models

specified by their manufacturers with maximum undistorted level (typically for 1 percent total harmonic distortion) of 170 dB SPL or greater and 12 models with 165 dB or greater.

- If the above options have been taken into consideration so that the microphone output is undistorted, then its maximum output voltage needs to be matched to the maximum input of the following preamplifier, if high-level sounds are encountered. Not all situations require this, because not all situations are at sound pressure levels that will clip the input of the preamplifier, but it remains a "gate" to performance that should be understood. With experience mixers find that they know going into a particular situation what to expect, and having experienced it before, they can arrange the items that affect dynamic range (preelectronic pads for electrostatic mics, post-mic pads, mic input stage gain controls) properly. However, facing a new situation may require some guesswork. It is helpful to know the maximum sound pressure levels to be encountered in various situations, and some are given in Table 6.1. The L_{eqA} column represents the long-term average and is most useful for low levels, to compare them to microphone noise floors. For example, a microphone for recording in a quiet room, but not a recording studio, should have a noise floor reasonably below 20 dBA equivalent not to add noise to the recording. The measurements marked "Flat, peak, impulse" are the fastest way that a professional SPL meter reads, so it catches close to the true peak level that must be reproduced undistorted. However, it still requires 35 msec to respond, not the 2 msec in which distortion can be heard, so some number of decibels still needs to be added to be sure that the sound is undistorted. One thing that the table demonstrates is the very large difference between the average level and the maximum level, exceeding 30 dB in some cases.

Take for example a high-quality electrostatic microphone plugged into a DV camera microphone input. This example is for the Schoeps CMC641Ug supercardioid plugged into a Panasonic AG-DVX100Ap DV camera. The camera's mic input overloads at 48 mV. The microphone has a sensitivity of 13 mV/Pa and maximum undistorted sound pressure level capability of 132 dB. We need to make a calculation of what voltage comes out of this microphone at 132 dB. The reference level of 1 Pa equals 94 dB SPL. So 132 dB − 94 dB is 38 dB greater, so the mic output is +38 dB re: 13 mV.[6] This is almost 40 dB, which is a factor of 100, so the mic can put out a little less than 1.3 V, or 1300 mV, a long ways above the camera's 48 mV clip point. To calculate this more exactly, 38 dB is a factor

---

[4]This may be found at http://www.microphone-data.com.
[5]See http://www.freehearingtest.com/hia_gunfirenoise.shtml.

[6]See the web addendum Calculating Decibels at http://booksite.focalpress.com/Holman/SoundFilmTV/.

TABLE 6.1 Some Sound Pressure Levels Found in Representative Situations

Item	L_{eqA}	SPL[a] Flat, peak, impulse
Quiet room suitable for recording without extra precautions	20	68
Fairly quiet room but would require careful manipulation of background in both editing and mixing	28	62
Apple G5 measured at user's head, with computer on the floor below, fan at low level	36	60
Quiet Hollywood street, peak represents car pass by; suitable for shooting at least close-ups especially if traffic could be controlled	46	65
Quiet model dishwasher at 1 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 due to city bus drive by across the street	66	105
Interior Los Angeles delicatessen (Langer's Deli) at lunchtime	70	104
Normal speech measured at lavaliere position	75	95
Urban street including bus passing observed from bench at side of street	82	107
Actor screaming at 0.5 m		128
Live music performance, in audience		129[b]
Inside the bass drum head of a rock band (REO Speedwagon)		138
Large orchestral bass drum mic'd at 6 inches off the drum head		138
0.357 Magnum pistol measured at the ear of the shooter		165

[a]The reference level for SPL is 20 μN/m².
[b]Fielder, L. D. (1985). Pre- and post-emphasis techniques as applied to audio recording systems. *Journal of the Audio Engineering Society* 33, 649–658.

of $10^{(38/20)} = 79.4$. Then 13 mV/Pa × 79.4 = 1.03 V. A pad must be used to bring this down to 48 mV, to record the loudest sounds undistorted. We will not want this pad in all the time, because the microphone has much more dynamic range than the recorder, and we'll find with it in

that we have to advance the mic level control so much that noise will result, but the pad will be necessary for really loud sounds. To find the value of the pad, make the following calculation: dB = 20 × $\log_{10}$ (1.03/0.048). The answer for the value of the pad is 26.6 dB. Note that a simple pad will in this case effectively block the camera's phantom power to the mic, so a separate battery-operated phantom power supply will be needed, wired between the microphone and the pad.

- Another way to look at this is that without any pad the mic preamp overloads at 132 dB SPL − 26.6 dB ≈ 105 dB SPL. So the unpadded combination overloads at 105 dB and will probably distort occasionally with just a bus passing by, as can be seen from Table 6.1.
- A method for building pads to specific values is given at http://booksite.focalpress.com/Holman/SoundFilmTV/.

CASE HISTORY

The first thing to consider is whether the expected sound pressure level at the microphone might clip the microphone's own electronics in the case of electrostatic microphones. For a New World Symphony recording that included a close-mic'd large bass drum I used my memory of a measurement of another bass drum from another time. I measured the inside of a bass drum head belonging to REO Speedwagon in my studio in the environs of Champaign, Illinois, in the late 1960s by putting a Shure SM-57 dynamic mic[7] inside the head and feeding it into an oscilloscope directly. By calibrating the scale, I was able to find that the peak sound pressure level was 138 dB.

The Schoeps MK2 omni capsule and CMC6U electronics to be used clip at just over 130 dB, but they don't reach 138 dB. So we needed pads, ones that screw in between the capsule and the electronics. We used 10 dB pads on the percussion spot mics, including on this bass drum, as these were the ones that were close enough to instruments loud enough to cause potential problems, and the signal-to-noise ratio was not harmed as the instruments were so loud that their spot mic contribution was well down in the mix.

To calculate the mic preamp gain, we use the microphone sensitivity, the pad value, and the input sensitivity of the analog-to-digital converter in the recorder for full scale to determine the unknown in the overall equation— the gain setting of the microphone preamplifier. The mic sensitivity is 15 mV/Pa, the pad is 10 dB, and the input sensitivity of the recorder used for full scale is 18 dB over +4 dBu, namely +22 dBu. Let us take 138 dB SPLpk as our level that must be handled cleanly. Then the rms level is 135 dB, and the rest of our calculation can be in rms.

[7]Dynamic mics can have an overload too, when their diaphragm bottoms, but it is typically significantly above 140 dB SPL in most types.

- With 15 mV at 1 Pa (which equals 94 dB SPL), 135 dB SPL is 41 dB greater in level. The 10 dB pad takes this down to 31 dB hotter. Thirty-one decibels up from 15 mV is 530 mV (calculated by dividing 31 by 20 and then raising 10 to the power of the remainder: $10^{(31/20)} = 35.5$ times).
- Multiply 15 mV (the sensitivity) $\times$ 35.5 (the factor that the SPL found is above the reference SPL) to get 530 mVrms. This is the undistorted output of the microphone that we must capture cleanly.
- Now the input overload of the A-D converter is $+22$ dBu $= 9.76$ Vrms (from $10^{(22/20)} \times 0.775$ Vrms (the reference level for 0 dBu) $= 9.76$ Vrms).
- Then calculate the difference between the input of the A-D overload and the microphone output to find the maximum permissible gain of the mic preamp: $20 \log (9.76/0.53) = 25$ dB. So the maximum preamp gain we can use is 25 dB, and to leave a little headroom beyond 135 dB SPL in case this drum is louder, let's make it 20 dB. We did. It worked. The maximum recorded level hit about -3 dB FS, and with 24-bit recording, we had a huge dynamic range captured.
- Because we hit -3 dB FS with only 20 dB of mic preamp gain and $+22$ dBu FS A-D input for 0 dB FS, the actual peak sound pressure level we got off the drum head was 138 dB SPL (peak measurement). Interestingly it matched the inside of the rock bass drum measured some 30 years earlier.

QUIET SOUNDS

Up to now we have considered mainly keeping the microphone, its associated preamplifier, and a corresponding analog-to-digital converter undistorted, and this is most important. However, in certain circumstances it is the other end of the dynamic range that most interests us. For instance, in Foley recording of sound effects to match picture the whole idea is to capture mostly low-level sounds and reproduce them at higher levels than in real life, as this tends to make things sound real. In this case, and in the case of ADR, in which there's plenty of silence between and around words, the noise level comes to the fore.

So that the microphone preamp will add negligible noise to that of the microphone itself, it must be quieter than the microphone. If the microphone and its preamp happen to have the same noise floor, the overall noise will be increased by 3 dB. If the preamp is 10 dB quieter than the microphone, the preamp adds essentially no noise to that of the microphone itself, so we have captured everything that can be gotten from a given situation. Of course if the level is still low, and the noise level due to the microphone alone is high, then the mic will need to be moved in closer to the source to capture higher sound level, allowing

the microphone's contribution to noise apparently to be reduced, or a quieter microphone may be necessary.

The noise floor of most microphone preamplifiers is in the range of an equivalent input level of -128 dBu. This equals 0.4 µVrms, a very small number. One of the best microphone noise floors is the Neumann TLM-103 at 7 dB A-weighted acoustic noise floor equivalent. How does it compare to the input noise floor of typical preamps?

- The mic's sensitivity is 23 mV/Pa.
- 7 dBA is 87 dB below the reference level of 1 Pa, which equals 94 dB SPL.
- 87 dB below 23 mV is $10^{(-87/20)} \times 0.023 = 1$ µV.[8]

This calculation is a bit sloppy because the mic noise is A weighted[9] and the preamplifier specification is the 20-kHz band noise, unweighted. Actually this improves things, because if we were to A weight the preamp noise for an apples-to-apples comparison it would mean that the mic preamp noise is even less of a problem, as it would be lower. A typical difference between audio band noise and A-weighted noise is 3 dB, so that would take the noise floor of the preamp down to 0.28 µV, which is 11 dB below 1 µV, and thus the preamplifier noise is swamped by the microphone's self noise, and at this very low level of noise, it means sound can be recorded cleanly and without added hiss down to a very low noise floor. So for Foley recording, for example, use of a low-noise microphone with high sensitivity in a quiet space close to the source provides the ability to "reach" to extremely small sounds. Because Foley is often exaggerated in playback in a mix, to make things seem hyperreal, it is a good idea to have a very quiet recording in the first place.

A common misnomer about microphone noise is the thinking that because electrodynamic mics contain no electronics they make no noise. This is actually true at absolute zero room temperature, but at real temperatures the Brownian motion of electrons caused by heat leads to Johnson noise. Any real impedance at room temperature exhibits Johnson noise. Once calculations are made for this noise, compared to the sensitivity of dynamic mics, it may be seen that dynamic mics are no quieter than electrostatic ones.

IMPEDANCE

Microphone impedance is technically the relationship between its output voltage and its output current. Today the standard is to make microphones with very low output

---

[8]Tables to do this for various combinations of microphones and preamplifiers are at http://www.rane.com/note148.html.

[9]A frequency-response equalization curve that makes the result more like human hearing, peaked up in the 2–4 kHz range and rolling off below and above there.

impedance, such as in the range of 25–200 ohms, and to make the preamplifier's input impedance relatively high, such as 2 kohms (2000 ohms). The microphone is barely loaded at all by the input impedance, so the voltage is kept practically constant, even when connected to the preamp. An analogy is plugging light bulbs into a power circuit: the power line has a very low impedance, and the light bulbs have a high impedance.[10] The voltage across them does not change much at all as you add more bulbs, until you reach a limit at which a circuit breaker opens.

This system allows for several things:

- The mic cable can be long without concern over response variations due to the impedance of the microphone cable. Several hundred feet of cable is not usually a problem.
- Mic "splitters" can be used. By using a multiwound transformer, one microphone can be used to drive one winding, and two or even more preamplifiers can be fed from different additional windings. This provides isolation among the parts while driving the various preamplifiers with virtually all the output voltage from the microphone because one low-impedance mic can drive multiple high-impedance loads, and the transformer provides isolation. Such a setup is especially useful in complicated projects such as the Academy Awards, in which one mic must feed the main broadcast output as well as the stage monitor, for instance. One item to note in using transformer splitters is that transformers will not pass phantom power, and thus it has to be provided on the mic input leg of the transformer.

An older method was to "match" the output impedance of the microphone with the input impedance of the mic preamp. This was really only applicable to dynamic mics, for which it does result in the greatest transfer of power rather than the highest voltage from mic to input, but this method has passed into disuse today.

Some preamplifiers usually used in music studios today provide adjustable input impedance. With electrodynamic microphones including ribbon types, response changes may occur with various degrees of "loading" or input impedance of the microphone preamplifier, so such adjustable impedance may make audible response changes that could be beneficial in a given circumstance. However, equalization later can do the same thing, so the use of adjustable input impedance is film and television production in which there are so many added variables is not likely to be productive.

DIGITAL MICROPHONES

Today most microphones produce voltages proportional to sound pressure[11] that need amplification by external microphone preamplifiers and then conversion to a digital signal for recording via an analog-to-digital converter. The conversion to digital is needed because almost all recording today is digital. However, some models have appeared that directly produce a digital output stream to the Audio Engineering Society standard AES42-2006,[12] thus eliminating external microphone preamplifiers and analog-to-digital converters. An example of a digital microphone is the Neumann TLM 103D, a large-diaphragm studio microphone especially useful in Foley for its low internal background noise floor. Digital microphones are not yet in common use in film and television sound production, but they hold out the promise that in the future a major problem addressed in this chapter may be eliminated.

A wide-dynamic-range mic such as the analog version of the Neumann TLM 103 has a range of 131 dB or a ratio of 3,550,000:1, and all signals in this range ought to be handled by the subsequent equipment so that no capability is lost. This is a range from 1 µV (one-millionth of a volt) to 3.5 V, a very formidable design challenge for the preamplifier. In fact, almost all microphone preamplifiers have to be hand adjusted in some way to handle this wide a range, and virtually none can handle both the smallest and the largest signals simultaneously.

"Digital" microphones are easier to use than analog ones, because their output bitstream can be designed by the manufacturer to handle both signal extremes simultaneously. By recording to a digital medium with an adequate bit depth (such as 24-bit recording, equivalent to 141 dB dynamic range[13]), the full dynamic range is preserved. If the level from the microphone is "too low," it can subsequently be "turned up" in the digital domain, with only the theoretical noise floor of the microphone to limit this process by adding audible noise.

> The manner in which digital microphones attain their extraordinarily high dynamic range is by converting from analog-to-digital simultaneously at several levels. For instance, a high-level and a low-level converter may be used and the output stitched together digitally, sample by sample, to extend the dynamic range of the converters. When the low-level converter is overloaded, its output is ignored, and a digital signal processor algorithm switches over to the high-level, undistorted converter, seamlessly splicing together high and low levels.

---

[10]Light bulbs are specified indirectly for impedance by being rated in watts. A 100-W bulb has an impedance of 144 ohms ($R = E^2/P = 120^2/100$).

[11]Or sound velocity, in the case of the figure-8 bidirectional pressure-gradient mic types.

[12]Available at http://www.aes.org.

[13]Note, not the more commonly seen calculation 144 dB because every linear PCM system requires dither to linearize the low levels of the system. See Chapter 3.

There are two connection "modes" for digital microphones, called in the standard Mode 1 and Mode 2. Mode 1 microphones contain their own sample rate clock and are thus fairly simple. However, a problem is that the following device, such as a recorder, must do one of two things: it must follow the clock in the microphone or it must sample rate convert to account for the tiny difference between the clock rates of the microphone and the mixer or recorder. Mode 1 microphones work without any additional complexity only in the case of having only one microphone connected to a freestanding recorder that can follow the microphone clock. If two mics are used, their clocks will be every so slightly different, and the result will be clicks or snats occurring from infrequently to frequently, depending on how closely the clocks match, unless each input has its own sample rate converter. Because good sample rate conversion is complex, it requires a noticeable amount of power, something not easy to get in battery-powered location recorders.

At the cost of greater complexity, Mode 2 digital microphones are fed information from the mixer or recorder back to the microphone over the microphone cable to "lock it" to the mixer's or recorder's word clock. No clicks should occur. Mode 2 also provides the capability for other microphone control functions remotely, such as inserting a pad between the capsule and the electronics to extend the maximum sound pressure level, adding a high-pass (low-cut) filter in the path, changing level or gain, and even potentially changing the polar pattern.[14]

Digital microphones are currently connected to the microphone cable with standard XLR connectors. This could lead to confusion, mixing up the signals from analog and digital microphones, which could potentially even result in damage. So a memory device such as using different color cables between analog and digital should be used to prevent errors. Using special AES3 digital cable rather than conventional microphone cable for the mic interconnect extends the range of the microphone cable to more than 1200 ft, without sonic problems, a large advantage in some situations.

DIGITAL MICROPHONE LEVEL

The advantage of digital microphones is their very wide dynamic range. Nonetheless the maximum sound pressure level limit may be exceeded. For Mode 1 digital microphones, the same remedy of a preelectronics pad applies. For Mode 2 microphones, with the correct external hardware such as the Neumann DMI-2 interface and an associated computer and software, a preelectronics pad can be inserted remotely via the computer for some models.

THE RADIO PART OF RADIO MICS

Selecting Radio Mics

The term "radio mics" used in film and video sound usually means lavaliere microphones attached to small belt-pack transmitters, often used with similar-sized receivers either on a sound cart or placed directly on the camera. However, there are other instances, such as hand-held microphones with the transmitter internal or external 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 transmitter/receiver pair, so what distinguishes them across this range? You might think that it could be the transmission power, because more power will reach further, but although that certainly makes a difference, there are other factors involved.

FIGURE 6.1   Several models of radio microphone transmitters. Photo courtesy Lectrosonics, Inc.

FIGURE 6.2   A radio microphone receiver for use with the transmitters shown in Fig. 6.1, a Lectrosonics UCR411A. This model includes a spectrum analyzer function described in the text. Photo courtesy of Lectrosonics, Inc.

---

[14]Some variable-pattern microphones accomplish the variation mechanically, so these are excluded, whereas others perform the variation electrically, and these may be remote controlled for pattern adjustment.

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 could use a wired microphone for best quality and reliability but you would be giving up the freedom of movement that is so cherished among actors and subjects, and with a good wireless system you can come close enough to the wired condition for practical purposes. Actually, in practice, the position of the microphone on the body—and the covering of the lavaliere microphone in the case of fiction films—ultimately limits the quality more than does a good radio transmission link so long as certain recommendations given below are followed (and these limitations apply to a wired lavaliere microphone as well as radio ones), and measures to mitigate the audio problems are described in Chapter 12.

Here are some qualities that make radio microphones have such a wide price range:

- Better radio frequency filters that are used in more expensive systems will work under more difficult situations. This is because the filters that separate the desired signals from undesired ones will be more elaborate in the more expensive ones and do a better job of rejecting other radio signals in the vicinity that could result in interference.[15] So although a cheap one may do in rural areas with no competition from other radio frequency sources, a more expensive one may be necessary to separate out signals in an urban area with lots of transmitters.

- 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 if 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 one on which no one else is broadcasting. If you had bought a fixed-frequency transmitter/receiver then you could try some of the suggestions under Radio Mics in Use, discussed later, or if you had rented them they could be exchanged for ones on different frequencies, but both of these are clumsy solutions.

- Better systems make use of true diversity reception. In the real world, a phenomenon known as multipath results in audio distortion of usually intermittent "fritzing" noises or even complete signal dropouts. Under multipath conditions the receiver gets one principal signal from the transmitter, but also a reflected signal

off a building, for example. 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 listening to FM in a big city while driving around in a car. The momentary noisy dropouts are due to multipath reflected off the buildings. You can even find that if you drive up to a red light there may be no signal at all, or a very distorted one, and then you move a foot or two and the signal is restored. This is the result of multipath.

- The way to minimize multipath problems is called diversity reception. Here two antennas physically spaced apart are fed to two receivers and their output is scanned, and the output that is chosen is 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. "Full diversity reception" is as described, with two full receiver systems used, in one or several chassis, and obviously adds cost. "Antenna diversity" systems attempt to add signals from two antennas together to achieve a similar goal, but such efforts are quite limited in quality. Directly adding the output of various antennas together electrically is a bad idea because then you are creating great potential for multipath.

- Because of the nature of noise on FM, which is used for most radio mics—and which without aid is very noisy—transmitters and receivers each use complementary companding noise reduction, with the transmitter variably boosting the highs before sending and the receiver cutting them back in an equal and opposite way. The attempt is to "hide" the noise of the channel behind the program content. The companders used are of various levels of quality as this is a tricky process, varying in time, and can add artifacts to the voice, so a more expensive one is likely to be a better one, although that might not always be true. One classic acid test for compander problems is to try to transmit the sound of jangling keys: poor companders screw up on this test.

- Audio performance varies from model to model. Because the dynamic range of the radio frequency channel is limited, it is commonplace to employ an audio limiter, putting a cap on the maximum level and thus keeping distortion due to overmodulation low. These limiters can range from benign to quite obvious in use, depending on their design and how hard they are pushed. Sometimes they may even need to be shut off. To be the most benign on loud speech, limiters turn the gain down very quickly but restore it slowly. If an actor fires a weapon in the middle of speaking, however, the limiter will turn the gain down a lot upon the firing of the gun, and the "duck" in the speech following the gunshot will be obvious. It is better to simply distort during the gunshot and then an

---

[15]This is akin to the difference between a poor FM radio and a good one; greater selectivity in the filters that pick out one station from among others is a distinguishing feature.

editor can cut it out and replace it, with little effect on the following speech.

- Some newer systems employ a hybrid of digital coding and analog FM transmission, or all-digital solutions instead of FM transmission, with claimed benefits in reliability of the channel, which is what radio mics are all about. To do this the audio must be strongly "coded" to fit in the channel, and this coding may or may not be audibly transparent, so it is a good idea to audition these before needing them. Also, the analog-to-digital conversion in the transmitter and the digital-to-analog conversion in the receiver add audio delay. When such mics are added to boom mics in postproduction mixing, the result may be audible comb filters, sounding like the voice is originating in a barrel. One model has 3.2-msec delay, so that tends to make up for the difference in spacing between boom mic and lav, but problems can result because of the time delays of such microphones.
- The most elaborate hybrid or digital systems may employ encryption, to prevent someone from "tapping" the transmission. With high-profile movies this may be a consideration, but it probably does not affect most producers.
- So far all known radio mic systems use FM or digital transmission assigned to one particular frequency (for the "carrier"; the audio-modulated radio frequency signals extend out around the carrier). In the future, "spread spectrum" systems might come on the market after certain difficult problems are solved. 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 one 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, during World War II. Kept as a government secret for many years and finally first employed during the Cuban Missile Crisis, Lamarr's method of communication was to use something very like player piano rolls synchronized in both transmitter and receiver to determine their carrier frequency moment by moment (there were 88 frequencies in the original, which equals the number of keys on a standard piano) and thus to "frequency hop" the carrier so the transmission could be hidden. Large uses today of related technology dating back to Lamarr's World War II invention are GPS, the global positioning system, and cell phone technology.

Radio Mics in Use

The U.S. radio microphone market was significantly changed by FCC actions with the introduction of digital television and the shutoff of analog services in 2009.

Frequencies above 700 MHz are no longer legal for radio mic use for film and television production, and further complications exist for both VHF and UHF radio mics below 700 MHz because the "white space" shown as holes between the rearranged television channels may be used by other services in the future, such as "Wi-Fi on steroids."[16] However, the FCC has recognized the importance of radio mic uses in its documents, especially for film and television production, so interference in the future into radio mics should be prevented by the measures imposed on manufacturers of new products that go "in between" television channels.

Frequency Coordination

Frequency coordination is necessary in many practical uses of radio mics. It is the selection of a frequency for each radio mic in use such that no interference is received or caused by each transmitter. If you were to use one radio mic in a rural area well away from sources of radio frequencies, then an inexpensive one may do, even a fixed-frequency one, and then getting the right frequency is not an issue. However, if you need multiple radio mics in a city with digital television transmitters, then you've got a much bigger problem, perhaps an insurmountable one for a given budget. Use of frequency-agile transmitter/receivers may be absolutely necessary, and higher quality ones with better radio frequency filters to prevent interference from becoming a problem may be necessary. Some radio mic receivers such as the one illustrated in Fig. 6.2 have a built-in spectrum analyzer that shows transmissions across its range of frequencies so that one frequency can be selected for each transmitter in that band (frequency range) to produce minimum interference in the channel.

Software is available by means of a web database with known fixed transmitters in the FCC database, such as those for digital television, so that frequencies can be chosen for radio mics at a given site that should minimize interference from other sources of radio frequency energy. See www.professionalwireless.com and follow the link to their Intermodulation Analysis System. This site points out the complicated interaction between frequencies of transmitters and possible distortions that can result in interference and helps you choose the best set of frequencies for a given application. Sennheiser has similar software for their models.[17] The software costs around $250, and constantly updated FCC information is accessed by this software, making it a good value.

---

[16]Wi-Fi operates on much higher frequencies, but those frequencies are blocked more effectively by building mass; placing a Wi-Fi-type service below 700 MHz would be more penetrating of structures, which leads to the desire of providers to use this spectrum.

[17]Available at http://www.sennheiser.com/sennheiser/home_en.nsf/root/professional_wireless-microphone-systems_sifm-software.

A broad-band radio frequency spectrum analyzer can be used to measure all the radio transmissions received at a given location across a wide band of frequencies. Such a device is the best arbiter of what frequency to use, as it can even select among the various ranges of frequencies available for various transmitter/receiver systems, because it will catch everything on the air, licensed or not. Kaltman Creations makes a portable battery-powered spectrum analyzer accessory for a laptop PC called a white space finder (http://www.kaltmancreationsllc.com/invisibleWaves.html) for just this purpose and it is well adapted for use by audio people. It costs in the range of $1500, which is a good value considering that radio frequency spectrum analyzers can cost up into the tens of thousands of dollars.

The application employing the most radio mics and other radio frequency devices, such as walkie–talkies, is probably the political conventions held every 4 years in which many hundreds of radio transmitters must be coordinated across government and media. The Super Bowl itself uses over 200 radio microphones.

Minimizing Signal Dropouts and Multipath

The idea is to get the strongest link at the radio frequency in use between the transmitter and the receiver for the best reliability. Here are some things that can be done:

- All other things being equal, radiated power from the transmitter counts. Radio mic transmitters range from 5 to 250 mW, and 100 mW is typical of better units.
- If the transmitter has an antenna separate from the microphone cable itself, usually a whip antenna, see that it is installed without twists and bends and is vertical. If the microphone cable is used for the transmitting wire, consult the operating directions for orientation.
- If the receiver has a short wire antenna (a whip), also see that it is installed without twists and bends and is spaced away from other metal and kept vertical.
- Minimize the distance from transmitter to receiver. In some cases this means you may place the receiver on the set close to the subject and run a long mic cable, which causes no problems with professional gear. But in so doing see to it that the receiving antenna is vertical. Right-angle connectors are available to use with a whip antenna if the receiver must be horizontal.
- Use a receiver whip antenna custom cut to length to match the frequency range of the transmission, and extend it out to the set with low-loss radio frequency cable with the correct connectors on it to match the antenna and your receiver input.
- Alternatively, make a receiving antenna out of a long radio frequency cable and put the receiving end near

the transmitter. The loss in the cable at radio frequencies is overcome by the proximity of the receiving antenna to the transmitting one, and the receiving antenna can easily be hidden on the set. To do this, connect a piece of low-loss 50-ohm radio frequency cable (RG-8/U, Belden 9913F) long enough for the application to a short adapter cable allowing connection to your receiver mating connector, usually 50-ohm BNC or SMA. This low-loss cable has a relatively large diameter at 0.4 inch and may not be bent into less than a 6-inch radius, which is the cost of producing low losses. At the other end of the length of cable, strip the outer insulation back by 6 inches for 470-MHz through 4 inches for 700-MHz radio mics, linearly interpolating between these lengths for the frequency to be used (it is not terribly critical).[18] Carefully unbraid the outer shield and unwrap the inner shield and twist them together into one conductor, cover it in insulating tubing such as heat shrink, and fold it back along the outside of the cable insulation, taping it in place. Spread the inner conductor straight away from the end of the cable, and tape off its end, keeping the cable straight for the length that is folded back plus the length of inner conductor that is sticking out. This is a dipole antenna centered at the correct UHF radio mic frequency. This particular type of cable has a loss of 3 dB at 100 ft at 400 MHz, and moving the receiving antenna to one-quarter of the distance to the receiver has the effect of raising the level by as much as 12 dB, so it is well worth it, even neglecting some connector and adapter losses. Its directionality is just like that of a dipole microphone (pure pressure gradient), a figure 8 rotated around its long dimension. So it picks up best perpendicular to the axis of the antenna.[19]
- Use a directional antenna and orient it for best pickup. A directional antenna, like a microphone, has a "hot" side and a "cold" side. The hot side may be wide, and the cold side narrow. In such a case, it actually helps more to point the null of the pickup at an interfering source than pointing the hot side precisely at the radio mic transmitter. This works both for interfering other transmissions and for multipath.
- Normally the "shark fin" style antennas (they are called log-periodic designs by engineers) are to be oriented vertically to have their widest acceptable angle in the horizontal plane.

---

[18]For a web-based Java application to determine the length see http://www.csgnetwork.com/antennagpcalc.html?frequency=700&gp=0.152&gpf=0.498&gpi=5.974&radial=0.152&radialf=0.498&radiali=5.974. The "vertical" dimension given is the $1/4$ wavelength, which is what the inner conductor length and the folded back shield length should each be.

[19]See step-by-step photos at http://booksite.focalpress.com/Holman/SoundFilmTV/.

- Height counts. Raising the receiving antenna gets it away from bodies, cameras, set elements, etc. In some cases I have put a directional antenna on the mic boom, when no boom mics were in use, and had the boom operator point it at the source (or include pointing the null of the antenna at interfering sources).
- In the case of documentary and reality filmmakers, using an over-the-shoulder bag for digital recorders and radio mic receivers makes for a difficult receiving environment, as the digital audio recorders emit some small amounts of radio frequency energy but are located right next to the receivers. With a well-shielded receiver, extending the receive antenna to over the shoulder will probably resolve these interference problems. A low-cost method is to buy an interconnecting cable with the right connectors on it[20] and cut it in half, and then follow the directions for making the open cable end into an antenna given above. Alternatively, see http://www.lectrosonics.com/catalogs/UniversalCatalogPages/CoaxAntUse.pdf for a commercial product.
- An important potentially interfering source on film sets is the UHF television transmitter sometimes connected to a video tap on the film camera or the video out of a digital cinema or high-definition video camera to feed receivers around the set. Although this eliminates a cable from the camera, and is especially important in Steadicam uses of the camera, these are not the best filtered transmitters and may emit radio frequency energy on frequencies other than that to which they are tuned, thus causing interference into radio mics.

Added Gain Staging Complications in Using Radio Mics

It is important to start level optimizing by setting the level at the transmitter rather than at the receiver. This is because it is the transmitter audio gain control that best matches the range of the actor's voice to the available dynamic range of the channel. Overmodulation will run into limiting and possibly obvious distortion, whereas undermodulation will make for noisy recordings. Sometimes two LEDs are provided, perhaps marked SIG for signal, indicating some noticeable activity in the channel, and OVLD for overload, 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. A whole performance must be included, because overload of radio mic channels upon an actor's 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 lavaliere in use, due to the artifacts surrounding the gain changes. Some transmitters show the actual level in more steps, and some are labeled in decibels, but the attempt is always to keep the signal from the talker up enough to produce nominal modulation and down enough that it does not distort the radio mic channel.

The output of radio mic receivers is often switchable between mic level of nominally $-50\,$dBu, intended for microphone inputs, and line level of nominally $0\,$dBu, intended for line-level inputs. Setting the output to mic and using a line input will result in the levels being too low, and setting the output to line and using a mic input will result in bad distortion. So match the output range switch on the receiver and the input switch or jack on the mixer or camera first, then set the adjustable level control on the receiver and on the mixer or camera for nominal operation (in their usual range).

Radio Mics Conclusion

There are a great many radio mics on the market, and perhaps a simple system will do for your purpose, but it should probably be tested under the conditions of use to prove that. One way to tell what systems seem to work is by consulting film/video sound-specific rental houses for big urban areas if you want to use them in such places. Local companies should know their immediate radio frequency environment and help to prevent interference. Always let them know, however, when UHF transmitters on the camera will be in use, as these continue to be a headache.

---

[20]See arch site such as www.cablesondemand.com for CO-174BNCX200.

Production Sound Mixing

INTRODUCTION

Production sound mixing is a potentially confusing term because it is ambiguous. Does it mean the actual process of operating the mixing console or the general processes that go on in production sound, including logging and set relations? Here we will discuss both, starting with the choices facing the production sound mixer in selection of equipment for a particular job, moving on through set operations, and winding up with set politics. The term *mixer* is also ambiguous, because it can refer to the physical console or to the production sound recordist.

Production sound mixing involves microphone technique, recording, and synchronization and has an impact on editing and mixing. Thus all of the factors involved are spread out across the many chapters of this book. Here, though, we take up some of the specific issues facing production sound. The next chapter, Sync, Sank, Sunk, describes the items involved in synchronization, such as sample rates, time code, and the like.

SINGLE- VERSUS DOUBLE-SYSTEM SOUND

The first sound decision made on any production is whether the sound is to be recorded to a medium separate from the picture, called double system, or whether the two are to be combined on one medium, called single system. With film in the camera, the choice is limited to double-system sound, because sound-on-film systems are used only once prints are made for theatrical exhibition—there is no way today to record sound on film in production.[1] When using high-end digital cinema cameras, some of which look a lot like high-definition video cameras but are intended for original exhibition in cinemas and employ different standards compared to video such as a 2048 (horizontal) × 1080 (vertical) pixel structure with 24 frames per second,[2] a double system is also usually used.

High-definition video cameras have audio channel recording capability. Starting from high-definition video cameras, moving down-market through various grades of cameras and recorders, such as electronic news gathering ones, to consumer camcorders, single system is the most obvious choice, but there may be persuasive reasons to use double-system recording in some cases. These are:

- The audio quality of double-system recordings is likely to be better than that available from on-camera recording in many cases, in particular by having greater dynamic range. Audio is just not treated as important as video in many camera designs, so analog-to-digital conversion has a more limited dynamic range than is available in external mixer/recorders, for instance.
- The range of features possible by having more inputs and tracks and other features described below that are not available on cameras leads to greater productivity and a better result.
- The traditional clapperboard slate is not necessary with single-system recording. However, its use gives everyone on the set of a fiction work a moment to concentrate and focus. Long thought to be a nuisance, in this sense it may actually be an advantage. It was motion-picture director Andy Davis who pointed this out to me—not so obvious technically, but it does affect shooting positively according to Davis, despite the loss of film or video media that this entails.
- The need for a slate can be overcome when necessary such as in documentary and reality situations by transmitting the time code from the camera to the recorder, either by wire or over a wireless link, described below.

All that having been said, the convenience in particular of not having to sync up dailies often weighs in favor of single-system shooting where it is possible. On the other hand, nearly automated synchronization of separate sound and picture when both have the same time code is possible through software.

---

[1]Although there was in the past, particularly for news, in the days before portable video cameras; however, the quality was quite limited.

[2]The jargon for this is "2K" imaging. There are also less frequently used standards with the same pixel structure at 48 fps, and for 4096 × 2160 pixels at 24 fps, called "4K." Because of the difference between standard

2K and 4K imaging being based on area, the data rate representing the number of bits per second is not twice but rather somewhat over four times as much per second for 4K imaging, which has made it a specialized format at this time.

Some media, such as DVCPro50, offer four audio channels, and with a boom fed to one of them, three are available for radio mics, for instance. This is a great improvement on most digital video cameras that have only two audio channels, a great limitation compared to today, when four to eight or more separate audio record channels are available from double-system setups.

I measured the audio dynamic range across a range of video cameras selling from about $3K to $75K+. Among three cameras across this range of the same brand, the $15K class camera was best by having the most dynamic range, whereas the $3K and $75K+ camera were worse and about the same. It seems that audio was taken most seriously in the $15K camera where a high-quality single-system sound was the most likely to be needed. The $3K camera was significantly worse because of its low price and what could be done for the money, and the $75K camera was worse because here it seems the designers anticipated that double-system sound would prevail.

COMBINED SINGLE AND DOUBLE SYSTEM

Sometimes you can do both single- and double-system recording at one time. Such a setup provides the ability for each of its microphone inputs to be recorded separately and a mixdown sent to two tracks to the camera (possibly by two radio mic links to avoid cables). Then recording proceeds on both the camera and the mixer/recorder simultaneously. With this setup we enjoy both worlds: instant dailies because the sync sound on the camera may be used, but also the ability to separate out the microphone signals in postproduction so, for instance, a noise in one mic can be cut out in sound editing without disturbing the other mic channels.

This leaves the problem of sync, because getting rid of clapperboard slates is one reason to use single-system recording. How are the double-system recordings to be synchronized? Here another radio mic could be employed, one sending the camera's time code to the mixer/recorder.[3] The only remaining problems with this approach are that the recordist needs to know when the camera is rolling so that the audio recorder is also rolling, and the radio mic links to the camera may not be as transparent as a wired approach with cables from recorder to camera. Also, the recordist is not monitoring from the camera's monitor output and cannot guarantee that the recording is good on the camera. So another set of radio transmitters and receivers can be used just for monitoring.

With these caveats in mind, this is a good approach, getting instant results but also being able to perform more complex postproduction. It also provides a backup by having redundancy by means of the two separate recordings. There have been camera faults that lead to sound dropouts, and having a separate recording permits replacing the bad camera audio with good tracks. Setting the levels correctly and toning the heads of reels while the camera is recording color bars help ensure interchangeability of the recordings on the double-system recorder and the camera, described further later in this chapter.

NEXT DECISION FOR SINGLE-SYSTEM SETUPS: ON-CAMERA OR SEPARATE MIX FACILITIES?

For the simplest "run and gun" style of documentary videotaping and reality shows, having separate camera and sound mixers may prove to be too clumsy, so connecting microphones or radio mic receivers directly to the camera is common. Such matters are covered in detail in my companion book *Sound for Digital Video*. The issues are the circumstances in which one can employ on-camera mics or when external ones are necessary, among others. An argument in that book is that as a minimum crew moves from one to two persons, the first thing that a second person does for the production is sound, although he or she is likely to be the producer/director and have other hats as well.

FOR DOUBLE-SYSTEM SETUPS: SEPARATE MIXER AND RECORDER OR COMBINED?

Whether single- or double-system recording, the sound recordist may use a separate mixer and recorder, or these functions may be combined into one. In the case of over-the-shoulder work such as documentary, the choice is simple: the combined mixer/recorder is most appropriate. On larger productions, however, the flexibility and added features of a separate outboard mixer may be reason enough to employ one, usually housed on a sound cart.

It should be said at this point that the focus of equipment on the sound cart in the past has been a separate recorder and mixer, given more or less equal weight. Today with hard disc recording on small hard drives, adding a recording function to a mixer is not so difficult, so there are products emerging that do both, while retaining larger console features.

Considerations affecting the choice of mixers and recorder functions are:

- On the simplest mixers, such as the common three-input electronic news gathering (ENG) style, the microphones are assigned with a switch or pan pot to

---

[3]There are special, nonaudio requirements on this radio mic. It must preserve waveforms fairly well to be able to read time code.

two output tracks for recording to the two channels of the camera.

- For more sophisticated mixers and mixer/recorders, there is an ability to send each input to its own output or track so that each microphone channel can be recorded to a separate track of a multitrack recorder, typically from four to eight channels, as well as performing a live mix to two tracks that may be recorded on an internal recorder, an external audio recorder, or the camera. The output used for the individual microphone recordings is usually the direct output of the microphone preamplifier, or it may be after filters and equalizers (see below). This recording is typically made from a point in the chain after channel gain trim controls but before the microphone channel fader, so mixing mistakes are not recorded to the individual tracks, although they show up in the two-channel mix. Separately recording the mics improves editing capability, because a distorted lavaliere channel can be edited out, allowing for a boom to be used momentarily, for instance. If the offending mic channel were to be combined with that from other clean microphones and recorded to the same track, then the distorted source sound could not be separated from undistorted ones in postproduction.

- More sophisticated mixers have extensive monitoring facilities for the sound recordist to hear sound being sent to the medium or switch to sound returned from the medium, thus ensuring the recording has been made, with superimposed intercom to/from the boom operator. They also may provide the ability to produce mixes for other people on the set: director, script supervisor, and other viewer/listeners in today's Video Village (the term for everyone crowded around the video monitors showing what is being captured). These need to be kept free of the intercom traffic, for instance. A consideration is that if the monitoring off the medium is a bit delayed, that will throw off the perception of the director and others who must have an undelayed feed. Thus the recordist and the director need to hear different things.

- More sophisticated mixers have cue-send facilities for input from external playback machines to be sent to actors' cueing loudspeakers or earphones and possibly be rerecorded so that sync with a prerecorded source can be easily established.

- Mixers intended for production sound usually provide a slate mic.

- There may be a Solo function on each input to interrupt the monitor only and listen to one source channel only, to track down faults quickly. The Solo function may be fixed or switchable to a type called PFL, for prefader listen. With this mode, the signal in a mic channel can be monitored in headphones, for instance, and with the fader down, that mic channel is not contributing to the mix.

- There may be a Mute function on each input to allow audio to be hard-switched off when not desired. This is useful to cut out a gunshot in a mic channel that will overload, for instance.

- Rarely there may be an output for a studio ON AIR light, for broadcasting applications, or to the red rotating light outside shooting stages, for film.

- Many mixers today have a form factor that permits sliding faders, which some mixers prefer for their ergonomics over rotary ones.

PRODUCTION SOUND CONSOLES: PROCESSES

Each microphone is connected to an individual microphone preamplifier, either in the recorder or in an external mixer, which then delivers signals to the recorder. Microphone cabling is practically always balanced, using two signal leads and a shield, with three-pin XLR-style connectors, except for some miniature microphones designed to be worn by performers.

Different microphones call for different amounts of preamplifier gain and/or padding, because of:

- Different microphone sensitivity from mic to mic;
- Different microphone dynamic range capability from mic to mic, with some being more capable than others of reproducing high sound pressure levels without distortion;
- Different volume ranges encountered by mics in different locations (a distant mic receives less direct sound than a close-up one).

Also, different microphones use various methods to supply power. None is needed for dynamic or ribbon mics, and 12- or 48-V phantom power (usually marked with a "P" suffix in the model number) is needed for most electrostatic mics today. However, earlier film sound production microphones may have also used T powering, which is not interchangeable with P powering at all.[4] Because missetting a microphone power switch can easily damage some microphones, be certain to match the power switch setting on the mixer to the power requirement of the microphone.

Accommodating Microphone Dynamic Range

Probably more than any other single factor, this is the technical factor that leads to poor recordings that are

---

[4]Furthermore, there were two types of T powering: for studio use pin 2 was wired positive and pin 3 negative, but for Nagra production sound $1/4$-inch recorders the polarity was reversed. Such microphones were made by Sennheiser and Schoeps and were marked with a "Red Dot" to indicate their reversed polarity relative to studio mics.

either noisy or distorted. The huge potential dynamic range that can be produced by microphones is quite often greatly reduced because of there being many potential places to adjust the level and some of them being set wrong. The potential consequences of setting one stage improperly, and making up for the error by subsequent gain changes in another stage, are high distortion or high noise.

The methods to maximize dynamic range captured by the microphone preamplifier are covered in detail in Chapter 6 and briefly are:

- In the case of electrostatic microphones, a pad between the microphone pickup element and its electronics, supplied either built-in or as an accessory to the microphone;
- A pad inserted into the mic line, although this can interfere with powering;
- The gain of the mic preamplifier, which may be adjustable.

Note that all three of these devices affecting level are used before the channel level control, also called a *fader, potentiometer*, or *pot*. Overlooking these potential areas of problems by thinking that "the recording level control will take care of it" probably causes as many or more distorted recordings as actual missetting of the recording level to the medium. The distortion occurs before the signal reaches the record level control in these cases.

The detailed study in Chapter 6 covers only part of the chain, that from the output of the microphone through the microphone preamplifier. Many other points in the chain are subject to such "gain-staging" problems. For instance, in the case of a video camcorder, whether analog or digital, professional or consumer, when used with an external microphone mixer, there are many potential devices that affect the level between the microphone and the mixer's monitor headphones to consider (see Fig. 7.1):

- Possible pad between pickup capsule and microphone electronics, built in as a switch or as a screw-in accessory for modular microphones;
- Possible pad between microphone and microphone preamplifier input (but note powering requirements; some are constructed to pass phantom power);
- Mixer input range sensitivity switch: mic or line level on a per-channel basis;
- Microphone gain trim controls;
- Mixer individual channel level and pan controls;
- Mixer master level control;
- Mixer output range sensitivity switch: mic or line level;
- Camera input range sensitivity switch: mic or line level;
- Camera channel level control: with the other items set correctly, these are set by using tone from an external mixer to set to a standard level, −20 dB F.S., if such a scale is available;
- Camera headphone monitor output level control;
- Mixer headphone monitor level control.

Adding a radio microphone to the mix adds two level controls, one on the transmitter and one on the receiver.
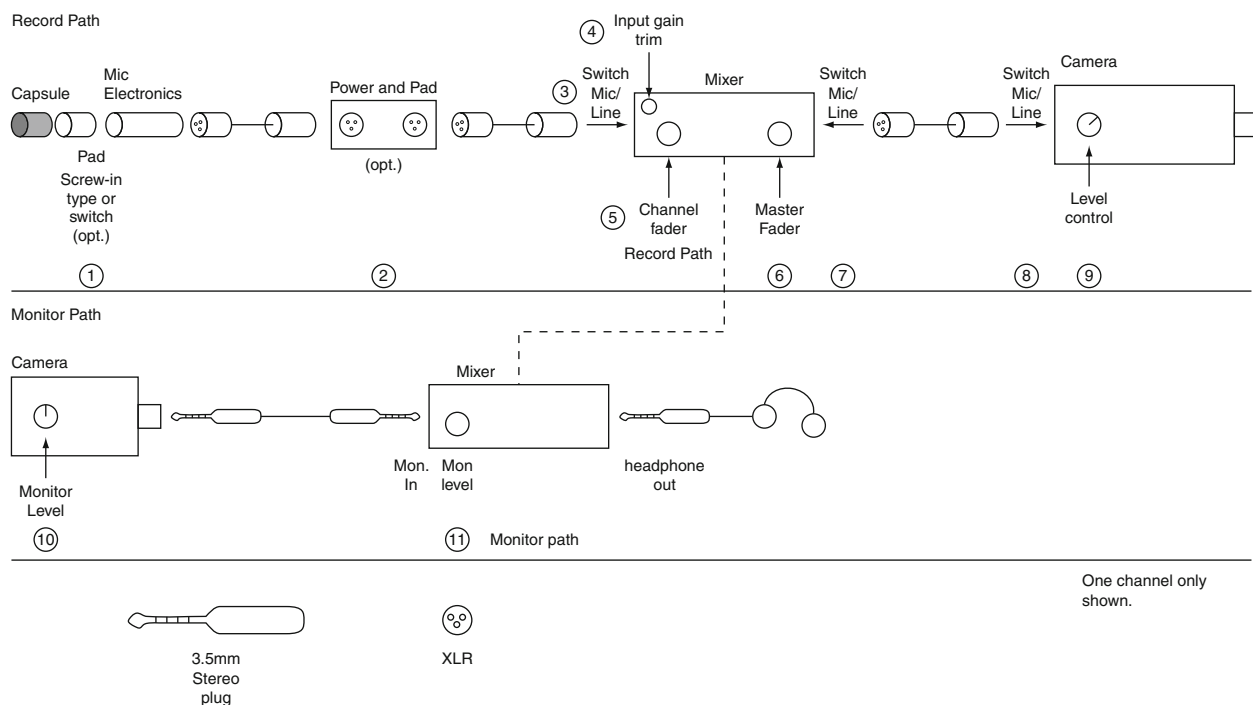
FIGURE 7.1 A single-system signal chain showing all the points at which level may be adjusted.

One not uncommon problem in this chain is for the output level range switch on the mixer to be set to line level and the camera input range switch to be set to mic level. In this circumstance, with typical equipment, a tone can even be sent and recorded at the right level (with the camera input level controls near the bottom of their range), but unless the scene is exceptionally quiet, the likelihood is that the line level output of the mixer will overload the microphone level input of the camera on signal peaks, with gross distortion resulting.

The way out of this complexity is to know what the nominal range is for each of the controls and to be certain that mic level and line level switches are matched between units. For instance, some wireless microphone receivers can output either mic or line level; the corresponding mixer input sensitivity switch must be set to the same setting. Also, knowing where the master level controls, the camera level controls, and the monitor level controls usually are set, and setting them there, increases the likelihood that the chain is free from gross problems. One way to do this is to "tape off" many of the controls in line. For instance, it would not be uncommon to find camera input switch and level controls taped off to standard settings when a camcorder is used with an external microphone mixer. White camera tape available from rental houses, which does not leave a sticky residue when removed, is useful for this task.

Other Processes

After the gain and power issues have been worked out, there are some other potential processing treatments to improve the sound quality. The difficulty with applying these processes in production are many:

- Headphone monitoring conditions do not permit hearing the program under conditions remotely resembling how a standardized sound system will sound at dailies, making judgments of sound quality very difficult on the set or location. However, sound personnel do come to understand the differences with training.
- Any change in the processing from shot to shot within a scene may not call attention to itself at dailies, but may well become a limiting factor on cutting the sound together.
- Any change from scene to scene in the processing may also call attention to itself as we come to learn the timbre of one performer's voice. The change from scene to scene may well be noticeable.

With the foregoing in mind, it is clear that minimum signal processing during recording allows postproduction mixing to produce the best-sounding track in the end. Everyone involved should take special note that good sound on the dailies does not always equate to the most editable tracks. Consistency is more important than one shot sounding great and the next shot not so great, as background noise, for example, intrudes. Of course, one seeks great-sounding dailies, but the cost of "overmixing" the production sound arises subsequently in postproduction, when some of the processes carried out in production must be "undone" to get to neutral territory, where more effective postproduction processing can occur.

Quite often in documentary filmmaking, because of the lack of control over the filming conditions, there will be a need for perhaps greater processing than there is in fiction work, but it is these producers who have much smaller budgets and can therefore not afford as complex a production mixing console. Another difficulty in documentary filmmaking is the need for extremely lightweight and portable equipment. Some of the following may help to ameliorate these problems:

- It is much more important to maintain consistency of sound quality within a scene than from scene to scene. Changes in timbre within a scene disturb auditory streaming and break up auditory objects, whereas each new scene offers a new opportunity for the formation of streams.[5]
- Unifying factors can be used to smooth the sound quality from scene to scene, such as a consistent voice-over narration or music.
- Recording *presence* (also called *room tone* or, in the United Kingdom, *atmosphere*) is an essential ingredient in being able to match background sounds. Its use is described in Chapter 11.

Audio processes are described more extensively in Chapter 12. Those available on field production equipment include:

- Low-frequency filtering: This is used to eliminate the often large amounts of low-frequency noise present on location, which would otherwise be recorded as rumble. A description of filters appears in Chapter 12. Called by engineers a "high-pass filter," and on consumer equipment a "low-cut filter" (or lo cut), on professional audio gear and cameras this might be called either, because the two are the same.
- Low-frequency attenuation (LFA): This is an equalizer that cuts a broad area of low frequencies, compared with the steeper cutoff associated with filters that is more abrupt. It is useful for overcoming the proximity effect in directional microphones used close to a source and the effects of increasing reverberation time at lower frequencies in large spaces. This can help to achieve greater vocal clarity. LFA may be used in addition to low-frequency filtering.

---

[5]See Chapter 2, Psychoacoustics, for a definition of audio streaming.

- Equalization: Two- or three-band equalizers are offered on some devices.
- Limiting: Documentary and even fiction film production sound recording often runs into unexpected, higher sound pressure levels than the recording level control was set for. To provide an undistorted recording, a circuit must "limit" the audio signal so that it is capable of being recorded without gross audible distortion. This is often necessary because performers frequently "play bigger" during takes with the camera running than they do during rehearsals. Poor limiters cause audible "ducking" as they reduce gain to prevent overload, but correctly designed and adjusted limiters can be highly useful.
- Time delay per input: When a boom mic and a lavaliere are recording simultaneously, the sound is picked up by the lavaliere before that same sound reaches the boom. When the two signals are combined in mixing, a strong potential for frequency cancellations occurs, called comb filtering. This occurs simply because of the time offset and the summation. It is worst if the levels are equal and the time is in a certain range. If the times are quite different, then echoes can arise. If the time differences are fairly small the repetition will not be heard as an echo but the sound will be as though the actor were speaking into a barrel. In fact, the mechanical-sounding voice of Darth Vader is produced by repeating James Earl Jones's original performance with a 10-msec, or about $\frac{1}{4}$-frame, delay. To prevent this comb filtering, time of arrival may be adjusted by delaying the earlier signal to match the later one. The Aaton Cantar X2 described later has a method to set the time delay by applying one mic signal to each ear of headphones. Because human perception of delay is fantastically good (10μ sec resolution), it is easy to set. In cases in which the time delay is constantly changing because of boom mic versus lav moves, the best thing to do in mixdown is to make one more prominent than the other as needed, because constantly riding the time adjustment is difficult.

PRODUCTION SOUND MIXERS: SIGNAL ROUTING

Production sound mixers, with generally from three to eight inputs, usually have a fairly simple block diagram. Studying the diagram will reveal how a signal makes it from input to output for recording and possibly separately to monitoring. There may be a Solo function, permitting listening to one or more microphones individually while the main recording channel remains undisturbed. Also, some mixers provide a Mute function, cutting off the output of a channel without having to move its fader.

EXAMPLES

Here are descriptions of four equipment categories and examples within each category. Twenty years ago this list would have been much shorter. Today the examples are by no means comprehensive, as the market is constantly changing as new products are developed, so this is only a survey of typical units. Features change even with software updates. For more up-to-date information than a book can provide consult http://booksite.focalpress.com/Holman/SoundFilmTV/.

Small Mixers

Mixers in this category are used for outboard mixing for recording on single-system cameras. They are small and lightweight and have a form factor and cable connections that allow easy over-the-shoulder use. Typical uses include ENG, documentary, and reality filmmaking. Classic features of small mixers include the following, but note that not all such mixers have all these features:

- "Bag" form factor for over-the-shoulder use;
- Three inputs, on XLR connectors on the left side of the box, switchable to mic or line level;
- Microphone power supply for electrostatic mics on all mic inputs, switchable among 48-V phantom, 12-V phantom, or 12-V T powering;
- Mic input gain trim;
- Channel input fader for each input—these are the controls that are used for the actual mixing;
- Pan pot or switch per input channel to assign the inputs to two outputs, at least switching among output A, output B, or output A + B;
- Channel limiters on each input;
- High-pass (low-cut) filter on each input;
- Master level control, available so that all the inputs can be faded out at once;
- Two outputs labeled A and B or L and R on XLR connectors on the right side of the box, switchable between mic and line level;
- Camera return input so the loop to the camera can be monitored;
- Headphone switching to various states, such as monitoring the output of the mixer or the camera, and in various modes, mono, stereo, decoding MS, and others;
- Slate mic built in, possibly with simultaneously recorded low-frequency cueing tone for quick search through source tapes or transferred files for slates;
- A method to gang units together.

An example is the Sound Devices 302 shown in Fig. 7.2.

Small Mixer/Recorders

The units in this category have a form much like that of small mixers and still may be used over the shoulder or

FIGURE 7.2   The Sound Devices 302 front panel. Photo courtesy of Sound Devices LLC.

on a sound cart. Because they are for double-system recording, however, they must have time code facilities to use them to accompany picture shooting.

 Typical features may include:

- Multiple microphone/line level analog inputs that may be recorded individually and also as a part of a mix-down, usually to L/R.
- Microphone inputs that provide power to 48-V phantom and possibly 12-V phantom and 12-V T powered microphones.
- Signal/overload indicator per input channel.
- Potential of support for AES42 digital microphones on inputs with requisite sample rate conversion for Mode 1 support and/or AES31 digital line signals on inputs and outputs.
- Inputs that may have polarity reversing switch, low-frequency filtering, low-frequency attenuation, limiting, adjustable delay, and solo monitoring.
- Input assignments to individual tracks pre- or post-fader, to main channels possibly with panner, and to auxiliary channels.
- Various recording media options including internal hard disc drive, external hard disc drive, or DVD-RAM drive or CompactFlash or other memory cards. It is an advantage if more than one of these can be recorded to simultaneously, for quick delivery to post-production and for simultaneous backup purposes.
- Slate microphone.
- Various sample rates supported such as 44.1 kHz (CD production), 48 kHz (conventional film and television production), 47.952 and 48.048 kHz (more specialized film rates), and possibly more. Word clock input/output at the frequency of the sample rate for connection to other digital equipment "locked up" to prevent clicks.[6]
- Time code generator with input and output facilities and support for various time code rates: 23.976, 24, 25, 29.97 fps in non-drop frame and drop frame counting sequences, 30 fps in NDF and DF types.
- Record .wav files either in multiple monophonic or in polyphonic (interleaved channels) format.
- File metadata wrappers may be iXML format.

- External keyboard connection for log keeping.
- Headphone monitoring of various combinations of inputs or tracks. The headphone monitoring may be performed prior to the fader called Solo or Prefade listen (PFL). This feature helps to track faults quickly and may work as a cueing function because the mic is heard without being recorded if the level control pot is down. Possible decoding of multiple channels for the MS microphone technique and other stereophonic techniques may be included.

An example is the Sound Devices 788T, which has an optional controller, the CL-8. It has eight inputs that may be recorded to 12 tracks, and other features, but the basic recorder may still be used in a shoulder bag for documentary and reality situations (see Figure 7.3).

A Production Sound Mixer and Separate Recorder

A production sound mixer and separate recorder will normally be used on a sound cart, along with radio mic receivers, source playback machines for prerecorded shoots, backup recorders for discs, IFB transmitters and housing for receivers,[7] laptop computer, and space for various duplex cables that run signals bidirectionally between the boom operator and the mixer and between the mixer and the camera.

The mixer will generally have the mix features described above and may add some, particularly in the way of set communications, with separate mixes possible for the boom operator and recordist on the one hand, and the director and script supervisor on the other, with different "feeds" for each of these. For instance, the boom operator and recordist may need an intercom function between them that they do not want to burden the director with listening to. A term for these functions is PL for private-line communications.

FIGURE 7.3   A Sound Devices 788T with accessory controller CL-8. Photo courtesy Sound Devices LLC.

---

[6]Sample rates, time code, delivery formats, and metadata are covered extensively in the next chapter.

[7]Covered later in this chapter.

FIGURE 7.4 The Sonosax SX-ST 8D production sound mixer. Photo courtesy Sonosax SAS S.A.

An example of a production sound mixer is the Sonosax SX-ST shown in Fig. 7.4. In addition to the features above, although it is basically an analog mixer, it may be ordered with internal analog-to-digital converters so that it is equipped for digital output for recording on a separate digital-input recorder or to an optional internal hard drive for recording directly. It also has such production-sound-oriented features as switchable gain range of the main faders so that it can have extra "reach" when needed to handle both low and high levels simultaneously.

The recorder of a mixer–recorder pair will normally have the recording features described above, including recording to various file formats and recording to one or more media simultaneously, and some do recording of file descriptions and metadata including iXML.

An example of a production sound recorder is the Fostex PD-606 shown in Fig. 7.5. It has both analog and digital multichannel input and output capabilities

Production Sound Mixer/Recorders

The Aaton Cantar-X has an interesting form factor, usable both as an over-the-shoulder and as a cart mixer and recorder (see Figure 7.6). It has the features of other systems above plus the following:

FIGURE 7.5 Fostex PD-606 multitrack portable hard disc recorder. Photo courtesy Fostex Company.

FIGURE 7.6 The Aaton Cantar-X2. Photo courtesy Aaton S.A.

- A system for automatically marking the time code of slates by listening for them (improved when a button is pushed within a certain time frame);
- The ability to play back and record simultaneously from various tracks so that only one machine is needed for shooting to playback;
- The ability to output iXML metadata reports with input from the recordist as PDF or as Avid log exchange format files;
- An extensive software accessory set for integration into postproduction work flows.[8]

PRODUCTION SOUND EQUIPMENT ON A BUDGET

The examples up to now have been of professional sound gear dedicated in design to the various styles of recording sound for picture. However, in some cases these solutions are just too expensive and a cheap way must be found to capture production sound. One way to save money is to piggyback on the home studio equipment trend. Mass manufacture leads to low prices, and in many instances these solutions may be adequate. They will not have the ruggedness needed for shooting under extreme conditions, and they will not have the depth of the various feature sets described above, and they will need a source of AC power, but you may be able to get by with them for specific shooting situations.

A typical contemporary setup of such gear, of which there are doubtless many examples, is to use a Mackie 1402VLZ3 mixer ($430, Figure 7.7), a Mark of the Unicorn (MOTU) 828mk3 analog-to-FireWire converter ($750, Figure 7.8), and a laptop computer that you may already own (but it needs a FireWire connection).[9] The Mackie has six mic preamps, the output of which can be intercepted after the input gain trim control and high-pass

---

[8]See http://www.aaton.com/files/cantar-post-chain-25.pdf.

[9]A book has a much longer shelf life than particular models of equipment. These are meant to be contemporary examples at the time of writing. Consult http://booksite.focalpress.com/Holman/SoundFilmTV/ for later examples if available.

FIGURE 7.7 The Mackie 1402-VLZ3 described in the text. Photo courtesy Loud Technologies Inc.

FIGURE 7.8 The MOTU 828mk3 analog-to-digital converter and FireWire interface described in the text. Photo courtesy MOTU.

(low-cut) filter and sent to the MOTU. The jacks for these signals are called channel insertion jacks, they are $^1/_4$ inch, and they may be used for a Direct Out function. A special cable needs to be built to obtain this signal and send it to an external recorder as well as to continue the signal continuity through the rest of the mixer. Figure 7.9 shows the wiring of a cable that intercepts the signal and sends it to the outside world while continuing it on to the mix busses, with the send signal on the tip from the mixer to the recorder, the return signal on the ring from the recorder to the mixer, and ground on the sleeve. The MOTU is connected to the laptop by FireWire and the supplied software can be used to record to the computer's hard disc, or their more sophisticated software Digital Performer may be used with added features.[10] The MOTU 828mk3 can accept time code from the camera to ensure sync. Time code details with particular software need to be worked

out as software features change with time. Any such system must be tested before shooting for maintaining synchronization from camera and sound acquisition through postproduction. Also, for our purposes the only sample rate suitable among those offered is 48 kHz.

This Mackie model has six microphone inputs that can be wired to the first six line level inputs of the MOTU, usually in order. The two XLR mix outputs Left and Right of the Mackie can be wired to channels 7 and 8 of the MOTU with an XLR connector wired to a $^1/_4$-inch phone plug. Then the six prefader mic signals are recorded on the first six tracks and the mix on tracks 7 and 8. Tracks 7 and 8 can be exported for picture editing, and all tracks can be imported into sound postproduction for greater editorial flexibility than provided by simply having the live two-channel mix. The MOTU itself has an additional two mic inputs with phantom power, a pad switch, and a trim control for those occasions when more mic inputs are needed, although these could not be included in the Mackie mixdown.

To use this combination, because there is no tone output available on the Direct Outs (see Toning Heads of Reels on page 113), first set the Precision Digital Trim controls of the MOTU to their nominal setting, and then set the Mackie "Gain" (input trim) controls for adequate level on the MOTU meters, such as occasionally lighting the −6 dB F. S. light, but not so much as to risk red light overload.

To use this system to record to playback, an external playback recorder can be connected to the line inputs and feed only the auxiliary output for connection to the system that will feed the performers. It can do this simultaneous with recording the mic channels and mix.

By the way, $^1/_4$-inch plugs and jacks, especially if unattended for a long time, may become intermittent. A spray of CAIG Laboratories DeoxIT DN5[11] and exercising the patch several times will help intermittent connections. For thoroughly oxidized connectors an abrasive cleaner may be needed before spraying.

Skimping on microphones or radio mics is a bad idea, so think about renting them per project until you can afford good ones. I have tried inexpensive ones only to find them clearly not as good as the types described on the page http://booksite.focalpress.com/Holman/SoundFilmTV/. However, you may run into a situation in which you need a "sacrifice" microphone, much like the sacrifice camera that is embedded in the ground for a stampede of buffalo: it is just expected to live through the scene, but not much else! For these circumstance, the use of an electret capsule and a little wiring will do, such as for planted microphones in animal nests. Figure 7.10 shows how to wire up such a capsule.

---

[10]At the time of writing, when both the 828mk3 and the DP are purchased together, a discount is available.

[11]See http://store.caig.com/s.nl/sc.2/category.821/.f. See a different material spray for gold-plated connectors.

FIGURE 7.9　The wiring diagram of a cable for Direct Out of a Mackie mixer to the input of the MOTU converter.

output

ground

Typical electret capsule
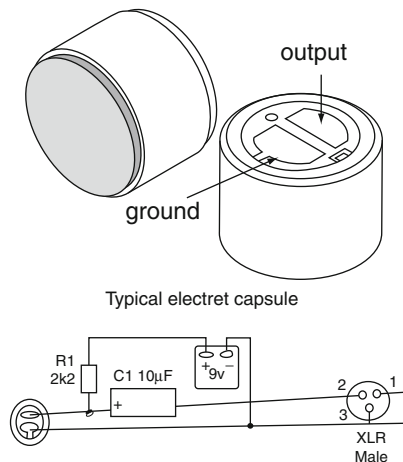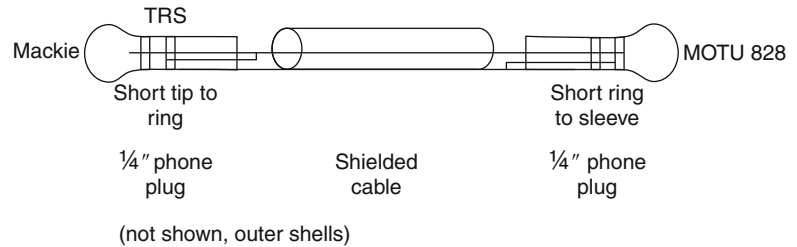
FIGURE 7.10　A "sacrifice" microphone consists of an electret capsule (available from www.digikey.com) and some external wiring and a battery.

CUEING SYSTEMS, IFB, AND IEM

In cases of working to prerecorded material, or to feed directions to a performer, a means has to be provided so that their work is in sync with a desired soundtrack. This cueing can take a number of forms:

- A "thumper," which is a special sound source that is fed to a subwoofer on set. It can set the beat so that dancing and/or singing can be performed in sync while recording the voice and even footsteps of the dancer. In postproduction the low frequency of the thumper can be filtered out to produce a clean track. A signal suitable to drive a thumper needs a distinct beat (which requires it to be fairly fast, and this can move its spectrum up into the audio frequency range, which is undesirable). A thump signal centered at 31.5 Hz that can be edited and looped at the required tempo is available as Track 6 of the accompanying DVD.
- IEM (in-ear monitor): Special small earphones can be inserted into the ear canal of the performer (some musician-oriented types fit the full concha too) and fed from receiver systems that use radio frequencies, magnetic induction, or infrared light. On most productions these systems are rented from professional sound houses because they are not normally needed continually.

IFB systems are a special case of cueing systems. IFB stands for interruptible foldback.[12] The term normally applies to live broadcasting in which a program feed is delivered to a field reporter through a separate channel called a backhaul or backstop. The interruptible part of IFB means that a communication channel may be opened to the talent superseding or ducking the program channel, to communicate privately to the talent. This function is also called a private line, or PL, function.

A problem with IFB systems is that when the reporter speaks, the time delay in the reporter-to-studio link and return (which can be very large, seconds if a satellite is involved) causes a very disturbing echo of their own speaking into the reporter's ear. For this reason the reporter will be fed back from the studio a "mix minus" program sound that does not include their own speech to prevent the echo from becoming a large problem.

EQUIPMENT INTERACTIONS

Radio Frequency Interactions

Digital audio equipment uses radio frequency energy in the form of clocks to perform various tasks. The equipment may thus radiate usually small amounts of radio frequency energy into the environment. Normally these do not cause a problem, but with sensitive radio mic receivers in close proximity, unfortunate interactions that seem mystical at first may occur, such as greater dropouts in radio mics than expected.

The method to prevent such problems is to follow the advice given in the previous chapter and move the receiving antennas (but not necessarily the receivers) to a point several feet away from the mixer and/or recorder. Also, it helps to make the receiving antennas vertical or to use separate higher-gain antennas.

A common source of interference heard often on television and which could occur on film sets or in documentary shooting is from cell phones or other mobile devices using radio frequencies. These are identifiable by their chirpy nature. A rule of thumb should be to ask everyone on the location to shut off their cell phones and other hand-held devices completely.

---

[12]Or sometimes, feedback.

Audio Frequency Range Interactions: Inputs

While mostly a thing of the past, loading of microphones by microphone inputs can cause frequency response variations and headroom limitations. In fact some studio microphone preamplifiers allow for the adjustment of loading conditions on the microphone for just the response changes that more severe loading (lower input impedance) can produce. Such interactions are minimal with normal production sound equipment today.

Also, phantom power must meet standards for both the source microphone's current requirement and the microphone preamplifier's current capability. This too is rarely a problem with modern equipment, but is a potential problem with certain combinations of equipment, especially of older models.[13]

Be certain that +48-V phantom power is not applied to digital microphones as it could cause damage.

Audio Frequency Range Interactions: Outputs

Production sound consoles are rarely called upon to drive long audio or telephone lines. Today it would be far more common to record a file and deliver it over an FTP (File Transfer Protocol) Internet service to postproduction. Certain live performances may need to be delivered in real time, and the output capability of the audio console may come into question. Normally today the output of the console would be fed to a device that would change it into optical fiber, microwave, satellite, or other means of communication rather than uncertain analog long lines. The best advice here as the range of conditions is so great is to test the system before it must be used in an important role.

INITIAL SETUP

Toning Heads of "Reels"

In the analog days, putting a tone at the head of each tape reel helped to make up for the tolerance of reel-to-reel variations in tape sensitivity, on the order of ±1 dB, and to standardize the transfer into postproduction. In the digital era, such a level change does not occur, although of course mistakes can be made. However, analog dubs may be made, or an analog interface may be necessary at some point in the chain, and so settings for those transfers can vary. Thus it is still good practice to deliver a reference tone for each interval of shooting, such as daily.

---

[13]See http://www.microphone-data.com/pdfs/The%20feeble%20phantom .pdf.

SMPTE specifies −20 dB F.S. as the reference level for digital recording. Besides the day-to-day variations in which an analog stage may be used, another reason to have a tone available that is traceable to the original recording is that not all digital equipment is yet aligned this way, especially when considering pressing semi-pro equipment into service. Some equipment may be aligned to, say, −14 dB F.S. for reference level, and if a −20 referenced source is recorded to a −14 dB F.S. reference level machine, then the maximum recorded level for the top 6 dB of the range will be overrecorded, resulting in distortion. If the source −20 dB F.S. tone is set to −6 dB compared to −14, then the top end of the dynamic range will be properly rerecorded.

The tone is normally a 1-kHz sine wave recorded at −20 dB F.S. and recorded for at least 30 sec. It should be recorded to all channels of a multichannel recorder and to the channels in use of a single-system video recording. Normally the tone will be recorded along with color bars on the head of videotape reels. However, some cameras have internal tone generators as well as accepting tone from the outside world, so care should be taken that the tone traceable to the original recordings from the mixer is the one recorded to the tape.

Slating

Clapperboard slates provide the means to find the start point for most film-originated material. In some cases, such as documentaries, slates are undesirable because they make subjects of the documentary feel like actors. Although methods were worked out to provide less obtrusive sync methods for film, today documentaries are customarily shot directly on video, and the single-system sound of video is employed, so no slates are necessary.

Why are slates important if they're all destined to be the first thing cut off in editing? Slates are the first to hit the screen in dailies, and it is a sign of good management of the activities of the crew if they come off well. Producers judge work on such matters. And even if technical means were developed to eliminate the need and the "wasted" footage that results in doing slates on every take, *The Fugitive* director Andrew Davis says that the slate would still be necessary: it is a part of the mantra that the crew chants together just before the take. It shows that everyone is alert and on board, and it gives the actors a moment to take a deep breath before stepping into the scene.

1. Slates must be legible to a telecine operator and editor. This means being:
   - all in frame and right side up for head slates;
   - sufficiently large in the field of view, say occupying one-third of the frame;
   - perpendicular to the line of sight of the camera;

- adequately lit to read;
- in focus;
- held still until closed; and then
- swiftly removed.

2. The slate operator and camera operator should collaborate to accomplish the above so that the camera needs minimal or no reframing and refocusing between slating and the body of the take, if at all possible. Sometimes this is impossible and the camera operator must say "Frame" so that the cue "Action" is not given prematurely, as shown in the following sequence.

3. The boom operator must "boom" the slate by orienting the mic in its direction. In some scenes, the mic is so far away and sometimes misaimed that the slate cannot be distinguished from other noises of the location. In this case a separate microphone, "opened" only for the slate, such as a radio microphone on the slate operator, may be used. If the slate is a long ways from the microphone, the speed of sound will be involved in getting the sync correct. At 46 ft from the microphone the sound will be 1 frame late at 24 fps, for instance. A notation on the sound report should tell the editor the facts.

4. Occasionally it is impossible to head slate because of special camera position or framing. In these cases, it is customary to tail slate, with the slating done upside down to indicate that it is a tail slate.

5. The slate must contain the correct information, especially scene and take. These must be legible, because they are the only place to refer to for this information (the can or box has the film title, reel number, etc.).

6. The slate must be closed smartly, without a bounce, and without any fingers in the way.

7. Follow this sequence:

- Director or AD: "Roll sound."
- Production sound mixer: rolls the recorder, observes that things are correct, optionally opens the slate mic, announces scene and take, closes the slate mic and says: "Speed."
- Director or AD: "Roll camera."
- Camera operator, when camera is up to speed: "Rolling."
- Slate person: "Marker" or "Sticks" and then bangs the slate crisply and exits. In the case of a close-up, in which the slate may be right in the actor's face, it is a courtesy to say instead "Soft sticks" and then close the clapper softly, so long as it can be heard clearly on the recording.
- Camera operator: "Frame."
- Director: a breath, then "Action!"

For tail slates:

1. For tail slated takes, the director must *not* say "Cut" at the end. Instead the director should say: "Tail slate."

2. The slate person steps into the shot and waits until the camera person says "Marker," indicating he or she is ready, the slate is in focus, etc. Then the slate person says "Tail slate" and bangs the sticks.

An alternative to the sound recordist stating the scene and take number by way of a slate mic is for the slate operator to do it.

The next decision made after "Cut" is whether to print that take. If the decision is to print the take, the director says "Print it," and the camera log and sound log have that take number circled, which is universally recognized as a take that is to be printed, have sound synchronized with it, and be shown at dailies.[14]

For double-system shooting the camera may not produce a time code output before it is rolling, so it is important for the sound person to observe that incoming time code from the camera to the sound recorder is being received and responded to as soon as the camera rolls, or else he or she must call a halt to the proceedings. This is for the case in which camera time code is in use, and not time code slates.

Many digital recorders designed for production sound have a function that records up to 10 sec *before* the Record button is pushed. This is accomplished by storing incoming sound all of the time and making the recording permanent when the Record button is pushed. Thus the time between "Roll sound" and "Speed" may be next to nothing. On documentary shows, this also helps not to miss important moments.

For synchronization of the sound and picture on a telecine, which is used both for film postproduction and for television, enough "preroll" time is necessary for the picture and sound to sync up. If the sequence described is shortened (usually to save film), there may not be enough recorded time before the slate occurs for equipment to synchronize picture and sound. The detail of how much time for picture and for sound is needed depends on the specific equipment in use, so it should be checked before shooting begins with the postproduction film to video transfer facility. It could be up to 15 sec.

The order "Roll sound" comes after "Quiet on the set." In the interval between production sound mixers listen carefully to the background, especially in exteriors. It is then they might interrupt and say "Airplane," for example. They are in the best position to hear this the earliest: they are concentrating on the sound, and they are listening over headphones that tend to boost the importance of background noise in perception. The rest of the cast and crew may groan over the sound mixer's meticulous behavior, but if the plane passes overhead and ruins the take there will be that wasted film and energy to pay for.

---

[14]In some cases today, all takes are "printed" because that involves copying into digital files, so the storage medium is not expensive. Still, the "Print it" helps because time at dailies screenings is limited, so it is still useful.

MIXING

The last chapter described the various microphone techniques at our disposal and also the consequences of mixing together various microphones. Here we want to describe just how much "mixing" is desirable at this point in the process of making a program. By *mixing*, we mean dynamically manipulating the level controls of the various microphones during a take for the purpose of emphasizing the desired sound, and the converse. The principles are:

- Produce the best sound within a take consistent with editability from shot to shot. Producing the best sound may call for a lot of active mixing, moving the faders a lot to follow the action; but making the shots consistent within a scene helps sound editing because bumps at edits will be less audible. So there is a trade-off between these two extremes, and there is no one correct answer in all cases. Covering a scene means filming the picture from multiple angles, and simply placing the boom mic in 1:1 correspondence to what is seen may not make for the smoothest edited scene. For instance, in one close-up on a pair of actors in a scene the boom mic could wind up aimed at a noise source behind one actor, whereas on the reverse (same close-up framing but of the other actor), the noise source would now be suppressed. Cutting between the two of them would make the noise pop at the edit, certainly undesirable for continuity.
- Microphones are rarely "potted all the way down,"[15] but instead are simply turned down enough that their contribution to the overall sound is negligible when they are not in use; this means they are already partially up to full value when they are needed, and less of a fade-up is required.
- Performers often act more strongly when the camera is rolling than in rehearsal. After rehearsing to find the correct recorded level, the wise sound recordist will provide a few decibels of margin for the adrenaline factor.

LEVEL SETTING

Level setting of recorded level depends greatly on the medium in use and the type of metering employed. Because this is so, these topics are covered in Chapter 6.

---

[15]Level controls have a variety of names, including *pot*, short for potentiometer, the actual circuit element that does the adjustment. Alternative names include *fader* and *volume control* (usually reserved only for the loudspeaker or headphone monitor controls). Thus, to *pot down* is to turn the fader down in level.

COVERAGE

Dialog Overlaps

Overlapping dialog from various actors speaking at the same time causes one of the most serious potential problems editorially. Because for all practical purposes it is impossible for the performers to overlap the dialog identically from take to take, allowing dialog overlaps on the set greatly reduces the possible edit points, sometimes to none! In a very rigorous form of direction, the dialog overlap is created in postproduction by showing the back of one performer while looking at another and dubbing in the overlapped lines of the actor whose back we see, as appropriate. Today it is probably more common to permit overlaps in master shots and to try to duplicate them roughly in close-ups, because of the naturalness that this brings to the acting, but the editorial problems are formidable and should be thought through by the director and sound crew before shooting.

The sound crew has a strong interest in dialog overlaps because a decision will be made about whether actors speaking off-camera should be heard directly on mic, off mic, or not at all. If the actors can reproduce the overlap from shot to shot adequately, then it may be useful to mic both of them (if there are just two) for all of the shots, even close-ups where the off-camera actor is recorded on mic. This permits editorial freedom, as the sound perspective will not change dramatically at the picture cut, but necessitates very careful control of the actors to permit edits.

A second possibility is to record the off-screen actor off mic as well. This means the actor will be heard but will be recorded at a high angle of incidence compared with the axis of the microphone and thus will be heard mostly as only a small amount of direct sound and potentially lots of reverberation. The advantage of this approach is that the actor on-camera has someone to react to, but the disadvantage is strong: The off-mic recording is very noticeable as being different from on mic, and if a cut is made to bring the off-screen performer onto the screen, the sound perspective will jump and destroy continuity.

The third approach is the most rigorous, and was described earlier. Have the off-screen performer just mouth his or her words and make no sound so the sound recordist can capture a "clean" track of the on-screen performer, then reverse the roles for the alternate close-up and build the dialog overlap in postproduction. This has many control advantages but requires more of the performers.

This discussion of methods for dialog overlaps has an impact on documentary interview technique as well, because the same issues are often raised. In a one-camera interview, it is common to shoot the interview first and then to shoot the interviewer asking the questions. If an overlap of speech occurs during the interview, the question is whether that should be on or off mic. Probably

the best solution is to record from two mics, on the interviewer and the interviewee, on two separate tracks, because this gives the most control in postproduction. A problem occurs if only the interviewee is mic'd and the interviewer asks a question that overlaps. In this case, the interviewer will be heard well enough for dubbing over a direct recording to be impossible, yet his or her voice will be heard grossly off mic. Of course, these problems are prevented if there are never any overlaps and both parties are shot individually. Then a natural-seeming interview flow can occur editorially, in postproduction, with constructed overlaps if desired.

Another consideration in such shooting is the need for cutaways. Cutaway shots are those of incidental occurrences surrounding the interview that, above all else, do not show the mouth of the person speaking. This permits editing the audio track to compress time in the interview without the discontinuity of a jump cut. A flaw that must be mentioned occurs sometimes in documentary production in which there is no cutaway to use—lip flap. This occurs when the picture shows a speaking person whose voice we do not hear. Although sometimes there is no way around lip flap, it shows very poor technique and gives away the mechanics of the process to the viewer.

Crowd Scenes

Another difficulty for production sound is crowd scenes. Generally we wish to focus on the principal performers and what they say in a scene, but what is supposed to be going on in the background may drown them out. Professional extras are good at simulating conversations while remaining silent, and an appropriate matching sound is produced either as "wild sound" during production or in postproduction during an ADR session, often with more than one actor, producing what is called a *Walla* track, that is, a track containing no discrete audible speech but providing the right sort of level of action to match the scene.

With less professional performers, the degree of the ability to simulate speech, without actually making sound, varies greatly. In some cases it may simply be impossible; this is surely true in documentary scenes played out in a restaurant, for example. There is just no practical method to control a scene in a working restaurant, so shooting should be scheduled outside of normal hours if a scene must take place under such conditions.

The principal actors must speak at an appropriate level and stress in their voices for the eventual situation. Confronted with a quiet stage and well-behaved professional extras, the tendency is to lower the energy level in the performance. However, they have got to "speak up" over the background noise that isn't there! Some actors are very good at this and develop reputations among sound professionals that they are; whereas other, perhaps more "intuitive" actors may start out with enough energy but drop over the takes. This leads to the question of who should tell the actor that there is a problem, covered under Set Politics, later in this chapter.

LOGGING

Logging the sound tapes is the responsibility of the production sound mixer. Usually the log will give:

- Production name/number;
- Shooting date;
- Reel number;
- Producer/studio/director, etc.;
- List of scene/take information recorded;
- List of any wild lines recorded;
- List of any sound effects or other wild sound recorded;
- List of any presence recorded;
- Takes that are meant to be copied from the production source media to be heard at dailies, with matching picture having their take numbers circled and the designation "Print circled takes";
- Track list for two or more channel media and for both audio-only and single-system video recordings.

At the end of a shooting day, the production sound recordist consults with the script person and a camera crew member concerning the three logs from the set—sound, script, and camera—rationalizing the list of scenes and takes shot that day to correct any errors.

SHOOTING TO PLAYBACK

Quite often it is essential to shoot musical numbers to playback, the logistics of recording music and shooting film or video simultaneously being too demanding for reasonable budgets and time constraints. To accomplish this, a playback tape is prepared, often a special mix emphasizing the elements to be lipsynced, and two recorders are usually needed, one for playback and one for recording. The recording machine records a slate from an open microphone and then rerecords the playback directly from the second machine, thus ensuring a reference for picture–sound sync. This also permits starting in the middle of a long number for a given shot, and a record of exactly what part of the song the picture is to match is recorded.

This method involves hand matching of the recorded portions with the original recording. Synchronization may also be accomplished by using time code on the prerecorded tape and recording it to a track of the audio recorder as a reference. Some software can be used then to sync up the shots to the original recording.

For shooting to playback, a number of considerations apply:

- The performers must be reasonably close to the play-back loudspeakers, say within 20 ft, so there is no time delay associated with the air path for their lipsyncing.
- For dance numbers, occasionally thumpers have been found useful, that is, low-frequency transducers that put vibratory pulses out into a dance floor so the dancers can follow the beat. The advantage of this approach is that ordinary recording can occur, and the low-frequency energy can be filtered out of the recording, leaving a synchronized dance number to music and also well-recorded direct sound.
- There must be a traceable path for the synchronization signal from the original music source, through the playback tape, to the recorded tape. Sync will be covered in more detail in the next chapter. If this "sync lock" chain is broken, then the editor may find it impossible to synchronize.

OTHER TECHNICAL ACTIVITIES IN PRODUCTION

The sound crew, usually being the most technically savvy on the set, is also given responsibility for such ancillary electronics as walkie–talkies and intercoms, especially when working on location away from the technical infrastructure available in major production centers. Battery charging, providing music systems for the director and actors, and many other such tasks often fall on the sound crew.

SET POLITICS

Among the factors that make a particular sound crew effective, and hired again, are not only the skills with which they perform their duties on the set, but also how skillful they are at set politics. Sound may often be considered to be an orphan on the set because the visual images tend to dominate and sound can always be looped, or replaced, albeit at high expense, later. That is not to say that the performances of the actors during looping will necessarily be as good as they are in front of the camera, with the added tension the live action setting brings about. Actors vary tremendously in their ability to loop in sync, and with great feeling.

A major factor in the performance of actors is their energy level. Modern-day acting technique often emphasizes naturalism and downplaying, because the camera seems to make everything bigger than life. "Theatrical" acting is discouraged for film actors, because it is "too big." However, it is also true that underplaying so much that the actor cannot be heard by the director listening on the set live (not over headphones) can indicate that the modern trend has been taken too far. This is especially true if one is shooting a scene in which special visual effects will be added later: it is simply hard to act in front of a greenscreen unless one knows something about what one is facing and can produce the right level of stress and loudness appropriate for the eventual conditions of the finished scene. However, it is not the sound people's role on the set to say something to the actor. A problem should be communicated to the director or assistant director, who may say something to the actor, instead of directly, unless prior conditions have been established allowing for this communication. It can happen that the sound crew, working with the same actor from picture to picture, will develop a rapport with him or her, particularly the boom operator, who is often the closest crew member to the actors.

Gerard Loupias, a French former boom operator, tells the story of an actor coming to him at the end of a day of shooting and thanking him profusely, as he was the only person visible to the actor, given the lighting conditions, and was the one person on the set who could engage the actor as another human in a scene where he had to break down. This example must be said to be abnormal, because it is not common for an actor to engage a crew member by acting to them. Boom operators typically work like the background actors in Japanese Noh theater; there, but not there, by not looking directly at the actor's eyes.

Set conditions may prevail, especially on big effects-laden movies, that essentially prevent the recording of high-quality sound. In these cases, the sound crew will revert to recording a *guide track* so that a record exists of exactly what was said, take by take, as guidance in eventual looping sessions. (The ADR or looping session usually does not occur until after the picture has been cut because there is no point in producing clean sync dialog that will not appear in the film.)

Probably the most contention between the sound crew and others on the set is over the use of the boom microphone. With the potential of dreaded boom shadows being thrown on the set, the boom is controversial from the point of view of the director of photography and gaffer, who may see no way to accommodate the boom and still light the set; it is just one more constraint that "breaks the camel's back." However, as already discussed, this is by far the most effective tool we have for recording production sound that is usable and creates the proper perspective. Cinematographers would rather we resort to body mics, but that just isn't the same. There are scenes in which nothing else will do because the shot is so expansive that no boom can be anywhere near the actors, and planted or body mics will have to do. In balancing the needs of the various departments, however, it must be said that the accommodation that the camera and lighting departments can give to the sound department in this area is usually well worth the overall impression of a scene.

The introduction of this book also pointed out that everyone on a set is actually engaged in capturing a picture *and sound* of a performance, so should be attuned to that end. Costumers can help in their designs by placing positions for mics and transmitters. They can provide booties for the actors so that they don't have to wear shoes when their feet won't show. Grips can oil dollies so they don't squeak. Carpenters can lay track that is stable and won't make noise. Generator operators can place their equipment around a corner instead of in the line of sight of the shooting area. And so it goes: insofar as everyone is aware that what is being done is to preserve the particular look and sound of a movie, the better off the sound will be.

Sync, Sank, Sunk

IN CASE OF EMERGENCY

Of all the issues in this book, sync is the thing that most frequently goes wrong, usually because not enough communication has occurred during preproduction about postproduction needs. Also, it is usually impossible to check sync on the set for double-system recording, so you will not know about it, whereas you can monitor audio, find problems, and fix them. There is little more disheartening to hear back from the post house than "the audio won't sync up"—and it is always blamed on audio until you can prove the camera was running off its stated speed! So this quick section offers advice on what to do, but it cannot cover all cases because there are so many work flows used today. The rest of the chapter provides a fuller background that can be used to sort out a variety of work flows.

1. Talk to the post house, post supervisor, or postproduction producer and ask "What sample rate and time code rate and type do you need?" Get the answer in writing if at all possible (perhaps by email) and follow the advice. There are many factors that go into making the choice of audio sample rate and time code:
   a. Picture frame rate (on film, digital cinema camera, high-definition video camera, standard-definition video camera), but CAUTION, see following, often the stated frame rate is not the actual frame rate!
   b. On-set picture monitoring requirements (high definition may be converted to standard definition for display, for instance, and that may dictate some high definition choices);
   c. Picture ingest method into postproduction (e.g., telecine-to-tape transfer, film scanner-to-digital file transfer);
   d. Where the sound is to be synchronized (in-the-camera single system, in telecine during transfer, or by picture editorial);
   e. Picture editorial frame rate due to the editing system in use, not just manufacturer but name of software, the version in use, and its settings;
   f. Need for copies of dailies and cuts for the producer, director, and others (these often call for the need to play on standard-definition playback equipment and this drives earlier decisions);

   g. Delivery formats (film, HD video, SD video, or both film and video) note that "I don't know" is sometimes a true answer at this stage.
2. If asked for input to the process, devise a work flow to minimize the number of conversions of digital audio files for sample rate, and minimize or make zero the need for digital-to-analog and analog-to-digital conversions, delivering as closely as possible the original files.
3. DO NOT BELIEVE the camera department when they tell you on a video shoot that the frame rate is 24 fps, or 30 fps. Regrettably there is great confusion and a lot of rounding off in this area that even manufacturers do such that their own data sheets may be wrong. Here are some common misnomers:
   a. 23.976, 23.98, 23.97, and 24 fps may all mean exactly the same thing! The "real rate" meant is usually 23.976 (which is 24/1.001 to five significant figures, for reasons explained later), but in some cases it is rounded up to 23.98, and still other usage truncates the number after two decimal places to result in 23.97. The worst is calling 23.976 fps camera operation "24," which is a very rough rounding and which causes great confusion.[1] It is most commonly done on lower-cost but professional video cameras to promote the fact that they simulate this aspect of the "film look."
   b. Actual 24.0-fps shooting is restricted to film cameras and some very high end high-definition video and digital cinema cameras (many of which are switchable between 23.976 and 24 fps, as well as other frame rates). Genuine 24-fps video is uncommon and would be a special case, generally associated with higher budgets.
   c. The need for copies that will play on standard-definition equipment in consumer formats (1.f above) often drives the decision making about production frame rates. Film and those few HD cameras operating at genuine 24 fps are generally slowed down when entering postproduction (the content is said to be

---

[1] At least one camera model's documentation, the Panasonic AG-HXP500, shows this issue in a footnote: "24p = 23.98p, 30p = 29.97p, 60p = 59.94p, 60i = 59.94i."

"ingested" into post) by −0.1 percent to run at 23.976 fps. If the sound is not slowed down correspondingly, it will run faster than the picture and come earlier and earlier the longer the video runs from the start of sync, such as a clapperboard that has been "synced" up. The difference of 0.1 percent is one frame in a thousand, or one frame out of sync for each 33.3 sec of elapsed time. If sound progressively precedes the picture by an additional frame for every 1000 frames played back, there may be a missing speed change for sound relative to picture: the sound needs to be "pulled down" to match the slowed-down picture.

4. All sample rate settings for sound accompanying a picture are centered on 48.0 kHz, and it is still the most common. The two others are 47.952 and 48.048 kHz, so-called pulldown and pullup sample rates, respectively. In the absence of instructions from postproduction, do the following:

a. Use 48.0 kHz to accompany an actual 24-fps film or video project if postproduction requires it and will compensate it on ingest. This is especially true if the final delivery will be to video. Compensation consists of sample rate conversion: if such a recording is played 0.1 percent slow to match the picture the sample rate would become 47.952 kHz, which would no longer match the editing system's 48.0-kHz sample rate. Note that some prominent editing systems do this sample rate conversion internally without necessarily informing you that they are doing it. In some cases what distinguishes software that will perform the needed sample rate conversion is price: lower end systems are generally completely video based, with little or no support for actual 24-fps production.

b. Increasingly common today is to use 48.048 kHz in production for 24.0-fps projects if the editing system is based on NTSC video frame rates and can slow down the audio to a lower speed with a sample rate of 48.0 kHz upon ingest. This is not a sample rate conversion but rather a 0.1 percent time stretch: the editing system simply plays back 48,000 samples of audio in the time it took to record 48,048 samples on set. Slowing down the audio makes it longer in time, which does change pitch by 0.1 percent, but has no other sample rate conversion artifacts. This is done to match what happens to the picture, which is actually running at 23.976 fps on the film transfer machine rather than the 24 fps at which it was shot, in the case of NTSC postproduction.[2]

c. Use 48.0 kHz when accompanying a video project shot at actual 23.976 (but see potentially confusing labeling in 3a, preceding), 29.97, or 59.94 fps.

d. Use 48.0 kHz for a 25-fps film or video shoot (for PAL countries).

5. The correct sound recorder and slate time code rate may not be the same as the camera frame rate! Instead the choice made is one that anticipates the needs of postproduction after ingest of the picture and sound. This leads to some surprises.

a. For film shoots at 24 fps, or video shoots at genuine 24 fps (remember these are rare), set the time code slate and time code generator on the recorder to 24 or 30 fps as required by postproduction. Jam sync one to the other by connecting the two together and forcing one to copy the other's time, usually every few hours, so that their clocks remain hard synchronized. (Remember that scene in all those WWII movies in the expositional warm-up to the action in which the officer says "Gentlemen, synchronize your watches. The correct time is . . . ." Well that's jam syncing.)

   i. The reason for using 30 fps time code despite shooting 24 fps on the camera is that in many cases film will be transferred to NTSC video for postproduction and wind up running at 29.97 fps during editing. A more detailed description of telecine operations in the NTSC world is discussed later.

   ii. The 24 fps time code went neglected for years because of postproduction requirements to work in NTSC video. More recently it has come into use as originally envisioned by the writers of the standards. Genuine 24-fps postproduction makes life easier as it does not require sample rate conversions or slowing down sound to match picture, which are both required by NTSC-based editing systems.

b. For video shoots at 23.976 or 29.97 fps use the time code 29.97 fps. That is because 23.976 material will have repeated half-frames in a sequence described under Telecine Transfer, or other similar ones that bring the frame rate up to 29.97 fps because that is normal for standard-definition television equipment used to edit both standard- and high-definition material.

c. Never use drop-frame time code on production unless specifically told to by postproduction. For live television, however, it is common. Non-drop-frame code is typical on productions that will have postproduction. The terms are defined later.

d. Occasionally 29.97- or 30-fps film may be used, typically on very short form high-budget material such as television commercials with a lot of motion.

---

[2]There is increasing use today of 24 fps editing, making life simpler: time code is as 24 fps, the camera shoots 24 fps, and editing is conducted at that rate. Video outputs to be compatible with NTSC video must have a 2:3 and pulldown process inserted electronically, described later.

That is because actual 30-fps capture renders fast motion more smoothly on television than 24-fps source material. In such a case, match the frame rate of the camera with the slate and audio recorder.

 e. Be certain that the slate is legible on the film and that a recorded word from the slate person such as "marker" or "slate" or "Camera A marker," etc., just precedes the slate in the audio.

6. *Sorting out mistakes*. There are two kinds of "normal" errors, in which sync is wrong but nothing is broken, and these are due to equipment errors.

 a. The first is an offset between picture and sound observed in the editing system. It does not drift, but stays constant. If it is one-two frames and the sound is early, one likely culprit is the video camera, which, in making a picture, scans out the picture information from the imager all at once at the end of a frame time and then puts it on tape or on another medium. This is a basic fact of life with these cameras, and a corresponding time delay of the audio is not done because the manufacturers do not consider a one-frame error important. This is easily checked with a slate—hit several of them over time to be sure and check sync in the video system. Quite often, the sound will have to be pushed later on the time line to be in absolute hard sync, usually by one frame or so.

 b. Another source of an offset is the position of the record head for the time code versus the record head for audio. This affects mostly analog machines but could affect digital ones as well. The "mistake" is easily corrected once found by simply moving all the audio along a time line according to that which gets a slate in hard sync.

 c. The second kind of problem is drift. A scene starts in sync but slowly drifts out of sync. Once the total reaches two frames out of sync, everybody can see it and they are annoyed. For one common error (described above in 3.c) this takes about a minute. The problem usually is that the picture has undergone a pulldown process by 1/1.001 or 0.1 percent slow. Thus the sound will progressively get further and further ahead of the picture. The best solution is to retransfer the audio with a pulldown to the correct speed. This will cause it to be at the wrong sample rate, and it will need to be sample rate converted (of which there a variety of qualities available in the market of both hardware and software products) or converted to analog and back to digital. Although neither is totally desirable, good sample rate conversion is probably the best, whereas poor sample rate conversion is probably worse than converting to analog and back to digital with good equipment. The analog system, though, has to be carefully aligned for level, whereas the sample rate conversion does not.

 d. I have seen rare occasions on which drift occurred, but it was not a missing pulldown. Sometimes it can be a camera fault, running off-speed. A way to see this is to run the dailies at high speed. If the picture brightness seems to be varying up and down in a regular way, what is happening is that the difference between the camera speed and the frequency of the lighting is being seen by the camera. Because real 24 fps is locked to the power line frequency of 60 Hz in a fixed ratio of 2.5:1, no variation is seen at the correct speed. But if the camera is running at 23 fps, for instance, then a 1-Hz "beat note" of brightness variation may be seen. More than once in my career I have been able to prove that a "sound sync" problem was actually the camera running off speed.

 e. There are known inconsistencies in the mapping of 24-fps film to 30-fps video from various cameras into editing systems. Only a preproduction test of the workflow with all elements involved can reveal subtle problems.

 f. Some of the most difficult work is done when productions simultaneously shoot film and high-definition video or with digital cinema cameras, usually for their various "looks." Such specialized cases absolutely need preproduction assistance from those who will handle postproduction. A straightforward way to handle mixed media is to make them all operate at the same frame rate. This is why at least one Panavision film camera, the PFX-GII Golden Panaflex, can be switched to run at 29.97 fps, for instance, so it can be used in a mixed environment with video-frame-rate cameras.

7. A pre- to postproduction test of sync is highly valuable when done during preproduction so that everybody knows that the workflow works. Because tape is cheap and reusable, and RAM cards are totally reusable, you are not even hitting the budget for stock. Simply record the medium from head to tail without stopping with slates every minute or so, export to postproduction, have a telecine operator or an editor (the one who will do the work) sync up the sound with the picture at the first slate, and check that all the others remain in sync. Then you know for sure that your work flow is all right and that any problems that come up are due to someone not following the established rules for that particular production.

INTRODUCTION

In the 1960s life was easy for synchronizing sound and picture. Film and television were separate except in one instance: films were shown on television—a one-way street to be described later. Film and television each used

their own capture medium and finished the process in media related in a 1:1 correspondence to that in which they were captured.

During postproduction of all but the very simplest productions, double-system sound was used, and for movies, sound magnetic film was used with the same dimensions as the picture film and with perforations. Teeth on playback equipment engaged the perforations, pulling it along and keeping it in sync with the picture across multiple machines, including a projector; sync was rarely a problem. For video, the original production sound was usually recorded on video-tape along with the picture and then copied (dubbed) to a multitrack recorder in a process called lay down or lay over. It was then "sweetened," that is, combined with other sounds laid down to other parallel tracks, and mixed down. The mix was then copied back to the edit master videotape in a process called lay back. These processes were kept in sync using SMPTE time code on the source tape and edit master and on the multitrack audio tape.

The frame rate for video is different from that of film, but this did not matter because the two were not inter-mixed. The case of films on television presented a problem, because the film frame rate of 24 fps was incom-patible with the nominal 30-fps rate of U.S. video. A method was found to transfer film material to the nominal 30-fps television rate and was universally practiced for NTSC countries, which is detailed later.

The intervening years have seen progressive use of varying methods for capture and postproduction such that almost all productions live in a mixed-media environment today, with a great deal of exchange among film, digital cinema emulating film, and video media. Although it is hard to argue with the productivity and cost gains that have been achieved, much of these gains are all too frequently eaten away at by the complexities of maintain-ing sync during transfers between media. You hear "The sound drifts!" all too frequently, even sometimes when it is the picture that is off the expected speed.

Getting good information in a rapidly evolving field is a problem too. Web sites help, but there is far too much misin-formation present on them. Did you know for instance that a 2 pop is called that because it is two frames long? Not! At the University of Southern California we post that on the wall as a reminder that all you read on the Web is not true.[3]

To understand from where the complexities originate, it is necessary to understand film and television, especially the NTSC and ATSC (U.S. video systems) in some detail. Regrettably, it is the lack of attention to detail in this area that leads to a great many problems in postproduction.

A LITTLE HISTORY

In the silent era it was commonplace to hand crank cam-eras in the range of 16 – 18 fps and to show the resulting movies in cinemas at 20+ fps. There are several explana-tions given for this, but the simplest is that it was possible to get audiences in and out of the theater faster at the higher frame rate, and thus make more money. Sound spoiled this method, because the pitch shift associated with playback at a higher speed than the original was obvious to all, whereas the picture speed-up had not been widely noticed. Movies were called "flicks" for a reason too. Low frame rates led to the perception of flicker on the screen. So when sound-on-film systems came into use in the late 1920s there needed to be standardization so that playback pitch was the same as the original, and sound quality and flicker perception dictated frame rates higher than those previously in use.

Twenty-four frames per second was the frame rate of the synchronous motors[4] used on cameras of the era and on magnetic film recorders running in parallel with the camera. By the 1960s lightweight portable equipment that permitted smaller and lighter cameras, and over-the-shoulder audio tape recorders, came on the market. Synchronization was maintained by using a number of schemes, winding up with cameras operated by battery-powered DC motors regulated for speed by crystal oscilla-tors like those in watches known as quartz crystal types and audio recorders that recorded an extra track from a separate quartz crystal oscillator. By regulating the play-back speed of the tape to match that when it was made, audio tape without perforations was given virtually perfect speed regulation, and copying the tape to perforated magnetic film and syncing it up was all that was necessary.

Television was a pre-World War II invention that had to wait out the war for its popularization. Humans perceive flicker due to frame rate more as the pictures get brighter. Because television pictures were brighter, albeit smaller, than those at the movies, the perception of flicker was greater. Films are seen in cinemas as 48 presentations of 24 fps to help alleviate flicker: each frame is shown twice (or occasionally, in some high-quality situations, three times). The image rate of television was selected to be 60 images per second (although each "image" or "field" as it is called contains only one-half of the lines; this is called interlace scanning). The corresponding "speed-up" of television over film of 25 percent was necessary to reduce flicker to a reasonable level.

However, when a few years later it was desired to add color, without taking up any more additional space in the

---

[3]This chapter has correct information insofar as the author knows, by restricting its range to only certain topics and by having it vetted by experts in the business, who are thanked in the Preface.

[4]Like clock motors, these "keep time," whereas some other kinds of motors, like those in washing machines, do not.
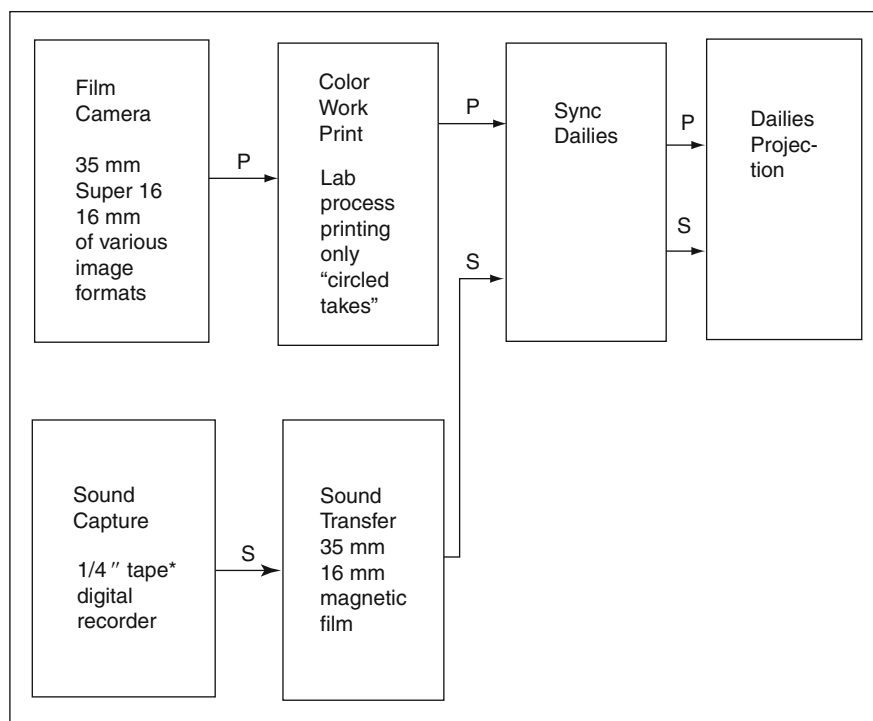
FIGURE 8.1   Conventional film capture, transfer, and dailies. Everything is at precisely 24 fps. The $\frac{1}{4}$-inch tape has a time code or other reference that matches the camera frame rate.

*If $\frac{1}{4}$-inch tape is used, it will be supplied with a synchronization signal that is "resolved" during playback to make "in-sync" film copies. This method has virtually succumbed to those that mix film and video/digital.

spectrum for a broadcast channel, a way had to be found to sandwich color into a black and white signal. Among other things, this led to a slight shift in the frame rate by a factor of 1/1.001 to 29.97 fps.[5] Meanwhile our European cousins suffered with a 25-fps frame rate based on their power line frequency of 50 Hz, and until recently this led to obvious flicker on their television systems. Only today, with double- and quadruple-rate sets and displays otherwise divorced from image rates, has the problem been reduced.

Videotape recording was commercialized in the late 1950s, but it was not until some years later that complex audio postproduction became available using the enabling technology of SMPTE time code. Essentially, time code "stamps" each "frame" with a number along a tape or time line of an editing system, even though there are no perforations, and it goes beyond the perforation system, because the stamp contains an absolute address to that exact frame, in the form of hours:minutes:seconds:frames, whereas the system of perforating the picture and sound materials keeps relative synchronization once it has been established (by use of a clapperboard, for instance).

---

[5]To be precise, repeating digits 29.9700299700....

TELECINE OR SCANNER TRANSFER

In most cases today, film shot for both theatrical features and television is transferred to video for editing. Theatrical films are often edited using video pictures digitally, and then a special cut list is produced for a negative cutter to cut the film negative for subsequent laboratory operations, or a high-resolution scan called a Digital Intermediate is used for copying to Digital Cinema or is output on film for prints. For films shot for television release only, either the original telecine transfer is used for the final output or a similar edit list is used to make a new telecine transfer at higher quality and in conformance with the offline video picture edit.

There are complications due to the differing natures of film and video that are necessary to understand to get and keep picture–sound synchronization throughout the various processes.

Most film in the United States is shot at 24 fps, whether for theatrical release or for television shows. But U.S. video operates at a higher frame rate, nominally 30 fps. If we were to simply speed up the film from 24 to 30 fps for the purposes of transfer to video, all the action would be much faster than normal, and the accompanying sound would have its pitch raised dramatically—hardly a usable result.

FIGURE 8.2 Conventional work flow for picture editorial, sound editing, and sound mixing, all on film. All elements are on 35mm film, picture and magnetic sound, separately. All elements run at 24 fps, which is 18 inches/sec on 35 mm. This is called "film finish," but is rare today.
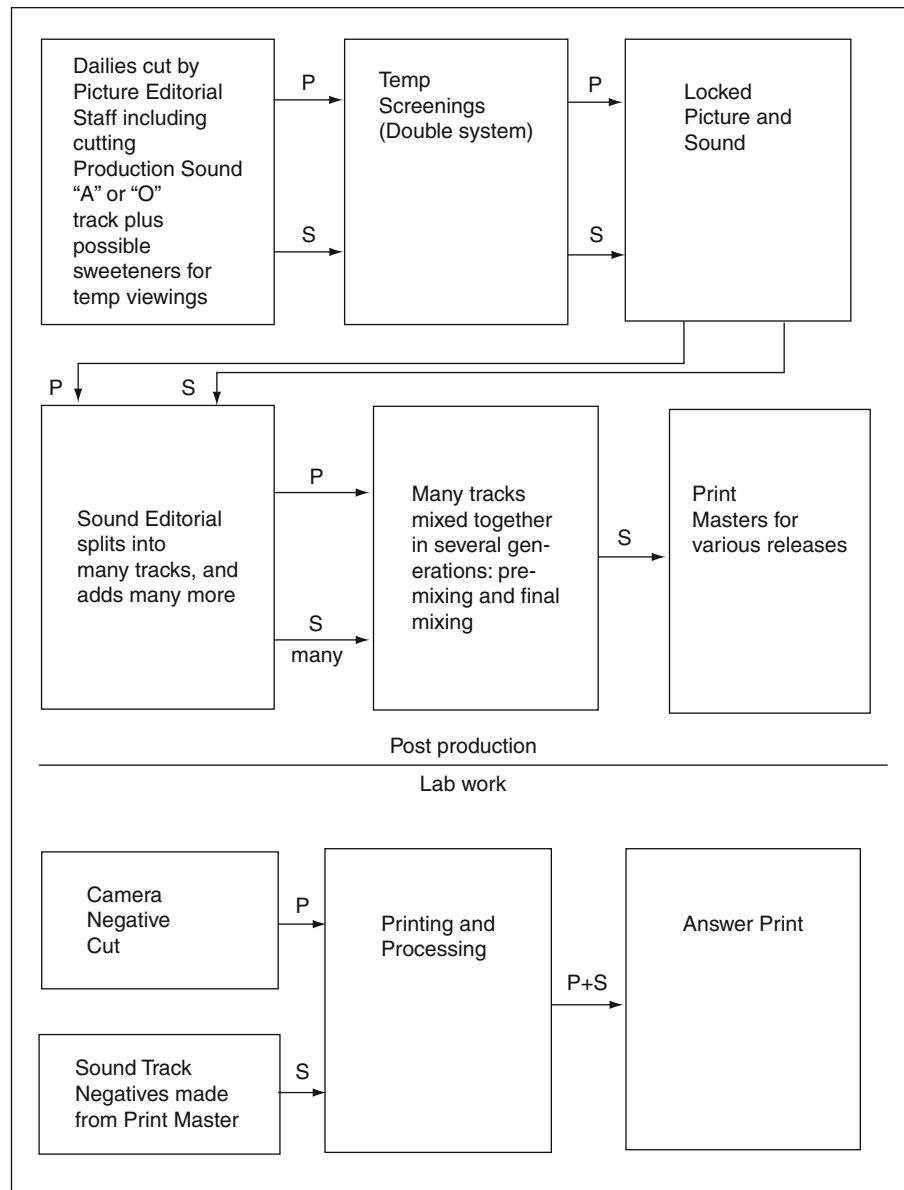
So what is done in the film-to-video conversion is to selectively duplicate some of the film frames into some of the video frames (actually into halves of video frames called fields), so that a ratio of 24 film frames to 30 video frames is produced, a ratio of 4:5.

This process is routinely called "3:2 pulldown." Technically a better name is "2:3 insert and pulldown," because the sequence starts with two video fields matching one film frame, then three video fields for the next film frame, and so forth—the 2:3 process constitutes the insertion of extra fields; the added process pulldown is explained later. The terms 3:2 and 2:3 are used interchangeably and mean the same thing. The 2:3 telecine process does the frame conversion from 24 to 30, and viewers typically do not perceive the artifacts produced by repeating the film

frames.[6] There are also more advanced ways of distributing film frames to video frames, such as 2:3:3:2, which permits easier extraction of 24-fps film frames.

If this was all there was to it, there would be just a few complications resulting from the transfer. However, NTSC[7] and ATSC[8] color video do not run at exactly 30 fps, but instead at 29.97 fps. All too often this is rounded up to "30" fps, when what is meant is actually this longer

---

[6]Although some specialized film shot for video, such as some music video and television commercials that feature a great deal of action, benefit from being shot directly at 29.97 fps on film.

[7]U.S. analog standard definition television.

[8]U.S. digital definition television.

Chapter | 8   Sync, Sank, Sunk   127
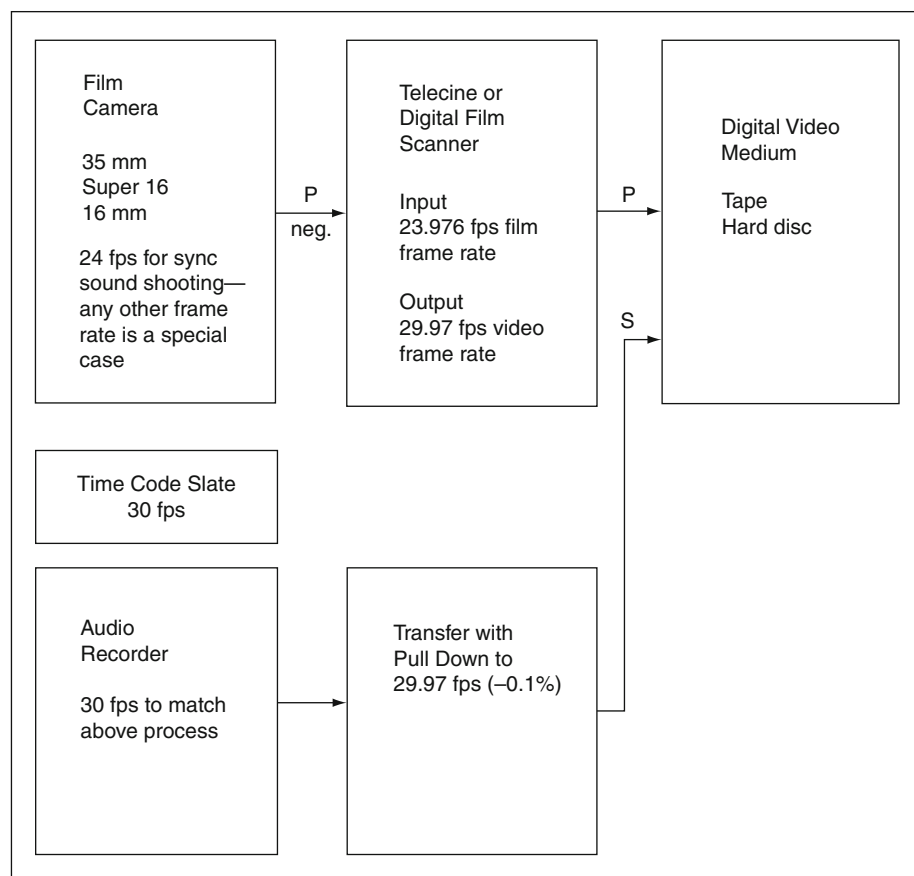
FIGURE 8.3 An example film–capture/video–finish work flow. There are many variations; however, this is one common method. Note that the slow down (by 0.1 percent) when picture material is ingested into postproduction requires a corresponding slow down in sound. The production sample rate may be 48 kHz and thus 47.952 kHz after slow down, which must then be sample rate converted back to 48 kHz. An alternative is to shoot at a 48.048 kHz sample rate, which is then slowed down to 48.0 kHz, the standard sample rate, for postproduction. The time code slate runs at the same rate as the audio recorder so that the time code photographed on film matches that recorded by the audio recorder (or there may be a fixed offset between the two that is easily accommodated once understood).

number. So after the 2:3 film frame-to-video frame/field conversion there is a remaining difference of practically 0.1 percent, with the video being slower than the film original, and the film is simply slowed down by this amount in transfer to video. This is the pulldown[9] part of the process and is responsible for the potential of sync problems arising, because the original sound recording must be slowed down to match. The process of slowing down the audio to match the picture is also given the name "pulldown." We say, "You must pull down the sound to match the picture," for instance.

The four film frames of the sequence in Fig. 8.4 are labeled A, B, C, and D. The concept of these frame labels is important to be able to generate a correct list to match back to film frames from a video edit decision list. The "A" frame in this standard scheme is defined as one that produces two fields in one video frame originating from the same film frame and is assigned a frame time code of 00, 05, 10, 15, 20, or 25 in 29.97-fps video.

THE EUROPEAN ALTERNATIVE

European users do not face quite so many complications as those in NTSC countries. Twenty-five-frames per second video is matched by shooting 25-fps film for television, and theatrical features shot at 24 fps are simply sped up by about 4 percent when played on local television. This does lead to a pitch shift that should be corrected, but often is not.

SMPTE TIME CODE SYNC

Of all the areas of film and television sound recording, probably none has simultaneously advanced the industry more and caused more problems than time code. Originally developed for videotape editing purposes, it was soon extended to keeping multitrack audio machines in sync with video machines for postproduction[10] and then to production sound. Planning time code usage in

---

[9]Pulldown has two meanings in the field: the −0.1 percent speed change described here and also that part of the cycle of a film projector in which the shutter is closed and the film is advanced.

[10]The companion use is for studio video recordings when multitrack is in use, such as studio audience shows.

FIGURE 8.4 The sequence of film frames to video fields and frames produced on an NTSC or U.S. high definition telecine. The film is run at 23.976 fps and alternating two and three video fields are produced from the film frames. The result is video frames running at 29.97 fps having the frames derived as shown. Film frame A and video frame 1 is the only combination that uniquely associates one video frame with one film frame, a fact that can be used to identify the A frame from the associated video. Video fields each consist of one-half of the lines of a video frame, interlaced with each other.
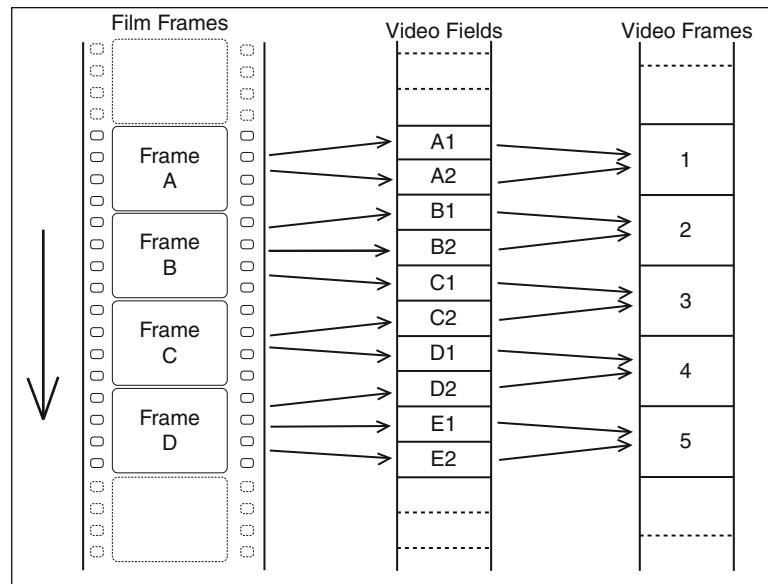
preproduction and having all parties involved carrying out the plan can greatly improve efficiency. The usual way problems are avoided today is to hold a preproduction meeting among the producers, the cinematographer and his or her assistants, the camera rental house, the postproduction facility, and the production sound personnel. There all the steps can be discussed, and the most expertise will be available to everyone.

It is commonplace to edit film shot for either film or television release with standard-definition video. This permits making copies that the producer can watch at home on standard television, for instance. The film uses time code slates running at the actual 30 fps frame rate and the same time code frame rate on the audio. The film runs on the telecine at 29.97 fps, to make an NTSC-compatible video output. The audio has to be slowed down 0.1 percent to match the video in length and stay in sync, and the resulting slight pitch shift is acceptable. However, there are several consequences of operating video at a rate slightly slower than clock time: one of them is that the time code numbers no longer match real time. Because real clock time is desired for some operations, like running a network, this problem leads to variations in types of time code.

Types of Time Code

There are a number of different time codes standardized by SMPTE, and several others that have come into use without being standardized (Table 8.1).

Each of them has a specific reason for existing, and they are very difficult or impossible to mix in some editing systems, whereas others may be capable of certain mixtures. It is not uncommon for a project to start out with one time code in production, use another in postproduction, and use a third for delivery on video, for reasons that we discuss later. In many cases today, though, the time code type needed for the final delivery master is used all the way from production through post to delivery master. Most problems related to time code have to do with misunderstanding the requirements for each stage of the chain or losing traceability somewhere along the way. To understand why a project may use three different codes during its production, it is important to understand a little about the function of each of the codes.

- **23.976 fps** code is used for certain high-definition video applications in which the HD camera is substituting for film. The rate is chosen so that direct conversion to NTSC by electronic 2:3 insertion like that in a telecine transfer is in a fixed relationship, without necessitating a pulldown.
- **24 fps** code languished for some years, as dominance by postproduction NTSC video-based editing systems precluded its use. Today, however, it is in increasing use in Hollywood as originally envisaged by the writers of the standard. It is extremely simple in implementation because the film frame rate is usually 24 fps, and if postproduction can handle it, one time code will do throughout. The complication comes when NTSC video outputs are necessary, for producers for instance. Then a pulldown must be inserted at the output of the editing system to record to a consumer format.
- **25 fps** code exists for the production, editing, and distribution of European television programs, whether originating from film or video. Films for European
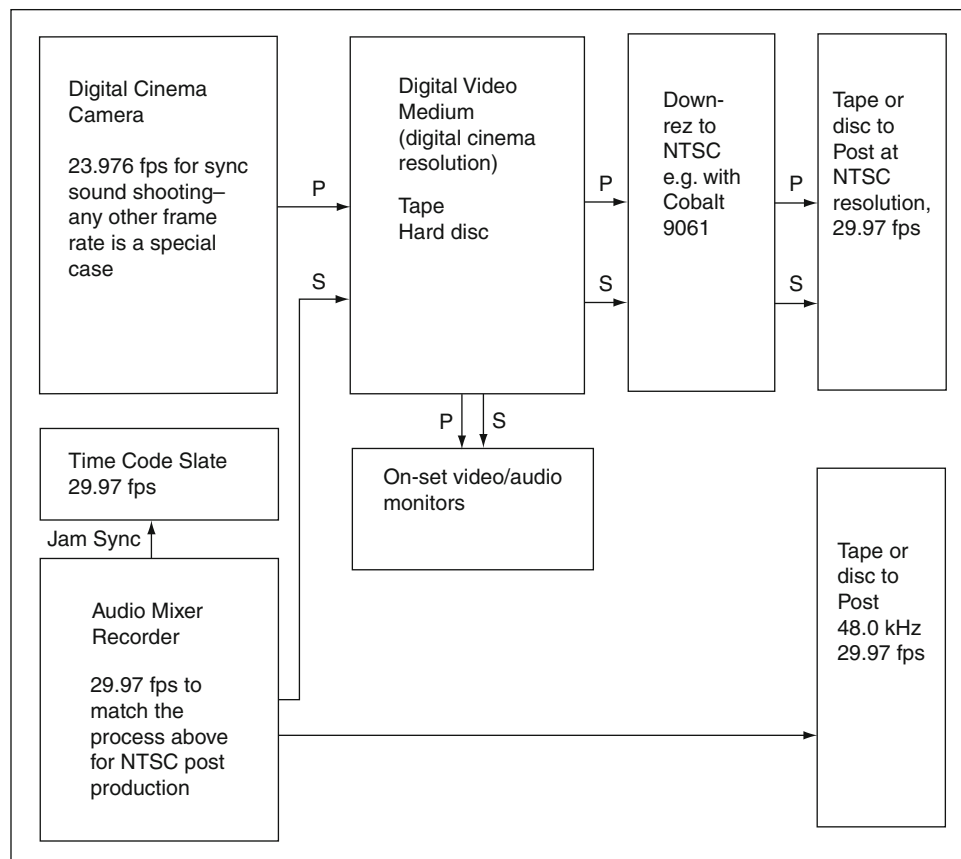
FIGURE 8.5 Note that the slow down (by 0.1%) when picture material is ingested into post production requires a corresponding slow-down in sound. The production sample rate may be 48 kHz and thus 47.952 kHz after slow-down which must then be sample rate converted back to 48 kHz. An alternative is to shoot at 48.048 kHz sample rate which is then slowed down to 48.0 kHz, the standard sample rate, for post production. The time code slate runs at the same rate as the audio recorder so that the time code photographed on film matches that recorded by the audio recorder (or there may be a fixed offset between the two that is easily accommodated once understood).

television are simply shot at 25 fps, and each frame of film transfers exactly to each frame of video, with the European video systems being based on the 50-Hz power-line frequency there. If U.S. films are shown on European television, a transfer is made at 25 fps, running 4 percent shorter in time and generally with the pitch raised 4 percent.

- **29.97 fps non-drop-frame** time code was developed to account for the fact that the frames in color television do not go by at 30 fps but rather slightly slower. The frames are numbered from 0 to 29 inclusive, so there are literally 30 frame numbers per "second." For every 30 frames, the time code *counts* 1 sec. The problem is that the real frame rate is actually slightly less than 30 fps. Because the rate is *slower* than real time, it takes *more* than one actual clock hour to increment the time code hours counter, an error that accumulates to 108 extra frames per hour of code. In other words, a program that measures 1 hour long using 29.97 non-drop-frame time code will run 1 hour plus 3.6 sec of clock time. This 0.1 percent error is unimportant in timing short segments of a minute, but is important in matching clock time for longer programs. Non-drop-frame code is used in many shooting and editing situations, partly at least because it has simpler arithmetic than the next code.

- **29.97 fps drop-frame** time code was invented so that the time indicated by this code could remain very close to clock time. By not counting certain frame numbers, drop-frame code keeps clock time. The counter frames 00 and 01 are skipped once every minute, 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 exact at many places in an hour but adds up to one displayed hour in one actual hour. Real frames of film or the corresponding video are not dropped or lost; it is the frame numbers that are skipped to keep on clock time. This results in jumps in the code numbers.

TABLE 8.1 SMPTE/EBU Time Code and Its Uses[a]

Frame rate (fps)	Frame count	Primary uses	Notes
23.976	Non-drop frame	Some high-definition 24P video for editing on NTSC 29.97 systems	Typical on some prosumer cameras
24	Non-drop frame	1. Film-only projects to be edited on specialty 24-fps video postproduction systems 2. High-definition 24P video (usually these are quite expensive cameras)	Now seeing increasing use, as originally envisaged, on slates and audio recorders
25	Irrelevant, but non-drop frame	European video and film shot for television	
29.97	Non-drop frame	Camera original production, postproduction editorial, delivery masters for some nonbroadcast uses such as DVD	One hour of time code runs 1 hour plus 3.6 sec of clock time
29.97	Drop frame	Delivery edit masters for broadcast, original video production on episodic television, live television production (so the time code can match the control room clock)	Skips selected frame numbers to stay on clock time; often represented with a semicolon between the seconds and the frames digits
30	Non-drop frame	Shooting film for music video or commercials for television at 24 or 30 fps	Becomes 29.97 non-drop-frame time code when slowed down in telecine
30	Drop frame	Used with 24-fps film origination of episodic or long-form television when the delivery master requirement is 29.97 drop frame and the editing system can handle it	Becomes 29.97 drop-frame code when slowed down on the telecine

[a]Do not confuse the time-code frame rate with the picture frame rate because the two can be different: Film can be shot at either 24 or 30 fps and will usually use the 30 fps time code on the slate and production sound recorder, for example.

Drop-frame code may be used in situations in which correspondence to real clock time is important, such as in live television. Broadcast networks universally use it and thus edit masters of programs delivered to networks must have drop-frame code. Drop-frame code is indicated by the use of a semicolon between the seconds and frames digits of the time code display, such as 01:23:59;15.

- **30 fps** time code was originally designed to be used with U.S. black-and-white television. Today it is widely used on sound recorders and time code slates as a choice for film shoots that are to be posted using video editing systems. When film shot at either 24 or 30 fps is put on a telecine for transfer to video, it is slowed down very slightly (0.1 percent) to accommodate the needs of color television and it is thus converted to 29.97 fps (with the 24-fps film also undergoing the 2:3 process described above). The sound is slowed down the same amount in transfer, so the two stay in sync. This rate is not used in "pure" video applications shot on video cameras, because even a black-and-white video today generally uses the color television frame rate.

- **30 fps drop-frame** time code has come into use for shooting episodic or long-form[11] television programs, for which the delivery master is to be at 29.97 drop frame. When the film and tape are run on the telecine equipment for conversion to video they are both slowed down by 0.1 percent, the time code becoming 29.97 drop-frame code in the process. The picture and sound editing systems must be capable of using drop-frame code.

A typical project for long-form television shot on film may use the following scheme:

- Shoot 30 fps drop-frame time code on the slate and on the production sound recorder, using a camera frame rate of 24 fps.
- Transfer the project on the telecine. Both the sound and the picture are slowed down by 0.1 percent to produce a 29.97-fps video picture, with the 2:3 insertion described above.
- Edit the project offline with a nonlinear system based on digital video, using 29.97 drop-frame code. Such

[11]A term covering made-for-television movies and miniseries.

offline editing makes use of typically a lower cost and lower quality video system than the original format. The output of the process is an *edit decision list* (EDL), a list of how to "conform" a high-quality master copy of the original video source to the decisions made by the editor in the lower cost setting. The EDL may be on paper but more likely will be delivered on a floppy disc or electronically from the offline editing system to the online system.

One difficulty with using video editing for 24-fps film projects is the necessity of using matchback techniques that cannot always be frame accurate in providing a negative cut list. For instance, it is possible to cut 30 video frames in 1 sec of screen time, but then there is no way to match that back to 24 frames of film. In fast-paced commercials and music videos, this is overcome by shooting 30-fps film as the original, but for television movies this is unlikely because of film consumption, and for theatrical films it is impossible because of universal use of 24-fps projection in theaters and no practical conversion means between the two frame rates being available. One way out of this dilemma is to use 24-fps video systems in which every frame of film is represented by a frame of video, exactly; but directly dubbed videotapes from such systems are unusable on standard equipment. Special tape copies must be made that employ an electronic process of 2:3 insertion and pulldown to run at 29.97 fps.

- The offline/online method is being challenged today by editing systems that store the full resolution of the original camera output or film telecine or scan and which output that resolution on demand. Online becomes a finishing process for color correction and certain high-end special effects. This is particularly so in lower resolution formats to begin with, such as DV with its 25 Mb/s data rate. With today's hard disc costs, it becomes quite practical to store hours of material and have it available at the full rate provided by the camera. Converting to megabytes, 25 Mb/s = 3.125 MB/s. Multiplying that times 60 sec in a min and 60 min in an hour yields about 10.9 GB/hour.[12] Using 1-TB drives (931 GB), about 85 hours of standard definition video at DV rates can be stored for about $100 at this writing!
- Conform the project from the original telecine transfer tape and possibly other sources to make the edit master tape or disk, usually under computer control in an online edit suite, with 29.97 fps drop-frame time code. An alternative is to go back to the film original and retransfer it at higher quality, under computer motion control so that scenes can be located quickly.

- Copy (dub) the master videotape to an edit master for broadcast distribution.
- Dub the edit master videotape to another delivery master for video distribution. The delivery contract will specify the type of time code required. In the case of networks, drop-frame code is typically required, whereas for DVDs, non-drop-frame code is standard.

This represents just one project, shot on film and delivered on tape, illustrating some of the issues regarding time code usage. There are many more, especially when dealing with mixed film and video projects for release in both media, in which shooting is on film, editing is on video, and release is on both film and video. Time code is the single part of production and postproduction in which planning is the most essential.

The term applied to the overall process of shooting on film, posting on video, and releasing on film is called a "film finish," and following the same preliminary steps and making a video master is called a "video finish." *Matchback* is the term applied to using a special type of EDL from a 29.97-fps source to locate and cut film frames to match a video edit. This process inevitably involves some error, but the best systems keep the error to within ±1 frame of film and maintain sync with sound. A *film cut list* is obtained from a 24-fps video system wherein each video frame matches a film frame, and editing can be exactly to the frame. This process also provides good sound sync.

Time Code Slates

In film sound production use of time code, a clapperboard slate showing a running time code is closed and at the moment of closure the displayed time code freezes. A time code generator, set to the same time and running parallel to the one in the slate, is included either internal to the sound recorder or as an external time code generator. The frame on which the sticks have just closed and banged is noted as the sync point for the picture and sound. See additional information about slating in Chapter 7.

Jam Syncing

Note that the camera is equipped with a quartz crystal speed-referenced motor and that the slate and recorder time codes must match and are both referenced to their own crystal oscillators. So their relative speed is given by the accuracy of their oscillators, but there must also be a way to set the time so that they have the same absolute starting point. This is accomplished by *jam syncing* the two together. Because the recorder and the slate may be disconnected for some time, much longer than the length of a reel, the quartz crystal oscillators in both of them must be much more accurate than for conventional crystal sync shooting.

---

[12]There are 1024 MB in 1 GB, not 1000.

FIGURE 8.6   A time code slate. Photo courtesy of Deneke, Inc.

It is customary to jam sync the recorder and slate at least a few times per day to ensure hard synchronization, although the requirement for this depends on just how good the oscillators are in the two devices.

Syncing Sound on the Telecine

Usually productions are immediately transferred from film to a video format for postproduction, even for projects that will ultimately be theatrical features. This is because of the convenience of editing on video, although certain editors still cut on film.[13]

For most projects a telecine transfer operator synchronizes the source file sound to the picture on a special playback machine that chases the telecine. The operator does this by reading the time code of the frozen frame on the picture when the slate just closes and typing it into a controller that synchronizes the sound and picture for that take and then makes the transfer to video in sync. The process has to be repeated for each take, but it is quick. There needs to be no other handling of the separate sound source material at this stage, unless a sound editor wishes to get outtakes for word substitutions and sound effects.

There is a trend to "print all" in today's world, because "printing" is no longer the expensive proposition that a film work print was. So "print circled takes," which was the method that the camera report, sound report, and script supervisor report used, is becoming more and more "just print everything."

Tip: An important feature that is required for the telecine operator to do this job is to have enough audio "preroll." That is time recorded before the slate, so that the equipment can be backed up, come up to speed, and lock sync before the slate occurs.

Latent Image Edge Numbers

Film manufacturers supply camera negative exposed with latent image edge numbers, including both human- and machine-readable types. Because these numbers are embedded at the time of manufacturing of the film, they are not referenced to the time code used on the set. During telecine, a database is made of the time code derived from the slate and the latent image edge numbers, for further editorial use.

Synchronizers

Time code synchronizers are devices that read time code and cause machine transports to come "into lock," thereby ensuring synchronization. *Chase synchronizers* are devices that accept code from a master source and cause a locally controlled machine to follow the actions of the master. The master and slave may lock to identical code numbers on the two media for dailies transfer as described, or there may be an offset of time code numbers between the master and the slave required.

There are at least two sets of time code used in a production. One is generated at the time of production and is used throughout the editorial processes. But this code would be discontinuous if copied to the edit master because picture edits would cause "jumps" in code numbers. Discontinuous code would confuse time code synchronizers, so a second set of time code numbers is introduced. Edit masters are "striped" with continuous time code, that is, time code for which the numbers increase monotonically throughout the tape. So there is one set of time code numbers for production and another for finished edited material. Other sets of time code numbers may be used for intermediate editing.

Machine Control

Synchronizers make use of shuttle commands, such as fast forward, rewind, pause, and play, that are passed from one machine, the master, to another, the slave. The common way to do this is by way of a DB-9 connector protocol standard called *Sony P-2 9-pin control* or colloquially just "9-pin" in the industry. Additionally, IEEE 1394 (FireWire) may be used to control cameras or decks by computer-based editing systems.

---

[13]See for instance the end credits of *Saving Private Ryan*, which proclaim that it was edited on the Moviola.

Time Code Midnight

One problem that occurs with time code is related to midnight. If the start of the program is at 00:00:00:00, rewinding 30 sec will yield the time code number 23:59:30:00. If a command is issued to locate 00:00:00:00, some equipment will go into rewind. It is trying to rewind by nearly 24 hours, rather than going ahead 30 sec! For that reason, it is customary to use 01:00:00:00 as the first frame of the program rather than zero time. The hours digit may be incremented to indicate reel numbers, with the first tape of the program starting at 01:00:00:00 and running for perhaps 20 min, followed by the second tape starting at 02:00:00:00, with unused code numbers in between. This avoids the midnight problem and provides simple reel numbering. In television, the "acts" of a show are defined by the hours digit.

Time Code Recording Method

SMPTE time code is recorded by a variety of means depending on the medium. Table 8.2 summarizes the methods and their placement on the medium. Professional formats allocate space on the tape to such a track. Recorded linearly along the length, this type of recording is called LTC, for longitudinal time code. Such code can be recorded on the audio tracks of consumer formats that lack a dedicated time code track.

TABLE 8.2 How Time Code Is Recorded by Medium

Medium	Placement of SMPTE time code[a] (LTC)	Other code
1/4-inch two-track analog tape	Center track	na
Analog multitrack	Track 24 of 24 tracks (typical)	na
Open-reel digital tape	Dedicated address[b] track	na
DAT	Special[c]	na
Analog videotape: 1-inch C format, Betacam, Betacam SP	Dedicated time code address track	VITC
U-Matic	Dedicated time code address track[d]	VITC

[a]All are longitudinal analog direct magnetic recordings unless otherwise stated.
[b]An address track is a dedicated area of the tape, usually used for time code recorded longitudinally.
[c]Time code recording is carried in data multiplexed with audio data called *professional running time* or Pro R time. On properly equipped machines, this means that the time code rate and type can be chosen upon playback as a calculation from the running time.
[d]Early machines lacked standardization in this area, so time code recorded on one brand of machine might not play on a machine of a different manufacturer.

In analog video formats that record the vertical interval,[14] in addition to conventional longitudinal recording, time code may be recorded as a video signal on an unused line in the vertical interval. The line to use for this vertical interval time code (VITC) is not standardized, but modern time code readers locate VITC, despite its placement. There are some other requirements regarding videotape use of time code, such as the alignment of the code to the color frame sequence of the video frames, that are beyond the scope of this book. This code in the vertical interval may be visible on some monitors.

In digital tape formats a corresponding vertical ancillary area of several of the serial-digital formats is available for time code.

For most media, one must choose the type of code from the list of seven available at the time of recording, and a dub to a copy will usually be needed to change the code type, with a special synchronizer capable of syncing to one type of code at its input while putting out a different type of code.

TIME CODE FOR VIDEO

Professional format video cameras with tape transports have built-in time code generators. For NTSC standard cameras, usually 29.97 DF and NDF are available alternatives.

In addition, there is an added concept of *Record Run* or *Free Run* for the time code generator. In Record Run, the time code generator runs when the tape runs and stops when the tape stops.

> When the tape is restarted in record, the machine reads the time code off the tape (or knows where it is), jam syncs the generator, and then begins recording.

The result is that the tapes made in Record Run have continuous code numbers, and this makes life easier for editing systems. Reels are usually numbered by setting the hours counter to correspond to the reel number, although reel numbers must be repeated when the number exceeds 23. The user bits of the SMPTE time code may also be used for numbering scenes and takes.

In the Free Run mode, the code generator counts all the time, and it may be set to clock time. For double-system video shooting, Free Run code is essential because the code must line up with the audio recorder, which has a separate generator that must also be set to Free Run and to the correct time of day. Also, for multicamera shoots, Free Run time code on each of the cameras, with them having been jam synced at regular intervals, produces matching code on all the pictures and simplifies postproduction.

---

[14]The lines that are above the picture, usually hidden from view.

An alternative to using Free Run code is to transmit the time code by wireless microphone technology to the audio recorder from the camera, in which case Record Run is an acceptable method of operation for single-camera shoots.

CONCLUSION

The bottom line on time code is that it can be very practical and efficient but its use requires continuity throughout a project and a clear understanding from the outset what is to be used at each stage of production. For this reason, interestingly, it is the producer of a television show who has the responsibility to deal with the vagaries of time code throughout the project, because it is the producer who follows the project through all stages of production and postproduction to delivery.

LOCKED VERSUS UNLOCKED AUDIO

Professional digital video formats have a fixed relationship between the frame rate and the audio sample rate: 8008 audio samples occur in five frames of 29.97-fps video, because this relationship produces the right video frame rate (by definition) and the right audio sample rate, 48,000 Hz, at one and the same time.

> To be exact, 30 fps/1.001 is 29.9700299700... fps. When five of these frames are timed out, they add up to 0.1668333333... sec. And 8008 samples at 48,000 Hz sample rate also equals the identical duration.

However, plain DV, often called mini-DV, has audio that is unlocked, without such a fixed relationship. (It turns out to be cheaper not to use video and audio clocks sourced from the same crystal quartz oscillator, but to have separate ones for audio and video.) Although this causes no problem in playback of a tape, with about $\pm\frac{1}{3}$ of a frame lipsync tolerance, it can cause a problem when sound is separated from picture in an editing system. The reason a tape will play back in sync is that there is basically no way for it to "slip" out of sync, and the playback equipment makes up for any record error by pulling the sample rate to where it must be to maintain sync—it may no longer be at 48 kHz, but this is no problem if watching or listening to an analog output.

When picture and sound are imported into a workstation, however, they are laid to separate tracks. Because the actual sample rate may be somewhat off, when it is used as a time reference for lipsync, long takes may go out of sync. Editing software today contains a feature, usually invisible to the user, that will synchronize unlocked audio samples by doing an internal sample rate conversion.

THE 2 POP

One of the requirements for synchronization remains the same, whether one is working on mag film, on the newest digital editor, or on anything in between. That is that the mixes carry a "2 pop," one frame of 1-kHz tone at reference level cut in editorial sync[15] with the "2" frame of the SMPTE Universal Leader (which indicates 2 sec before first frame of program) or the "3" frame of the Academy or SMPTE Projection Leader (which indicates 3 ft before first frame of program), cut into or recorded on the master. This is used for both analog and digital soundtracks to establish a head synchronization point. The reason it is still needed for digital formats that ought to be "perfect" is that there are various problems that can occur that the 2 pop will expose, such as missing pull-downs, delay compensation settings in the workstation, and so forth.

A corresponding pop should be edited in *edit sync*, also called *level sync*, with the Tail Sync mark on the tail leader. Then any sync issues with drift can be better understood and corrected rather than guessed at by looking at picture sync.

PRINCIPLE OF TRACEABILITY

In complex situations such as shooting to playback for a music video, a preproduction meeting should bring together all the parties, from the camera, the production sound, the editorial, and the post sound departments. Some of the principles to be followed are:

- At each stage, use word clock between digital audio sources and recorders so that there can be no synchronization errors at all. In the past, people have relied on the accuracy of sample clocks in various pieces of equipment only to find that in the end, there was sync drift. So if every transfer is "locked together" this cannot occur. For instance, for playback on the set, usually two machines will be used, one playing the source track and one recording. Connecting the word clock output of the recorder to the word clock input of the playback machine (assuming each has this feature) obviates any potential for drift.
- Use the correct time code at each stage, making the production recorder traceable back to the original

---

[15]Editorial or edit sync means the sound occurs at the same time as the corresponding picture. The alternative is projection sync, where the sound is displaced along the length of motion picture prints to accommodate separate picture and sound pickup.

source recording in the case of using prerecorded material for lipsyncing, for example. This is possible using user bits on the recorder to represent the source tape time code, whereas the "main" bits are used to synchronize with the picture. Specially equipped time code generators can read code from the source machine and place it in the user bits (this function is called reader-into-user), while simultaneously representing a main time code that matches the slate and/or camera. Thus postproduction has available the two sync sources it needs: one corresponding to the prerecorded content and one matching the picture.

This page intentionally left blank

Transfers

INTRODUCTION

Transfers are a necessary if not very creative part of the postproduction process. In fact, it is important to rule out creativity because the job of making transfers is to produce identical work, week in and week out, during postproduction so that transfers made weeks apart can intercut with each other transparently. To do this it is exceedingly important that tight standards be set and maintained for every kind of transfer.

It is worth pointing out, however, that for this "noncreative" job there are multiple Academy Award-winning mixers who started their careers in the transfer room, making transfers from original production sources, sound-effects sources, and compact discs; from film to film; and to and from digital media.

In today's world, knowledge of computer systems and how to interface them has become necessary to work in many positions in sound, because digital audio is so important. George Lucas's 1977 invention of a droid familiar with 6 million forms of communication today seems prescient. There are many file systems and formats, with transfers among them being an everyday process, that need to be understood. Sound transfer operators and sound editors in particular become conversant with words that only a few years ago were the sole province of computer sound people.

There are many proprietary digital workstations, recorders, dubbers, etc., on the market, and these may use custom formatting of digital audio within them, with certain import and export facilities to and/or from other formats. With the emergence of interchange standards at the beginning of the 21st century, transfers began to be simpler with less trouble as time went by, but today a great deal of time is still devoted to solving problems in transfers.

Although digital transfer operations are taking over, analog transfers, particularly into digital, must still be made. A web addendum to this book covers analog operations. It is at http://booksite.focalpress.com/Holman/SoundFilmTV/.

Note that most postproduction soundtracks such as premixes, final mix stems, and print masters (described in Chapter 13) will carry a 2 pop to permit finding the start point and that many will also carry a tail pop. Head and tail pops are described on page 132.

DIGITAL AUDIO TRANSFERS

Digital recording should make the task of transferring easy: As long as a bit-for-bit copy is ensured, then the digital-to-analog converter at the end of the chain sees the same digits as converted by the analog-to-digital converter at the beginning of the chain, and all the intermediate stages should be transparent, no matter how many generations are involved. However, various media are affected by different potential problems, and systems have varying susceptibility to digital errors (some are protected by error-correcting codes; others are not). We will describe the sources for potential errors along with the descriptions of the various kinds of transfers, so that they can be recognized and avoided.

TRANSFERS INTO DIGITAL AUDIO WORKSTATIONS

Postproduction audio workstations get their inputs in a variety of ways:

- By way of files exported from picture editing systems, delivered on removable magnetic or optical discs, memory "sticks," or hard drives or over a network;
- From digital or analog sync sound source machines containing production sound;
- From digital or analog source machines for wild sound or sound effects;
- From sound effects libraries in a variety of forms, most often originating on CD but supplied to editorial workstations over a network;
- From other audio workstations.

Types of Transfers

Two basic methods of making transfers need to be distinguished: *file transfers* and *streaming audio transfers*. These are quite different operationally, although the basic function when transferring digital audio is the same and may even be of identical bits. File transfers take place in a nonlinear environment and are inherently digital-to-digital transfers, whereas streaming audio transfers are inherently a linear operation, that is, they take place in real time and may be from either analog or digital sources.

File Transfers

The source for file transfers is the audio recorded within a file format on a magnetic or optical disc, tape, or memory card. This may be delivered as an original production recording or as the output of a picture editing system. The medium can be fixed magnetic disks (hard drives) that are transported among systems or removable magnetic discs, tape, or optical discs. The source may be an original or a bit-for-bit (cloned) copy.

File transfers may occur at far higher speed than normal audio rates, because bits can be copied at the maximum rate of the computer interface, which may be far faster than required for audio, so long as there is no need to monitor the audio at the same time. This makes file transfers potentially better than streaming digital audio transfers because of the greater speed. An example of this is that a USB 2.0 transfer (see later) of a 72-min audio CD may theoretically be done in 4 min and 20 sec, 16 times faster than real time.

For file transfers there are several basic items that receiving equipment must understand for interchange. The first item that must be known is related to the physical and/or electrical properties of the source:

- For removable media, the basic disc or digital tape type so the receiving computer media player can find the bits on the medium. The web site http://booksite.focalpress.com/Holman/SoundFilmTV/ gives information about removable media used for digital audio files.
- For fixed hard drives to be "mounted" on a receiving system, the basic drive protocols, such as ATA (IDE), SCSI, FireWire, or USB and the details for each of these for connection to the receiving computer. Web tables give information about some of the electrical and mechanical interface standards for contemporary disk drives. Although older drive protocols may be usable, and surely new ones are always under development, these are representative today.
- For networked systems, the network protocols to find and read files, along with necessary authorization and contention control (who can read or write particular files and when), for both local and wide-area networks, such as the Internet.

Next,

- The operating system and the corresponding file system in use so that files can be located on the medium. The web tables referenced above give examples of file systems.

Then, for all the source types above,

- The audio file format so that the receiving application software can understand what parts of the file are audio, how it is organized, and how to convert it into the representation needed within the receiving software. Table 9.1 gives examples of audio file formats in use.

And/or,

- In the case of systems in which the audio has been manipulated before entering the audio postproduction process, such as on a picture editing workstation, a list of instructions for how to duplicate what the former stage did to the audio, called an edit decision list or an audio decision list. In addition, there may be other information about the files transferred, such as scene,

TABLE 9.1 Some Digital Audio File Formats without Edit Data[a]

Shorthand	Full Name	Origin	Notes
IFF	Interchange File Format	Amiga operating system	A general file format that may contain sound. Grandfather of RIFF and AIFF, only occasionally used today.
RIFF	Resource Interchange File Format	Microsoft developed from IFF	A general file format with two common variations: .wav for audio and .avi for video.
WAVE, WAV, .wav	WAVE	Microsoft developed from RIFF	A variation of RIFF specifically for audio; nearly 100 bitrate reduction schemes are registered, but only LPCM is used in film and television production.
AIFF	Audio Interchange File Format	Apple developed from IFF	Standard Mac digital audio.
AIFF-C	Audio Interchange File Format–Compression	AIFF with provision for bitrate reduction	One of the export formats supported by Avid.
SDII	Sound Designer 2	Digidesign	Pro Tools, Avid's Media Composer, Film Composer, Media Express, and others.

[a]These are files of historical interest or currently used in film and television digital audio. There are many other formats for computer audio.

TABLE 9.2 Some Digital Audio File Formats with Edit Data

Shorthand	Full Name	Origin	Notes
BWF	Broadcast Wave Format	EBU developed for radio workstations from WAVE	WAVE files with added information such as origination time, time of first sample since midnight, description, etc.[a]
AES31	AES31	AES developed from BWF with extensions for time code, etc.	BWF with added edit instructions called audio decision list (ADL) in human readable form, with among other things sample accurate time code (SMPTE code plus sample number), clip name, source material file, track number, in time, destination time, out time.[b]
OMF1	Open Media Framework 1	Group of companies led by Avid, including Microsoft, Intel, etc.	May contain multiple media including picture.
OMF2	Open Media Framework 2	Group of companies led by Avid, including Microsoft, Intel, etc.	May contain multiple media including picture. Use with DigiTranslator to export Avid files to Pro Tools.

[a]The wav file extension is still used because then applications that make no use of the additional data in BWF can read the file.
[b]This is called a simple interchange format. Exact instructions for re-creating an equalizer, for instance, are deliberately left out of the ADL. Standards in development allow for more sophisticated project information to be transferred.

take, camera roll, production notes, clip as named and logged by the picture assistant editor, source time code, etc. Table 9.2 gives examples of audio file formats that include EDL/ADLs.

Confusingly, some picture editing systems may deliver files that have either:

- The digital audio samples *and* the editing/manipulation instructions in one file or
- The editing instructions in one file with pointers to other files containing the actual audio.

This point has probably resulted in more wasted hours in postproduction than nearly any other in recent years, because it is commonplace for people to export the second type of these from, say, the picture editor to the sound department, without understanding that they are sending only an edit decision list, and not the actual audio! A clue to this is the size of the file. Editing instruction files are rather small, whereas audio and video files are usually quite large.

Files that contain both edit instructions and the media needed for the edit of a project are called "consolidated" on the Avid workstation. Files wherein the audio has been signal processed by the editing system are called "rendered," a term that comes from computer graphics and is now applied to audio. Normally the desire is to export the raw audio plus the instructions for rendering rather than rendered files themselves. However, the export process may not provide enough information so that it can be duplicated in the audio editor for some processes. In this case, it may be useful to produce two versions of a given track, one rendered and one not, within a file for export from picture editing to sound editing.

Editing systems take various tacks on how much information they present to the operator about where files are located, and sometimes "automatic" systems place files in areas of the file system that are not understood by users. For instance, an editing system may place a newly generated audio file on the top directory of the drive that is least full, without the editor knowing this. When export to the sound department is necessary, it may be essential to know where this particular file is located. Overcoming this media management problem depends on understanding the types of files in use and their origin and destination locations in a transfer.

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 older formats, accomplished by placing new "wrappers" around the file recorded in older formats to add new information. In this way some audio file formats today bear a strong resemblance to Russian nested dolls—opening each one reveals a new 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-bitrate coded audio (see Chapter 3), but when used as an AES31 file, the coding is restricted to linear PCM.

One of the first of these formats was called IFF.[1] 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

---

[1]See http://www.ibm.com/developerworks/power/library/pa-spec16/; retrieved 1 June 2009.

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.

The second type of file contains the audio media and editing instructions, and additional *metadata*, or data about the data. These formats include OMF1, OMF2, AAF, and AES31. Audio file formats that may be embedded within AAF and 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.

Digital audio with a sample rate of 48 kHz using 16-bit linear PCM coding requires 5.76 MB per track minute (48,000 samples/sec × 16 bits/sample × 60 sec/min ÷ 8 bits/byte = 5,760,000 bytes). So 1 GB[2] of digital audio storage holds 173 min of monaural 48-kHz-sampled 16-bit audio. Other sample rates, word lengths, or numbers of channels can be scaled from this number. For 5.1-channel sound, for instance, multiply 5.76 MB × 5.005[3] to find the size of the required file by track.

Media are filled to various degrees depending on their use. For instance, although media for exchange 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.

Common Problems in Digital Audio File Transfers for Film and Television

Here are some of the common problems found with digital audio file exchanges, in particular those that come from the picture editing department. Some of the most crucial of the tasks above devolve to picture editing assistants, so sound people can help by knowing what the issues are and communicating clearly what is expected. In addition to the more strictly technical matters above, there are a number of ordinary editing tasks that fall to the picture editing department, but which can have a large effect on the sound department.

1. File operations
   - The software version of the system that generates the export data may be different from expected between source and destination or among the source, the translator software, and the destination.
   - The software version may change part way through a project.
   - File naming conventions may be unintelligible or illogical. A file named "tempdub" conveys hardly any information as there are a great many of these in a production.
   - File extensions, such as .wav, may be left off the end of file names.
   - Export media may be formatted improperly, including fragmentation caused by not starting from an empty file structure on export media.
   - OMF files may be transferred between different editions of an editor before export to the audio system.
2. Editorial operations
   - Track layout may be illogical, for example, sounds jumping among tracks, interchanging boom and lavaliere tracks at random. Consistency in naming and logging of tracks is necessary.
   - Audio may not be 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. Dialog may then be split across other tracks and its 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 very nearly in perfect sync. There is software on the market that helps to autoconform sound to sound, such as Vocalign and Titan from Synchro Arts.
   - Wild sounds may be inserted 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-coded file before insertion into the picture edit, and then importing into the edit file with time code, provides a means to get precisely the same sync in sound editing. Note that such sound is exported, but the sound editor may need, for instance, 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.
   - Overreliance on exporting systems has led to less accurate production of human readable EDLs, but these are the only backup if the export fails.

---

[2]Computer RAM memory size counts have traditionally been binary, in which the count is by units of 1024, not 1000. However, hard disk drives have been marketed by units of 1000. So within one computer there may be two counting schemes at work, one for RAM memory and the other for hard drives. To help reduce this confusion, SI standards now state 1024 bytes = 1 KiB, 1024 KiB = 1 MiB, 1024 MiB = 1 GiB, a GiB thus consisting of 1,073,741,824 bytes. This counting scheme is applied to random access memory chips. The standard kB is now 1000 bytes and so forth without confusion, and this is applied to hard drive sales. It is the reason that the Apple System Profiler reports a drive sold as 1 TB as having a capacity of only 931 GB. Apple is using the strict definition, which today should be called GiB.

[3]5.005, not 5.1, because 5.1 was gross rounding of the actual value to one that could be understood more readily. The LFE channel is sampled at 1/200 of the system sample rate of 48 kHz, yielding the 0.005 number. See Chapter 13 for a full explanation.

- There may be text information on a sound clip that confuses the system to which the audio files are being imported.
- Subclips/group clips in multicam shooting also may confuse the receiving system.
- Start times for a session may not match the sequence time into which sound has been laid, thus causing an offset in export/import functions.
- Noisy editing rooms with bad monitoring lead 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 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[4] in the monitor speakers were burned out.

3. Digital audio problems
- There may be different sample rates between source and destination. This is a problem particularly when music interests are involved, as 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.[5]
- Audio may be digitized on the picture editing station in "draft" mode at lower than normal sample rates. Less common than it once was, as the capacity of hard drives has grown exponentially and it is less used nowadays, it could still be a problem.
- There may be sample rate conflicts due to pulldown on a telecine. Being shot with the audio at 48-kHz standard sample rate, and then pulled down on the telecine, results in the picture running at 23.976 fps rendered with 3:2 field sequencing to 29.97 fps. If the audio has been inserted into the picture editing directly without a corresponding pulldown, it will drift in sync by 0.1 percent; the original audio is running in clock time, but the slowed-down picture is off-speed by being 108 frames per hour long (one hour of non-drop-frame time code time takes one hour plus 108 frames of clock time). This can be checked by measuring the length and calculating the difference. If it is 0.1 percent, which is 1 frame in 1000, the likely source is the lack of a required pulldown. What needs to happen is for the audio to be pulled down to a sample rate of 47,952 Hz, to correspond to the picture pulldown, and then sample rate

converted to 48 kHz exactly so that the audio will be on the standards for the rest of the process.

4. Ordinary audio production item-caused problems
- The level may be too low or high.
- The channels may be mixed together or interchanged.

5. Improper setup for export
- Incorrect consolidation (in the case of Avid, leaving Audio Suite plug-ins activated) of group information, pitch change, time compression/expansion, fades, and levels may occur.
- The sound department needs "handles," that is, sound for each region of audio used in the picture edit before the beginning and after the ending, providing the sound editors with 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 editors. For long-form work, handles that are the length of the take are provided, so the maximum chance is available to find presence that intercuts.

6. The media may be improperly or not labeled. A web label at http://booksite.focalpress.com/Holman/Sound FilmTV/ suggests content for a label for interchange, with examples for hard disks and removable media.

7. 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. For finished files, this is one of the purposes of the head 2 pop and tail pop.

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.

Streaming Digital Audio Transfers

More traditional streaming digital audio transfers usually occur over an electrical interface called AES3. This electrical format carries two channels of audio of up to a 24-bit word length at a 48-kHz sampling rate on one wire, with the audio playing in real time. The corresponding consumer version of this electrical signal format is called S/PDIF (the two versions have incompatible electrical standards, and some bits are used differently between the two), and S/PDIF is what appears on the digital audio output jack of a CD player. The professional version is balanced, of 110-ohm impedance, and uses XLR connectors, and the signal occupies frequencies up to 10 MHz. A variant used alongside professional video systems called AES3id substitutes unbalanced BNC connectors and video levels and impedance (75 ohms) so that audio pairs can be treated just like video in routing around a facility. None of these signals should be plugged into analog equipment inputs.

---

[4]High-frequency radiators in a loudspeaker.
[5]But many effects may be all right without correction.

Problems Affecting Streaming Transfers

Inevitably there are digital bit errors with tape or optical media, so there are error codes used that are capable of recovering the original digits for the assumption of transparency to be true. If the error code is capable of recovering the original bits during each generation, then all is well despite the number of generations. But should the error coding break down and start substituting interpolated data into missing portions, then each of these interpolations may accumulate over generations, and what might not be noticeable in one generation could be quite noticeable after several generations.

The difficulty with errors leading to interpolation is that the machine may not indicate to the user that it is interpolating. No consumer digital machine, for example, has indicators of how much correction is occurring, and many professional machines make this an obscure matter, with the user having to hunt for the correction indicator, sometimes hidden inside the machine!

Audio Sample Rate

The audio sample rate of either file transfer or streaming types of digital audio is affected by the targeted delivery medium. Film and television video is standardized for release at 48-kHz sample rate, and this is widely employed on most machines (Table 9.3). However, audio from production recorders shooting double system for film are destined to be slowed down in transfer. The accompanying slowdown in sample rate would yield a nonstandard rate, unless measures were taken in advance to prevent it. Thus double-system production recorders shooting with film often use a sample rate of 48,048 Hz, exactly fast enough so that when slowed down on the telecine, a 48.0-kHz sample rate is produced that matches the 29.97-fps video. Otherwise either a sample rate conversion or a conversion to analog and back to digital must occur to keep sound in sync with picture. See Chapter 8 and the next three sections for additional information.

Locking Sample Rates in Streaming Transfers

Another problem, especially in mixed audio–video environments, is synchronization. When sound editors talk about *synchronization*, they are typically concerned with audio–visual synchronization to within a fraction of a frame, close enough to satisfy the human brain that dialog and effects are in sync, but not sample accurate. For transfer operations, on the other hand, we use *synchronization* to mean sample-accurate sync between the source machine and the recorder, a resolution much finer than the frame-based lipsync provided by time code. The variance here is as small as 1/48,000 of a second.

Each recorder must synchronize either to the incoming data stream or to a separate sync signal that the incoming stream also is following, or else clicks or dropouts may occur. Normally this is simple if both machines are digital audio and are locked to a common external sync source.

> The input for this sync signal may be called *word clock, word sync, sample clock, reference clock, color black*, or *line reference*. Some of these are at different rates, but all have integer relationships to the audio sample rate. For instance, word clock, which is a signal at 48 kHz, has a precisely fixed relationship to 29.97-fps video of 8008 audio samples per 5 frames of video. Interconnection details must be worked out with specific equipment to make use of one of the references.

TABLE 9.3 Frame Rates for Film and Video with Time Codes and Sample Rates

Frame rate	Where used	SMPTE time code frame counter	Audio sample rate (Hz)
23.976	24 p video	0–23, then 0	48,000
24	Film; 24 p video alternate	0–23, then 0	48,000 or 48,048[a]
25	Film or video for PAL television	0–24, then 0	48,000
29.97 NDF	NTSC video, some production	0–29, then 0	48,000
29.97 DF	NTSC video, some production, and network delivery masters	0–29, then 0, dropping frame numbers at the start of each minute except for minutes 0, 10, 20, 30, 40, 50	48,000
30 NDF	On production sound recorders for film to be posted NDF	0–29, then 0	48,048
30 DF	On production sound recorders for long-form television and film projects	0–29, then 0, dropping frame numbers at the start of each minute except for minutes 0, 10, 20, 30, 40, 50	48,048

[a]See Chapter 8 for an explanation of which to use.

Without both sender and receiver locking to the same source for sync, or the receiver following the sender, the two may drift, if even slightly, with respect to each other in speed and eventually cause a click or dropout as the digital audio words cross the boundary of available buffering in the receiving recorder. Normally in a production facility, the recorder will be connected to word clock or its equivalent, so that it runs in complete synchronization with the facility. The source machine must also be "locked up." However, not all sources of digital audio are capable of locking to an external sync signal. For instance, CD players very rarely have this facility, and consumer DAT machines also do not. Thus for an error-free digital transfer from a CD or consumer DAT, the recorder has to be taken off the facility's word clock and made to run with synchronization derived from its input signal. Then, when playing back such a recorder into the rest of the facility, word clock must be reestablished.

In mixed audio and video environments, the audio machines must be synchronized to the video studio master generator, just as all the video machines are, so that audio can be exchanged between the machine types. When all the studio's audio and video machines are not locked to the same source, the result will be errors.[6]

Wrong Sample Rates

The CD runs at a sample rate of 44.1 kHz, nonstandard for film and video use. To import a CD file into a session for picture by either the file transfer or the streaming audio methods, it is necessary to undergo sample rate conversion, for which hardware and software programs exist. These vary greatly in quality, mostly based on the available resources to do the mathematics involved. Sample rate conversion can also provide clickless transfers in the case of near, but not quite equal, sample rates. This would apply, for instance, to a consumer DAT machine with recordings at 48 kHz. Using their digital output and connecting it to the digital input of a workstation may or may not work, that is, be clickless. It depends on the ability of the workstation to be set to follow the clock of the incoming source, rather than running on its own or being locked up to a facility's master clock. The inability of inexpensive machines to be locked up to external word clock means that transfers could contain errors. Such problems can be overcome by an appropriate sample rate converter used on the input. Professional facilities have DAT machines with word clock inputs, overcoming this problem because the DAT clock and the workstation clock can be made to march in lock step.

For file transfers, sample rate conversion software may be necessary for pulldown, even with "drag-and-drop" applications, to come out with the correct sample rate, or the file must be played out in real time through an AES3 port on the equipment into a hardware sample rate converter and thence be rerecorded.

The method to get around this is to shoot the original at 48,048 Hz for 23.976- or 29.97-fps video, then upon pulldown the rate becomes the standard 48,000 Hz and no sample rate conversion is needed, just a speed slowdown.

Sample rate conversion may occur in one of two modes, synchronous and asynchronous. Synchronous sample rate conversion is done when both the source and the destination machines are locked to the same reference, but are at different rates, such as 44.1 and 48 kHz. Asynchronous sample rate conversion is done when one machine is unlocked, such as a consumer DAT machine feeding a professional recorder locked to house sync, where the source is at 44.1 and the destination at 48 kHz, but the two are not locked to each other. In this case the sample rate converter continuously changes its multiplication and division ratios so that clicks or dropouts do not occur.

Revert to Analog

In some cases, digital transfer is impossible because the conditions required for locking sample rates cannot be accomplished, for instance, if sample rate conversion is not available. In such cases, it may be necessary to go through digital-to-analog conversion on the playback machine to analog-to-digital conversion on the recorder. In such cases, the biggest issue is perhaps level setting. Head tones on source reels recorded at a known digital level, such as −20 dB F.S. (see later), can help to make a 1:1 transfer despite having "gone back" to analog.

Digital Audio Levels

Level change across a digital-to-digital interface is rarely used, with most copies being bit clones of the source.

> If a level change downward is needed between the source and the copy, then the level change should be accompanied by the proper amount and type of dither (deliberately added noise to eliminate quantization distortion) to maintain low distortion at low levels, as described in Chapter 3. Level changes upward across a digital interface will probably result in the dither present in the source properly dithering the output, so no special measures need to be taken.
>
> In the case of some computer digital audio boards, there may be multiple level controls, such as one provided by the application software and another provided by the operating system software. If levels recorded digitally across an interface to or from these systems seem especially low or high, it may be due to there being multiple "hidden" level controls that must be set.

ANALOG TRANSFERS

Analog transfers are covered in a web extension to this book at http://booksite.focalpress.com/Holman/SoundFilmTV/.

---

[6]The clicks or dropouts may occur as infrequently as once in a program or as often as many times per minute, depending on the difference in sample rate clock between the source and the receiver.

ANALOG-TO-DIGITAL AND DIGITAL-TO-ANALOG SYSTEMS

Any conversion operation between analog and digital brings with it the potential for level shifts, but these are simply removed through calibration. If all machines in a facility are aligned in the same way, then there will be no unexpected level changes throughout the generations required by postproduction processes.

In an effort to make digital machines more like analog ones, some manufacturers have placed a 0 dB reference level at anywhere between −12 and −20 dB re: full scale,[7] with −12, −14, −18, and −20 dB being the most popular. It should be noted that these are purely arbitrary references. They have little or no meaning for any real program material because the program is dynamic, moving continuously both below and above the reference. These references should not be treated like some maximum for program material, because to do this would be to lose 12 to 20 dB of dynamic range. For this reason, most professional digital audio metering has 0 dB at the top of the scale, corresponding to exercising all of the bits in the system. This is sensible because it ends arguments about how much headroom to leave above 0 dB: The simple answer is "none"! Still, there is reason to leave some of the top of the dynamic range unused, such as 2 dB, for two reasons. One is that meters may not in fact read the true peak of the signal as the actual requirements because the potential of intersample clipping has only recently become understood, and the other is that the antialiasing and other potential filters downstream in the system may have some "overshoot," in that their output level exceeds their input level by a small amount.[8]

TABLE 9.4 Reference Electrical Studio Bus Levels Used for Reference Digital Coding Level[a] on Digital or Reference Fluxivity[b] on Analog

Reference level	Reference voltage (Vrms)	Users
200 mV	0.200	CD, DVD, Blu-ray
−10 dBV	0.316	Home, semi-pro recording equipment (where it is often abbreviated "−10")
+4 dBu	1.228	Most music and postproduction studios, modern broadcasters

[a]Such as −20 dB F.S.
[b]Such as 185 nW/m.

Using system equipment for a purpose different from what it was designed for may lead to unintended level changes. For instance, a consumer digital tape recorder with a reference level of −10 dBV is brought into a professional studio, where the reference level between devices is +4 dBu (Table 9.4). The 12-dB difference (not 14 dB because the references are different) must be made up in transfer by an amplifier called a *match box*. Also, when using analog sources to be copied to digital media, the analog source should be decoded for any companding noise reduction before recording digitally.

---

[7]Written −20 dB F.S., meaning −20 decibels referenced to full scale. *Full scale* refers to the maximum possible undistorted level of a sine wave.

[8]See AES-R7-2006 (AES standards project report): Considerations for accurate peak metering of digital audio signals, http://www.aes.org/publications/standards/.

Sound Design

Sound design according to one definition is the art of creating a coherent soundtrack that advances the story and the picture, and it demands an overarching conception of a movie's sound as well as a capacity to solve aesthetic and technical problems at the level of 1/10 of a second. "Design" is used to emphasize the creative larger conception of a movie and the capacity of sound personnel to create imaginative sounds that advance the story. Thus sound design is the art of getting the right sound in the right place at the right time. The right sound means that the correct aesthetic choice has been made for that moment in time. The right place relates to the high degree of organization that is necessary over the process, combining sound where possible within a premix for simplification of the final mix, but also keeping the various sound elements separate enough so that necessary changes can be made late in the process. It is the balancing act between these two issues that is at the core of the mechanical part of a supervising sound editor or managing sound designer's work. The right time refers to the correct position in editorial sync. So sound design can be seen to embrace both aesthetic issues and "manufacturing" details, from the inception of temp sound for temp mixes through the preparation of print masters for release on a wide variety of media.

The term "sound design" arose in the 1970s to describe a difference from how normal postproduction was done. In Hollywood, a team of sound editors led by a supervising sound editor prepared tracks, and rerecording mixers finished the process. Typically each sound editor was assigned a reel (10 min long) or more, and the supervising sound editor kept everything hanging together: making sure the Ferrari screaming by in *Grand Prix* was the same one in reel 2 as in reel 7. Although there was interaction between sound editors and mixers, editing on the dub stage was frowned upon because of the time it took to do when cutting mag film. In fact, some dub stages were equipped with ping–pong tables and pinball machines, and at least one had a basketball hoop, to placate the producers, directors, and mixers while re-editing.

A different conception of soundtrack completion arose in this period, with Northern California roots. Unhampered by strict union rules governing what union local had jurisdiction over which jobs in Hollywood, in particular Walter Murch was called upon, for *Apocalypse Now*, to handle all the postproduction sound chores, from conceptual work on the track months before intensive postproduction began, through sound editing and completion of the final mix. The inspiration for use of the term "sound designer" seems to have been that live theater had used the term for some time, and Francis Coppola[1] had directed at the American Conservatory Theater in San Francisco in 1972, where there was a resident sound designer. One of the features of Murch's approach was to assign sound editors to specific kinds of sounds and have them cut them throughout the movie, such as backgrounds. This could help improve the unity of conception over the reels and lower the burden on the supervisor.

Related to this was the fact that Ben Burtt had done more than a year's preparation of special sounds for *Star Wars*, leading the Academy to give a second Academy Award related to sound that year. Up until then the Academy Award for Best Sound had gone to the production sound and rerecording mixers. In 1977 a Special Achievement Award, an Oscar, was given to Ben Burtt for "the creation of the alien, creature, and robot voices featured in *Star Wars*." Subsequently Sound Effects Editing became its own separate award in 1982, also going to Ben for his work on *E.T.: The Extra-Terrestrial*, along with Charles L. Campbell.

> In fact, he might have been given the Academy Award for splicing for that movie. The voice of *E.T.* was provided by a number of nonactors, the main one of whom read completely flatly. Ben had her mimic him, moment by moment, and then cut the performance together with an amazing number of splices, which I personally witnessed.

In the meantime he'd also been voted a Special Achievement Award for work on the soundtrack for *Raiders of the Lost Ark* along with Richard L. Anderson in 1981.

One broad definition, then, of sound designer is one who works on or supervises all of the sound work from the beginning of postproduction, through rerecording, to the final print mastering, to provide an overall conception of a soundtrack. The second definition is one who provides special sound effects specifically created for a particular part of a film, such as a processed voice of a character or device or other especially creative use of sound, often involving original recording.

---

[1] The director of *Apocalypse Now*.

The term was first applied in the film business to people like Walter Murch, Ben Burtt, Randy Thom, Gary Rydstrom, Leslie Shatz, and Richard Beggs, all Northern Californians, and a few others. Soon people who were doing car commercials and the like started calling themselves sound designers, because no one had a trademark on the term, and it regrettably became diluted, with a Hollywood backlash. On the air at this writing is a car commercial with a downright bad "homage" to the great sound designers: a big cat growls loudly and obviously just before a high-end automobile crashes through the window of the brand's car museum to join its place in the pantheon—a really tacky overuse of anthropomorphizing sound. A good example of the use of animal sounds to heighten a sound montage is the opening surf of *Jurassic Park III*. It contains various animal growls that you probably would not consciously notice until they were pointed out, in contrast to the overly obvious car commercial.

Today, a few academic institutions have sound design curricula, but these programs are mostly aimed at live theater and include aspects of music composition. A few books have begun to emerge from critical studies, such as William Whittington's book *Sound Design and Science Fiction* (2007).

In fact, the ingredients of sound design as we know it today have roots even before the sound film. Early 1900s motion picture exhibitor magazines include ads for kits for devices to make sound effects to accompany silent films. The traditions of half-coconuts clopped on a table for horse's hooves, canvas stretched over a barrel of wood slats for wind, and a sheet of suspended metal shaken for thunder are all examples of pre-sound-era "sound design." With the coming of sound films, some experts from other fields such as music and radio were attracted to the movie business. One in particular had a long and significant career, with his inventiveness on display in 1933's *King Kong*. Murray Spivack used recordings of lions slowed down and reversed to make the "voice" of the title character, and these are techniques still employed today. Further details, including a sound effect list and cost estimate, are available at http://booksite.focalpress.com/Holman/SoundFilmTV/, dated 19 July 1932. Spivack ran the sound effects department at RKO Studios and had a long career in film, but capped even that with another career teaching drums until at least the age of 85.

WHERE DOES SOUND DESIGN COME FROM?

Sound designers report several sources for inspiration in their work:

- The history of sound effects in movies. I was struck reading about how most novelists spend a great deal of time reading so that they know what context their work is going to be in—a logical flow that includes the history of the novel. Expert sound designers too seem to share this interest in hearing other people's work, both historic and contemporary. There is even a running inside joke among sound designers that at this point must be said to have had a good run, but is ready to be put to bed, the Wilhelm scream. First used in 1951's *Distant Drums*, it was later employed in *Star Wars* and many subsequent movies for a baddy falling to his doom.

- Aural memory plays a significant role in their work. Ben Burtt explains that the sound of the arrow striking the piece of wood that Indy uses to trigger the deadly mechanism in the cave scene of *Raiders of the Lost Ark* came from listening to a girl at the next desk in grade school twang her ruler on the edge of her desk. In a parallel idea, Walter Murch, while editing both picture and sound for *The English Patient*, used aural memory for transitions to the flashbacks that the title character has.

- Gary Rydstrom reports as an opening joke in lectures that the sequence sound designers use is get idea, reject idea; get idea, reject idea; get idea, reject idea; . . . run out of time. This is ever more true in today's world of digital central libraries and digital audio workstations because the pace of work has picked up dramatically. For the 1984 television movie *The Ewok Adventure*[2] I had an intern count all of the sound clips on mag film in all of the units. I also had available the time cards for each sound editor and knew what they were responsible for. Doing the division of how many sounds were found, cut, and documented by cue sheets per working hour the number was around 4, changing little by discipline. So one of the primary factors in today's sound design as opposed to 15 years ago is that the digital postproduction has resulted in many more sounds per hour being inserted and the ability to make many more trials of sounds.

- Ben Burtt[3] likes to say that each of us carries around in our heads an emotional dictionary, associating certain sounds with certain emotions. The low-frequency rumble equals threat equation is one of the simpler manifestations of this. Doubtless this is true because it is known that a smell can evoke a memory, and certainly a sound can. But even below the threshold of producing a specific memory, there is subliminal feeling, for example, the feeling that something just isn't right when a low-frequency rumble is present. A sound designer uses the sound to tap into this emotional dictionary. Three examples of this use of low-frequency sound are:

---

[2] I use this as exemplary of average films, not blockbusters with large postproduction budgets.

[3] Academy Award winner for special sound effects editing of *Star Wars* and *E.T.: The Extra-Terrestrial*.

- The off-screen posse chasing the leads in *Butch Cassidy and the Sundance Kid* provides a threat that comes closer and farther away as a leitmotif, edited by Don Hall.
- Four low notes in the score on a perfect ocean day indicate the presence of the shark, in *Jaws*, provided by the composer John Williams.
- As described in Chapter 1, B-52 bomb strikes at a distance were indicated by low-frequency sound in *Apocalypse Now*.

• Substitution of similar sounds for the object being represented in the picture is often used. For instance, a believable sound for a dinosaur may be needed, so by recording all kinds of animal noises, and then combining them, an extinct creature can be given a voice. In *Jurassic Park* some of the dinosaur sounds designed by Gary Rydstrom originated from penguins and another from a baby elephant trumpeting; many other animals were involved. Perhaps the most surprising is a recording of a koala bear's unique coarse breathing, which up close sounds remarkably fearsome; but of course the Australian native looks like a teddy bear.

• A variant on substitution is that sound effects can be made to seem like something other than what they are by recording technique. Whereas many practitioners concentrate on the very best quality recording methods to capture the actual sound of a source, others are more interested in what variations can be achieved with invented techniques. For example, the sound of Luke's land speeder in *Star Wars* came from another transportation sound—the Harbor freeway in Los Angeles[4]—by recording that sound through a vacuum cleaner tube. The tube acts as an organ pipe, emphasizing one frequency range to the exclusion of others, and makes a vaguely "transportation-like" sound, without revealing the source identity. Other methods include contact microphones used, for example, on bridge structures, hydrophones used underwater, and many more experimental ones. Some of these are aided by sacrifice microphones, referenced in Chapter 7.

• Perhaps the main place that sound design comes from is the story. Sound effects can be put to narrative use; for example off-screen sound (incidentally not added until postproduction) can motivate an actor's actions and literally tell the story. In Kubrick's *Killer's Kiss* the heavy, having assaulted his girlfriend, waits in a stairwell while the hero runs down the stairs to save her . . . . Sometimes off-screen narrative effects can simply be a cost-cutting measure. When asked by the studio exec how to show a police helicopter, an expensive addition to the budget for 1991's *Boyz n the Hood*,

director John Singleton replied that he could do that with a sound effect and a light—he credits that at least partially for getting his first directing job. On the other hand, a little less literally or directly, sound effects that are more ambience oriented can help set the stage for the action. For the classroom scenes in *Boyz n the Hood* Singleton remembered attending a South Central LA grade school in the flight path of Los Angeles International Airport, so he added airplane passovers to the effects track—an effect that portrayed for him that this scene occurred in a poor neighborhood.

• Spotting sessions, in which the director and the supervising sound editor or sound designer, and possibly the composer, come together and run through the picture, literally "spotting" where sound effects might go. Gary Rydstrom tells the story of working with Robert Redford on successive pictures. In early ones Redford would say, "There's a car going by," and Gary finally got up his courage to ask Redford for a less literal interpretation—how he wanted to *feel*. On *Quiz Show* then, Redford asked Rydstrom for a morality tone. Rydstrom was at first flummoxed, then rose to the occasion.

• Having the composer present at the same session as the sound designer before the score is composed helps, because areas in which sound effects are going to have large amounts of low-frequency energy, like the rolling boulder in the first reel of *Raiders of the Lost Ark*, are regions for which scoring large numbers of double basses, for instance, would be wasted. In *Raiders*, John Williams's score concentrates on higher brass instruments in this area, avoiding the masking that the rolling boulder provides. By the way, the rolling boulder sound effect was principally the sound of a small Honda station wagon's tires, close mic'd, rolling down a gravel hill.

• Films have styles, like realistic or more ethereal, and so do sounds. Point of view is important, potentially swamping other concerns. In the opening D-Day beach scene in *Saving Private Ryan*, the listener is switched among various points of view of reality: the beach in full-on assault, underwater with bullet zings, and again above water from the vantage point of men on the beach who have temporarily lost their hearing because of the extremely noisy conditions. All of these are true and real, yet they are differing points of view of the same event, driven by the story.

SOUND STYLES

Films and videos have sound styles, which can elevate the story line. By style we mean the aggregate of sound perspectives, methods, correlation or lack of it with the picture, and other factors such as the degree of reality versus constructed space that is used.

---

[4]One block from the USC campus where sound designer Ben Burtt went to school.

Sound style may vary within a given program—and usually does over longer programs—to keep things interesting. While listening to all programs we encounter sound style continuously, and it is a more or less conscious choice on the part of the film- or videomaker throughout the piece. Although the field encompasses all decisions about what sound is heard when, and thus is difficult to quantify, a few organizing principles can be given.

- Musical score heard alone usually distances us from the picture content, because there is no synchronous sound (dialog, cut effects, or Foley) to make things seem real. This is why many, perhaps most, films start with music and then add more real effects as the action gets started: it is a way to start with an abstraction and then to draw the viewer/listener into the story.
- Likewise, a break in the middle of action to hearing just music tells us that we are in montage. The old vaudeville plea from the stage for the orchestra to produce "a little traveling music please" still works in film. For instance, *Days of Heaven* opens on score underneath black-and-white photos with titles and moves on to the interior of a steel mill, with such loud sound effects that we can't hear the language building up to a fight that loses Bill (Richard Gere) his job. We have gone from filmic montage to reality. Next there is a scene change to a very dark interior and we hear some distant train effects and voice-over narration as the characters travel west. Then we hear music and see the train in a long shot—the little traveling music—in pure montage again. One definition for montage is, "A process in which a number of short shots are woven together in order to communicate a great deal of information in a short period of time,"[5] and the end of this sequence certainly qualifies. Another definition might be: a period in a film in which the picture and sound are less associated than usually, such as when the sound-track is music and the picture combines shots spaced over time.
- The debate on the dubbing stage for *Out of Africa* was interesting, as reported by rerecording mixer Chris Jenkins: just how much of the airplane should we hear while it flies through the flamingoes underneath the great big string score in a huge wideshot? The answer chosen was just a little bit. It could have been none—pure montage—but that would have been a distance from reality that was too great at that point in the film—it wasn't the beginning or ending, but rather a pause in the middle, and the decision was to hear the plane, only just a little.
- Foley sound effects seem to make things more real. Their "hyperreality," achieved through close mic'ing

of small-scale events, helps make this so. A movie that combined ADR sound unmodified in mixing, so that it seemed disembodied, and passages with no Foley sounds is *A Room with a View*. The overall effect is one of an undesirable abstraction—it just doesn't seem real.
- Ambience is the connective tissue of film soundtracks. Its constancy across picture cuts provides an anchor for the visuals that means we are in the same space, with a different perspective on the action. Conversely, its abrupt change at a picture cut means we have changed scenes or, at the very minimum, point of view. This idea is played upon by prelap edits wherein we hear the sound change before the picture change, in effect "warning" us of the scene change. This effect can even be heightened by changing the source for sound sent to the surround channels before that sent to the screen channels.
- Besides its connective nature, ambience has a particular storytelling effect. In the language of semiotics, the study of signs and how they work (which includes aural experiences), certain effects, many of them ambient ones, are "signifiers." That is, such sounds have near-instantaneous accepted meaning, shorthand for describing the "signified." For instance, after the lid falls back onto the ark in *Raiders of the Lost Ark*, and the ruckus it has made subsides, we are left with one sound, that of a cricket. The signified is peace and quiet, that the storm is over, the climax has occurred, and dénouement begins. All from one sound effect.
- Dialog has its own set of conceits. Quite often we are able to hear sound through walls or windows, or at a very great distance, just as if the actor speaking was in a close-up. Although this is neither good nor bad necessarily, it certainly does not match reality. In some cases it can be disturbing to an audience when they don't know who is supposed to be speaking, and so it may be useful to provide this information before such a sequence.
- There are several juxtapositions at work in selecting sounds. First is the juxtaposition with picture. Critical studies thinkers have been at work on this since the introduction of sound, although much less critical work has been done in sound than in picture. Some writers have said sound operates in counterpoint with the picture, using the musical term for parts having independent motion of pitch and rhythm played simultaneously. But this would be true of any juxtaposition of picture and sound. Michel Chion argues for the term "harmony" instead, with picture and sound different, but related. In the end, it is the production of meaning (or wonder, or fright, or . . .) in the mind of the viewer/listener that is sought in selecting sounds, whether simply illustrating the picture or providing an extension to it into other aesthetic domains. If anything, sound professionals

---

[5]James Monaco (1981). "How to Read a Film," pp. 183–184. Oxford University Press, Oxford.

complain that directors tend to be overly cautious and conservative in this area, using sound to match the picture rather than provide extensions to it.

- Picture and sound have mutual subjective effects. Picture without sound runs slower subjectively than with it. The presence of a picture makes sound slightly less loud and bright. Color pictures sound slightly clearer and softer than black-and-white ones of the same material.

- There are also juxtapositions within a soundtrack. Some examples are: (1) In *Band of Outsiders* Godard has a character bang her bicycle up against a railing, ending her ride and the scored music abruptly at the same time. (2) Sound designers pay attention to the effects of frequency and temporal masking, leaving, for instance, "little holes"[6] for dialog. In *Platoon*, the crucial battle near the end of the picture has lots of shooting and artillery, but none of it is over critical lines of dialog necessary to the story. Lines such as "Get out of the hole!" stand out against the background because of the space carved in the sound effects for them, not some mix manipulation. (3) At the beginning of *Apocalypse Now*, the sound of a helicopter and that of a ceiling fan are juxtaposed, by blending them in a cross-fade that takes us from one reality to another.

- The sound can be very clean and bright, or it can be more obscure. Comparable to the filters that cinematographers use on lenses to soften the image, it may be undesirable to have everything completely clear on the soundtrack all of the time.[7] However, convention demands that all dialog be intelligible, even when the director says the line is of no importance and need not be heard. If it is not, it will provoke wonder in the audience: "What did he say?" must be one of the things most said among patrons in movie theaters.

- Each film or television show has the freedom to establish its own conventions, but generally it must pay attention to them. *Ally McBeal* has meditative sequences in which we are taken inside a character's thoughts to what is occurring, and we are jerked back out to reality by a sound that is like that of a phonograph tone arm skated across a record. It becomes a convention through use that this sound has this meaning, for that show. Used in another context, the same sound could be one of reality: say, in a disco club it could indicate frustration on the part of the record spinner.

- Not all films use the conventional DM&E (dialog, music, and sound effects) breakdown because some use virtually no dialog or effects. An example is a film of the ballet *The Nutcracker*.[8] It would be possible to have only a music track for a film of a ballet, although in this particular instance, the filmmakers chose to add a Foley soundtrack of the dancers' movements to heighten the "reality" of the film, rather than to use the more abstract music with no other sound, "distancing" the audience from the work.

- Beginnings and endings are important. As we have already said, music often begins a program, which then cross-fades to realistic sound. The ending is often signified pictorially by a crane shot, starting on the level of the characters and then moving up to an extreme wideshot. The equivalent for sound is also usually music, moving us "away" from the reality of the picture.

- The style of a film is important in determining how realistic the track must be. Is it strictly to be what is seen is to be heard, or is there another dimension to the track? More than one sound designer has explained to me that many directors are very literal, wanting to hear what they see, and that's it. Movies do evolve over their length, and alternating methods, from producing a full and complete sound world to making a highly abstract one in all-out montage, may help a given story in changing time and place and even, frankly, in preventing boredom.

- Theme music is especially important as what MBAs would call a "corporate identity program" for television shows. Although I don't remember too many plots from *Cheers*, I can't get the line "where everybody knows your name" from the theme song out of my head, years after the show has left prime time.

EXAMPLE OF SOUND DESIGN EVOLUTION

One particular event common in movies draws attention to how sound design[9] has evolved since the introduction of sound to the movies in 1928. A phone call to or from a character on screen always leads to interesting possibilities, for how is it possible for us to eavesdrop on the call when we are "standing" across the room with the camera's point of view?

The early sound era answer to this is that we don't hear the other side of the phone call at all, and the actor on screen has to repeat the salient points aloud, so that we as an audience can hear them, and thus further the narrative. This is such a clumsy method of dealing with the problem that in a parallel case it became the subject of a sitcom joke. Roseanne Barr on her show *Roseanne* made a tongue-in-cheek reference to an episode of *Lassie* in which the dog of the title enters the scene and barks, and

---

[6]Ioan Allen's term for this effect.

[7]An idea expressed by Randy Thom.

[8]The 1993 edition, sound design by Randy Thom.

[9]Many would call this sound editing and rerecording. In this case, however, it extends from principal photography throughout postproduction, because a coherent means of shooting a scene including a phone calls for the attention of the filmmakers at all stages.

poor June Lockhart has to play second fiddle to Lassie and interpret. To one of her kids, who's being mute at the time, Roseanne says "You say Timmy fell down the well?" She got a big unaided audience laugh for that.[10]

In the early 1930s the "telephone filter" was invented, and it became a convention that we would hear the other end of the call, only filtered to sound like the restricted frequency range of a telephone. This modus operandi became more or less standard and has been in use up to the current time. However, there are several challenges brought about by changing technology and taste: what happens if at first we don't see the caller, but in a split-screen optical effect we then do. Do we at first put in the filter and then remove it? Probably. However, this draws attention to the change at the edit, which may be undesirable in keeping constant the dialog stream of the at first off-screen, and then on-screen, actor. What if the distant caller slides into a split screen while talking: should the filter progressively be removed? This is where art comes in: full bandwidth sound for actors heard over a telephone seems wrong, so a filter is required, but then when they come on screen it is definitely not desired, so we are left with a conundrum, solved by individual films in different ways through time.

More recently, a method has been used to establish the far end of a call as filtered, but to slowly remove the filter over time, as the fact that the off-screen voice is over the telephone has been established. In one case, it is used as a method to increase the tension of a blackmailer's control over his victim across time. In 2001's *The Deep End*, the blackmailer played by Goran Visnjic makes his early demands of the character played by Tilda Swinton over a telephone, well filtered. By the end, over multiple calls, the filter has been removed, and his presence is felt more heavily by the mother trying to protect her son.[11]

In a similar vein, *Vertical Limit* (2000) uses the same method, only for walkie–talkies. A sister played by Robin Tunney is trapped at the top of a mountain with diminishing air and knows she is about to die. Her brother played by Chris O'Donnell wants to come to her rescue, but there is insufficient time. They both know her fate and have an engaging conversation over radio that at first starts out filtered and becomes less so as one is "projected" into the space of the other. In this case, the scene cuts back and forth between the locations, and the filtering is removed over time.

The history above shows at first a clumsy representation, supplanted by the use of a technical innovation—the "telephone filter"—in use so long that its supremacy has been challenged by more daring filmmakers in the past 10 years. Still, in the modern cases cited, the off-screen voice had to start out filtered so that the audience could

figure out that what they were hearing was over the phone, not from someone standing off-camera.

Today some television commercials alternate between full-bandwidth and heavily filtered and/or equalized sound, with no apparent rhyme or reason, i.e., not related to telephone usage at all: the filter is simply used as an attractor to sound different from all the other commercials. In one commercial, when a character is supplied with a ghost image of himself fading in, his voice is doubled with a short time delay, about ¼ frame, producing a comb filter[12] to indicate the presence of a second, related voice. The commercials are doing their best to stand out from the crowd by deliberately and substantially changing the sound timbre from line to line.

SOUND DESIGN CONVENTIONS

In most cases of editing and mixing dialog, the attempt is to get a smooth and continuous-sounding stream so that the attention of the audience is not distracted by the technique. But there are several conventions that seem to contravene this idea:

- We often hear dialog completely clearly that we shouldn't be able to. In the opening shot of *American Wedding* (the third *American Pie* movie), we dolly in to an exterior of a college dorm. We hear the conversation from within the room, even though it is through a solid wall—this is a film sound convention like hearing battles in space.
- In many pictures, even though we do see the characters in long shot we hear them as unnaturally close, to promote intelligibility and storytelling. For example, Stanley Tucci and Meryl Streep as Mr. and Mrs. Paul Child in *Julie and Julia* are seen from across a busy street sitting on a park bench in Paris, and yet we hear them close up and dominant, whereas otherwise the city sounds would dominate. No one complains about this completely artificial convention, with the notable exceptions being some French new wave filmmakers such as François Truffaut. He believed that smoothing over edits was concealment, and so there should be discontinuities, and thought one boom mic hung high would do—the opposite of Hollywood tradition.
- A taxi arrives at its destination. Passengers get out and we cut to the interior looking outward. We don't see the taxi anymore, but we assume the driver has been paid, delivered the luggage, and is driving off. Actually time is compressed by the tactic of "merely" hearing the taxi drive off without actually having to follow all the events. This occurs in *My Own Private Idaho*, and is a common tactic for time compression.

---

[10]Television series *Lassie*, 1954–1974; *Roseanne*, 1988–1997.

[11]Supervising rerecording mixer Mark Berger.

[12]See pages 64 and 72.

The table at http://booksite.focalpress.com/Holman/SoundFilmTV/ gives some common film sound conventions, along with their shorthand meaning.

In a possibly apocryphal Hollywood story, a producer runs out of money. An important scene is of a kidnapping, with the victim placed in the trunk of a car and driven around to confuse him as to locale. The scene is played out in black, with sound effects, and therefore no shooting budget at all.

OBSERVING SOUND

When I was a boy going to the movies in the tiny Oregon Theater (in Oregon, Illinois, where my father had been a projectionist years earlier), everything on the screen seemed real. I well remember *Forbidden Planet* in 1956, with its theremin on the soundtrack and tales of another world. I couldn't yet recognize it as Shakespeare in modern dress, nor how special effects were made; it was all just real, even though I was listening over a mono Academy optical soundtrack in a small-town theater. As I grew up and into a professional role, my critical facilities were raised by orders of magnitude: I could take apart what was technically wrong with a soundtrack with the best of them, basically in order to tackle one problem after another so that others had the clearest path to expression—I had developed an ear. When I went to work at Lucasfilm, I delineated what I did from what Ben Burtt did as: "I make it sound good; he makes it sound interesting."

I learned only through teaching that my boyhood and my professional lives were at opposite ends of an age-old dichotomy: the romantic versus the classical—I was a romantic as a boy, dealing in superficial appearance and inspiration, driven by the creative and the intuitive; and I grew into a classicist, concerned with reason and rules, which in my case is a thorough perception and understanding of soundtracks. (It was reading the classic *Zen and the Art of Motorcycle Maintenance* that brought this transition home to me.) For many years, I saw this shift as growth. I was better able to do the job. However, perhaps I enjoyed movies less as the classical took over. Oh, a very good one would suspend my ever-increasing level of disbelief, but that was increasingly rare as time went by.

It was the experience of watching Steven Spielberg viscerally enjoy a dubbing stage screening of reels of one of his movies that showed me how wrapped up in the professional I was. He could suspend his classical side to enjoy a movie he'd worked on for several years and saw only in segments—why couldn't I come to do that?

Increasingly I have.

And what I'm left with is a profound appreciation for what the movies do for us. Since the beginnings of drama, people have come together for a few hours in a crowd to join the characters in their journey, to have the catharsis, and to go home refreshed. I minored in theater at university and learned about Epidaurus, the best-preserved ancient Greek amphitheater. I had learned that the words "drama," "character," "catharsis," "comedy," and "tragedy" were Greek. I knew the stage and seating layout of the amphitheater, another Greek word. What I found incredibly important and did not know until I visited there is what else lay on the grounds of Epidaurus: a hospital, where the terminally ill came to experience the drama, to be taken out of their skins for a few hours while they empathized with the characters and followed their character arcs to their conclusions.

Frankly, at the end of the day, it's magic. How else to explain the fact that some directors make brilliant, and then stupid, and then brilliant movies again? How else to explain that there's a certain mood and frame of mind that I get into, in which a key fits right into a lock, and I burst with joy over how wonderful a film is, and then looking at it 20 years later think, "how did I ever think that?" At the end of the day, it isn't clear how it all works, and maybe it shouldn't be. Orson Welles had this to say in a tribute to "The Greatest of All Directors," Jean Renoir, upon his death, and it bears repeating. The quotes inside the reference are ones that Welles attributed to Renoir.

> Knowing him as I did, I know there was nothing of self-pity and only a dry and impersonal bitterness in his statement to Gilliatt[13] that, "The money men think they know what the public wants, but the truth is that they know nothing about it—any more than I do." And when he said it, as he often did, that the most dangerous mistake of all was "to be afraid the public *wouldn't understand*," he was not defending the intelligence of the public (no, "the public is lazy"), but rather proclaiming the virtue of a certain degree of deliberate ambiguity.
>
> When we strain for perfect clarity, what we finally achieve is perfectly banal. That, he was sure, was the real trouble with Hollywood: Not that it worshipped money, but something much worse—that it worshipped an ideal of so-called perfection.
>
> "They double-check the sound, so you get perfect sound, which is good. Then they double-check the lighting, so you get perfect lighting. But they also double-check the director's idea—which is not so good. In the case of the physically perfect—the perfectly intelligible—the public has nothing to add and there is no collaboration. A silent film was easier to make than a talkie because there was something missing. In the talkies, we have to reproduce this missing something in another way. We have to ask the actors not to be like an open book. To keep some inner feeling, some secret."[14]

---

[13]Penelope Gilliatt, film critic for, among others, *The New Yorker*.

[14]Orson Welles (18 February 1979). "Jean Renoir: 'The Greatest of All Directors,'" in the *Los Angeles Times*, pp. G1 and G6.

This page intentionally left blank

Editing

INTRODUCTION

In this chapter, after a general introduction to digital work-station editing and the overall scheme of sound editing and mixing, the specifics of editing for feature films are discussed. This is described before other types because feature films generally have the most elaborate soundtracks, and other forms use many of the same techniques at various levels of intensity. Next, a television sitcom and a documentary/reality production are described. Note that the methods used for each of these are meant to be indicative of certain working styles, which in fact are used across multiple genres. Information given in the documentary section on file management is just as valuable to feature film production, for instance, so read all of the chapter to get an overall feel for editing. Also, although people tend to specialize in particular sound jobs, dialog editing for instance, they may do this job one week on a feature film and the next on a sitcom. What is transferable is the particular skill set of dialog editing, despite what style of production it may be performed on. For those starting out, it is most useful to understand file management, operating systems, networking, and so forth, as the first job one can get is usually bit slinging, defined later.

Today digital audio workstations dominate sound editing, because their productivity gain over cutting on mag film is overwhelming, and their multigeneration potential for high sound quality is not possible using analog methods. They use a nonlinear paradigm explained on page 41 (of Chapter 3) wherein what you are looking at on an edit page is just a series of pointers on how to reconstruct the audio in real time from a random-access hard disk. The silent areas between the sound regions to space them out and maintain sync are just moments when the computer outputs nothing for that track.

Besides being more productive, one person at a digital audio workstation can both edit sound and process it with numerous plug-ins, tasks that traditionally had been performed by specialized transfer room processes or by mixers on dubbing stages. This changes the whole process of postproduction, with implications for personnel, equipment, and buildings. The most important of these processes are covered in Chapter 12 on mixing, because this is where such devices have traditionally been employed. A guide to available plug-ins is available at http://booksite.focalpress.com/Holman/SoundFilmTV/.

To understand sound editing in context, it is necessary first to understand the various stages used in making a soundtrack. This is because the sound editing process is sandwiched in between so-called picture editing[1] and sound mixing, with free flow of material back and forth among picture editing, sound editing, and mixing as revisions are made.

OVERALL SCHEME

In overall form, the process of making a soundtrack follows an hourglass shape. At the top are source sounds and their media. These are edited into what have traditionally been called *cut units* or *elements* or, more likely today, called *tracks* if they are on a digital audio workstation. Sound editing principally works at the top layer of the hourglass model, choosing sounds and placing them into logical order within tracks. See Fig. 11.1.

The following issues affect where a given sound will be placed:

- Like sounds are first grouped by discipline, for example, whether they belong to dialog, music, or sound effects. Although for most sounds this is a simple decision, there are subtleties that arise even within such an apparently straightforward set of choices, as described later.
- Within each category sounds are further broken down to a finer level of detail. For example, sound effects are broken into cut (foreground) sound effects, Foley effects, background or ambience effects, etc.
- Sounds are next assigned to individual tracks to be edited. In this process, the main consideration is the layout of tracks for the mix. For example, all of the sound units to make a given car effect, such as car steady, car braking, car screech to a halt, and car door open and close, will be cut into adjacent or the same rather than widely separated tracks. The decision on whether to intercut them into a single track depends

---

[1] So-called because cutting picture virtually always involves cutting sound as well, and the two interact.

on the experience of the editor taking feedback from the mix as to whether elements need to be separated for different mix treatment or whether the simplicity of combining them on one track is preferable.

- Within each cut soundtrack, there is a certain complexity that is optimum. Cutting too many sounds into one track makes it too difficult for the mixer, who must be constantly adjusting for the changes from cut to cut. On the other hand, spreading out the sounds across a great many tracks, with only a few effects per track, will strain the facilities, for example, taking up too many console input channels. This is the principal reason some Hollywood consoles now sport up to 1000 input channels! It is a big arms race to keep up with editorial complexity.
- The sound editors cut the tracks with a view to making the mixing easiest on the rerecording mixer and on the facilities, prepare cue sheets, and deliver the tracks on the medium of choice and cue sheets for the first stage of mixing. The units are then mixed together into *premixes*, also called *predubs*. In the premix process, typically one premix is made at a time. Thus, for example, all of the various Foley elements of a reel would be mixed together in this process. These might include footsteps, clothing rustle, creaking doors, and the like, but all would be consolidated on the one Foley premix.

The various premixes are then mixed together to produce the *final mix*. The final mix represents the waist of the hourglass, because this stage has the minimum number of tracks needed to represent fully the overall soundtrack. The final mix is divided into *mix stems*, parts such as dialog, music, and sound effects, and each of these is likely to be multichannel, representing directional information such as left, center, right, left surround, and right surround. The final mix process has combined as many tracks as possible to get down to the minimum number needed to:

- Contain all the required sound;
- Minimize the number of tracks for simplicity, consistent with the following;
- Keep separate those parts that may be needed separately later, such as keeping separate English-language dialog so that foreign-language dubbing is simplified.

The advantage of keeping the various mix-stem component parts on tracks separate from one another is the convenience with which a variety of different types of output mixes, called *print masters*, can be prepared for different purposes. For example, it is relatively straightforward to substitute a foreign-language dialog track for the primary-language dialog if the dialog stem has been kept separate. It is also simple to produce an *M&E mix*, which

contains only the music and sound effects, for dubbing to foreign languages in countries where such dubbing is done locally. Another purpose is to fit the dynamic range of the mix stems into various media, emphasizing dialog for an airline mix, for instance.

Remember that once sounds have been mixed together, they are "taken apart" only by human perception, so there is an advantage to keeping tracks separate in the mixing process to provide the maximum flexibility at the final mix. In theatrical film mixing, the director may or may not sit in on the premixes, and if not, then the sound professionals working on the show must make decisions about the balance of sounds that is appropriate. If the director subsequently disagrees with some choices, then it is best if little combining has gone on so that new sounds can be substituted, and new premixes made, on short notice. On the other hand, if no combining of sounds ever goes on, the final mix becomes a nightmare, because all decisions have been postponed until that time. So an orderly progression of mixing together like sounds is the best method.

One principle that grows out of this approach is *checkerboarding*, which means that one premix might occupy the red squares of a checkerboard and another occupies the black squares. Each is active only a part of the time and is otherwise silent when the alternate premix is playing. However, note that the lines on this checkerboard are not hard and fast: Premixes may also overlap one another, playing simultaneously. The advantage of checkerboarding is the momentary silence between the alternating sounds, permitting the editing of premixes. This is an advantage for last-minute alterations because if the director does not like one particular effect at the time of the final mix, it is possible to remove just one portion of one premix and substitute a different sound. Then a new premix can be made and resumed into its final mix stem. That is, the whole process can be unwrapped back to original elements if need be, so long as there are some edit points available, especially silent portions of a premix.

From the mix stems a series of print masters is made, representing the sound to be copied in 1:1 correspondence to each medium of release. A moderately big show might have many print masters:

- 5.1-channel English-language digital stereo master;
- 2-channel English-language LT RT matrixed stereo master;
- 5.1-channel M&E;
- 2-channel M&E;
- 5.1-channel French;
- 2-channel French;
- 5.1-channel German;
- 2-channel German;
- 5.1-channel Spanish;
- 2-channel Spanish.
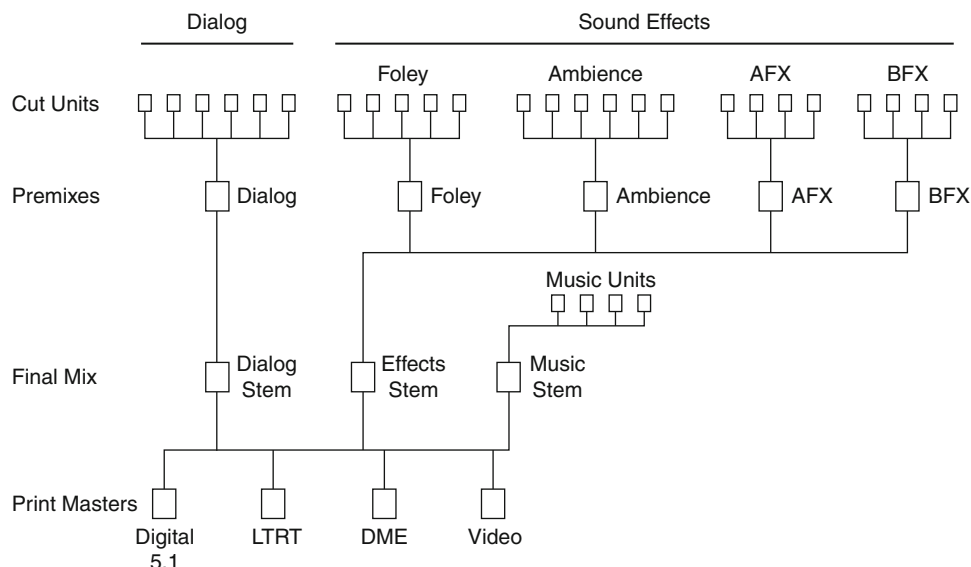
FIGURE 11.1 A block diagram of the overall mixing process for a feature film. Each row represents a generation, so units are mixed together into premixes, premixes together to form a final mix, and then the final mix stems are mixed together to produce the print masters that are the output of the sound postproduction process. Note that each of the premixes, the final mix, and the print masters is likely to be multichannel.

This list is long but by no means exhaustive.[2] These types are more fully described in Chapter 13. Suffice it to say that the various print masters fill out the base of the hourglass because of their number and variety.

So there are seven generations involved in typical film work:

- Original source recording or library effect;
- Cut units, also called elements or tracks;
- Premixes, also called predubs;
- Final mix, composed of mix stems such as dialog, music, and effects;
- Print masters;
- Optical soundtrack negative;
- Prints.

Television production, having tighter schedules than film production, and also virtually always working with a "locked"[3] picture, may proceed somewhat more simply:

- Original source recording or library effect. Copy camera original into the postproduction picture editorial system in a process called *laydown*. After editing export the editor's work from picture editing to sound editing.
- Digital audio workstation edit.
- Mix stems.
- Make master mix.
- Layback to video master to make edit master.

COMPUTER-BASED DIGITAL AUDIO EDITING

Digital Editing Mechanics

Figure 11.2 shows an editing screen from a digital audio workstation (DAW).

Traditionally, tracks are represented horizontally, flowing from left to right. They allow "cut-and-paste" sound editing to be done rapidly, following a visual track metaphor. Interestingly, this is in contrast to cue sheets, shown later, for which the work flow is from top to bottom.

Edit screens show the progress of playback along a time line, the individual tracks with their associated sound regions displayed, and potentially many other items such as input and output assignments of signals sent to and from the tracks, "mode" of the tracks such as Play or Record, Solo functions described in Chapter 12, and where sounds are located on the media so they can be quickly accessed and dragged into a track in the desired sync

---

[2]Italian is missing from the list because Italian dubbing must go on in Italy by law, so the M&E masters, along with a mono English track for reference, are sent there. Asian, Dutch, and Scandinavian languages use English-language soundtracks with subtitles traditionally, although some films might be dubbed, probably locally. The largest international dubbing occurs for Disney-animated features, because a major market for them is kids who cannot yet read, so dubbing is done into more than 30 languages and release is simultaneous throughout the world to lower piracy, a truly remarkable feat.

[3]Although sound work starts in earnest after picture lock on a feature film, the picture is still subject to change, even after picture lock. There is no such luxury of time in television postproduction, for which the time frame is measured in working hours, not weeks.
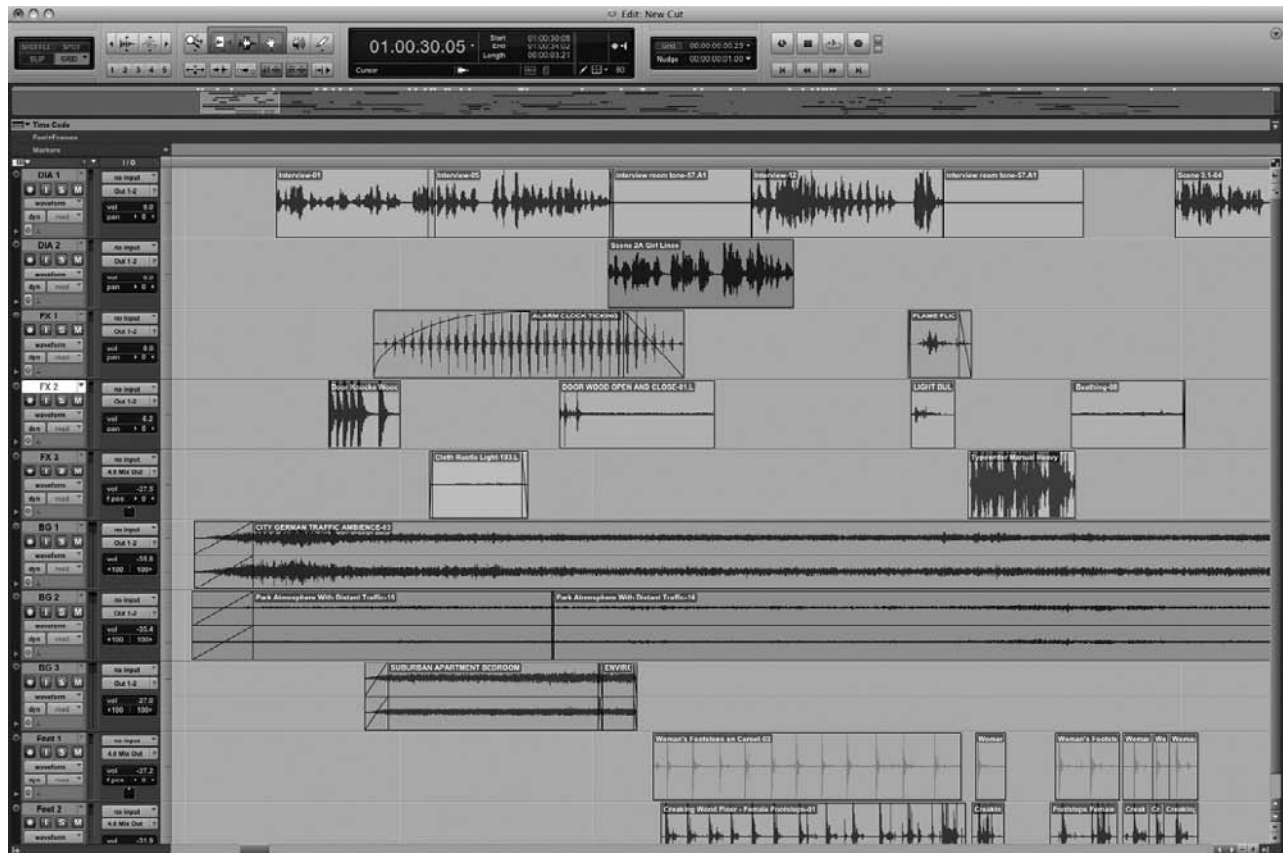
FIGURE 11.2   Editing screen of a digital audio workstation.

along the time line. Other screens are more mixing ori-ented, described in the next chapter, and they may include "overlays" for transport control, time/footage readout, etc. Individual details vary and are a point of competition among workstation manufacturers.

Perhaps one of the most remarkable things is the wave-form display of the audio within the active regions. Not every workstation has these, but those that do may include as much as a sample-by-sample display of the amplitude through time, the actual digitized waveform. This is remarkable because it turns a task that for cutting mag film was done completely audibly into a visual task.

> The amazing thing about this is that it is a reversion to pre-1952 film sound cutting, when work was done on optical soundtracks that were visible. Sound editors at the time thought mag would never catch on because it wasn't visible, therefore you couldn't edit it!

Types of Cuts

Sound editors use a mixture of types of splices or cuts. Butt splices are instantaneous ones. On film and tape, these are straight across the mag or tape, and on worksta-tions they are simple vertical cuts. Although they may work in several situations, the use of butt splices is

actually a fairly low percentage of audio edits. The prob-lem is that the audio waveform is being chopped on or off instantaneously, usually at nonzero amplitude. Cutting instantly to or from zero to any other nonmatching level will probably result in an audible click.

Diagonal splices (known as fades in the DAW environ-ment) are therefore favored for audio cuts, with a typical cross-fade time equal to that of one perforation of film passing by, about 10 msec. Longer ones are easy to per-form on digital audio workstations and are routinely used when a smooth transition is needed. Where butt splices may be used are: (1) in silence, where it doesn't matter; therefore Foley editors, cutting between effects, for instance, make use of butt splices; and (2) when backward masking is needed, like cutting in the middle of a word on a "hard" phoneme, to prevent hearing the edit.

Fade Files

We have already described the necessity for making diago-nal versus butt splices in audio. A diagonal splice is a fade-in, a fade-out, or a cross-fade. Fading involves level change, which in digital audio means doing multiplications, in fact, thousands of them per second. To have the equiva-lent of a diagonal splicing capability on each edit, the

workstation signal processing requirements are formidable. Whereas some high-end workstations do such cross-fades in real time, on all the channels, others use a different method, the idea of *fade files,* to accomplish the same thing. Fade files are derived from those regions of audio files where the sound is being changed in level. The computation of the fade can be done slower than real time and the corresponding snippet of sound stored away as a fade file. Then, in real time, all the workstation has to do is play the fade file at the correct time, splicing it instantaneously to the main file once outside the affected audio region. The instantaneous splice is not audible because the waveforms are perfectly matched at the edit point. Some systems divide their media recordings into *audio files* and fade files, with the fade files having the potential to be regenerated from the audio files if they are lost.

Cue-Sheet Conventions

Sound editors prepare two things: sound elements and cue sheets. Cue sheets are the principal means by which sound editors communicate the layout of their work to the human mixer, essential to mixing. They are necessary for premixing, and subsequently they are used to find individual sounds and what premix they are assigned to, so that corrections can be made. Cue sheets are organized with columns representing what is going on in each element. Within an element, they indicate whether a sound is to cut abruptly or fade in and out. They also give the critical footage or time code numbers corresponding to the start or, if needed, an intermediate point in a sound effect. See Fig. 11.3 for an example. One curious fact is that cue sheets are organized with tracks represented vertically, with time flowing down the sheet. This is the opposite of editing systems, in which time flows horizontally. Each of these is due to the history of editing and mixing: editors worked on benches where the tracks flowed horizontally, but consoles are organized with channel strips aligned vertically, one per track.

FEATURE FILM PRODUCTION

Syncing Dailies

Production double-system sound may be acquired on a variety of analog or digital recorders, either of which may be with or without time code. After resolving of non-time-coded tapes and copying to the editorial medium, the picture and sound need to be synchronized. Often this will be done as a part of telecine operations, as described in Chapter 8.[4]

Whatever the source, either time code slates or in-camera time code, the time code is used either manually or

---

[4]Traditional syncing of dailies on film is described at http://booksite .focalpress.com/Holman/SoundFilmTV/.

with an autoconform program to sync the production sound on the telecine, as described in Chapter 8. Alternatively the syncing may be done after telecine transfer by a picture assistant in postproduction. Video is most often used for editing today, whether the source is film or video. On higher budget shows, camera capture may be by means of a digital cinema camera, and with digital cinema dailies projection a complete chain may be of high picture quality. Dailies are synchronized often with an autoconform function, called AutoSync by Avid, and by other trade names (Table 11.1).

A single-track linear representation of the production sound in sync with the picture is called an *A track*, and it may be left intact throughout picture and sound postproduction to form a *guide track*. A guide track is always left in sync with the picture to have a reference soundtrack known to be in sync with the picture. On a workstation, the *regions* or *clips* corresponding to the shots may be copied and pulled into separate tracks so that overlap and cross-fade editing from shot to shot can be performed for smoothing dialog edits. The guide track provides a continuously available check that this process of moving copied regions to different tracks has occurred in sync.

Traditionally film picture editors would cut just the A track in sync with the picture while editing. However, more elaborate editing in the picture suite soon occurred, as the importance of sound to pacing the action, driving the story, and so forth, was noticed, and B and even C tracks came into being. Today, with picture cutting workstations like Avid Media Composer, more elaborate sound editing is done in picture editing than ever before.

However, the primary focus of picture editors has to be getting the story into coherent shape, and sound inevitably takes a back seat, especially sound quality. For this reason, among others, retransfers of the original production materials may be needed once the picture is delivered to the sound editors, although when working with digital workstations and with files correctly transferred into the picture editor and exported to sound editing, as described in Chapter 9, one should be able to avoid this step for most production sound. Most often, sound is exported from the picture-editing system to the sound-editing system, transparently. If retransfers are needed because of lost material, for instance, then the source camera time code and software are relied upon to reconstruct the soundtracks.

Dialog Editing Specialization

Sources for Dialog

Dialog editors depend on several sources for their tracks:

- Production soundtracks (probably retransferred if the production A track has been modified).
- Outtakes of the production sound from alternate takes, even outtakes that were never printed, to cover miffed words.

FIGURE 11.3 A cue sheet showing conventions such as length of an effect and continuation across pages.

Title: Ghost of Christmas Past — Sheet 2 of 7 across — FX & Music

Supervisor: James Brown Date: 10/10/2011 Format: 35 mm ft/fr   Page **1**

Editor: Tim Brown

Effects 7	Effects 8	BGs 1	BGs 2	Music 1	Music 2
		12:00 base wind			
				55:09 plaintive solo flute (mono)	
			72:00 ghost wind		
	150:22 weird wind 3		160:00		
				250:23	
272:08 door squeak	280:05				

Page ends 300:00

- Wild lines. These are dialog lines that the actor records under the director's guidance, often at the end of a production day. Recording wild lines on the set after hours, or even at the end of particular setups, gives a quieter place having the same acoustics as the rest of the production sound and may avoid the added expense of an automated dialog replacement (ADR) session. For instance, noisy lighting equipment, generators, wind machines, and the like, can all be shut off. The situation is optimized for sound recording because no cameras are operating, so a boom mic can be employed in all cases. The job of the actor and director is to repeat the performance that has occurred during shooting, not to "improve" on it, at least insofar as sync goes, so that there is a chance that the dialog editor can get it into sync by manipulation of the track, cutting in between words and even syllables to synchronize the wild lines with the original production sound for the takes that are used in the picture.

- ADR recordings. The ADR editing supervisor or an assistant does the preparation for the ADR session by producing a log of all the required lines by footage or

TABLE 11.1 Methods of Syncing Dailies

Source material	Basic method
Analog double-system production sound with neopilottone or FM sync resolved and transferred to mag film	Assistant picture editor matches slate clapper close on pix and snd on edit bench with squawk box and synchronizer
Analog double-system production sound with neopilottone or FM sync resolved and transferred to digital audio workstation (DAW)	Assistant picture editor matches slate clapper close on pix and snd on DAW
Digital double-system production sound with time code, but no camera recording of code nor time code slate	Assistant picture editor matches slate clapper close on pix and snd on DAW
Analog or digital double-system production sound with time code, with time code slate[a]	Telecine operator reads time code visually at clapper close, finds corresponding sound, and slaves playback audio machine to make in-sync transfer, shot by shot
Analog or digital double-system production sound with time code, with in-camera optical time code	Autoconform by chase synchronizer on telecine
Digital double-system production sound with time code	Autoconform with AutoSync function or similar
Analog or digital single-system sound	Self synchronized

[a]This method is probably the most common today for theatrical films, television movies, and the like.

time code for a given actor. This allows quick work on the ADR stage, where the equipment will fast forward to a line, then play the line over and over and record it when the actor is ready, as many takes as are needed, and then fast forward to the next line by that actor. Actors are typically brought in one at a time to loop their lines, although there are certainly exceptions to this rule. In particular, where overlapping dialog may occur, it is difficult for the actor to perform alone.

● There are three major considerations in getting good ADR: sync and performance—mostly in the hands of the actors and director—and recording, mostly in the hands of the ADR mixer. Sync can be accomplished by good acting, by skillful editing, and with help from software that slides ADR performances into sync with production sound (Vocalign). Performance is another matter; some actors can perform ADR well, but others are not so good at it. It must be said that if you have ever tried it, you would find it to be a highly artificial thing to do and to require a special skill. When a particular actor is known from the outset not to be a good "looper," the production sound recordist can be told to concentrate on him or her. The ADR mixer attempts to mimic the perspective of the original, in particular with the correct amount and type of reverberation, if there is to be any attempt to use any of the production sound from a given scene.

Track Breakdown

Dialog editors make use of the principal on-screen production dialog, off-screen lines, wild lines, outtakes, and ADR recordings in pursuit of the best combination of performance and technical quality. Whether dialog and wild lines can be intercut depends on the background noise on the set present during shooting, the microphone perspective, and the actor's performance, but it is unlikely that they can be intercut all of the time, and the source tracks will have to be split into more edited tracks when no match can be ensured. This is so that the rerecording mixer can easily make separate adjustments of level, timbre, reverb, etc., in favor of better matching. The objective is to make the most mixable soundtrack, so if all sound can be cut perfectly so that the foreground says the right thing and the background matches perfectly at the edits, there would be no reason to split the sound segments into separate tracks at edits.

Some mixers, however, prefer different characters to be split out to separate tracks. A potential problem with this is that then, when the tracks are filled with presence and two actors are playing, we have two presence tracks, or twice the noise, than there would be if we stuck to just one track. Such mixers are asking for the ability to process each actor's voice separately, and this could be important if one were to be more on mic and the other off.

Dialog editors say that they listen "through" the dialog to the backgrounds to choose edit points. Of course they are not ignoring the dialog, because saying the right thing comes first, but they choose their edit points more on how well the backgrounds match at the edit point than perhaps any other single factor.

Among the methods to match cuts are:

● To cross-fade at the edit (virtually all edits are cross-faded, but the timing and length of the cross-fade can be used to smooth the transition);

- To delay the sound cut compared to the picture cut until just at the beginning (head) of the next word to allow backward masking to cover up the discontinuity in the background at the edit;
- To find a smooth, albeit noisy, background on the noisier side of the edit and loop and extend it across the cut, splitting the track, and allowing the greater background noise of the first track to dominate that of the second in the interest of smoothness.

On large Hollywood features, another set of sound editors works on ADR. ADR must always be cut on tracks separate from production dialog, because it will virtually never match. This is so the tracks may be separately available for different kinds of signal processing, such as equalization and reverberation, during mixing or, possibly, even on the editing workstation. By separating them by source, one maximizes the possibility that the rerecording mixer can produce a smooth-sounding track, because, for example, it will probably be necessary to add reverberation to ADR recordings to get them to sound like production sound recordings, so the processing on the two has to be different.

With the source of the sound being the first level of splitting off dialog tracks, things get more complicated on the next level. If there is a two-shot, and both characters are recorded well on mic, then there may be no need to split the track. If, on the other hand, one character is off mic in one shot, then there is the possibility to improve the sound by using sound from an alternate take, placed on an alternate track. By the way, this off-mic sound example shows up another difficulty with dialog overlaps: if an off-mic line overlaps an on-mic one, to replace the off-mic one with ADR would result in "doubling" of the part of the line during the overlap, and this is usually painfully audible. That is why it is important to exercise discipline on the set, particularly over off-camera lines, because they can be so easily constructed by dialog editors and so difficult to take apart should the need arise.

In the end, the objective of making the most mixable soundtrack is the ruling principle. Choosing how to split tracks is mostly achieved by careful listening, and through experience.

Presence or Fill

To smooth dialog tracks, it is essential to have a recording to fill in between the gaps when things must be cut out. The difference between having this presence or *fill* and not having it is like the difference between seeing an unpopulated set and going to black. Black equals silence in this example, and cutting to it would be just as obvious as missing the picture—so with sound. Material for fill is often obtained by copying the background sound between the actor's lines and looping it to lengthen it as required. The advantage that this has is that it is sure to match,

at least at the portion from which it was lifted, to the production sound.

Another source for presence is the space between the various things going on at the beginning and end of a take, as described in Chapter 7. For example, if the director waits a moment between the exit of the slate person and calling "Action," there will be some sound recorded that will exactly match the start of the take and thus can be used as presence. Even as little as 2 sec is very useful because that sound can be looped and extended as needed.

Another recording made for this purpose by the production sound mixer is *room tone*, or in England, *atmos* (short for atmosphere).

Obviously, the ability to intercut presence into a scene depends on the exact nature of the background noise present on the set or location. Intermittent background noise will make it hard to intercut presence. For a relatively noisy New York street scene there may be difficulty in cutting between presence and production sound, because there can be quite audible components of the two that do not match. We could not cut, for example, in the middle of a taxi passing by the microphone to presence recorded without the taxi without drawing attention to the lack of continuity of this stream.

Dialog editors thus usually "clean" individual soundtracks of such intermittent background noise as well as they can by cutting, and sometimes provide another track of matched, continuous fill called a fill track or a fill loop that the mixers can use so that the "bumps" that remain at the edits in the dialog track can be masked. They may also cut hard effects from the dialog track, such as a door closing, into a *PFX track*, for production effects, that may be used as one choice along with others such as Foley at the mix.

A difficult dialog editing situation occurred on *Mosquito Coast*. Harrison Ford has lines while lying down in a boat surrounded by lapping water. The dialog editor Laurel Ladevich had to manually clean out the "laps" between each word of dialog, hoping that laps under dialog would be masked by it, and simultaneously match a second track with laps matching those seen so that the rerecording mixers could have the capability of adjusting the relative levels of dialog and water sounds, all in sync of course.

For an illustration of how dialog editors cut when there is significant background noise, play Tracks 36–38 of the DVD.

Presence should be cut into dialog tracks where it matches well. To produce a separate continuous presence track may seem to make sense at first glance, but then you realize that the presence would be doubled up underneath the dialog lines, and would be singular in between the lines, and this could easily be audible as bumps. The rule is, if presence intercuts well, then cut it into the dialog track; if it doesn't match too well, split it out into a separate track.

Handles, sound recorded before and after the actual dialog lines, are important in production sound. For dialog editors, handles are very important as they provide a means for the mixer to make smooth transitions and provide an important source for fill. Picture editors do not typically deal in handles, but they should export them so that dialog editors have adequate material with which to work.

Sound-Effects Editing Specialization

Because of their complexity, sound-effects editing is often broken down into subspecialties, along the lines of the premixes: cut effects, ambience (backgrounds), and Foley. Often on a large film production, these will be the subdepartments of sound editing.

Hard or "Cut" Effects

The simplest definition of hard effects is "see a car: hear a car," a trademark of the craftsmanship that goes into filmmaking. It is the expectation of the audience that everything we see on the screen that makes a noise should be heard and thus covered by a hard sound effect, unless a sequence is a montage. *Hard effect* in this context means that the sound was obtained from a source other than production sound, Foley, or ambience, and the sound was cut in sync by a sound-effects editor to match the picture.

The relationship between the picture and a hard effect can be one-to-one correspondence, such as the car example, or it may be more tenuous, such as the high-pitched processed insect sound effect in the jungle scene in *Apocalypse Now* that acts as a builder of tension.

The sources for hard effects, in order of their likelihood of use, are as follows:

- Commercial or private sound-effects libraries. Virtually no filmmaking activity has enough backing to record all new sound effects for each film, so libraries are relied on to provide many of the basic effects. Commercial libraries are made easily accessible on series of compact discs. Still, to make use of CD recordings, they must first be copied to an editorial medium, such as film or a digital audio workstation, which is likely to involve a sample rate conversion. This is because the sample rate for CDs is 44.1 kHz, and for all sound media accompanying a picture it is 48 kHz. Thus all sounds copied directly without conversion would be raised in pitch by a ratio of 48/44.1, about 8.8 percent, a significant amount. Distribution of some sound-effects libraries is also available today of 24-bit, 48-kHz effects, some of them even up to 5.1 channels.
- Custom recordings. When working on *Top Gun* the sound-effects editors probably ran out of jet recordings from libraries rather quickly, using up the whole recorded repertoire from sound-effects libraries in a few minutes of screen time. Although it is possible to reuse recordings, if done too much the pattern of usage becomes audibly obvious. The overuse of particular effects seemed clear to me in the 1983 helicopter cop movie *Blue Thunder*. Many new recordings are needed for such a specialized soundtrack.
- There are at least two ways to find a source for given sound effects: the literal source and those that bear a more tangential relationship to what is being portrayed. Table 11.2 gives some of each of these types.
- Built-up elements. Many sound effects are built up from multiple recordings layered together to achieve the needed level of complexity. Today this layering can be accomplished quickly using a digital audio

TABLE 11.2 Some Sources of Sound for Movie and Television Sound Effects

Sound effect	Source
Emperor's blue lightning, *Return of the Jedi*	1930s tesla-coil-based "Jacob's ladder" machine used in James Whale's *Frankenstein*
Pile of rats in cave under the church, *Indiana Jones and the Last Crusade*	Petaluma, California, free-range chicken ranch, standing among the chickens, pitch shifted up to multiple frequencies, plus horses moving gently in corral, pitch shifted up
Biplanes, *Indiana Jones and the Last Crusade*	Kenosha, Wisconsin, annual air show of antique planes
Distant thunderstorms on Dagobah, *The Empire Strikes Back*	Blind Midwestern farmer who records all passing storms, layering together multiple storms
Flybys of alien craft, *Independence Day*	Screaming baboons
Footsteps on grass	Foley stage walking on unspooled audio tape
Footsteps on ice and snow	Foley stage walking on rock salt
Device for killing aliens, *X-Files*	Producer close-mic'd saying "pfffft"

workstation, at reasonable cost and investment in time. The sound of a face punch in an Indiana Jones movie, for example, has the sort of "Kabamm Pow!" character that cartoons have. This is accomplished by layering together different recordings, including throwing a leather jacket onto the hood of an old fire engine and dropping overly ripe fruit on concrete—together they sound like neither, but like something completely new.

Hard effects are organized according to their type into what premix they will be in, to make mixing more logical. Sound editors are looking ahead to the needs of mixing, whether that mixing occurs on their workstation or in a dubbing stage. Like sounds are usually grouped into the same premix. They could be grouped by event, for instance, one car door close, start, and drive away may be assigned to one premix, if that makes sense as a unit. On the other hand, a single event might not be so grouped, but rather the grouping could be organized by the kind of sound, such as Wind, Metal FX, Water-A FX, and Water-B FX for a given film, say a sea-faring one, whereas the breakdown for another type of film could be quite different. In this case, Water-A might be more or less continuous water sounds, but not necessarily having a 1:1 relationship with picture, and Water-B could be specific water events that "sell" the veracity of the track by having what sound editors call *sync hits*, those effects in hard sync that convince you everything you hear is real.

An example is that of the Imperial Snow Walkers in *The Empire Strikes Back*. Seen in a long shot, three of them menace the rebel forces. With 4 feet each, that's 12 big footfalls to cover—a lot of cutting. This can be simplified by cutting two tracks, one an effects *bed*, with the sound of multiple feet falling, in this case a slowed-down punch press from a machine shop, at about the correct rate. The second track, run in parallel, "sells" the shot by cutting single footfalls in hard sync with the most prominent visible feet falling. The two work together, and all the individual feet do not have to be cut.

Another method used to sell the verisimilitude of a shot is to change the perspective by changing the effect. In the case of a closer shot of the Imperial Snow Walkers, the sound of a bicycle chain dropping on concrete was added in sync with the punch press sound, because, as we know from Chapter 1, closer sound is louder, and brighter, than more distant sound.

The editing is performed so that not only do the source tracks checkerboard in time, but the premixes are also organized to checkerboard the sounds in time, making some of the combined effects at the premix stage first appear in one premix and then in another.

One consideration in assigning sounds to premixes is their place in the frequency and dynamic ranges. This idea is based on frequency masking. It would be typical to have the ambience premix contain low-frequency sounds, whereas Foley provides higher-frequency ones, thus helping to discriminate the two despite their similar levels. Quiet effects would not typically be intercut with loud ones, as that would make mixing too difficult.

Combining sounds together to create denser sound effects is a principle often used. Special manipulations may be done either on individual elements of a complete effect or on the whole, mixed effect. These include speed change, used to produce a corresponding pitch change, making a sound "bigger" by slowing it down; lengthening by mechanical or electronic looping; and pitch shifting to make a sound seem to move past a point of observation by faking Doppler shift. Other manipulations are available today in a vast array of plug-ins for digital audio workstation software.

Foley Sound-Effects Editing

Foley sound effects are those made in a recording studio called a *Foley stage* while watching the picture and performing the action more or less synchronous with the picture. More than any other single part of a sound mix, it is the Foley sound effects that often make the sound seem real, because Foley recording exaggerates real-life sounds to make them audible. What effects to record in Foley and what to leave to hard effects is a decision made between the supervising sound editor or sound designer and the Foley artist, called a Foley walker in England, perhaps because the most salient Foley effect is footsteps. Many small sounds typically dominate Foley: footsteps, clothing rustle, pouring a glass of water, moving a chair, and even body hits could all be Foley effects.

Early in the history of film sound, Foley recording was invented by a man named Jack Foley, working at Universal Studios. The coming of sound had brought international distribution headaches to an industry that had enjoyed simple means for foreign distribution. All that had to be done to prepare a foreign-language version of a silent film was to cut in new title cards translated into the target language. The coming of dialog spoiled that universality. At first, to solve this problem some films were shot on a 24-hour-day schedule, with three casts working in three shifts recording three languages. The casting for the foreign-language versions had two concerns, that the actors could speak the language and that they could fit into the costume made for the English-language star! It wasn't long before foreign-language dubbing was invented, lip-syncing a foreign language to the English original.

The difficulty with these foreign-language dubs was that they lacked all the low-level sounds of the actors moving around the set, sitting down, pouring a glass of water, etc. Although these sound effects could be provided by a sound editor cutting in effects from a library, this was an exceedingly tedious process. Thus, the stage was set for the invention of Foley recording. The idea was that many

sounds could be recorded "to fit" the time that they appear on the screen by simply performing the action in sync with the picture and recording it.

Today, Foley recording is likely to involve a workstation so that various record passes can be used to add layers of different effects, building up to a complete whole. There are a number of people involved in producing a Foley track:

- The Foley artist's job is typically done by one or two persons, who spot the picture with the supervising Foley editor (that is, assign what sounds are needed from Foley as opposed to hard sound effects), gather props, and perform for the recording.
- The Foley editor prepares the cue sheets, attends the recording session, works with the Foley recordist in track layout and aesthetic recording issues, and prepares the Foley units for rerecording.
- The Foley recordist chooses the microphone technique and operates the equipment, layering sound to separate tracks and monitoring the ongoing work as needed to be certain everything needed is covered.

After the recording has been made and monitored for completeness, the recording is handed off to a Foley editor, who adjusts each of the tracks on a digital audio workstation for sync. This *fine cutting* consists of moving the recorded sound with respect to the picture by usually just a frame or two to put it into *hard sync*, which is certainly one of the most important things to do to achieve verisimilitude.

Ambience Sound-Effects Editing

Ambience, also known as "backgrounds" (BGs), is sound that produces a space for the film to exist in. Although superficially similar to production fill, there are some distinguishing factors. Fill or presence is sound that can be intercut transparently with production sound. It is mono, restricted to the center channel with the production dialog. Ambience is most often multichannel, as it involves the space sensation. Presence is a record of the background sounds present on the location or set. Ambience is a separately built soundtrack, selected by sound-effects editors. Ambience is artificial presence in the sense that it provides a "space" that wasn't there during shooting.

Ambience most typically consists of more or less continuous sound, often with a low-frequency emphasis we associate with background noise of spaces. Thus, in reel 1 of *Raiders of the Lost Ark*, the scene in the cave has low-frequency rumble acting as a more or less distant threat once Indy triggers the ancient mechanism by appropriating the idol.

Ambience plays a significant role in scene continuity. If ambience stays constant across a picture cut, it says subliminally to the audience that although we may have changed our point of view, we are still in the same space.

Conversely, if there is an ambience change at a picture change, it says the opposite—we are in a new scene. Ambience may even be overlapped across certain scene transitions, either to create an effect of the former scene lingering into a new one or to anticipate a cut to a new scene.

In the cave scene in *Raiders* there is an ambience change accompanying a picture cut that is significant in explaining the story at that point. Indy leaps the chasm to get away, and as he scrambles up the other side of the hole, the picture cuts to a view from the other side of a portal that is starting to close. Although scientifically speaking the level of the ambience rumble would be just about the same through an open door as in the first space, in fact the level changes abruptly downward with the picture cut. The subliminal indication to the audience is that by making it through this opening before the stone comes down into place, Indy will have reached a place of safety, away from the threatening ambience rumble.

Ambience is one of the most interesting soundtracks from a spatial point of view, using stereophony to achieve its goals, whereas presence (fill) is never stereophonic because it has to intercut with mono production sound. The stereophonic nature of ambience tracks is also one thing that makes them different from presence/fill in production tracks and a reason for including them in a mix even if the (mono) production backgrounds sound great.

One crucial spatial question about ambience is whether it should include sound in the surround channels. The difference between ambience on the screen and ambience that includes surround sound is related to the degree of involvement of the audience. A good example is from *Apocalypse Now*. In the jungle scene, we first see the boat on the water with the jungle in front of us. As two characters get off the boat to gather mangoes, we follow them into the jungle. The sound of the jungle remains present on the screen channels, but also creeps up in the surround channels, enveloping the listener. This use of surround sound creates greater involvement on the part of the listener by breaking the bounds of the rigid screen edges and brings the audience into the action. Then when the tiger jumps out, the action is much more frightening because we accompanied the characters on their search, rather than observing them from afar.

Another way that ambience may be used is to smooth over small changes in presence that may otherwise draw attention to the artificiality of building the scene from different shots. Suppose the presence changes because the microphone is forced to shoot a different direction from shot to shot and a directional noise source is involved. As long as the level of the presence can be kept quite low and the discontinuities in it masked by a continuous ambience track, all is well. So the use of ambience varies from providing a sonic space for the scene to exist in to the practical covering up of presence discontinuities, auditory "perfume."

Music-Editing Specialization

Music scored for film is composed to fit the time given it by the film. Although this might seem to be a limit to creativity for the musician, perhaps the following story is illustrative.

In producing a ballet from a work by Stravinsky, the stage director had a problem. There was a certain amount of stagecraft that had to go on in order to change the scene, but there was no music to accompany the change. This was the premiere of the work, and Stravinsky was the conductor. With fear of what might happen when he did so, the stage director approached the famous composer with the problem.

Stravinsky replied, "How much?"

"Eight seconds. That's no problem is it? You can have as much time as you want."

"Of course not. I'm delighted to have a requirement!"

The first step in the process of composing the music is that the composer, the director, and the supervising sound editor or sound designer *spot* the picture. *Spotting* refers to going through the picture and noting where music should be present and what kind it should be. (The term is also applied to any process of matching required sound to picture.) This process has to take place after the picture is locked and before the composer can begin to work in earnest on the film or television show (although he or she might be able to write some musical themes before picture lock, to write the actual music cues the exact length of cuts, available only after picture lock, is needed).

It is at the stage of spotting that the composer finds out the general outline of the sound effects for a given sequence, so that attention can be paid to producing music that not only matches the mood of the picture but also is kept relatively free from frequency-masking effects. Thus, if the sound designer says that a scene will have a loud rumble to indicate a threat to the audience and there isn't any dialog to be heard, the composer may concentrate on higher instruments, say, violins and brass, rather than double-bass and timpani, to accompany the scene, maintaining the most separation possible.

In theory, the practice of music editing should be straightforward: Just cut the music in where it was written. In practice, things are much more difficult. The single largest problem is that it takes weeks to write the music and record it, and this must be started early enough to ensure that it will be finished well in time for mixing. The start date on composing the music is as early as possible, usually right after picture lock. The problem is that picture lock is often not all that firm, and the picture continues to be edited while the music is written. By the time the music is recorded, the version of the picture has often changed, and sync is no longer ensured. This is where the art of music editing comes into play. The music editor cuts the music tracks with great sensitivity to both the integrity of the music and the needs of the picture. There is potentially a required trade-off between these two issues because the music must, for example, pay off at the end of a scene and yet be cut to make musical sense within the scene.

One trick to accomplish music editing is to make cuts that will be masked by other sounds on the soundtrack in case "perfect" cuts cannot be obtained that are inaudible. Although the general outline of the music, its orchestration, key, tempo, etc., probably needs to be the same before and after an edit, a difficult edit is hidden by masking by putting it "behind" a loud sound effect.

Music editors for film are generally trained musicians who read music, help the composer prepare for the scoring session, sit through the mixdown from the multitrack original recording to the music tracks, and then cut the music tracks so that they may be played during the final mix. Table 11.3 gives the track formats for music delivered to dubbing stages.

One technique that should not be overlooked is the possibility of cutting the picture to the music instead of the other way around, which is frequently used in montage scenes. In this case it would be typical to be using either existing pop music (such as in the babysitter scene in *Risky Business*) or "source" music,[5] which is music present in the scene, heard by the characters, rather than "scored" music. In the documentary *The Wonderful Horrible Life of Leni Riefenstahl*, the subject explains with utter delight how she cut a scene in *Triumph of the Will* to the music and how that tactic helped to move the audience (in all the wrong ways, as it turned out!).

In filmmaking that cannot afford custom-written music, or for which there is a need to adhere to existing music as a storytelling method, there is a danger in using existing music sources that should not be overlooked. Each audience member brings a lot of potential baggage to bear, such as associations from circumstances of hearing the song formerly. This makes it difficult for the filmmaker to stay in control of the storytelling process, because there is an outside influence that

---

TABLE 11.3 Track Formats for Scored Music

Name	Tracks
1-track	Used for source music within a scene and for recording to playback
2-track	L, R for television use only
2-track	LT, RT[a] usually for television use
3-track	L, C, R (the most common)
5-track	L, C, R, LS, RS
7-track	L, LC, C, RC, R, LS, RS (rare)

[a]See Chapter 13

---

[5]Also called *diegetic* by our critical studies colleagues.

is out of control. On the other hand, known music can bring with it known emotions, which, being old and familiar, may just fit like an old pair of shoes.

Music also imposes its own order on a scene, whether the picture is cut to the music or not. It is surprising how many times music will be laid underneath a scene of someone walking, for example, and the walk appears to be "on the beat." Somehow our brain searches for order out of chaos and imposes it, finding order where none was intended. Unusually, the director Sergio Leone played pre-recorded score on the sets of his movies *Once Upon a Time in the West* and *Once Upon a Time in America,* and for these the order was intentional.

There are two basic music issues regarding rhythm— beats and downbeats—that impose their own temporal order on a scene. The beat is the "atomic clock" of music, propelling the music through time. The downbeat is one beat of a repeating sequence that has an accent—4/4 time emphasizes the first beat of each measure of four beats over the others. Each of these can affect the match to picture. The tempo of the beat is quite important to motion. For instance, "andante" is said to be a walking tempo, and compositions at andante tempo could best be expected to match walking on screen. Key signature has an impact too. Major keys are said to be bright and happy, minor ones sad or contemplative. But this idea is a vast oversimplification of a very sophisticated field.

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 points 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 lavaliere microphones, so long as that stays constant within its own scene. It is when these methods, along with ADR, become mixed up within one scene that there's more trouble in maintaining continuity. A scene change is like erasing a blackboard—the lecturer gets to start over. Of course, this is not meant to say that a given character should not be recorded 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 a number of 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 a number of different characters' stories (which

come together at the end). They are, with citations from the DVD of the film:

- The straight cut. Although ambience may change, it is at a normal level before and after the cut. The cut illustrating this is from one dialog scene to another. In actuality, it is probably a very quick cross-fade to prevent clicks, perhaps ¼ frame in length. Chapter 3, 15:07.
- The hard or "bang" cut. This breaks the ongoing scene and slaps you in the face; we have made the change. 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 helps indicate 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 that something happens in the audio track before the video) in some circles. Sound is heard before the picture change of the incoming scene. This helps propel the story forward 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 being seen.
- Postlap 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, an actor speaks the tail line in the scene, "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 . . ..." Chapter 3, 20:07.

A table containing the references to the scene edits in *Love Actually* is at http://booksite.focalpress.com/Holman/SoundFilmTV/.

Premix Operations for Sound Editors

The output of the three subspecialties—dialog, music, and sound-effects editing—are the cut units or tracks delivered to postproduction mixing, along with the cue sheets prepared by the editors, organized to explain the placement of sounds in the units to the rerecording mixers. The reason for sound editors to so concern themselves with track layout is to make the mixing process as simple as it can be. If sounds are organized inexplicably across tracks, the soundtrack is much harder to mix.

One likely breakdown of the premixes is:

- Dialog
- Foley
- Ambience
- A FX
- B FX
- C FX
- Music

Foley, ambience, and the cut-effects premixes A through C constitute the sound-effects premixes. Consolidating them in the final mix results in the sound-effects stem. The most commonly found breakdown of the tracks at the final mix stage is:

- Dialog
- Music
- Sound effects

Called *DM&E*, these are known as the mix stems.

Another possible breakdown, used on sitcoms is:

- Dialog
- Music
- Effects
- Audience reaction (laughs)

Finally, special films call for special breakdowns. For instance, *Return of the Jedi* used:

- Dialog
- Music
- Effects
- Creatures

The reason for keeping creatures separate from dialog is that the creature utterances are to be used in foreign-language versions. They cannot be mixed into sound-effects units because if a

different balance between dialog and the rest of the mix is needed in a language other than the primary language (which is quite often the case), then the creatures need the same adjustment as the dialog, so they need to be separate. Let us say that the Italian mix has dialog some 6 dB louder relative to music and effects than it was in the English-language mix, which is often the case (the Italian mixers trying to be certain that the dialog is intelligible in poor venues and given that lip reading on the part of the audience, a part of the cocktail party effect of improved intelligibility, is unavailable to the audience in a dubbed version); then creatures will have to be raised 6 dB as well.

This brings up one large problem facing sound editors, the "purity" to apply to the breakdown along dialog, music, and sound-effects lines. Let us say we are shooting a pirate movie and there is to be a battle scene in which a set of pirates boards a ship. Is all the shouting that is heard dialog or sound effects? Is it intelligible or not? The advantage of all of the utterances being dialog is that all the sounds having a given principal language are grouped together, foreground "dialog" as well as background shouting. Then the effects tracks remain "clean" of dialog.

The problem with this approach comes when it is time to dub the picture into a foreign language. The process has been enormously complicated by the requirement that all of the background shouting must be looped. In point of fact, this is often not done because of the time and expense involved. Therefore, the foreign dubs seem surreal to the producers, because so much winds up missing in this case.

The alternative is to put the shouting-crowd noises into sound-effects units. Then the sound will appear in the M&E for the foreign-language dub, with the problem that the background language may be recognizable. The audience has the problem that the principal acting is in the dubbed language but the fight in the background is in the original language! A factor that can ameliorate this problem is to divide the background shouting into the lines that may be recognizable, put them in the dialog stem, and put those that are mutters or indistinguishable shouts in the sound-effects stem. There is no easy solution and the problem is routinely faced by productions that have foreign distribution.

TELEVISION SITCOM

Let us say we are shooting a four-camera television show. Although shot on film in the camera, a *video tap* also provides a simultaneous video output from the camera that is sent by radio to receivers in the studio and thence to videotape machines. There will probably be five video recorders in use, one for each camera and one to record a live, switched composite. A live audio mix is sent to each video recorder. Additionally, a multitrack audio machine records a separate track for each mic on the set. For instance:

- Boom mic 1
- Boom mic 2
- Radio mic 1
- Planted mic 1
- Planted mic 2
- Audience mixed mics L
- Audience mixed mics R
- Live mix

To maintain sync, the same time code is sent to all five video recorders, as well as to the multitrack audio recorder. Because of the nature of the production, this will be drop-frame 29.97 fps SMPTE time code, starting at 1 hour for the first reel, 2 hours for the second, etc. In this case the motion-picture cameras ("Filmed in Hollywood before a live studio audience") operate directly at 23.976 fps and no pulldown is needed. A 2:3 insertion is carried out electronically to make a 29.97 video directly.

The production delivers six sets of reels of tape or files to postproduction, five video and one multitrack audio, as well as production logs. In the case of film original, a telecine transfer is used to produce each tape or file. Each of the camera tapes is dubbed to a file format if it is not already in that form and sent to the picture editor. The edit produces cut video and an EDL. Once the EDL is done, the production multitrack audio may be dubbed into a digital audio workstation, or an export of the picture editing system may be used. The EDL is used by the sound editor to conform the dubbed audio to the picture, maintaining sync (in some systems, this mechanical synchronization process may be automatic, called *autoconform*). It includes handles from before and after edits so that the sound editor can do his or her job effectively.

The production soundtracks are first cleaned by having extraneous noise removed and then sweetened by having additional tracks of sound effects and music added. Once the sound editor is satisfied with the cut, the multichannel output of the digital audio workstation is provided to the mix stage. The mix occurs from tracks of a separate workstation, to other tracks of the same workstation or to a synchronized multitrack. Once mixed into stems, the stems are simply summed together to make a master mix, and this is then dubbed back to the videotape master in a process called layback.

The video quality of the video taps is low, and what is produced by the editing system is an edit decision list that can be used to telecine the footage from film that is necessary for a high-definition release. An online editing system edits together the source telecine files from disk or tape, to produce the picture on the edit master.

This description has been of a hybrid system, involving original sound capture on a multitrack, editing digitally, and then dubbing to a new multitrack linear format digital machine. One advantage of this work flow is that there is a physical product, a master tape, at the end of the day, not amorphous audio files, which must be stored somewhere permanent. This is not an unusual situation, because each technology is used in its strong suit: multitrack for massive capture and storage and random-access digital for its editorial capability. Historical trends favor increasing use of nonlinear digital technique over time, but the current cost/performance trade-off favors hybrid systems.

DOCUMENTARY AND REALITY PRODUCTION

Documentary and reality production most often today originates on single-system video media, with sound recorded to a removable memory "stick" or to the digital audio tracks of a digital videotape format. Time code on the camera "stamps" each frame of video with a unique identifying address for use throughout postproduction, including steps in which the audio is edited separately from the video, to maintain sync.

There are two ways to do postproduction picture editing (really it should be called story editing, because documentaries are built in postproduction and because editors cut both picture and sound). In the first method, the full resolution of the source medium is copied into the editing system and employed throughout the editorial process. This means that nearly finished programs can emerge from the editor, with only a need for color correction and a few special effects to produce a finished show. One reason that color correction is needed in a separate step is that it is unusual for editing systems to be equipped with color-accurate and calibrated video monitors. Although modern editing systems provide many of the effects formerly seen only in specialized equipment, there are still some elaborate effects that are performed after conventional editorial processes. The output of the color correction step, with the addition of the mixed audio in a layback step, is an edit master of the show.

In the second method, the source medium is ingested into the editing system with video compression, to save space. An offline edit is performed. The edit itself becomes notional, with an edit decision list being the main output of the editing system. A subsequent online edit session draws once again from the original source materials, steered by the edit decision list, to produce a finished edited program. The piece is then color corrected, special effects are added, and so forth, to produce the finished edit master.

The full-resolution method is both widely used and increasing in market share today. There are two reasons for this. Cameras routinely contain video compression between their light pickup sensors and their media, which limits the amount of storage needed. Table 11.4 provides examples of the required storage for various video media. Also, the cost of storage has dropped very dramatically in

TABLE 11.4 Data Rate and Disk Capacity for Some Video Formats

Format	Mbps	GB/hour	1-TB drive capacity in hours
DV tape	25 (pix + 2-ch sound)	11.25	88.9
DVCPro50 tape	50 (pix + 4-ch sound)	22.5	44.4
DVCPro100 HD tape	100 (pix + 4-ch sound)	45.0	22.2
XDCAM HQ mode memory card	35 (pix + 2-ch sound)	15.75	63.5
HDCAM SR HQ mode tape	880 (pix + 12-ch sound)	396	2.52

This is a simple calculation, not using GiB (see page 138), and not counting the need for overhead. For instance, the capacity of 32-GB SxS cards calculates to 121.9 min in HQ mode, but they deliver 100 min.

the past 25 years, by a factor of over 200,000:1.[6] The video compression factor, along with the dramatically falling prices of hard disk drives over time, has made full-resolution editing the method of choice, and it works if the editing equipment has sufficient storage and data-rate capacity. The method has the advantage of being able to reuse the relatively expensive camera memory cards, but reuse brings with it the risk of losing content altogether should something go wrong with the edit computer, as it is the only place where the precious camera original copies may exist.

Bit Slinging

The cowboy movie had its gunslingers, and some of them were very good at what they did. Modern editing rooms have their assistants, and I call them bit slingers, because they move data around, a lot, and many of them are very good at what they do, too. Although many of the jobs performed by assistant editors of an earlier film-based generation are now done by the editing computer, such as saving and subsequently finding head and tail trims of a shot, the modern-day editing assistant must be familiar with computer operating systems, application software, and networking and have many other computer skills. Backing up is one of the most essential skills, and it usually occurs after an already long day. It takes considerable time, but is essential.

The best backups are to separate drives or media than the main editorial one, are verified (compared to the original in a separate step from recording) on both the device that made it and, sometimes, a second device of the same kind,[7] and are stored in a different geographic location.

The information in Chapter 9 on Transfers, Chapter 8 on Sync, and other information in this book and at http://booksite.focalpress.com/Holman/SoundFilmTV/ should be useful to editing assistants.

Back to Our Story

The nonlinear editor is used in an *edit bay*, also called an *edit suite*, or, if applicable, *off-line room*. In doing this, the typical edit is in a form called *audio-follows-video*, that is, each picture cut is accompanied by a simultaneous audio cut, without intervention from the editor.

There is a critical feature of such nonlinear editing systems that makes them especially useful in this case: they record *nondestructive* edits. This means that what an editor is doing in cutting is really just making a list of pointers to the media files, with discontinuous jumps in the media for edits. These appear continuous (if they are intended to be!) because the machine is so fast that the jumps around the media by the computer are not seen.

The way that sound editing works in this environment is as follows. First, sound is exported from the nonlinear editing system to the sound editing system using a high-level file transmission scheme that sends the required media files along with the editing instructions that have been produced by the picture editor.[8] Intermediary software may be needed to get the translations to occur correctly, and the details vary. See also Chapter 9 on Transfers.

In addition, the sound editor may want access to all the material, including full handles and outtakes, as a source for presence, for example. So the sound editor needs access to the full original recordings.

The sound editor treats the sound delivered by the picture editor just like feature film does their A track. Leaving that track alone to provide a sync reference should

---

[6]This is not from a market survey but rather personal experience. I paid $1150 for a 60-MB drive in 1982, and at this writing a 1-TB drive costs $90.

[7]I have personally had this problem with the normally highly reliable MO drives. One drive recorded and the disk played back on itself, but would not play back on another MO drive.

[8]The media files may be embedded in the edit file, or separate, as described in Chapter 9.

anything go wrong in sound editing, copies are made and split at times where it is appropriate. Just where depends on when differences need to be made for the mix. If two people sound fine on a boom mic, there would be no reason to split the tracks, but if one is on mic and one is off, then the track would be split for each talker. This idea, however, is held differently by different mixers, some of whom would like the sound editor to split out everything by character, so some pre-postproduction is needed.

What is actually going on is that no new recording is being made by the "copying," but rather, new pointers are being written to the media file; this is an alternate definition of nondestructive editing. In the example above, in which we split the track just for perspective difference between subjects, each of the split tracks can be "pulled out" to reveal the audio content that existed in the original recording before or after the split point. This content is useful for making smooth transitions by fading one track out quickly as the other fades in. DAWs have a variety of shapes of fades, and all may be useful at one time or another. However, the most likely fade is one that maintains constant power across the fade by being down 3 dB in the center of the fade. Any other shape is fine so long as it sounds good, but power fading (wherein presence stays constant across the cut) is a good starting point. It is important in such track splitting not to allow any content to be doubled, that is, heard from both tracks simultaneously, as this would be obvious in a mix.

If the export has been done correctly, the sound editor also has sound available from before the beginning of a picture edit to after the ending of the shot, and it is these handles, extra time at the beginning and ending of each shot, that provide a means to smooth transitions across the edits.

Then dialog *cleaning* is done by editing the dialog tracks to improve them and substituting "room tone" in those places that need to be filled. In documentary production, usually there is no available "retake" from which to grab a word, should one not be clear, but we live with a style in documentary work that includes people talking as they do, with "flubs," more readily than in narrative filmmaking. On the other hand, there is usually enough space between words somewhere in a take that can be grabbed and looped and extended to provide matching presence to use as fill.

Many times documentaries are built upon interviews. Supplementary visual material is used, both of the live scene, called *cutaways*, and of stock or vintage material. Voice-over editing is a little different from when we see the subject on camera, as there is no need to provide sound to cover what we see. So "ums" and "ahs" of subjects can be cut out, for instance, making them easier to understand and follow. However, these cannot be cut out when we see the subject on camera, as it would lead to lip flap, or seeing a person's mouth moving but hearing no accompanying sound.

In many ways, because of the lack of alternate dialog sources that are available for fiction films, documentary sound editing is quite challenging. Here is a suggested outline to follow to edit documentary or reality dialog tracks:

- If the track sounds smooth and continuous, despite, say, changing from one subject to another, you need not split out by person, which would be more common to fiction films. The reason is that the same mixing processes will apply later to each of the subjects, leaving no reason to split them out.

- If there is a reason to split out various subjects within a shot, for instance, if one is on and one off mic, in fiction films the blank parts of the track remaining after a track split would often be filled out with presence. However, this also results in doubling the amount of presence, because we have two tracks full with alternating dialog but with continuous noise. The trade-off is the difference between smoothness and lowest background noise. In feature filmmaking, in which there is time, filling the tracks is more common so that the mixer can perform cross-fades as necessary to make a smooth and continuous sounding track. In documentary and reality-based work, filling out the tracks is far less common.

- If one word is needed to fix a flub, for instance, fill one track and cut the new word tightly into a new track in parallel at the right time. By cutting tightly around the word, the double presence for the duration of a word may go unnoticed.

- In cutting from track to track it is important that the timbre match. This is usually achieved by equalization with an editing system plug-in. At this stage, final decisions on overall timbre are hard to make because of monitoring conditions, but editorial monitoring is good enough to match an insert to a main track using equalization.

- There are circumstances in which one inserted word is clearly "drier" than the "wetter" main track, that is, less and more reverberant. Again, filling out the main track with presence and cutting the word into a second track allows putting a reverberator on just the inserted track, and this produces a better match.

- Another aid in matching an insert is to use pitch shifting. Sometimes quite small pitch shifts will help match a voice, and occasionally a statement can be turned into a question by automating a pitch plug-in and raising the pitch at the end of a phrase.

- Badly distorted, that is, clipped, sound can sometimes be "undistorted," by one of several means. Drawing in the assumed peaks of a clipped waveform on a workstation may be helpful, albeit tedious. Software specific to the task is available, called declipping or unclipping generically. Sometimes these devices seem to work well, and at other times, such as on radio mic overload, not well.

- Timbre is elusive to match. The order to proceed is to match level, declip, match timbre with equalization, match perspective with reverberation, and match the timbre of the insert by pitch shifting. It is not in my experience necessary to do all of these to any one sound, but any of them, and combinations of them, can be useful.

One of the biggest problems of documentary production comes from the huge amounts of material that are often available. To make the editorial task easier, transcripts of interviews are typed. It is tempting to build the story by editing the pieces of the transcript into a logical whole. The problem with this is that it ignores sound quality and the huge shifts that may take place for different locations. By recording under a variety of conditions, completely different patterns of reverberation and background sound are present in each of the venues used for interviews. It may be tempting to cut the interviews together as a voice-over for other footage so that the picture does not appear to pop back and forth from one scene to another, but to do the same thing with the sound-track should be seen as equally troubling as changing the picture.

After cutting the dialog the program is then *sweetened* by adding additional tracks of sound effects and music. Mixing proceeds along the same lines as for feature films, with some simplification. Because no foreign language dubbing is done, there is no reason to keep things separate, and because there are fewer tracks to begin with, and console automation to help out today, it is common to mix from all the available tracks directly to the final mix. Still, premixing is not unknown in documentary sound.

Mixing

INTRODUCTION

The process of mixing soundtracks together has slightly different terminology applied to it depending on the field of use. In film work, mixing is called *rerecording* or *dubbing*. In television production, the words mixing or sweetening are more commonly employed to name the process.

The term rerecording is perhaps the most self-defining of these terms. The term mixing, on the other hand, can be applied to a wide variety of tasks, from live sound mixing at concerts, to combining the tracks of a multitrack recording into fewer tracks for release of a music compact disc, to making adjustments in level in transferring one two-track master to another. Rerecording is a more limited term, meaning taking something already recorded and distilling it down by mixing processes to a more convenient representation from many units, elements, or tracks to premixes, or from premixes to final mixes. Dubbing is synonymous with rerecording when applied to the mixing process, but it is also applied to any copying process that may occur in audio or video recording.

Lay down, lay over, and *lay back* are terms that essentially mean precise equal-level copying at the various stages of a television production. Laydown means capturing from the production sound medium into an editing environment, usually a workstation. Layover is synonymous with dub (used as the second definition above), and with x-copy or exact 1:1 copy, but these two terms are more associated with film than with video. Layback is the copying from the sound master to the composite video master, usually the last step in the process.

Rerecording is typically done on a dubbing stage by one to three mixers working simultaneously on the program material. The stage is arranged to be like a theater, with the mixers occupying the best audience seats, the basic idea being to combine the tracks of the relevant stage of processing, simultaneously manipulating them for the best sound quality and the desired effect, while hearing the movie under the same conditions as an audience would in a theater.

A top rerecording mixer brings a lot of experience to projects. The lead mixer generally is responsible for dialog and possibly for music, too. The complexity of sound effects means that if only two mixers are available, one will mix dialog and music and the other sound effects. The lead mixer in doing dialog must blend it from disparate edited sources, through the use of the techniques explained in this chapter. But he or she also has another important role: to stand in the shoes of the audience. A common problem in postproduction rerecording is that the director is so familiar with the dialog, having heard it hundreds or thousands of times at this point in the process, that he or she can perceive the dialog, even if you shut it off! The experienced rerecording mixer knows how to shape the dialog and other tracks so that the dialog in particular is intelligible (if not intelligent) throughout the picture. Even throwaway lines that important directors say are unimportant during mixes will get audience members to turn their heads to their companions and ask: "What did he say?" It breaks the concentration of the audience on the story. An excellent example of a dialog-heavy mix that is very well balanced and blended is *Julie and Julia*, mixed by Lee Dichter.

In postproduction, it is necessary to have a system that can "rock and roll," that is, respond to commands and go forward or backward, at high or normal speed, while maintaining sync. In fact, one of the distinguishing factors of film dubbing compared to most digital audio workstation playback is the ability to play in reverse, which is revered by mixers who work on the sound processes during reverse play and so maximize their productivity.

Film has traditionally been projected for film dubbing, but its slow speed access to different parts of the reel has given way in recent years to dubbing while showing a projected video image, which may be available from a nonlinear editing system and thus able to "reset" to another time instantly. Entering the mixing room has also been the digital audio workstation, usually of the same type used by the sound editors. In this way, a sound editor can be on standby to make rapid editorial changes if called for. Also, digital film dubbers offer offset capabilities, so sound can be slipped in sync relative to other sounds during dubbing without tying up editorial facilities.

Sweetening for television programs is typically done in a relatively small studio, by one or two mixers working simultaneously, which is arranged to be closer to a living room than to a theater in size; however, more expensive TV dubbing is done in larger rooms.

Sound Source Devices Used in Rerecording

Traditional dubbers played magnetic film across virtually any number of machines interlocked by a signal called *biphase*. Magnetic film could carry from one to six tracks per strand of film, and interlocking multiple machines meant a very large number of audio tracks could be played simultaneously. Large studio sound departments had machine rooms with up to about 100 machines, each of which could theoretically carry up to six tracks, but this centralized room served multiple dubbing stages simultaneously, and none of them could handle 600 tracks, although 200 was not unknown. Devices were built to provide a slip sync function for one machine at a time, so sound effects could be quickly resynced if relatively simple changes were needed. Edit/change rooms were usually made available so that sound editors could quickly revise units more thoroughly as needed. Television postproduction has relied on 2-inch 24-track recording for many years as standard, until recently.

Modern postproduction has mostly done away with the magnetic film dubber in the past few years, although they will be in use for a very long time to come to play back legacy recordings. The number of 2-inch machines in use has also begun declining. At first the replacement was a generation of multichannel digital audio tape recorders called MDM as a class, for *modular digital multitrack*. The colloquial name for these machines is "DA-88," named for the first machine of the class, but the tape format is more formally named DTRS. These eight-track machines cost about one-tenth the price of an analog dubber, so were quickly adopted. They could be ganged by control functions, so that a large number of tracks could be supported. The sound quality was decent, although the machines were 16-bit linear PCM and did not have as much dynamic range as the magnetic film or tape with Dolby SR that they replaced. Combining many 16-bit tracks results in a dynamic range of less than 16 bits, so this is seen as a limitation. Also, the first machines lacked confidence head recording, although this was added to later models.

Modular digital multitracks were a way to store a lot of digital content cheaply, but access to that content was limited to the linear domain. To go from the end to the beginning of a reel, the tape had to be rewound and then, when Play was pressed, would have to sync up. Although the machines were relatively cheap, faster operation could be achieved if a random-access-based system were to be used. Also, MDMs could not play in reserve as dubbers could, so mixers considered them to be a step backward.

Thus digital dubbers were developed, characterized by random access, quick lock up, play in reverse, and control integration features for dubbing and control over record insertions called punch-ins, described later in this chapter. The format of storage is hard disk or magneto-optical drives. With this development, software was made available to read some standard DAW file formats directly, so putting sound up on a digital dubber usually means simply plugging in a drive mounted in a carrier today.

Digital dubbers are on the order of double or triple the cost of MDM machines, but, being purpose-built with the features desired for the film and video postproduction market, were successful. Meanwhile, digital audio multitrack machines based on hard disk drives have come on the market too. Usually in a 24-track format on hard disks, they are used more widely than dubbers for all kinds of recording. Their large-scale manufacture makes them cost about what an MDM machine costs, with the added benefit of random access. However, they have limitations in dubbing applications, such as no play in reverse function, etc. Their low cost makes them attractive, though, for those willing to give up on features.

DAWs are the form of playback and rerecord most often found on dubbing and mixing stages today. An advantage is their instantaneous ability to re-edit. This is also a disadvantage if the process is not disciplined. Using an expensive dubbing/mixing stage as an editing room is not a productive use of the producer's cash. DAWs may deliver their output over digital audio interfaces such as AES3, in which case the workstation is tied up supplying sound to the console and cannot be used simultaneously for editing. Some dubbing stages have been built recognizing this trend, with booths or balconies for DAWs. Editors sit on the dub stage and take their workstations offline (that is, not supplying sound to the console) and re-edit on the fly (using headphones to monitor) for quick edit changes.

Mixing Consoles

Mixing consoles used for dubbing are often large and intimidating, with hundreds to thousands of controls. Luckily, there is a great deal of duplication among the controls, so by learning just one area of a console one learns nearly all of the areas. Fundamentally, what goes on inside consoles can be broken down into two ingredients, processing and configuration. Sound processes are the devices used to affect the sound, including all the way from simple internal level controls to sophisticated outboard reverberation units. Configuration issues are about signal routing from the input to the output of the console through the various processes. Of the two, processes are easier to study, because they are represented by the knobs, switches, and other controls on the console. Configuration, on the other hand, is more obscure, because modern consoles are so dense with controls there is no room to draw diagrams

on them showing the electrical order of the controls. Given this state of affairs, we will look into processes in some depth and then give the general principles of organizing them by configuration.

It is worth pointing out that each new console faced by a professional is just as much a sea of knobs to him or her as it is to the preprofessional until the array can be broken down into logical units and addressed singly. Whereas a professional will recognize a great many of the knobs for the processes they represent, everyone inevitably needs training on the configuration of the processes.

Before describing processes, one organizing feature is worth noting. In many consoles, the construction is such that a series of processes that are associated with one input are arranged vertically in a console *slice*. This means that a primary issue in configuration is accounted for by this fact: An entire column of knobs is likely to be associated with processing the signal from one source. With that in mind, we will first look at various processes that may appear in an individual slice and then at variations from this standard.

Obscuring the classical distinction between editing and mixing is the fact that DAWs today have many mixing features and may even have more potential different processes available as software plug-ins than major consoles. Console control surfaces that operate the functions of DAWs are becoming popular. The distinction between DAWs equipped with a control surface and large consoles is usually that, if the console is digital, it will have dedicated digital signal processors for each channel and thus may be designed not to "overload" under the burden of

signal processing and possibly crash or lose signals. DAWs are more likely to dynamically assign resources like digital audio signal processing power and so could run out if a great many signal processes were in simultaneous use. This can often be solved by plugging more hardware into the DAW, but then its cost may approach that of a console.

PROCESSES

Level

Setting the level of each of the elements of a mix is surely the single most important item to be done in mixing. Even the simplest of equipment has a means to adjust the relative level (also called gain) or volume of the individual elements by way of faders, also called potentiometers or pots. The reason is simple. If each recording has been made to make best use of the medium, then a Foley recording of footsteps, for example, will be recorded about as loud as a dialog recording. When rerecording, though, it is necessary to get these various elements into balance with one another, so inevitably the Foley element will be turned down relative to the dialog element to assume its proper relationship in the mix.

The main level control for each input is given more weight than any other console process by the placement and type of control. On rerecording consoles, the main channel fader is always the control that is largest and closest to the operator and is usually a vertical slider type of control with markings for resetability.

FIGURE 12.1 A Harrison MPC4-D console installed in the Hitchcock Theatre at Universal Studios. Photo courtesy of Harrison Consoles.
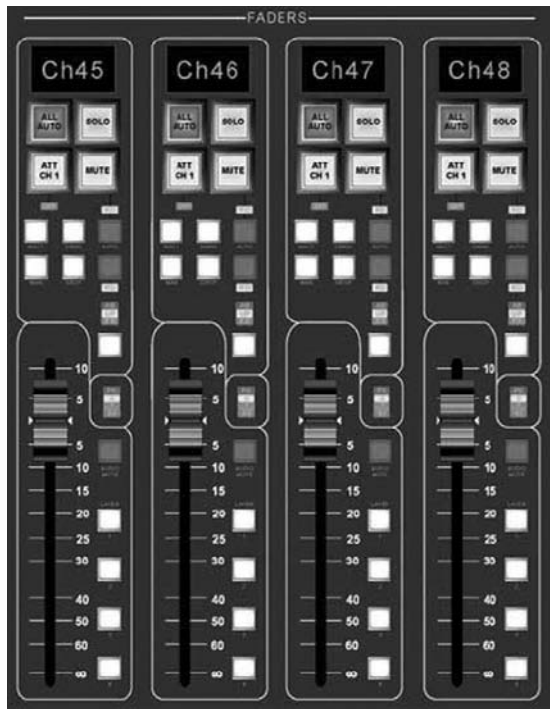
FIGURE 12.2   Level controls. Photo courtesy Harrison Consoles.

Another related primary control is called a *mute*, which is simply a switch that kills the signal altogether, allowing for a speedier turn-off than turning the fader all the way down rapidly. Mutes are probably more commonly used during multitrack music recording than during film mixing because in music all tracks are on practically all of the time, whereas workstations produce silence when there is no desired signal, thus accomplishing muting right at the source. Music mixes may mute individual channels for whole sections of the mix, say, the string channels during a brass solo, to prevent audible cross talk into the unused channels from the open, but unused, mics. The mute function also provides a means of identification of where a sound might be. By activating the mutes in turn during a trial run, the mixer can learn where the various sounds are, in case the cue sheets are faulty or unclear.

In cases in which there has been too much sound cut for a sequence, and not enough time to change it editorially, a mute function may become valuable. Here, a computer follows the action of switches throughout a reel, keeps track of what is muted and what is unmuted, and performs the mutes on subsequent passes. Automation of mutes can be built up into a complex pattern over many passes.

These two functions, level and mute, are so important that they were the first functions to be automated in more elaborate consoles. A means to mute all other channels and to monitor only what one channel is contributing to the mix is very useful. This function is called *solo*. Pressing the Solo button on a channel will make the monitor

mute all other channels. Pressing more than one Solo button will produce a mix of only those channels.

> There are variations on the solo idea. Most solo systems offer only a "cue" or "audition" function, so the signal processing, such as the position in a stereo mix, is lost when solo is activated. Some offer "solo in place," representing the sound being soloed correct spatially. Many solo systems are "destructive," that is, the output mix of the console is affected by the solo function so it cannot be used during principal mixing, but others are "nondestructive," affecting only what is monitored, not what is recorded. So individual consoles vary greatly in their possible solo functions. Photo courtesy of Harrison Consoles.

Multiple Level Controls in Signal Path

On its way through a modern console, a single signal may well pass through a large number of level controls—individual channel fader, subgroup master fader, master fader, and monitor volume control. This multiplicity of controls, while offering high utility and flexibility, also creates problems. The problems are similar to those of recording on a medium: If a tape is underrecorded, when the level is subsequently restored by "turning it up," noise will become evident. Conversely, if overrecorded, the resulting distortion is permanent and will not be removed by turning it down at a later stage. Although consoles generally have a wider dynamic range than recorders, hitting the dynamic range of each of the intermediate stages correctly is an important issue to avoid excessive noise or distortion.

On the best professional consoles, with their multiplicity of controls, attacking this problem of the correct setting of the variety of controls is accomplished relatively easily. The scale on them is the clue, with 0 dB the nominal setting of the controls. Many of the controls have "gain in hand," which goes above 0 dB, that is, one can turn it up from the nominal to reach for something underrecorded as needed, but the nominal setting is clear. On consoles that lack this feature, it is necessary to determine which settings of all of the controls are the nominal ones. It is usually the channel fader for each slice on which most of the actual mixing is performed. The other controls, such as submasters or master level controls, are used for slight trims to the overall section-by-section balance or for the main fade-ins and fade-outs of the overall mix. On the other hand, because the individual channel slice gain controls can be used to set the "balance" among the parts of an effect, a submaster can be used to set the overall level of an effect.

Dynamic Range Control

Compression

Each track used for rerecording has a volume range. When tracks are combined in mixing, the problem of unintentional masking of one signal by another arises. Let us say that we

have a dialog recording, with a volume range, and a music recording, with its own volume range. Starting at the beginning of a show, we have music as the foreground, but it is not theme music in this case that fades out before the dialog begins, but rather source music, fading under the dialog. The problem is that although most of the time the music will lie underneath the dialog, there may be a point in time at which the peaks of the music correspond to the minimum level of the dialog and the dialog is obscured. On the other hand, there will also be times when the music is faded under so that it seems to go away altogether. The alternating presence and absence of the music is distracting.

To solve this problem, we could "ride the gain," turning the level of the music down during its higher level passages and up during its softer ones, to maintain a more even level behind the dialog, but this would be tedious and time consuming. The process can be automated by a device called a *compressor* that does just what has been described. A compressor is equipped with a number of controls to vary the volume range over which the action of the compression occurs, the amount of the compression, and how fast or slow the compressor acts. Each of these devices is fairly idiosyncratic as to control functions, so the number of knobs associated with them varies.

A typical compressor may have the following control knobs:

1. A "threshold" control, below and above which the compressor exhibits a different "transfer function." Usually below the threshold, the compressor acts as a linear amplifier, such that each decibel in yields a decibel out, and above the threshold each decibel in results in less than one decibel out.
2. A compression ratio control with markings such as 2:1, 4:1, 20:1, or more. This is the ratio between input and output in decibels above threshold. A 4:1 compression ratio would mean that a 4-dB change in the input produces a 1-dB change at the output, above threshold.
3. An output level control. Because compression often has the effect of lowering the overall level, this control is used to make up the gain and to raise the overall level after the compression process.
4. An attack time control. This control modifies how quickly the controller circuitry responds to an increased level. If it changes too quickly, short sounds that do not reach full perceived loudness may control the gain excessively and audible gain riding may result. If it changes too slowly, then loud attacks may be heard followed by a level change downward, also leading to audibility of the action. A typical starting value for this control is 80 msec.
5. A release time control. This control modifies how the controller acts when the signal decreases. If the control function is made too fast, the gain will change within one cycle of the signal, which leads to harmonic distortion. If the gain change is set too slow, then soft sounds following loud ones could be lost completely. A typical starting value is 0.5 sec.

Of course, technically speaking it would be equally possible to compress the dialog into a narrow volume range to keep it above the music continuously, but even with the music in correct overall balance, it may seem to come and go, which can draw attention to it. The principle of "least treatment" for this scene would say that the mixer should process the background sound first rather than the foreground sound, to leave the fewest artifacts present. On the other hand, I have found it useful to use small amounts of compression on dialog to make it sound more natural. The reason for this may be that we are recording dialog typically at one point in space, but we hear at two points (our left and right ears). The spatial averaging of level that occurs by averaging sound at two points tends to mitigate the more extreme level differences observed at just one point, which for us would be a microphone. Thus slightly compressed sound may actually be a better representation of what we hear than "pure" uncompressed sound. But note that I am speaking of small amounts of compression, on the order of 6 to 8 dB of maximum gain reduction from the linear condition.

If the dialog and music tracks have already been combined, then compression is not an option. The louder track dominates the "thinking" of the compressor, moment by moment. A likely outcome of this condition is that the level of the music would audibly go up and down with the level of the dialog. Assuming the dialog is practically always above the level of the music, when the dialog is soft the compressor will turn it up, bringing up the music as well; and when the dialog is loud, the dialog and music would be turned down. This is an effect known as *pumping*, when the level of one element of a mix audibly affects another element in level and is generally undesirable. It is thus best to compress each individual source alone and then to combine sources, rather than to try to compress the entire program.

Keeping music "under control" while faded under a dialog source is an aesthetic use of a compressor that would be made regardless of any other consideration in the overall system, but there is another reason to use compression that does not involve the aesthetics of a mix. Generally, postproduction works with media that have a wider dynamic range capability than that of the ultimate release format, and a compressor may be used to reduce the overall dynamic range of a program to fit it within a specific release format.

One nonaesthetic use for compression is to reduce the large dynamic range of a theatrical presentation to the smaller range thought necessary for home video. If all users are expected to listen to a program over a television set internal loudspeaker, then that may be given weight in a compression decision. For instance, children's VHS video

and airline and hotel video copies are compressed compared with the theatrical version of the same movie. So, there are different uses for compression—in rerecording to control certain tracks to make them more manageable and subsequently to "fit" the program material's volume range to the range needed by individual media or users.

Expansion

The opposite of a compressor is an expander. An expander increases the volume range of a source and may do so across a wide dynamic range or may be restricted to a narrower region by control functions. Restricting expansion to just low-level sounds helps reduce noise. Called *downward expansion, noise gating*, or *keying*, this function turns the level down below a threshold set by a control. For example, all sound below, say, −40 dB could be rapidly turned down to essentially off. The advantage of this is that there is often low-level noise on each of a number of tracks, which is undesired and would be a problem if all of the noise sources were mixed together continuously. With use of a noise gate (which would be more properly called a program gate because it essentially "turns on" for signals above a certain level and off for ones below that level), such additive noise can be reduced because only tracks above a certain level will get through the gate.

Noise gates have a number of audible problems. Let us say that we have a dialog recording with some air-conditioner noise in it. The threshold of the noise gate can be set to distinguish between the dialog and the air-conditioner noise because the air-conditioner noise is lower in level than the dialog. The problem is that we hear air-conditioner noise behind the dialog when speaking is going on, and we hear its absence in between lines. The dialog is pumping the level of the air-conditioning noise. The exaggerated difference between the noise being on and it being off may well draw more attention to it than just leaving it alone, unprocessed. This is related to the reason that editors cut presence or fill in between lines of dialog—to make a smooth and continuous whole.

For this reason, traditional "all-or-nothing" noise gates are not used too often in critical rerecording tasks. They are used, however, in multitrack music studio work. For example, suppose we have recorded an orchestra and have placed the instrument groups on different tracks. Upon playback, we find that the string track is polluted with acoustical cross talk from the brass section. To eliminate the brass from the string track during passages in which the strings are not playing, we can use a noise gate, set so that when the strings are playing the gate is on, and when they are not it shuts the signal off. This operation depends on the string level being higher than the cross talk of the brass and on setting the noise gate so that it can discriminate between them. In this application, the noise gate is used on playback, in which the control function can be

changed and repeated if necessary, rather than on the record side, in which any mistake could not be corrected subsequently.

An advancement on the all-or-nothing approach is frequently used in dubbing. Some noise reduction devices work like gates but do not turn the signal fully off, so changes are less abrupt. A more sophisticated approach breaks the audio spectrum up into multiple frequency bands and applies a downward expansion below a certain threshold separately in the different ranges. Also, the expansion is not made as dramatic as turning the signal all the way off. These two strategies have the effect of producing much less objectionable side effects. Audible pumping is greatly reduced or eliminated by these approaches. The generic name for such a device is a *multiband low-level expander*, although because that is such a mouthful, the actual units go by their trade names. These operate in both the analog and the digital domains, with potentially many parallel frequency channels. All of them attempt to distinguish desired program material from background noise content and decrease the level of the background noise without affecting the program material.

Limiting

A variation on the idea of a compressor is a limiter. A limiter acts on signals above a certain threshold, as a compressor does. Above that threshold, the level is controlled so that for each decibel of increase on the input of the limiter, the gain is reduced by the same amount. Thus, above the threshold, the level simply stays practically the same despite any increase in level.

Limiters are useful in production sound to "catch" occasional events that might not otherwise be controlled as to level, to bring them into a range in which the recording medium can handle the signal linearly. They are routinely included in production recorders and in camcorders for the utility they offer in keeping unexpected signals under control.

> Limiters are very useful for keeping unexpected high levels from distorting, such as an actor "going over the top" and shouting more loudly on a take than in rehearsal. With varying program material, though, there may be a problem. Let us say that there is a gunshot in the middle of a scene. With a limiter in use, the gunshot will certainly "trigger" the limiter and the gain will be dramatically reduced. The difficulty is that it will be reduced for sounds coming after it as well, and the "duck" in level may well be noticeable for speech coming immediately after the shot. In such a case, it may be better to proceed without using a limiter, letting the tape distort on the gunshot briefly. In that way, a postproduction editor can "clean out" the gunshot from the production soundtrack by cutting it out, leaving no artifact other than a hole, and cut a clean gunshot into the sound-effects track. The limiter must reduce the gain quickly and keep it that way for a while, because if it did not, the result would be distortion of low-frequency sound, so the duck after a very loud sound is a natural artifact of limiting.

Limiters are also useful in postproduction in several ways. They can:

- Put an upper limit on one track, so that it cannot rise in level so much as to interfere with another track. An example is Foley recordings of footsteps in which one or a few footsteps sound like they stick out of the mix, but lowering the overall level of the Foley track is not satisfactory, because then the overall impression is too weak. So applying a limiter set so that it catches the highest level footsteps and keeps them under control is a useful function.
- Probably the most common use of limiters is to control the highest level signals on a soundtrack so that they can be recorded on a particular medium without distortion. Examples include limiting for analog optical soundtracks, which have 9 dB of headroom, that is, 11 dB less than the digital tracks used as their source, which have 20 dB headroom.
- Another use for a limiter is to raise the overall loudness without affecting the maximum level. By limiting the highest peaks of the program to a lower level, the average level can be raised without exceeding a particular maximum. With the proper sort of limiter, 5 to 6 dB of limiting can be practically inaudible, and this amount is a large advantage in fitting into the requirements of, say, an optical soundtrack or any other limited-headroom medium.

De-essing

One particular specialized form of limiting is de-essing. This refers to the fact that many media are sensitive to reproduction of the "esses" in speech. By limiting strongly only on such high-frequency sounds in dialog tracks, the resulting soundtrack can be easier to record to certain media, such as optical soundtracks. Sibilance distortion is the result of imperfect waveform reproduction, when the high-frequency sounds in speech, especially esses, are affected.

A de-esser is a limiter with one difference—the control function for limiting is made sensitive only for high-frequency sound, so most signals are unaffected. Should an ess sound cross the threshold of limiting, then the level will be ducked for the duration of the ess. You might think that this would make the sound dull because high frequencies are being reduced in level, but de-essing can be surprisingly benign, having little audible effect except to reduce the sensitivity to such distortion at subsequent stages.

Conclusion

All of the items so far discussed affect the level of audio signals. By far the most commonly used of these processes is level control, which is used even on the simplest of mixers. Muting and solo are also found on most rerecording consoles. Dynamics processing, including compression, expansion, limiting, and de-essing, are also frequently used in rerecording, but more rarely than level controls.

Some rerecording consoles include all of these processes in every input path of the console, whereas others do not have dynamics controls in each input, but rather route signals from the inputs to separate dynamics processors, either built into the console or external to it.

Processes Primarily Affecting Frequency Response

Processes that affect principally the frequency response of the signal are probably second in importance only to level control. These processes can clean up the audio signal, make it more interchangeable with other signals (for instance, adjusting the timbre of production sound and ADR recordings to be more equal), adjust for the loudness effect (by adding bass for sound portrayed at lower than its original level), and generally make the sound more intelligible or pleasant or, for effect, deliberately worse.

Most of the ear training necessary to become a rerecording mixer is involved with level and frequency response processing, with other factors such as dynamics playing a subsidiary role. The reason for this is that the maintenance of good continuity depends much of the time on the soundtrack not drawing attention to itself through unexpected changes in level or frequency response. The changes must be smoothed out so that the audience is not distracted by them.

The two principal frequency response determining processes are *equalization* and *filtering*. Although either of these processes may affect any frequency band from low bass through high treble, there is a fundamental difference between them. Filtering is done essentially to eliminate certain frequency ranges from the output, and thus the action of a filter changes abruptly across frequency. In fact, filters are rated for their frequency and their slope in decibels per octave. The slope of a practical filter, how much its output changes across frequency, is generally greater than that of equalizers.

Equalization

Almost everyone has some experience with bass and treble controls on a stereo system. Equalization is the professional name for the process of altering the frequency response in a manner similar to what tone controls do on a stereo system. However, only the simplest professional equipment devotes only two controls to this important task. For professional equalizers, often found in each input channel slice in rerecording consoles and possibly in other parts of the chain, the audio spectrum is more frequently broken up into three or four parts, which might be called low-bass, mid-bass, mid-treble, and high-frequency controls. Alternatively, the two middle controls could be labeled midrange 1 and 2.

Each of the main equalizer controls, called EQ, are rated for how much variation they produce in decibels when set to their extremes, such as ±12 dB. These main controls may be supplemented by other continuously variable knobs or switches to provide more flexibility in use. If provided, these extra controls affect the parameters of frequency and curve shape, resulting in the name *parametric equalizer* for the device.

The first of these subsidiary controls usually available is one that changes the frequency range in which the EQ control is most active. Note that the frequency range of the various controls may be set to overlap, producing the possibility of up to twice as much boost or cut as for one control on analog consoles, or up to potentially four times as much on a four-band digital equalizer with a range of 20 Hz to 20 kHz in each section.

The second most likely control is one that changes the general shape of the curves produced by the equalization control. The two shapes offered are bell and shelving. Bell-shaped curves are centered around a specific frequency and at maximum boost or cut look like the outline of a bell, right side up or upside down, respectively. Shelving curves are like conventional tone controls: Once boosted or cut, the whole frequency range from the center frequency of the control to the audio band extreme is equally affected. Thus, this type of control is usually found only on the lowest and highest frequency bands of a multiband equalizer (see Figure 12.3).

The use of the two depends on why the equalization is being done. For example, a shelving characteristic is used

FIGURE 12.3 The curves given by (A) represent commonly found bell-shaped equalization curves, with the "Q" as well as center frequency varied between the two sections. Curves (B) and (C) represent low- and high-frequency shelving curves, respectively, most commonly found as the bass and treble controls on a stereo system, but also useful in rerecording.

to overcome a muffled sound associated with too much cloth over the mic, whereas a bell-shaped curve may be used for precise equalization of musical overtones of particular instruments. The shelving equalizer is a broad stroke, and the bell-shaped equalizer is more specific.

The third most likely control is called Q. This relates to the "sharpness" of the control function with respect to frequency. Two controls can have the same center frequency and boost, say +8 dB at 2 kHz, but they may get to that boost in a manner that is either very narrow, having little effect on neighboring frequency ranges, or wide, having effects well away from the center frequency of the equalizer. A narrow equalizer has high Q, and a wide one has low Q. Because low Q affects the response in more critical bands (pg. 38), it is generally more audible. Low Q is usually used for program equalization unless a particular problem in a narrow frequency band is the trouble, and then the high-Q version is valuable for having little effect away from its center frequency.

Professional equalizers could then have as many as 4 knobs or switches per frequency section of an equalizer, and 4 sections is not uncommon, so 16 controls affect frequency response in this scenario, and these program equalizers are usually found in every channel. On a large postproduction console with many inputs, controls just for equalization run into the hundreds, demonstrating the importance that equalization has for the postproduction process.

Another type of equalizer is also found on dubbing consoles, but usually not in each slice. *Graphic equalizers* consist of a row of multiple knobs that can be used to "draw" a desired frequency-response curve, offering a very easily grasped human interface. The number of knobs across frequency varies with the model, with 6 to 8 being common. The curves are usually bell shaped, and most graphic equalizers offer no means to become parametric—the frequency of the controls and the curve shapes are fixed. This type is patched into the channels as needed, either into individual channels or, more likely, into groupings of channels. Because of the lack of space on the console operating surface, the use of graphic equalizers has declined in recent years.

Perhaps the most sophisticated example of equalization is the matching of dialog from one source to another, such as lavaliere to boom, or ADR to production sound. The principal tool is equalization. We know the timbre of voice so well through internal memory that we can recognize a voice from just one word spoken, even over a telephone. A great deal of timbre is controlled by equalization. By measuring the long-term spectrum of the voice on a boom and a lav simultaneously, the frequency response of the lav in situ is developed. It is shown in Fig. 12.4. The equalization needed to get the two to match is the inverse of this, so perhaps the most prominent things to treat are the 630-Hz peak with a dip and
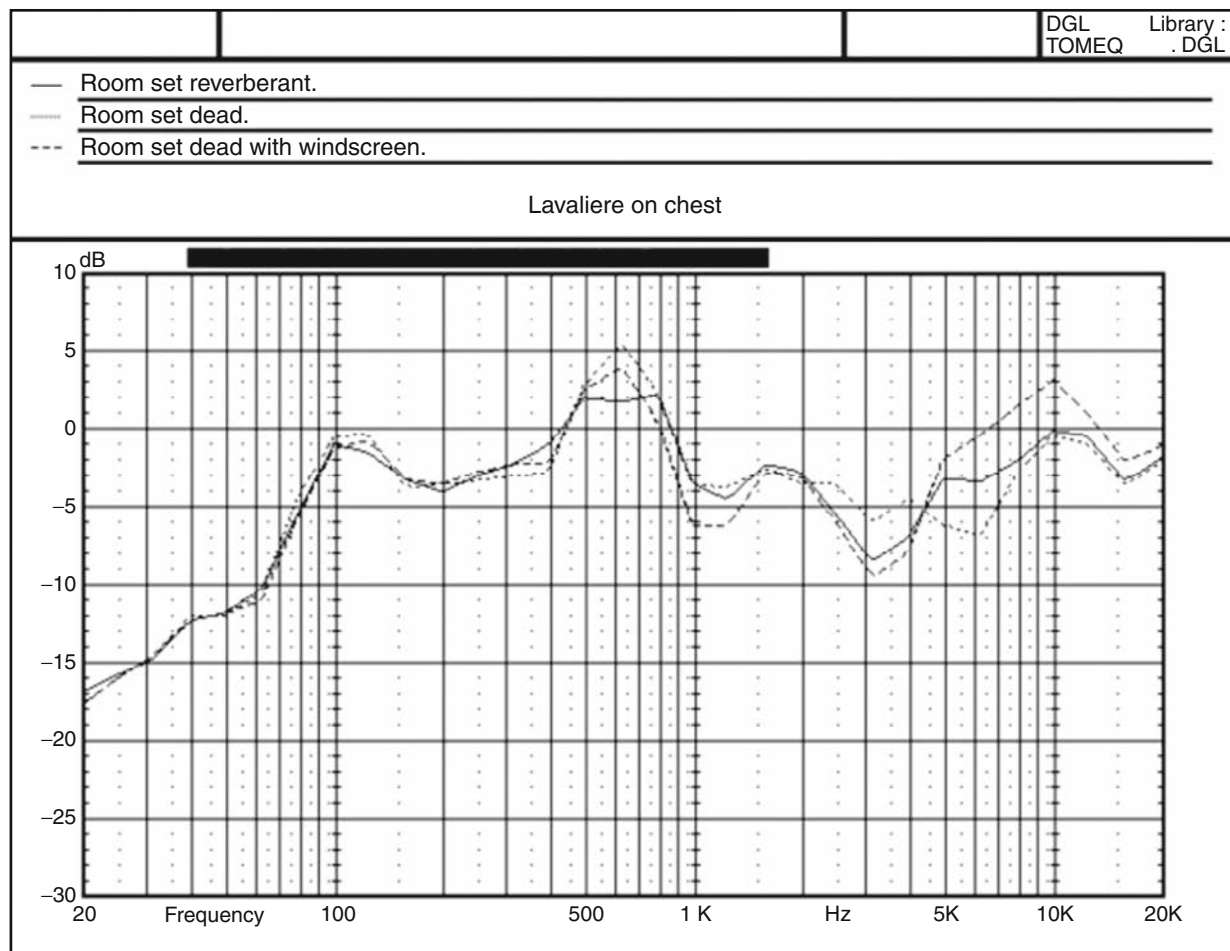
FIGURE 12.4   The frequency response of a lavaliere compared to a reference boom mic under various conditions.

the 2-kHz dip with a peak. With a 10-band parametric equalizer, the results can be as good as that shown in Fig. 12.5.

Filtering

Filters are distinguished from equalizers by being more brute force in their action. They are intended to essentially eliminate certain frequencies from the output. The utility of filters is in corrective action generally, compensating for noise in the recording, especially low-frequency noise.

Filters may strip off any part of the audio spectrum. Probably the most commonly used filter attenuates low bass and passes the rest of the spectrum essentially unchanged. Such a filter is called a *highpass filter* in professional circles because it passes highs while attenuating lows. On consumer equipment, on the other hand, exactly the same device is called a *low filter* or *low-cut filter*. So-called *prosumer* equipment uses either the term highpass or the term low-cut to mean the same thing.

The various filter types and a typical use for them are as follows:

● *Highpass (low-cut) filter:* used to remove excessive room noise, which is often concentrated at low frequencies.
● *Lowpass (high-cut) filter:* used, for example, in music recording to help isolate a low-frequency instrument playing in a recording studio along with others. Isolation from cross talk is improved by stripping off the highs from other instruments in the studio that are leaking into the open mic in front of the bass instrument. It also may be used in speech channels, to strip off high frequencies that are noisier than normal, such as an 8- to 10-kHz lowpass filter.
● *Bandpass filter:* a combination of high- and lowpass filters. One use is as a "telephone filter," so called because restricting the audible frequency range in this way sounds like one of the primary things that a telephone does to sound.
● *Notch filter:* a filter that greatly attenuates only a narrow frequency range. It is useful for removing tonal noises such as certain whines or whistles. Notch filters
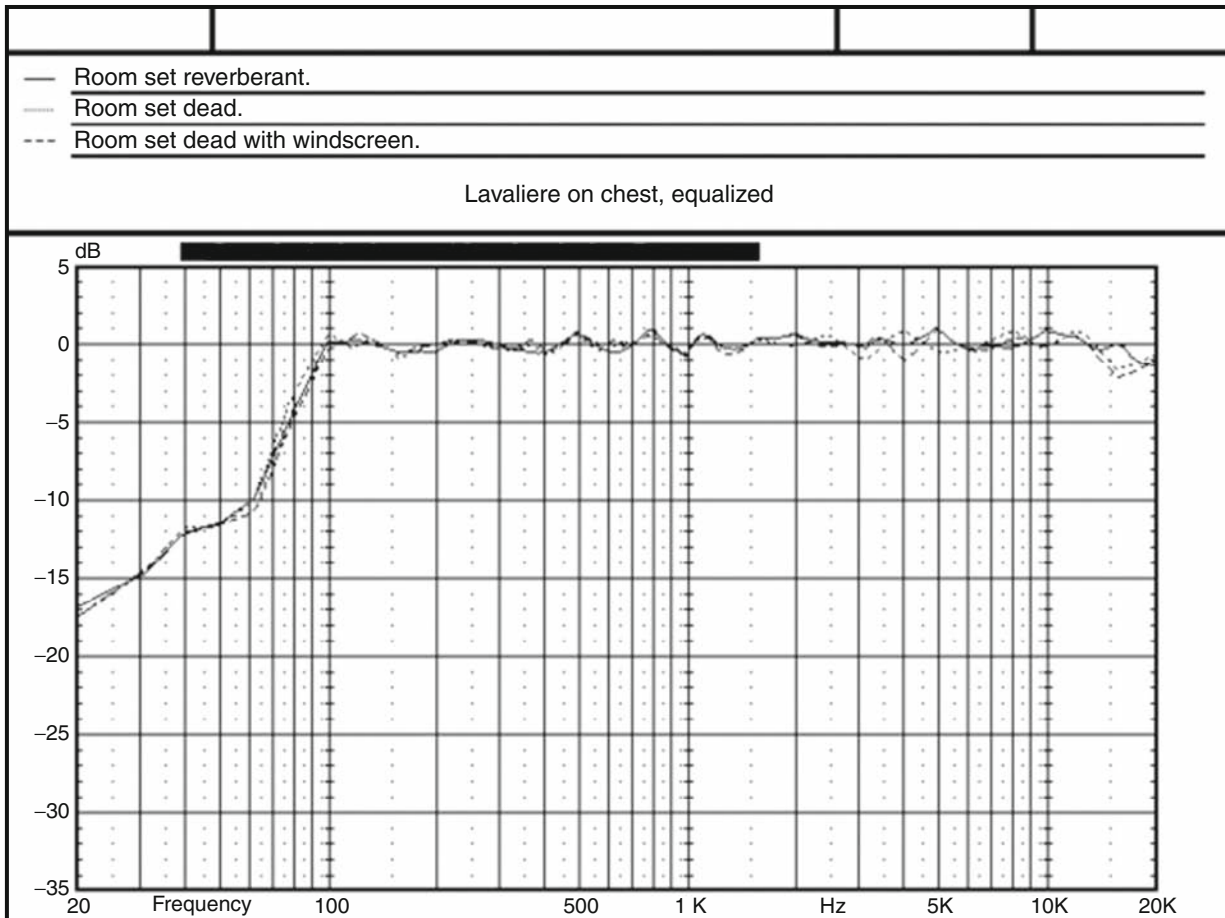
FIGURE 12.5   When compensated with a 10-band parametric equalizer, the response matches between the lav and the boom reference, and they sound remarkably alike. Below 100 Hz, not much useful content occurs from voice, and boosting this region could lead to excessive handling noise, so it is not compensated.

usually can be adjusted for center frequency and the width of the notch.

- *Hum eliminator:* a filter that has a notch at the power line frequency (60 Hz in the United States) and its harmonics, for use in reducing recorded hum.

Most of these filter types are rated by the frequency at which they attenuate the signal by 3 dB and by their slope versus frequency in decibels per octave. Typical filter slopes are 12, 18, and, more rarely, 24 dB/octave. The notch filter is not usually rated in decibels per octave, because its slopes are extremely steep (see Figure 12.5).

Developing an Ear for Timbre

Perhaps the most important issue in training for mixing is developing an ear for timbre. This is quite complex on program material because it is constantly changing, and so it takes a lot of accumulated impression to hear. One way to short circuit the time needed to learn to hear timbre differences is to listen to equalized pink noise[1] and to match the equalization with a program equalizer at hand. An unknown can be arranged by sending pink noise through a console and using one channel's equalizer to make a particular sound and then, through switching, arrange for a second equalizer to be available for matching the first by ear. Of course, the first equalizer should be covered up and shown only after a "solution" is found. Pink noise has two advantages over program material: it is constant in time and it has all frequencies present. These two combine to simplify the experience as a starting point.

Processes Primarily Affecting the Time Domain

The former processes work practically instantaneously, in real time. Some processes, however, work deliberately to change the time characteristics of the signal, in particular, adding reverberation or deliberate echoes and echo-like effects.

Reverberators

Reverberators are very useful, like equalization, in matching the difference between production sound and ADR

---

[1]Broadband noise having equal energy in each octave across frequency.

FIGURE 12.6   Various filter curves.

FIGURE 12.7   One popular reverberation device and its controller. This is a TC Electronic System 6000. Photo courtesy TC Electronic.

categorized together by human hearing in a process called auditory streaming (see Chapter 2). This is an important feature in layering sound in depth from "in front" of the screen to behind it.

In the early days of filmmaking, shooting stages were made very dead acoustically, in part because theaters were reverberant and any added reverberation in recordings detracted from speech intelligibility when heard in these theaters. In the same film, however, it was more pleasant to hear music with added reverberation, so scoring stages were built with moderately high reverberation times. This dichotomy started a feature that is still present today: Dialog is generally less reverberant than the orchestral score underlying the scene, partly for speech intelligibility and partly for the aesthetics of music listening. If a film's music was recorded in too dead a space, then reverberation was added to the recording by playing it over loudspeakers located in a reverberant room, picking up the reverberation with a microphone, amplifying it, and adding it back to the direct recorded sound in a rerecording console. Thus, the reverb chamber became a part of film sound technique as early as the mid-1930s.

Although these lasted well into the 1970s, the real estate they took up became very valuable and they were not very flexible (one could change the reverberation time only by changing the absorption in the space). Starting in the 1970s, mechanical and then fully electronic reverberators came to dominate the scene. There are many types of reverberators available today, most of which are based on digital electronics. Most reverberators are designed for music enhancement, and have "good sounding" reverberation, but these often do not have an adequate range of reverberation types for film sound because the spaces that need to be synthesized to make scenes sound realistic range from acoustically good to downright bad and from short to long reverberation times. For synthesizing smaller and less reverberant spaces, a *room simulator*, which is a variant on a reverberator using many of the same techniques, may be useful.

After Orson Welles created mass hysteria in the United States with his Halloween radio broadcast of H. G. Wells' story, "War of the Worlds," of a Martian invasion, he was welcomed in Hollywood as a storyteller and, thus, a filmmaker. He brought with him techniques from radio, such as the use of artificial reverberation. In *Citizen Kane*, when Kane enters his large living room, crosses, and sits far distant from his paramour, what is widely known is that a very great depth of field photography was used to keep the two, one in the extreme foreground and the other in the extreme background, in focus. What is less well known is that Kane's distant voice is very reverberant, an unusual thing to do in 1941, because theaters themselves were reverberant, which didn't help intelligibility. However, as a storytelling device, the reverberation added increases the distance between the character, the point of the scene, and the shot.

recordings. They are also used to sweeten music that may have been recorded in too "dry" a space or has been recorded with closely spaced multiple microphones. Another use is to help distinguish among auditory objects; all sound having one reverberant character will be
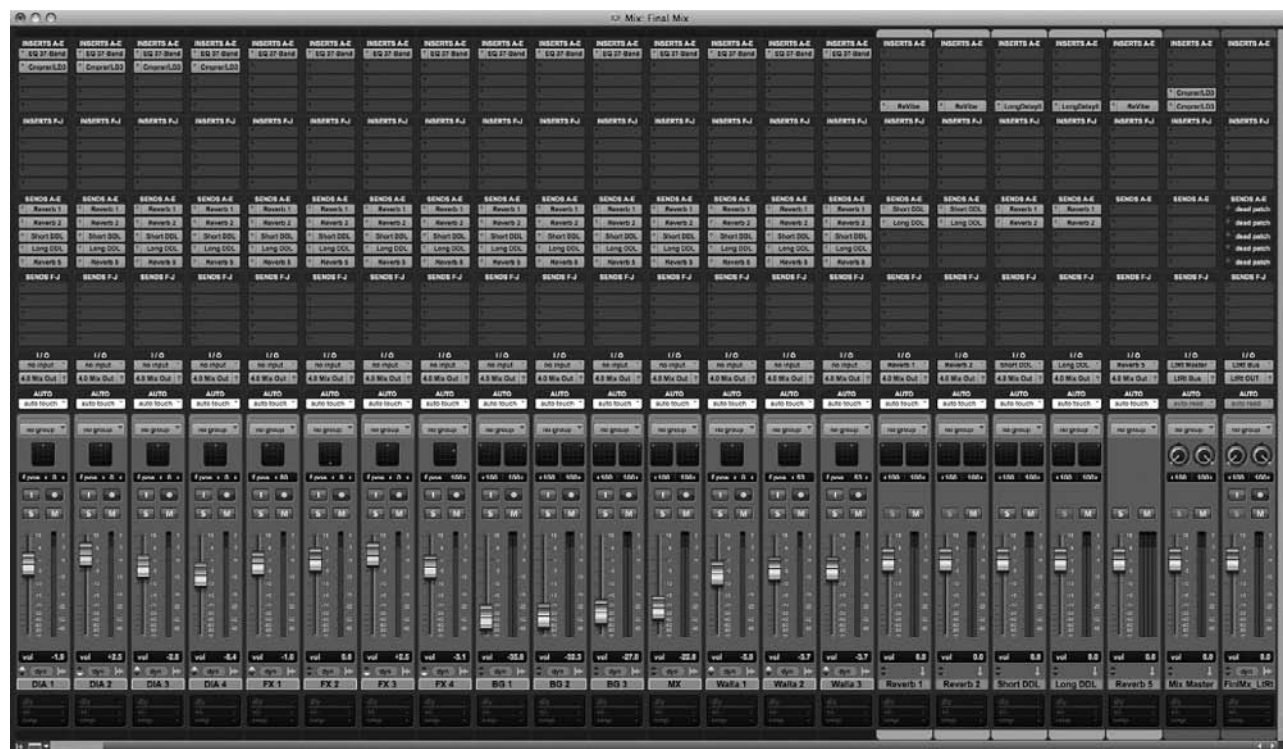
FIGURE 12.8 The mixing screen of a digital audio workstation. The topmost set of boxes represents plug-ins, *processes* associated with each channel. The second set represents signal *configuration*, direct sends to outputs or to buses for additional processing, similar to aux sends. The third set represents automation conditions and the Solo–Mute function. At the bottom come the track faders and designation strips.

Echo Effects

A digital delay line (DDL) is a device that simply delays sound by converting audio into samples and storing the samples in a digital computer memory and then withdrawing the samples at some later time and converting them back into audio. *DDLing*, as this has come to be known, adds a variety of effects depending on the delay time and the relative strengths of the direct sound and the artificial reflection. At relatively short times, between 1 and 20 msec, strong timbral effects, like speaking into a barrel, are heard. This "thickening" of the sound is what makes the distinctively metallic voice of C-3PO in *Star Wars*, for example. Longer time delays and stronger artificial reflections sound progressively more like separate echoes.

FIGURE 12.9 Mix-in-context block diagram.

Pitch Shifters

Pitch shifters are usually digitally based units that take in sound of one frequency range and translate it up or down to a different frequency range. Such pitch shifters are sometimes known by the trade name Harmonizers. They are useful in affecting voices that are meant to be in a range that is different from the actor's natural voice and yet maintain sync and to change the quality of sound effects.

This is an area in which plain old analog tape recorders can outperform digital methods in cases in which the pitch and duration can be altered together and sync considerations do not occur. The method of pitch shifting digitally involves interpolating samples that are not actually present in the original, and this limits the range over which a digital pitch shifter can work and the quality in that range—poor digital pitch shifting sounds like a stutter. An

ordinary analog tape machine equipped with a wide-range speed varier works well as a pitch (and duration) shifter.

> Videographers say that you only have to get the colors of the skin, grass, and sky right to have people "believe" the colors, and any other color can be completely distorted and still work (this really bothers sponsors of commercials, who insist on getting their product colors right). Pitch shift is the audio equivalent: We cannot afford pitch shift on recognizable items, but lots of pitch shift can be used on the less recognizable ones.

This factor alone is enough to keep many old analog tape machines creaking along, at hand for a sound designer to perform major pitch shifts.

Subharmonic Synthesizers

Subharmonic synthesizers are devices that find the fundamental in program material and synthesize subharmonics, as described in Chapter 1. They are useful for adding desired "weight" to effects.

Combination Devices

There are devices on the market that combine two or more of the previously mentioned processes for doing one task, such as equalizing and compressing a vocal track to be heard in a mix with other music tracks. These are given names by their manufacturers that emphasize the purpose for which they are built, such as Vocal Stressor.

CONFIGURATION

Each of the previously mentioned processes may be needed to manipulate a channel during the mixing process, but no practical console has all of the processes present on each and every input channel. Thus, the configuration of the console becomes important for organizing the various processes and to reduce the number of control functions to a (barely) manageable number.

Early Rerecording Consoles

Early film sound consoles had relatively simple signal paths. Each console input channel was equipped with a fader and patch points to insert processing equipment (which at that time included all equalizers, filters, and the like). Processing equipment was patched into input channels as needed. The input channels were summed to produce the output and were sent to a recorder. A loudspeaker monitor system was switchable between the signal sent to the recorder and the return from it. Surprisingly, perhaps, the name used for this function in film rerecording has persisted from the 1930s. It is called PEC/direct switching, for listening from the photo electric cell (which observed the modulated light of an optical recorder, see

Chapter 13) and the direct output of the console. Alternatively this may be called tape/direct or tape/source.

If reverberation was needed, the summed output of the channels was additionally sent to a reverberation chamber, and the amplified output of the chamber's microphone was summed together with the dry sound in an additional summing stage before being sent to the recorder.

Adding Mix in Context

One difficulty with this arrangement is that it is hard to make predubs[2] having all of the necessary fades because the mixers can hear the elements belonging to only one predub at a time. Thus, if a sound-effects premix needs a fade-under to accommodate narration, and only the effects can be heard while making the predub, it is difficult to judge the timing and the amount of the fade that is needed. This is the origin of a technique called *mix in context*, which uses two consoles in effect, although they are typically today in one housing.

The actual mixing for recording is done on the first console, with its output sent to the recorder. This output is also sent to the second console, along with the other existing predubs. The second console is for monitoring purposes only, and little actual mixing is performed on it, because its output is not recorded but rather is sent to the monitor loudspeakers. The second console is usually set for "unity gain" on all of its active inputs, so the predubs are represented in 1:1 level correspondence to each other. By such a setting, the mixing that occurs on the first console is "in context" with the rest of the existing material. If the dialog predub is played through the mix-in-context inputs, then the premix that occurs on the first console can be done with respect to all of the dialog.

Busing

To account for the separation of DM&E, the absolute minimum console must contain three signal buses, one for each part. A bus is an electrical connection that brings together and sums multiple inputs, much like a milk tanker driving among farms, picking up milk and delivering it to the processing plant. Note that once mixed together, we can no longer separate out the individual contributions—summed is summed, and reversing the process is practically impossible. This puts summing at the heart of the sound-mixing process, because it is necessary to combine sources to simplify mixing, but going back on a mix is very difficult, usually requiring work to be redone.

To account for stereo production, each of the three *stems* (DM&E) needs multiple channels to represent at

---

[2]Synonym for premix.

least left (L), center (C), right (R), and surround, with two surround channels (LS, RS) being the norm today, as well as a possible extra low-frequency-only enhancement channel for effects. Each of the tracks for each of the stems is thus recorded, typically to a digital audio workstation, possibly the same one supplying the source tracks or, for more complex mixes, to a separate one. Each of the three sections of major rerecording consoles usually has a minimum of six output buses to produce multichannel outputs capable of being panned anywhere in the sound field: left-, center-, and right-screen channels; left and right surround channels; and the auxiliary low-frequency effects channel (that would not be panned to but rather "hard assigned," with its own fader).

The term "tracks" applies in several ways and may be ambiguous here. Technically, a track is a space on a medium assigned to carry one signal. So a tape recorder capable of recording more than two signals[3] is called a *multitrack recorder*. The term also applies to the individual time-streaming channels of a digital audio workstation or to the overall soundtrack of a film.

The term "channel" refers to a conceptual model of a signal path. We would think of a microphone channel as originating with the microphone and winding up on a track of a multitrack machine. An *input channel* of a console includes all the signal processes from the input to the buses, which may also be called *output channels*. The assignment of an input channel to buses is set by hard switches and/or by panning among the output channels. For example, a given input channel might be assigned to the dialog stem and then panned to center. (Channel is the more global term, allowing for signal processing after the summing accomplished by a bus.)

Patching

Because all facilities are not generally available to all input channels directly, there needs to be a means to *patch* specialized processes into the signal flow within a channel. All large consoles provide a way to do this, usually by way of *patch bays*, with many jacks, permitting insertion points within an input channel, or an output channel, so that processing may be applied to an individual track or to a sum, such as to all the dialog. Alternatively, on digital consoles specialized software programs called plug-ins may be patched into the signal flow. Alternate terminology is *insertion point*, into which processes may be patched.

Each physical console has its own specific rules governing patching. Some general rules that usually apply are as follows:

- Inserting a piece of equipment into a single console input channel involves patching the equipment from a source signal (called *insert send*) to the input of the process and from the output of the process back to the signal path (called *insert return* or *receive*), usually to the point in the chain immediately after the point from which it was originally detoured.
- Patching that connects two outputs together is impermissible: They short circuit each other. To hear simultaneously the effect of two devices, normally they would be placed in series or, in some unusual cases, they might have their outputs summed together in a spare summing bus.
- Patching that forms loops around equipment is not permissible: The signal must always progress, not backtrack. The consequence of forming loops is the potential instability called oscillation, which audibly or inaudibly (ultrasonically) is feedback, much like the acoustic feedback we hear when a public address system has too much gain and it feeds back. This can be at ultrasonic frequencies and overload and blow up equipment, especially high-frequency loudspeakers.

Panning

A fundamental part of configuration in a multichannel rerecording console is panning. *Pan pots* are devices that place sound among the channels described earlier: L, C, R, LS, and RS. There may be several knobs, such as three panning among L, C, R; front/back; and LS/RS. A joystick somewhat like that used with video games may also be used. Each input channel of the console is typically equipped with a panning section or, at the very least, with switches that hard assign an input channel to an output bus. In some cases specialized panners such as joysticks are provided on a console separate from the input channel slices and without a particular channel assignment; in these cases they are designed to be patched in as needed. In digital consoles fitted with joysticks, the stick is a controller that can be switched to take over the pan function for a slice to make the adjustment of the panning control more visceral. When the setting has been made and the moves, if any, recorded, the joystick can be switched to another channel. Some digital consoles use trackballs or graphics tablets instead of a joystick.

Many source sounds are monaural, single-channel recordings. When rerecording these sounds in a multichannel production, there are rules governing placement. These arise from aesthetic considerations based on psychoacoustics and practical considerations based on the nature of film production and viewing. One of the psychoacoustic issues was described in Chapter 2 in reference to dialog panning under Speech for Film and Television.

---

[3]Mono recorders are one-track; stereo are two-track.
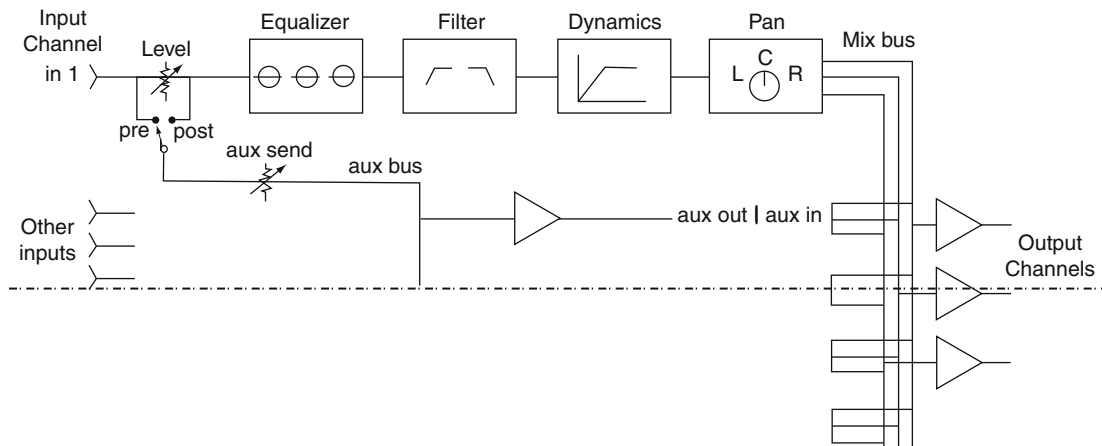
FIGURE 12.10   A single-line drawing for a simple rerecording console. A main signal path routes the input channels to the output channels by way of various signal processors, and the auxiliary system provides a means to send a proportion of the signal to an output processor, such as a reverberator, and return it to the main mix bus.

Some of these rules follow:

- Dialog in ongoing conversations is usually either centered or kept close to center because sound edits that match picture edits cause the sound to noticeably "jump" around the screen.
- Off-screen lines are usually panned hard left or right, as makes sense; panning them to the surrounds in the auditorium breaks the "box" of the frameline too much.
- Lines that are isolated from others in time may be panned.
- Foley is routinely recorded in mono and panned into place to match.
- Ambience is most often from original stereo recordings that are placed from two or more source channels into two or more output channels. The principal aesthetic concern of ambience panning is whether to include the surrounds, depending on whether the audience is supposed to be sharing the space portrayed on the screen.
- Cut effects could be either mono or stereo source recordings. Mono cut effects are panned into position to match their on-screen image. Stereo effects are usually two-channel recordings that are also panned in place, such as into left and center for a stereo effect located left of center. The danger of using this technique in movies is that only people seated exactly on the centerline of the room hear the center of the sound image in front of them. Because of the precedence effect, strong in theaters because of their size, anyone seated left of center will hear a left–right stereo sound biased to the left, and vice versa. It is better to derive a center using one of a number of techniques, including decoding with a professional LT/RT decoder that produces a center output when the two channels have content that is in phase.

Auxiliary and Cue Buses

So far the only kind of bus discussed has been an *output bus*. Destined ultimately to reach a loudspeaker monitoring channel such as left, center, or right, such buses may be further separated by discipline, such as a left sound-effects bus or channel, as we have seen. The idea of these output buses does not, however, cover all possible purposes for which we may need to combine channels. For example, we may wish to send signals from multiple input channels of the console to the same reverberator, especially for sounds that share the same space, and because reverberators are generally more expensive than other processing equipment, it is useful to share resources. For this reason, *auxiliary buses* have been developed. Auxiliary buses have two primary purposes, effects send and cue send. Effects-send auxiliary buses are used for the purpose already described, to gather signals from multiple inputs and to deliver them to a processor. Effects-return modules, similar to input modules,[4] then direct the return signal, such as reverberation, to the main buses. See Fig. 12.10.

The second purpose is cue send. Cueing, in general, in film and television production, means any activity meant to alert an actor, newsperson, or musician to a timed event, so that he or she can start on time and perhaps even maintain time. In a multitrack music studio, it has a more specific meaning. There, the first track recorded is often a *click track*, the output of an electronic metronome, used for keeping musicians on time. Then, during recording the cue-send feature of the console is used on the click track to send the clicks to headphones on the musicians. Because the musicians cannot play well without hearing

---

[4]But, it is hoped, lacking, above all else, effects sends themselves, because if an effects-return module could send to an effects unit, an unstable feedback loop would probably be formed.

themselves, some cue-send level is added from their own instrument tracks so that they can hear themselves. Because different parts of the orchestra may wish to hear different balances among instruments (usually with themselves louder, of course), there may well be multiple cue-send buses so that separate mixes can be made for different groups of musicians.

AUTOMATION

Increasingly, mixes have gotten so complex and are under such time pressure that it is impossible to perform the mix by operating the controls in real time, even with more than one mixer doing the job. Although some complex mixing can be done during premixing because of the multistage nature of film and video sound rerecording, nevertheless it is important to have sophisticated control over all of the premixes at the final mix because this is the stage at which the producer, director, and others become most involved.

There are several levels of automation possible. The most fundamental automation is over level control, because this is by far the most active part of the console typically. Moving-fader automation has come to dominate film and television mixing. In these systems, the rerecording mixer performs moves on the faders during one pass of the film, and a computer memorizes the moves and reperforms them during subsequent passes. On these later passes, more faders can be brought into play and the computer continues to move already established faders. In this way, one person can do an extremely complicated mix, by adding more and more sophistication over time.

Fader automation is accompanied by automation of mutes, as described previously. Consoles also exist wherein the control surface is not directly connected to the circuits it is controlling, but rather operates as a user interface, sending all control functions as messages to a rack of equipment located in a machine room. These consoles may use "rotary shaft encoders" for their control functions, which means that the setting is no longer tied to the physical setting of the controls but rather to changes in the controls. Full automation is possible, with all the control functions affected. On fully automated digital consoles, it is possible to work on a program without committing it to being recorded, because the console will continue to reproduce all the fader moves, equalization changes, etc., in all subsequent passes through the material. Once the sound of the program is finalized, then it can be recorded out to a medium.

The move toward full automation, plug-in parameters to digital audio workstations, and outboard gear such as reverberators and noise reduction units, although not accomplished 100 percent today, is on its way. Once the transition has been made to full automation of everything, and the systems become 100 percent reliable, then film mixing will consist of recording automation and recording

FIGURE 12.11 Paddle switches on this box control punch-in/punch-out recording.

out when needed. One problem with this is in coming back to the mix months or years later and finding that the software has moved on and now makes errors in playing legacy material, so complete systems, including operating systems, application software, plug-ins, etc., have to be maintained to be able to come back to a mix months or years later.

Before this nirvana of full automation with lifelong backup is accomplished, film dubbing currently uses a system developed on mag film mixing in the 1960s, and still well known, punch-in/punch-out recording.

PUNCH-IN/PUNCH-OUT (INSERT) RECORDING

Punch-in recording is an important concept in postproduction, because it allows updating mixes without remixing whole reels. However, its utility is decreasing in an era in which console and outboard gear automation is capable of reproducing all the changes that a mixer makes and can reproduce them all while be recorded off to a stem recorder. Although still widely used, punch-in/punch-out recording, alternatively called insert recording or update recording, may be seen today as something of a throwback to the past. It is done because some mixers do not yet rely completely on console automation. Punch-in recording relies on the ability of a mixer to achieve console settings that are identical to those used during a prior mixdown. This is ensured by switching back and forth between the source (the mixdown from units produced by the console) and playback off the recording (PEC/direct switching).

Once throwing the direct-film switch back and forth reveals no sonic difference, the mixer can punch in, that is, activate timed erase and record circuitry so that a new recording is begun seamlessly. Then the mixer can proceed to remix a portion of a reel and, coming once again to a place where he or she can match the console settings with the original, can *punch out*, yielding a newly mixed section without abrupt changes at the transition between the new and the existing mixes. This process saves a great deal of time and money in rerecording.

Punch-in recording is also used in music studios working on multitrack recorders or DAWs. Let us say that all tracks are all right, except one, which contains a wrong note. What is done is to punch in and out just before and after the note, having the musician play the right note at the right time. The musician is "cued" by listening to sound from the other tracks, and he or she performs continuously (because it is hard to play just one note at the right time). The engineer punches in and out at the right moments to substitute a new performance of the one note.

One difficulty occurs in both of these cases due to the use of three-head mag or tape machines. With separate record and playback heads, if one listens to playback and plays in sync with the music, the musician will be in sync with the sound at the playback head, but not at the record head where sync is needed. This problem was solved early in the history of multitrack recording when guitarist Les Paul devised a method to listen to tracks *played back* from the other channels of the *record* head for cueing purposes. Called *sel-sync, sel-rep*, or by other trade names, this method of recording ensures sync is maintained despite the three-head configuration. In TV sweetening it is commonplace to use machines in this mode, so that all the signals are always coming from or going to the record head, to prevent any sync errors.

This page intentionally left blank

From Print Masters to Exploitation

INTRODUCTION

Print masters are the final output of the film rerecording process and are the delivered movie soundtrack in the eyes of the producer and studio, because subsequent processes are not aesthetic, but manufacturing ones. Following various manufacturing steps to be described, motion-picture release prints, television masters, and video disks and tapes are made. Then, the prints or videos are shown on theatrical, commercial, home theater, or computer systems for the enjoyment of the audience.

Audio master is the more common term in television and video mixing for the same thing. An *edit master* is the final prepared picture and sound master for a television program, with the audio from the audio master laid back to it.

The word "exploitation" seems to most people like an evil thing. In the topsy–turvy world of the film business, it is considered a good thing. One "exploits" a film property through releasing it progressively into movie theaters and then for the home on DVDs and for download, on pay-television services, and finally on free television. Programs made originally for television see a first and perhaps more prime time runs, then potentially go to syndication, then onto DVD as well, in a similar exploitation path.

PRINT MASTER TYPES

Print masters are made for each medium of release and possibly in multiple languages, so there are many different print masters of a theatrical feature. We will examine print masters in the order in which they are typically made for a large theatrical release. Print masters may be recorded on disk drives, read–write optical computer discs, mag film, or digital multitrack. Permanence is important, as print masters represent the value of the product to the producer and studio, but the problem of what constitutes a permanent digital medium has not been solved. Until it is, formats for storage of the print masters proliferate.

Each release medium has two primary factors that together determine the parameters of the associated print master: the number of available audio channels and the

dynamic range capacity of those channels. Each print master must be tailored to the capability of the specific target medium. There is simply no point in sending to the analog optical recording camera a print master that exceeds the dynamic range capability of the associated optical track. The best way to proceed is to have the original mixers of a show prepare separate print masters that respect the limits of the individual release formats because they are in the best position to make such compromises.

Another factor in making print masters for both film and television shows involves foreign-language distribution. For dubbing purposes, it is commonplace to supply M&E masters, containing mixed music and effects, without dialog. The dialog is usually supplied as a separate track for translation and synchronization purposes at the foreign-language dubbing site. In some instances, the foreign-language pre-mixed dialog stem is returned to the original mixers to make new print masters in the various foreign languages; in other cases, the dubbing site prepares the print master.

Audio master work on television entertainment programming parallels that of film, with some simplification due to the faster postproduction schedules. Usually a television show will have one primary master, the English language composite, and one secondary master, an M&E, if foreign-language dubbing is contemplated.

All masters must carry a 2 pop. This is one frame of 1-kHz tone at reference level recorded on all the tracks of the original print or audio master, and in *edit sync*[1] with the 2 frame of the SMPTE universal leader (time countdown, more often used on television) or the 3 frame of the Academy or SMPTE projection leader (footage countdown, for film). Both of these are 2 sec before the first frame of picture and sound. Note that although the picture may start in black, or the sound may start in silence, the pop is 2 sec before the first frame intended to be projected and heard. If the filmmaker wants a long slow fade-in of both picture and sound, it should start fading in at 0 feet

---

[1]Edit sync, also called "level sync," means that it occurs at the same time. This is universally used throughout postproduction, but editors should know that there is another kind of sync, projection sync, in which the sound is displaced along motion picture prints from its associated picture frame to allow the projector to show the picture and play back the sound at separate positions.

or time code, 2 sec after the 2 pop. The 2 pop is necessary to synchronize the soundtrack negative with the picture negative, or to the edit master for video described below. A tail pop is also very useful, although often forgotten. It can be used to sort out various sync issues.

Print Masters for Various Digital Formats

These are the first to be made, because they involve little compromise to headroom or to the number of discrete audio channels. The preparation of such a print master is therefore relatively simple because little or no tailoring to the specific requirements of the medium is needed. This is because the dubbing stage console is aligned to the SMPTE reference level of − 20 dB F.S., with 20 dB of headroom, and all the digital sound systems have 20 dB of headroom.

There are three prominent digital sound formats for film in practice today, after some initial shakeout. They are the Dolby Digital, DTS Digital Cinema, and Sony Dynamic Digital Sound (SDDS) systems. These are fairly similar in their print-master requirements, except in the number of channels. The minimum number of channels for digital sound on film was determined by a SMPTE standards committee in 1987 to be 5.1. *Five point one* channel sound, as it is called, includes left, center, right, left surround, right surround, and low-frequency enhancement channels. Dolby Digital and most DTS installations provide 5.1 channels. The SDDS system and some rare DTS ones add the potential for two more screen channels, for a total of 7.1: left, left center, center, right center, right, left surround, right surround, and low-frequency enhancement. Many installations of SDDS, however, are simplified to 5.1 channels. In all cases, the "0.1" channel designation means that a separate low-frequency channel has only a small bandwidth requirement compared with full-frequency range channels. See Fig. 13.1 for the positions on film release prints for the various sound tracks.

Although the 5.1-channel SMPTE requirement was for sound-on-film systems, DTS Digital Cinema carries only a custom time code track and not digital audio on the film and uses a CD-ROM follower with low-bitrate coded audio on it to meet the requirements—it is a double system. Missing footage from the print is cleverly handled by the custom time code system, which reestablishes sync quickly.

The amount of headroom available in each channel of the release print for all these formats is 20 dB, which is the same headroom that is available on digital stems. However, note that adding the multiple final mix stems together—dialog, music, and effects—could add up to more than the headroom available on the print master because of summing. This occurs when the peaks in one stem coincide with the peaks in either one or both of the
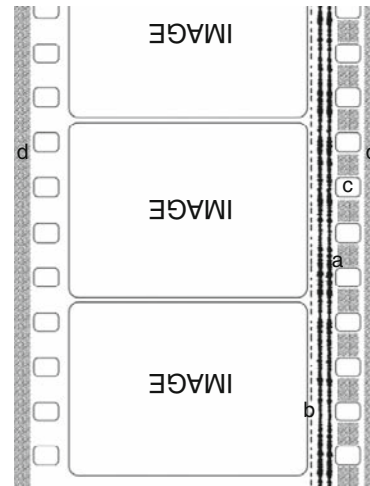
FIGURE 13.1   A representation of a release print showing the relative position for four types of soundtracks: (a) conventional stereo variable area in the standard analog soundtrack area next to the picture, (b) DTS time code located between the picture and the conventional soundtrack areas for synchronization of an external disc, (c) Dolby Digital between the perforations on the soundtrack side of the film, and (d) Sony Digital Dynamic Sound located outside the perforations on both sides of the film. Frame lines of a Cinemascope picture are shown for reference. A positive release print is the inverse of the negatives.

other stems. The result is potential overload of the digital print master, with attendant hard clipping. Some peak limiting may thus need to be done, usually in audibly benign amounts, to "fit" the dynamic range of the mix stems together into one print master channel and prevent audible distortion.

The preparation of print masters for these three formats thus involves what is more or less a transfer operation, mixing together the final mix stems to a 5.1- or 7.1-channel master, as needed. In the postproduction schedule, this usually represents a small fraction of the mix time, at the very end.

Low-Bitrate Audio

If you have 5.1 channels of digital audio coded by linear PCM coding, with 20-bit word length and at a 48-kHz sample rate, they require a data rate of 4.8 Mbps as calculated in Chapter 3. The space available on the film can accommodate only between 584 kbps (in the space between the perforations on one side of the film in the case of Dolby Digital) and roughly six times as much (outside the perforations on both sides of the film in the case of SDDS and using a smaller bit size). Furthermore, besides the digital audio, strong error coding to protect the audio bits is needed because there will surely be dirt and scratches on a motion-picture print run in exhibition, and the audio payload comes down to 320 kbps. Thus a means to reduce the number of bits per second by a factor

of about 12:1 is needed. Although normally we might think that throwing away $^{11}/_{12}$ of something might cause it severe damage, a clever solution was worked out. By using psychoacoustic knowledge, particularly of frequency and temporal masking, systems were devised to transmit just the audio signal that is relevant to human listeners and to discard information that would be masked, achieving great "coding gain," equivalent to bitrate reduction. Although such systems are not necessarily perfectly transparent all the time, they are on most program material, and the audible "cost" in terms of audio quality is relatively low. The three main systems, only two of which store the digital sound on the film, compete based on different trade-offs among bitrate, quality of the bitrate reduction coding, unrecoverable error rate, and flexibility.

Print Masters for Analog Soundtracks

Whereas digital playback equipment is at near saturation in first-run theaters around the world, conventional analog optical soundtracks are always also recorded on all theatrical release prints, because *all* of the world's theatrical 35mm film projectors can play them. Also, the analog track serves as a backup in case the digital track should become unreadable. There are some severe constraints in making conventional optical soundtracks that must be factored into the print-mastering process, however. The two primary factors are limitations in the number of recorded channels and limitations in dynamic range.

Because of size limitations of the soundtrack area on film (it occupies only about one-tenth the area of the picture), practically speaking there is room for only two physical audio tracks to be recorded within the area assigned to the soundtrack on the 35mm medium. To carry the multichannel presentation prepared in postproduction through this limited capacity medium, a method of encoding four audio channels worth of information into two tracks was developed. Called generically *4:2:4 matrix technology*, this process has several trade names, including Dolby Stereo. Four channels of content can be stored within two audio tracks with some compromise by use of *matrix* technology, and this is one process that occurs in preparing stereo analog soundtrack masters. The name given to print masters in this format is *LT RT*. The T stands for "total," meaning a left and a right track carrying information that may be decoded into left, center, right, and surround.

The other primary limiting factor is that the headroom of an analog optical soundtrack is quite limited compared with the postproduction generations that have come before it. Today's analog optical soundtracks employ Dolby Spectral Recording. This involves the application of the signal processing system around the soundtrack channel, with the SR encoder being part of the preparation of the soundtrack at the laboratory and the decoder being located in the theater's equipment. This system as used has 9 dB of headroom in the midrange, with somewhat more at low frequencies. It also produces 23 dB of noise reduction compared to unprocessed analog tracks, a very large amount, but the 9 dB of headroom is a large limitation for some program material compared to the 20 dB of digital formats. On the other hand, there are programs that fit completely within the 9 dB capability, and these are less affected by the limitations of the analog system. Although the term *Dolby SR* applies to the dynamic range processing system, the 4:2:4 matrix technology is also used on these prints.

The job of the postproduction mixer in preparing print masters for each of these release media is to do the best job possible of "fitting" the often wider volume range of the final mix stems into the capacity of the media. This can range from being simple to nearly impossible, depending, to a large extent, on the original soundtrack. *Driving Miss Daisy* is easier to fit into a medium with less dynamic range than is *T2 Judgment Day*.

Specialized multiband limiters and clippers are often used to get the best representation of the wide dynamic range master into the print master.

> Multiband clippers are called *containers*. The idea behind a multiband clipper is that although clipping of audio signals is certainly a bad distortion, constraining the distortion components to the narrow frequency region causing the clipping produces a greater likelihood of the distortion being masked. These devices consist of a series of bandpass filters abutting one another in frequency, followed by individual clipping stages for each band, a second series of bandpass filters, and an output summer. The whole arrangement keeps frequency response flat, while containing clipping artifacts at one frequency to the frequency band around it, thus not producing the higher order harmonics, which would be more audible.

For the digital formats it is typical to record the print master onto a digital recorder in custom formats supplied by the companies involved. Some of the formats, such as Dolby's use of magneto-optical recordings for print masters, contain both multichannel and 2-channel LT RT masters on the same medium.

Interestingly, this brings up the issue of what delivery to a studio means. Because digital delivery formats are a constantly changing environment, just what constitutes delivery to a studio becomes an item of more than passing interest as it becomes a contractual matter. What is needed is a permanent medium that is transparent for the underlying bits, with a very long shelf life, and not too costly. This medium seems to be a holy grail for archivists, unfound today. My personal opinion is that the industry should develop its own methods, rather than relying on the transient nature of computer formats and companies, and that the emerging standard would be photographic rather than magnetic, stored on an impervious base, with large

bits, and with codes that will be readable for 500 years. Silver-on-glass photo negatives have already been around for over 100 years, so could be taken as the basis for a new digital medium.

Other Types of Delivered Masters for Film Uses

To produce foreign-language dubs, it is customary to make intermediate masters that are very much like the English-language print masters, minus the dialog. By mixing these together with dialog premixes in the various languages, final print masters can be made in a variety of languages for various types of prints (Table 13.1).

With multiple languages, each with its own print master requirements, the number of "masters" can grow to be very large. Usually, the postproduction house prepares one or more of the types shown in Table 13.2 for these purposes.

Digital Cinema

Digital cinema servers and projection are finally becoming important commercially, with a market share approaching 20 percent of cinemas at this writing. Sound for digital cinema is based on SMPTE and Digital Cinema Initiative (DCI) standards. The most commonly used of these today is 5.1-channel 48-kHz sampled 24-bit resolution audio. Thus print mastering for digital cinema is the same as for digital film, without the bitrate-reduction schemes needed to get audio onto release prints. The most common method of distribution is by shipping hard disks to theaters. Elaborate protection ensures that the content cannot be stolen.

TABLE 13.2 Other Types of Delivered Masters

Name	Purpose	Number of tracks
Multichannel M&E	Preparation of digital foreign-language dubs	5.1–7.1
M&E LT RT	Preparation of stereo foreign-language dubs	2[a]
DM&E	Preparation of mono foreign-language dubs	3
M&E	Preparation of mono foreign-language dubs	1
Dialog	Translation and synchronization purposes in the preparation of foreign-language dubs	1

[a]If supplied on mag film, typically recorded on tracks 1 and 2 of the 3-track format.

Masters for Video Release

Once theatrical print master types have been made, sound masters are typically prepared for the video market in one of several formats:

- 5.1 channel for multichannel DVD, with the same general characteristics as the theatrical print master (but see following), recorded as 6-tracks.
- LT RT for two-channel PCM, used for delivery to many video services not yet equipped for 5.1 and with headroom limitations such as 10 dB. This includes much cable television and satellite delivery.
- Mono, used most often for historical or documentary works, which may be supplied as a DM&E, recorded as a 3-track.

TABLE 13.1 Print Master Types and Their Characteristics

Name	Print technology	No. of audio tracks on print master and print	No. of audio channels delivered	Midfrequency headroom[a] of target medium
Dolby Digital	Digital sound on film	5.1	5.1	20 dB
Dolby Digital Surround EX	Digital sound on film	5.1	6.1	20 dB
DTS	Digital sound on synchronized disk	5.1	5.1	20 dB
DTS ES	Digital sound on synchronized disk	6.1	6.1	20 dB
SDDS	Digital sound on film	5.1–7.1	5.1–7.1	20 dB
SR LT RT	Analog optical with 4:2:4 matrix stereo[b]	2	4	9 dB

[a]See http://booksite.focalpress.com/Holman/SoundFilmTV/ for more information on headroom versus frequency and signal-to-noise ratios of the various media.
[b]The LT RT print master tracks are recorded on the same MO media as the Dolby Digital tracks and are destined for the analog optical tracks on release prints.

For the 5.1-channel market, although an exact copy of the theatrical print master will do, in recent years more manipulation has been done for home video. At the dawn of the DVD era, between 1997 and circa 2003, there was so much vault material to transfer in a rapidly growing environment that the job became making audibly perfect copies of the theatrical master. Home sound systems could employ techniques described later to make the transition from theatrical to home environments. In the world of 2008, however, once the rapid growth rate of DVD had abated and even fallen back some, 57 percent of the retail dollar volume of sales was in DVDs and Blu-ray and only 43 percent theatrical. Furthermore, it is likely that the studios make a considerably greater percentage on video sales than they do on theatrical sales, so the balance is tipped even more in favor of video. For this reason, studios have established home video departments to remix movies for video release, using typically only the 5.1 theatrical print master. The methods employed vary from studio to studio, including equalization, low-level compression, and so forth. Regrettably this has meant that today there is actually less consistency in the end product than there was 5–10 years ago and less likelihood that the original filmmakers are involved. In some notable cases though, the original rerecording mixer is allowed to make the DVD release master; this occurred on the movie musical *Chicago*, for instance, and this is a far better procedure, as the original rerecording mixer has had extensive contact with the director and the vision of the film mix.

For the LT RT matrix market, which is shrinking, a copy of the theatrical print master will serve, but the limitation on dynamic range used for the theatrical master may be misplaced. For example, DVD digital tracks have 20 dB of headroom, and if the LT RT has been limited for optical soundtracks having 9 dB of headroom, 11 dB of headroom of DVD's two-channel capacity is unavailable to be used. It is not uncommon then to make a separate LT RT for video release, using a full 20 dB of headroom.

Not all video-release market channels can handle this 20 dB headroom though, so on transfer to a digital master videotape with four digital audio channels, two pairs of LT RT soundtracks are sometimes recorded. One pair uses the full 20 dB headroom and the other just 10 dB headroom, accomplished with a combination of limiting and reducing the level, each typically by 5 dB. In this way, the producer winds up with a "one size fits all" video master with two stereo soundtracks, one with a large and one with a restricted dynamic range. Thus far this discussion has considered the primary release media with differing headroom, such as theatrical release, pay-per-view, cable, satellite broadcast, conventional broadcast, etc., and the master described with two pairs of recordings can be used for all of these purposes (one or the other pair will service any of these markets). Subsequent exploitation of the work in ancillary releases, such as airline viewing, requires even less dynamic range than any of the types already discussed, so compression and limiting are liberally applied to specialized copies made for these purposes and possibly a new print master made from the final mix stems emphasizing the dialog.

One of the largest problems in film and tape dubbing—the coexistence of media separated by large differences in headroom capability—means that it is simply not possible to capture the full dynamic range of one medium in another for many instances. For example, to copy a digital video master to a U-Matic cassette could mean that as much as 20 dB of level reduction and limiting would be needed, a severe amount. This alone is a reason for anyone using such old analog video formats to retire them.

Television Masters

Typically LT RT or mono Print Masters are copied precisely onto the master videotape, which already has the video program recorded on it in the process called *lay back*. An M&E LT RT may also be prepared for foreign-language dubs and supplied in one of a variety of formats.

SOUND NEGATIVES

For 35mm and 16mm film, a soundtrack negative is photographed on a special sound camera, processed, and then used to print the release prints. Unlike picture prints, no intermediate stages can be used, so if many prints are to be made, multiple soundtrack negatives will be prepared. The reason that a separate sound negative is needed from the picture negative is that the sound negative requires very high contrast, and is monochrome, whereas the picture needs to have "tone," all the region from dark to light displayed correctly, and color. The soundtrack negative may contain one or more of the different types of tracks, distinguished by their position on the negative and subsequent print. It is possible for one 35mm negative to contain a conventional analog optical soundtrack and all three of the competing digital film sound formats. This is called a *quad format negative*.

Conventional analog soundtracks today are of the variable area type, in which what is varied to carry the sound waveform is the width of clear area within an otherwise dark soundtrack on the print. In the sound camera, a light source and a condensing lens system concentrates light onto devices that vary the width of a beam of light in accordance with an applied signal. These light variations are picked up in the projector by a photovoltaic cell[2] located just behind the film from the light source. The area

---

[2]The same type of device as a solar cell used to make electricity from sunlight, although much smaller.

of each recording is symmetrical about its centerline. This is called *bilateral recording*, and its use reduces distortion caused by uneven illumination across the width of the soundtrack. In 35mm use, the two tracks needed for recording the LT RT signals are placed side by side within the area devoted to the original mono soundtrack, an arrangement called *stereo variable area*, or SVA. In addition to the conventional analog soundtrack, one, two, or three of the possible digital recording formats may also be recorded by the sound camera in one or several passes.

Quality control of soundtracks starts with correct exposure and processing of both the soundtrack negative and the release print. If any of these four stages is incorrect, then sibilant distortion (distorted ess sounds) is usually the first problem heard with analog tracks, whereas dropouts or no sound at all could result with the digital tracks.

> Traditionally analog soundtracks have been "redeveloped" in the soundtrack area so that they exhibit high contrast, especially in the infrared, where the incandescent exciter lamp bulbs used have their highest output. This meant that during processing of the prints, the film came up out of the chemical baths and had a wheel applying chemicals only to the soundtrack area to turn the color dye image into a black-and-white one using silver. Environmental concerns over the effluent of film processing due to this process led the industry to make a massive change-over from silver tracks to dye tracks in two stages over the period of the late 1990s to the middle 2000s. When completed, the exposed areas of the analog track are no longer black but cyan, and the exciter lamps necessary to illuminate the track had to be changed from incandescent bulbs to red LEDs. If played with ordinary white-light bulbs (with the pickup solar cells most sensitive in the infrared) the output will be extremely low. Thus most of the world's projectors have been converted to red-light readers at this point.

THEATER AND DUBBING STAGE SOUND SYSTEMS

Motion-picture theater sound systems consist of two basic parts called the A and B chains. The *A chain* consists of the soundtrack recording on the print itself and all of the equipment needed to recover audio from the print and process it to the point at which it may be interchanged with signals from other formats. The sum total of the recording plus processing together constitutes a film sound format. In this way a format is like the combination of a compact disc and a player, which produces a standardized output but no sound by itself. The end of the A chain is just after a switch that selects the format of the soundtrack to be played, such as stereo variable area or digital.

The *B chain* consists of the rest of the sound-reproduction channel, from the volume control through to the room acoustics of the auditorium, including equalizers, power amplifiers, and loudspeakers. Note that although a theater may switch sources by changing the A chain that is in use at any given time, the B chain is the part of the

theater sound system that remains constant. In this way it is like the rest of a home stereo used with the compact disc and player given in the above-mentioned example.

A-Chain and B-Chain Components

The precise elements contained in the A chain depend on the format and are shown in Table 13.3 and Figure 13.2. The table is arranged from top to bottom in the order the signal is processed, and the bottom of the table represents the output to the B chain.

The B chain consists of the following elements:

1. Multichannel volume control;
2. Room equalization and level setting to set the sound system to frequency response and level standards;
3. Power amplifiers;
4. Loudspeakers;
5. Screen, which exhibits a high-frequency loss and also some spreading of high frequencies;
6. Room acoustics of the theater.

The combination of a selected A-chain format and the B chain produces the final overall impression of the soundtrack to listeners. Each of these chains is standardized to promote interchangeability from theater to theater and from print to print. The idea is that once the production has left the filmmaker, who approves it on the dubbing stage, all the following downstream processes are technical ones, meant to reproduce the approved sound as it was made. Another way to look at this idea is that, as a filmmaker, if you want to record with two tin cans and a string to get an effect in *production*, that's perfectly all right, but once you buy off on that on a dub stage, then it is everyone else's job downstream to reproduce that sound: their job is *reproduction*.

A story told by Warren Beatty about the London premiere of *Bonnie and Clyde* illustrates the point. After the screening he headed for the projection booth, because the gunshots, especially of the shoot 'em up ending the lives of the lead characters and the movie, seemed muted. The projectionist, upon seeing the star of the movie he'd just shown (and probably not knowing that Beatty was also its producer), said to him that he was glad he'd been able to help out, because the movie was mixed so badly, by remixing it on the fader in the booth! Take this as a salient point: the last creative job in filmmaking is *not* projection. However, it is an important job. The projectionist for the New York press screening of *2001: A Space Odyssey* was Stanley Kubrick.

Theater Sound Systems

Sound systems in theaters range from the simplest mono system with one loudspeaker centered on the picture through ones equipped with matrix LT RT decoders playing

TABLE 13.3 A-Chain Format Devices

Trade or common name	Dolby SR	Dolby Digital	DTS Digital Cinema	SDDS
On release print	Stereo variable area optical soundtrack with SR-type signal processing and 4:2:4 matrix encoding	AC-3-coded 5.1-channel sound recorded between the perforations on the soundtrack side of the film; variant is Dolby Digital Surround EX, which adds a matrix process to the surround channels to derive 3 or 4 surround channels from 2	Custom time code recorded between the analog soundtrack area and the picture. The DTS EX variant provides 6.1 channels.	ATRAC-coded 5.1- or 7.1-channel sound recorded outside the perforations on both sides of the film.
On the projector	Optical sound head with stereo pickup, 26 frames ahead of (below) the picture gate	Proprietary digital sound head, mostly near the analog reader today, the position called a "basement" reader	Proprietary time code pickup head placed anywhere in film path	Proprietary digital sound head, placed above the projection gate in the "penthouse" position
A-chain decoding	Stereo optical preamplifier	Proprietary electronics	Proprietary time code electronics with CD-ROM follower using APT-X bitrate reduction	Proprietary electronics
+	Dolby SR decoder			
+	4:2:4 matrix decoder			

FIGURE 13.2 The A and B chains of a motion-picture theater sound system. The A chain consists of the print, the projector and its sound head(s), the electronic decoding of the format, and the selection of the format. The B chain constitutes that part of a theater sound system that is common for all of the various formats, from the volume control, progressing through room and speaker equalization, power amplifiers, loudspeakers, and the effects of the motion-picture screen and room acoustics of the theater.

over left, center, right, and surround array loudspeakers to more than 100,000 installed 5.1-channel systems (left, center, right, left surround, right surround, and subwoofer). More than 5000 theaters are also equipped to split the surround channels into three, left surround, right surround, and back surround. A few specialized theaters also add two intermediate front channels, left center and right center.

Note that no motion-picture theater sound system uses two-channel stereo played directly. Two-channel stereo was a simplification of foregoing stereo systems for the home market, introduced in the late 1950s. It is nearly useless as a theatrical format because of the precedence effect described in Chapter 2, wherein the earliest arriving sound localizes the sound. So those off the centerline get centered material in a two-channel system from the closer loudspeaker.

Dubbing stage and theater sound systems have traditionally been calibrated to a reference level of 85 dB SPL.[3] With 20 dB headroom standard, the maximum undistorted sound pressure level of each of the five main channels is 105 dB SPL. The low-frequency enhancement channel, however, is calibrated to a higher reference level, yielding a maximum undistorted sound pressure level of 115 dB SPL in this channel.

> The reason that the low-frequency headroom is made greater is to provide more equal *perceived* headroom across frequency, because more low-frequency level is required to sound equally as loud as a more midrange sound. This is in accordance with the equal loudness contours discussed in Chapter 2.
>
> Also, note that the headroom here is given one channel at a time. With all channels operating, the instantaneous peak sound pressure level could approach 120 dB SPL.
>
> A picture such as *Forrest Gump* contains both high- and low-level passages, well exercising the dynamic range of the medium. Played at the reference sound pressure level (the standard theater fader setting), the average SPL over the whole length of the film measures 80 dB SPL, with an A-weighted, fast-reading sound-level meter.[4]

Today it is recognized that an identical sound pressure level played in a smaller room sounds louder, because of psychoacoustic scaling effects. Thus the reference levels in Table 13.4 have been standardized to account for this variation.[5]

Theater Acoustics

Motion-picture theaters are unlike other large venues as they are meant as spaces for *re*production of a soundtrack

---

TABLE 13.4 Reference Sound Pressure Level for Various Room Volumes

Room volume in cubic feet	SPL in dB re: 20 μN/m^2
>20,000	85
10,000–19,999	82
5,000–9,999	80
1,500–4,999	78
<1,499	76

---

[3]With a C-weighted, slow-reading sound-level meter, for an audio level of −20 dB full scale.

[4]Although this measure is different from that used for calibration, it is common practice to measure program levels with A weighting to better account for the equal loudness contours. A weighting roughly reflects the midlevel frequency response of hearing. The fast reading is usually used with program material, and the slow reading is for setting the calibrated level on pink noise.

[5]ATSC Standard RP-A/85. See http://www.atsc.org/standards/practices .php.

rather than for *pro*duction of sound. A concert hall plays an important role in the sound of an orchestra, but if a cinema added its own sound to that of the film, then all scenes would have the same added sound, and that would not make sense. In other words, cinemas should be acoustically neutral so that the filmmaker can place the listener in the correct setting. This basically boils down to control over reverberation, discrete reflections, background noise, and intrusive noise. If these factors are brought under control, then the capability of the sound system to deliver the soundtrack as it was prepared is greatly improved.

One system that sought to standardize sound quality in theaters was the THX Sound System. It started with the premise that the room acoustics are equally as important as the sound system. Prerequisites for installation of THX were meeting room acoustical requirements for background and intrusive noise and reverberation. Specific, patented installation methods were used for the sound-system components. When properly applied, the result was greater uniformity of performance from auditorium to auditorium in sound quality. THX was deployed starting in 1983, and by now the patents have run out, and many of the ideas used in THX are now widespread throughout the industry.

The factors subject to standardization that can improve uniformity from venue to venue are as follows:

1. Calibrated reference sound pressure level, which helps to standardize loudness for each channel playing alone at reference level for the medium.
2. Calibrated frequency response to match a standardized response curve, the X Curve, typically accomplished by means of loudspeaker design and one-third-octave room equalization, which helps to standardize timbre perception. The X Curve standard include an adjustment to it to accommodate various-sized auditoriums.
3. A high ratio of direct-to-reflected and direct-to-reverberant sound for the screen channels, which helps dialog intelligibility in the midst of interfering sound (music and effects) and helps to correctly localize sound in multichannel.
4. An array of multiple loudspeakers for the surrounds, split into left and right sides for 5.1-channel systems, promoting a diffuse sound field that remains differentiable into left and right sides and sometimes back surround.
5. Low-frequency system to accommodate the *x*.1-channel formats.

> Interestingly, film sound has some distinct differences compared with the design of other large-venue sound systems. This occurs primarily because the standards are so explicit as to maximum level versus frequency and so forth. Whereas large-scale music systems may have to play at any level demanded of them by the occasion, film sound systems just have to reproduce correctly standardized film, so the range of variations presented to the designer of the system is less.

Most equalizers meant for program equalization are one of the types already discussed, but there is another important use for equalization, for setting up a sound system in a room to a specific standard, so that sound in that room is more interchangeable with other rooms adjusted to the same standard. For such equalization, a device having a control for each critical band of hearing is useful, so that fine equalization can be carried out. One-third-octave bands approximate the critical bands of hearing (which are broader in the low bass and narrower in the treble), so one-third-octave band equalizers are routinely used to equalize dubbing stages, screening rooms, and theaters to achieve the same balance across the audible spectrum (Table 13.5). Although details vary, the typical range is from a center frequency of 25 Hz to 20 kHz, at one-third-octave spacing, for a total of 30 bands.

> This large number of bands makes such equalization unwieldy for program equalization. There are simply too many parameters to deal with. Technicians equipped with measurement equipment can make good use of such equalization by producing a signal containing all frequencies at equal strength for equal-frequency divisions on a log scale, called pink noise, and by measuring the pink noise electroacoustically with one or preferably more measurement microphones and positions, thus *equalizing the room*.

SOUND SYSTEMS FOR VIDEO

Home Theater

Today home theater has become a widespread trend, with over tens of millions of installations of multichannel equipment worldwide. This trend had its beginnings in the widespread acceptance of video media, VHS videotape and laser discs, as a consumer product (Table 13.6). Although tape started with only mono sound, and both tape and disc started with limited dynamic range, ongoing technical development, such as the introduction of Hi-Fi tracks on tape and especially 5.1-channel digital sound on DVD, has brought the media to a high level of performance.

With thousands of movies released on video this way, there was a "hidden" capability when these media were first introduced to the market. That is, most people played them over two-channel stereo systems, recognizing them as stereo, but without knowing that there was an embedded surround-sound capability. When 2:4 matrix decoders such as Dolby Pro Logic were subsequently introduced to the market, the home theater movement got its start. Virtually all of the tapes and discs of "stereo" movies were in fact surround-sound capable. Combined with a large-screen television, good surround sound systems can create a highly engaging experience in the home, which is attractive to millions of people.

The home environment is very different from the theatrical environment, however, often requiring some changes to ensure good translation of sound from one venue to the other. Whereas the goals of theater systems are well

TABLE 13.5 One-Third-Octave Band Center Frequency by Band Number

Band No.	Center frequency (Hz)
1	25
2	31.5
3	40
4	50
5	63
6	80
7	100
8	125
9	160
10	200
11	250
12	315
13	400
14	500
15	630
16	800
17	1,000
18	1,250
19	1,600
20	2,000
21	2,500
22	3,150
23	4,000
24	5,000
25	6,300
26	8,000
27	10,000
28	12,500
29	16,000
30	20,000

known and at least sometimes met through the application of widely respected standards, home systems are not well defined by standards. Also, equipment for the large and fractious consumer market shows a greater range of quality than does the equipment designed for theatrical use.

Most of the same acoustical issues that have effects on cinema reproduction apply in the home, but the scale of

TABLE 13.6 Video Release Media Characteristics[a]

Format name	No. of audio channels	Method of recording	Headroom (dB)	Approximate signal-to-noise ratio (dB)
VHS linear	1 or 2	Longitudinal edge track, mono or stereo, possibly with Dolby B noise reduction (rare)	6	40
VHS Hi-Fi	2	AFM buried layer recording with Hi-Fi companding noise reduction	16	76[b]
DVD PCM	2	16-bit linear PCM digital	20	72
DVD AC-3 or DTS	5.1 (7.1 with matrix)	Low-bitrate-coded multichannel digital	20	>75
Blu-ray	5.1 (7.1 with matrix)	Multichannel digital in a variety of formats including PCM and lower bitrate solutions	20	>75

[a]See http://booksite.focalpress.com/Holman/SoundFilmTV/ for more information.
[b]This raw number is subject to reduction during modulation due to leakage of audio into video, which causes a varying buzz behind the audio, in cases of imperfectly aligned head-to-tape contact.

the space makes different problems take on more or less importance than in the theatrical environment.

> The native background noise of home rooms used for watching television or home theaters (not dedicated home theater rooms) as measured in 27 of them by Lewis Fielder and Elizabeth Cohen, is quite low. The average background level of the homes was NC-17, well below the NC-25 of a good cinema. NC curves are a method of rating indoor acoustic noise levels, with a psychoacoustic basis of allowing more low-frequency sound energy as it is less audible than higher-frequency energy. However, individual variations in home rooms are great: air conditioning in the summer, the dishwasher running, etc., may make individual rooms quite a bit louder.
>
> There is less likelihood of acoustically absorbing material being present in a home theater than in a cinema. Although really excessive reverberation is not likely to be a severe problem once the room is equipped with normal finishes and furnishings, nonetheless other acoustic defects cause problems. Standing waves cause a lack of uniformity around the space in the mid-bass, and often an overall mid-bass boost to the sound energy in the room, which tends to make male speech rather muffled sounding. Flutter echoes between hard parallel surfaces can also be stimulated by transients on the soundtrack and lead to audible defects. Installations of loudspeakers in cabinets that may be resonant cause problems. Light-weight room surfaces and cavities that are resonant may also cause problems.

A recent trend to ameliorate the frequency-response problems caused by home listening-room acoustics is to employ autoequalization to the loudspeaker/room system. Now deployed in many products, autoequalization helps to make the experience of the program material much closer to that intended by the original program producer, just as the manual equalization in theaters has helped the industry enormously.[6] It is a one-time setup using test signals emitted by the loudspeakers and picked up at a number of positions with a calibrated microphone that is moved around the seating area to produce a good spatial representation of the sound field.

Home THX was a system designed to provide better translation of program material into the home environment and to set minimum standards of performance for the hardware of the system. Among the developments for home THX is re-equalization. This uses a high-frequency equalizer to rebalance the spectrum for the home given the different standards between theater and home listening. Unfortunately it can no longer be relied upon because home theater DVDs have today often been remixed for the "home" environment by the releasing studios, regrettably not under standardized conditions.

Desktop Systems

Computer-based workstations are already important as professional tools for manipulating sound. Also, computer users are increasingly involved in video games and other forms of entertainment offering a combined audio and video experience. Computer sound, however, has grown out of a background that started when the computer could only make beeps, which then progressed to an environment with a small, internal loudspeaker just able to distinguish sounds from one another. Today, some CD-ROM games with 8-bit sound are still available, as are inexpensive stereo computer loudspeakers of dubious quality to play them back.

Home computer audio took a giant step ahead when 16-bit audio (at least) became standard compared to the earlier 8-bit and very limited audio dynamic range on which many games were built. This accomplishment means that higher quality sound systems for desktop

---

[6]In full disclosure, I am the Chief Scientist for Audyssey Laboratories (www.audyssey.com), which licenses an autoequalizer called MultEQ, along with other developments with which I have been involved, some for many years.

systems are useful to reproduce the capability of new media. Two loudspeakers placed in front of the listener, to either side of the screen, can be supplied with signals that can audibly place sounds in three-dimensional space, called *3D sound*. This involves computing the sound that should be present at the two ears for a source to one side, for instance, and compensating for the acoustical cross talk from the left loudspeaker to the right ear, and vice versa. With signals that represent the inputs to each ear independently, a sound can theoretically be generated at any angle. Practical implementations of this have limitations in timbre reproduction but can make a convincing display of sound that breaks the bounds of the screen.

In the late 1990s, I built a system for the desktop that could be used to make professionally useful sound decisions in an edit room. Called MicroTheater, it reproduced the full frequency, dynamic, and spatial capabilities of a theatrical film program. Used with a digital audio workstation, it was a powerful tool for making films and video on the desktop. The design used audio processing technology, at that point mostly analog, to solve the problems posed by small room acoustics and psychoacoustic principles to scale theater sound to the desktop. Using this system and a digital audio workstation, sound personnel could design, edit, and mix sound in an inexpensive environment. It was used on *Titanic* in the editing room to attend scoring sessions and dubbing remotely, on *Contact*, for which it was used for temp dubs that went directly to theaters for test screenings, and on a documentary on the history of electronic music, *modulations*, which was completed on a desktop and went from there directly into theaters, lowering costs to make such a production practical. A problem for the introduction of the system at that time was the long distance between editors and mixers, because it allowed editors to do what had normally been reserved for mixers. Today, with many more people trained in the use of plug-ins on the desktop, such a system could provide the monitoring that could lead to better sound being delivered to dubbing stages, or even directly to theaters.

TOWARD THE FUTURE

The upper limit on expression is set by the characteristics of media and sound systems. Today that is most often 5.1 (or occasionally 7.1) channels, nearly 24 kHz bandwidth, and over 100 dB dynamic range. In digital cinema installations these are done with linear PCM coding so there is no theoretical quality limitation. However, the industry has reached the limits of its audience in cinemas as to loudness capability, and it is commonplace to play movies some 5 dB lower than reference levels. I tried playing *Sea Biscuit* at reference level, not a particularly loud film, in a full house on a Saturday night in Memphis, Tennessee, and within the first 10 min about six people came

out and asked for it to be turned down. Also, movie trailers pack the climax of movies into 90 sec, whereas the feature starting after the trailers usually has 90 min or so to go before the climax, so trailers were very loud compared to the openings of movies. To solve this problem the loudness level of trailers was limited for MPAA-rated films to the levels in a Trailer Audio Standards Agreement (TASA).

Because the frequency range of human hearing and theatrical conditions is already covered, and the dynamic range is more than necessary, growth in the capacity of the medium will be in the number of channels. Already upward pressure can be seen in the marketplace, with non-standardized 7.1 systems. Extracting even more channels from fewer, called upmixing, is a challenge, but some implementations are appearing in the home market now, such as Audyssey's DSX.

For some years my colleagues at the University of Southern California and I have been at work on a 10.2-channel system. It starts with the premise that the 5.1 channel system is well established and standardized, so that if new channels are to be added, the question becomes where to add to the already standard systems for them to have the best advantage. There are three reasons to add more channels: to reproduce more faithfully the acoustics of real spaces, to stimulate psychoacoustic hearing mechanisms better, and to fulfill the desires of sound designers and composers. Considering these, added channels in the future may well be:

- Wide channels: Off the screen at $\pm 60°$ in plan view (looking down from on top) from center. These two channels wide left and right reproduce the first sidewall reflections in real rooms like concert halls, where such reflections are known to be important.
- Height channels: The perception of height is strong for elevations in front of the listener and becomes less good overhead and to the high sides, so left and right front heights become useful, both for physical acoustics (the first ceiling reflection) and for psychoacoustics.
- Center back: Already in use in some systems and called rear surround in 6.1-channel formats, it helps differentiate side from back and adds to envelopment.
- To these are added a difference between direct-radiating and diffuse-radiating surround speakers, to make physical acoustical differences according to localizing or enveloping. Also added is an extra LFE channel, just to keep up with the headroom necessity of more channels, if nothing else.
- These all add up with the original 5.1 to 14 electrical and loudspeaker channels, although some share locations, so we call it 10.2-channel sound.

Progress in 10.2-channel sound has been installing multiple permanent fixed and temporary installations across a wide range of room volumes, making more than 20 items

of custom-mixed program material and becoming standardized in advanced digital cinema and QuickTime systems for distribution.

NHK, the Japanese national broadcaster, has shown a 22.2-channel system.[7]

Another system seeks to do wavefield synthesis. Using a large number of electrical channels and loudspeakers, such as 96 in a cinema, sounds may be synthesized in space, "captured" by an array of microphones or their software equivalent, and played as expanding bubbles of sound. Called Iosono, it is backed by the Fraunhofer Institute.

All in all it seems likely that growth in the artistic abilities of audio accompanying pictures over the next decades will be in the growth of the number of channels and their exploitation by filmmakers to make sound art.

---

[7]See http://www.nhk.or.jp/digital/en/technical_report/pdf/nab200502.pdf.

Working with Decibels

The decibel is a means to express logarithmic relationships across levels. For voltages, or sound pressure levels, the equation for a dB relationship is

$$DB = 20 \log_{10}\left(\frac{Vmeasured}{Vref}\right),$$

where V_{measured} is the measured voltage and V_{ref} is the reference voltage. There are several reference voltage values used in audio, shown in Table A1.1.

The most common reference for sound pressure level is $20\ \mu\text{N/m}^2$ rms. Voltage and sound pressure level are analogous.

To use these, let's calculate the voltage level for the specified input noise voltage of a microphone preamplifier having a rating of -127 dBm, a rather commonly seen number. The first thing we need to know is that dBm equals dBu in voltage. dBm is decibels relative to 1 mW of power, an older method of measuring level that originated with the telephone company (the villain in one of the Flint films, called TPC there).

> The dBm measurement method used matching impedance conditions, but today we virtually always use bridging conditions, explained in Chapter 3. We hang onto the tradition of referencing to the voltage that corresponds to 1 mW in 600 ohms though, and that is 0.7746 Vrms. Root mean square is a method of measurement of AC waveforms that is equivalent to their DC value in terms of heating up a resistor. It is not the peak, nor peak-to-peak voltage, but instead for a sine wave is 0.707 times the peak voltage.

So -127 dBm $= -127$ dBu.

$$-127\ dBm = 20 \times \log_{10}\left(\frac{Vx}{0.7746}\right),$$

and solving for V_x means:

TABLE A1.1 Reference Voltage Levels Used in Audio

Level	Voltage (Vrms)
+4 dBu	1.228
0 dBu	0.7746
−10 dBV	0.316

1. Dividing -127 by 20,
2. Taking the antilog (10^x),
3. Multiplying by 0.7746.

The answer is $V_x = 0.35\ \mu\text{Vrms}$.

When it comes to power, whether it is electrical power or acoustical power, a different equation is used,

$$dB_{power} = 10 \times \log_{10}\left(\frac{Pwr_{measured}}{Pwr_{ref.}}\right),$$

where the most common P_{ref} is 1 W. Watts are not rated in rms (although you will see that happen every day, it is incorrect), but rather in average watts.

So twice as much voltage is about 6 dB (actually 6.02 for pedants), twice as much power 3 dB (3.01). Psychoacoustically, however, to sound twice as loud the sound pressure level must be increased on the average by about 10 dB, although answers can be gotten anywhere from 6 to 11 dB, depending on the experiment.

Table A1.2 allows you to estimate voltages and powers from decibels and vice versa quickly.

Many dB ratios can be determined from the table by adding and subtracting decibels and multiplying or dividing correctly. For instance, a voltage gain of 46 dB is $20 + 20 + 6$ dB, so it is $10 \times 10 \times 2$ or a factor of 200 in voltage.

The ultimate noise floor of a system is controlled by the impedance at its source. That is because there is an irreducible noise due to the random motion of electrons in any impedance that occurs at any temperature above absolute zero. This is called Johnson noise, after its discoverer at Bell Labs. A typical electrodynamic microphone has an impedance of 200 ohms. The Johnson noise of a 200-ohm impedance is

$$e_n = \sqrt{4kTBR},$$

where k is Boltzmann's constant (1.38×10^{-23}), T is the temperature in degrees Kelvin ($293\ ^{\circ}\text{K}$ at room temperature), B is the bandwidth in Hz (20 kHz), and R is the impedance in ohms (200). This voltage is $0.25\ \mu\text{Vrms}$.

Interestingly, now we can calculate how much greater the microphone preamplifier noise specification calculated above is than the thermal noise of this microphone, in decibels. Such a comparison is used widely in radio frequency

TABLE A1.2 Decibels versus Scale Factor for Voltage and Power

Power multiplier	dB	Voltage multiplier
100	+20	10
10	+10	3.16
4	+6	2
2	+3	1.41
1	0	1
0.5	−3	0.71
0.25	−6	0.5
0.1	−10	0.316
0.01	−20	0.1
0.001	−30	0.0316
0.0001	−40	0.01
0.0001	−50	0.00316
0.00001	−60	0.001
1×10^{-6}	−70	0.000316
1×10^{-7}	−80	0.0001

engineering, but not much in audio, and it is called the noise figure:

$$dB = 20 \times \log_{10}\left(\frac{0.35\mu Vrms}{0.25\mu Vrms}\right).$$

The noise figure of this microphone preamplifier is 3 dB; that is, when you turn it up, the noise you hear as hiss is within 3 dB of the theoretical noise floor that you could possibly have for this microphone.

CALCULATING dB GAIN

A calculation can be made of how much gain is needed in a mixing console so that the loudspeakers produce the same SPL as was picked up by the microphone. Let us say that a performer is speaking at an average 65 dB SPL, 1 m from the microphone. The microphone sensitivity is 6 mV at 94 dB SPL, and its output voltage will thus average 0.2 mV, because this is 29 dB (94 − 65 = 29) below 6 mV. In order for the loudspeaker to produce 65 dB SPL at our listener's ears, we need about 0.4 V (85 − 65 = 20 and 20 dB below 4 V is 0.4 V, from the preceding performer example). Thus the gain needed in the electronics is 0.4 V/0.2 mV = 2000 = 66 dB.

A more typical case would call for more amplification, because movie soundtracks for speech are on average 10 dB louder than face-to-face speech. Thus 76 dB is necessary, overall, divided among the microphone preamplifier, the rest of the mixing console, and the loudspeaker power amplifier. Because a typical power amplifier gain is 30 dB, the console must provide about 46 dB gain in this instance.

Filmography

A personal list of films that have affected me. While they are ranked from the most important on down, this task is nearly impossible since they are so different. This is a Robinson Crusoe "lost on a desert island" list—cut from the top if you must. They may not have anything special about the sound, or they may be sound tour de forces, but they merit attention in either case.

Lawrence of Arabia
Who's Afraid of Virginia Woolf
Forbidden Planet
To Kill a Mockingbird
Brokeback Mountain
The Godfather, Parts I and II
2001: A Space Odyssey
Casablanca
Dr. Strangelove: or How I Learned to Stop Worrying
    and Love the Bomb
E.T. The Extra-Terrestrial
Apocalypse Now
12 Angry Men
High Noon
If
The Manchurian Candidate (1962)
Bringing Up Baby
Ran
The Great Escape
The King and the Clown (Korea, 2005)
The Graduate
Ikiru
Midnight Cowboy
The Great Dictator
Tongues Untied
Modern Times

City Lights
Sunset Boulevard
It Happened One Night
Some Like It Hot
The Man Who Knew Too Much (1956)
Metropolis
Napoléon
On the Waterfront
Singin' in the Rain
Psycho
Rebel Without a Cause
Shane
Witness for the Prosecution

Films that are more recently made that one day may make the first list, but adequate time has not elapsed to evaluate them for inclusion.

The Boy in the Striped Pajamas
Capturing the Friedmans
The Hangover
Inglourius Basterds
The Lives of Others (Germany, 2006)
The Queen
Star Trek (2009)
Slum Dog Millionaire (India, 2008)
Tsoti (South Africa, 2005)
Invictus
Ratatouille

One television program stands out for inclusion, although not available on video at the time of this writing.

Nova episode (PBS 1986) *The Case of the Frozen Addict*

This page intentionally left blank

The Eleven Commandments of Film Sound

1. **Allow the sound crew on the set an overhead boom microphone.** The overhead position is usually decently far from the room boundaries so that directional microphones can work properly, and it is usually the best location to capture actors' voices.
2. **Always wait a beat before calling "action" or "cut" so that the sound editor has some footage that matches the scene for a presence track.** This is often overlooked in production, but a few seconds on each shot saves a great deal of time in postproduction. The few seconds can be copied to any length necessary to fill out the scene if it is "clean."
3. **Make sensible perspective choices in recordings.** Extreme perspective changes are jarring as the direct-to-reverberant ratio changes from shot to shot; only subtle changes are typically useful. Remember that it is always possible to add reverberation, but exceedingly difficult if not impossible to remove it in postproduction.
4. **In narrative filmmaking, exercise discipline and control on the set by minimizing all undesired noise sources and reverberation, and maximizing the desired source.** When you are making a fictional film, you have the ability to "pan off" an undesired object; use the same control for the sound.
5. **Make sure the sound is in sync with the picture.** Nothing is more amateurish than out-of-sync production sound: there is a need for traceability of sound sync and camera sync to a common or to matched sources.
6. **Organize tracks during editing with a strong eye to mix requirements.** Fit tracks to the available number of mixable tracks, leaving as much space between different sounds as possible. Keep similar sounds in the same tracks, and different ones in different tracks.
7. **Normally, provide a complete audio world, including adequate fill and Foley or equivalent effects.** Many poor films simply do not have enough effects: silence is rarely found in nature, and should usually not be found in films either. The lowest level sounds, such as background noise of rooms, must be brought up to such a level that they will "read" through the medium in use.
8. **In mixing, one principal job is to get the program material to best "fit" the dynamic and frequency ranges of the target medium**. If you are mixing for VHS, there are severe limits that must be observed, or else distortion and/or hiss will be audible. Also, given today's digital capabilities, mixing to the "top of the scale" is extremely loud when played back in a film-calibrated environment. Moving from desktop mixing directly to a theater is crazy without testing first the levels.
9. **Storytelling always comes first: if it works, break the rules.** Other than doing damage to people or equipment, all the "rules" given are breakable for artistic purposes, if breaking the rules results in art being produced.
10. **Separate strongly the requirements of production from those of reproduction.** The filmmaker is highly involved with the first, but the second should be practically a mechanical process. That is, once the mix is committed in a calibrated environment, everybody else's job downstream is simply to reproduce it, not to "improve" it. Remember the Warren Beatty story about the premiere of *Bonnie and Clyde* in London (how the projectionist remixed the movie on the fly!).
11. **Separate physical sound cause and effect from psychoacoustic cause and effect.** The advantage of doing so is that problem solving is best handled in the domain of the cause. Human perception of sound fields wraps together physical and psychoacoustic sound. Test equipment virtually always works in the physical domain, and thus may not show best what is perceived to be a problem.

This page intentionally left blank

Bibliography

Altman, Rick, ed., *Sound Theory Sound Practice*, Routledge, New York, 1992.

Berger, Eliot, ed., et al., *The Noise Manual*, Revised Fifth Edition, American Industrial Hygiene Association, 2003.

Bregman, Albert S., *Auditory Scene Analysis*, MIT Press, Cambridge, 1990.

Eargle, John M., *The Microphone Book*, Focal Press, Boston, 2004.

Eargle, John M., *Handbook of Recording Engineering*, Chapman Hall, New York, 2005.

LoBrutto, Vincent, *Sound-On-Film, Interviews with the Creators of Film Sound*, Praeger, Westport, CT, London, 1994.

Moore, Brian C. J., *An Introduction to the Psychology of Hearing*, 5th ed. Academic Press, 2003.

Murch, Walter, *In the Blink of an Eye*, Silverman-James Press, Los Angeles, 2001.

Pasquariello, Nicholas, *Sound of Movies: Interviews with the Creators of Feature Sound Tracks*, Post Bridge Books, San Francisco, 1996.

Pickles, James O., *An Introduction to the Physiology of Hearing*, 3rd ed. Academic Press, London, San Diego, 2008.

Taub, Eric, "Production-Sound Mixer, Boom Operator, Third Man," in *Gaffers, Grips and Best Boys*, St. Martin's Press, New York, 1995.

Whittington, William, *Sound Design and Science Fiction*, Univ. of Texas Press, 2007.

Yewdall, David, *The Practical Art of Motion Picture Sound*, Focal Press, Boston, 2007.

This page intentionally left blank

5.1 the prominent system for the use of loudspeaker channels to produce front imaging and surround sound, with an added Low Frequency Effects (0.1) channel. The principal channels are Left, Center, Right, Left Surround, and Right Surround.

A chain that part of a theater sound system including reproduction of the print with the sound head on the projector, preamplification, noise reduction or Academy filtering, digital decoding, and matrix decoding, each where used. The A chain is the part of a system that is identified with a particular format, such as *Dolby Digital*,[1] *DTS, SDDS*, etc. See *B chain*.

A track a term used in editing for the production sound that is cut in sync with the picture.

Absorption the property of materials to turn incident acoustical energy into heat. Absorption of the materials on the surfaces of a room is usually the only factor available to control reverberation time, because absorption and room volume (three-dimensional size) are the only classical factors affecting reverberation time. Also used to describe losses that occur during transmission through a medium, such as *air absorption*.

AC (alternating current) applies to the power derived from commercial power generators as a 60-Hz sine wave in the United States, as opposed to DC power derived from batteries or DC generators.

Academy mono a conventional monaural optical soundtrack format intended to be used with a rather strong high-frequency rolloff in playback. The rolloff is necessitated by the need to suppress audible noise due to grain of the motion-picture negative and print. The rolloff is accomplished in two possible ways: in a conventional monaural theater it is due to loudspeaker and screen high-frequency losses, as well as a possible *lowpass filter*, whereas in a stereophonic-equipped theater playing a monaural print, it is a (different) *lowpass* filter. The distinctions are explained in the standards SMPTE 202 and ISO 2969. The effects of the rolloff are partially overcome by boost *program equalization* applied by the postproduction mixer during dubbing, but this process has limitations because boosting high frequencies may lead to noticeable distortion.

Acetate (tricellulose acetate) basically wood pulp. A material, now obsolete, used for the base of motion-picture film. It was usually chosen for its ease of cutting in editing situations, despite the fact that it is not as stable as *polyester* base. Acetate base is hygroscopic (absorbs water from the atmosphere), so temperature and humidity conditions of storage are more critical than for polyester-based films.

ADR automated dialog replacement. A system of equipment and a controlled recording studio or stage that permits watching and listening to a performance through headphones and repeating the performance for recording without the background noise and reverberation of location shooting. Also known as looping.

AES see *Audio Engineering Society*.

AFM recording a method of recording by way of heads on the scanning drum of helical scan videotape formats. The recording is buried underneath the corresponding video and is at a different azimuth angle, which permits separation from the video. Cannot be separately recorded subsequent to video recording.

Air absorption losses in propagation beyond those expected from considerations of sound spreading out over distance; greater at high frequencies than at low frequencies, and changes with humidity.

Ambience generally speaking, ambience is widely used as a synonym for *ambient noise*. In film and video sound production, the term is used more specifically to mean the background sound accompanying a scene, whether present in the original production recording (for which a better term is *presence*) or deliberately added in *sound-effects editing* to provide an acoustic space around the rest of the dialog and sound effects. Ambience helps establish the scene and works editorially to support the picture editing by, for example, staying constant across a picture cut to indicate to the audience that no change of space has occurred, but rather only a simple picture edit. Conversely, if ambience changes abruptly at a picture cut, an indication is made to the listener that the scene has also changed.

Ambient noise the acoustic noise present at a location (which includes a set) without considering the noise made by the production. A preferred term is *background noise* so that ambient noise and *ambience* (which may be deliberately added) are not confused.

American National Standards Institute the U.S. supervisory standards body. AES and SMPTE standards may become American National Standards.

Amplification see *Audio amplification*.

Amplitude the size dimension of a waveform, usually represented graphically in the vertical plane. The size represents the strength of the unit being measured.

Analog audio any signal representation in which the signal waveform is transmitted or stored in direct 1:1 correspondence

---

[1]All italicized words are themselves defined within this glossary.

with the sound wave. The means of representation may be electrical, mechanical, magnetic, electromagnetic, or optical. The amplitude dimension of the audio *signal* is represented by means of a direct analogy between a voltage, displacement of position (such as for the phonograph or an analog *optical sound track*), or strength of magnetic flux (analog *tape recording*), for example, and the signal. Analog systems may also employ modulation and demodulation such as frequency modulation (FM), in which audio is imposed on a carrier frequency, such as a radio frequency, by means of modulation; despite the modulation/demodulation cycle, the audio remains in analog form because no digitization has occurred.

Analog audio is distinguished from digital audio by the fact that analog systems do not use quantization (see *Digital audio*). This means that the representation is intended to be continuous in the amplitude domain. On the other hand, analog audio may be sampled, which is generally considered to be a digital audio process, but if the audio remains unquantized it is not digital audio. An example of a device that uses sampling but not quantizing is the bucket brigade analog delay line, often used in the past for inexpensive audio delay.

General areas of advantage for analog audio over digital include some specialized applications, such as most transducer-associated amplifiers (for microphones and loudspeakers) and the fact that analog audio is a mature technology, with good economy for a given level of sound quality. Disadvantages include the fact that because each stage of analog audio involves making an analogy between the actual signal and its representation, inevitably noise and distortion accumulate across generations because the analogy fails to completely replicate the signal; the signal is not a clone but a more or less good representation of the signal. This process is manageable with care and has led to some great-sounding soundtracks. The capabilities of digital audio grow daily, especially considering that many of the possibilities are driven by a much larger industry, computer manufacturing. On the other hand, the viewpoint that film and television "would sound better if it were just all digital" is naive, because "sound better" depends on a great many factors, the most important of which is probably the sound design.

ANSI American National Standards Institute.

ASA 1. Acoustical Society of America.
2. (obsolete) American Standards Association, the forerunner of the American National Standards Institute.

Atmos (British usage) atmosphere.

Atmosphere (British usage) synonym for presence.

Audible frequency range usually considered to be between 20 Hz and 20 kHz, although these are not rigid limits. Below 20 Hz, sound is less well perceived as tonal than above 20 Hz, degenerating at lower frequencies into individual pulsations. Sound at high levels and infrasonic frequencies is more likely to be perceived as vibration than as sound. Above 20 kHz, the hearing of sound in air of even the best young listeners rapidly rolls off, although there are young people who hear out to about 24 kHz. Below 20 Hz, sound is called infrasonic, and above 20 kHz it is called ultrasonic.

Audio a broad term covering the representation of sound electrically or on a medium: "audio tape" is better usage than "sound tape," although soundtrack persists in the film industry to describe audio accompanying a picture.

Audio amplifier audio amplifiers take a number of forms. First are input transducer preamplifiers such as microphone preamplifiers, intended to raise the output of the transducer to a higher nominal level for processing by further circuitry. Specialized preamplifiers in this category include phonograph and magnetic film and tape preamplifiers, which have both a gain function and an *equalization* (definition 1) function to compensate for the transducers and the frequency response used on the medium to optimize dynamic range. Second are line-level audio amplifiers, which raise the signal level further. Third are specialized audio amplifiers for processing the signal usually having unity gain but offering features such as *equalization* (definition 2). Fourth are amplifiers used principally for buffering, that is, for preventing unwanted interactions among components. Fifth are summing amplifiers used to add together multiple signals, without interactions among them. Finally, there are audio power amplifiers intended to drive loudspeakers.

Major concerns with all audio amplifiers include the dynamic range capability, which means both the maximum signal handling capability versus frequency and noise versus frequency; linear *distortion*, such as *frequency response* errors; nonlinear *distortion*; and input and output conditions, such as nominal levels and impedances.

Audio bandwidth generally considered to be the range 20 Hz to 20 kHz, although these are not hard limits.

Audio Engineering Society (AES) a U.S.-based international group of professionals in audio that includes standards-making among its activities.

Audio frequency usually considered to be the frequency range between 20 Hz and 20 kHz, based roughly on the limits of human perception.

Audio mixer an ambiguous term applied to both audio mixing consoles and the persons who operate them. Generally speaking, the functions of an audio mixing console can be broken into two broad classifications: processing and configuration. Among audio processes are *audio amplification*, control of levels by way of *faders, equalization*, and control over program dynamics by means of *limiters, compressors*, and *expanders*. The configuration part of console design has to do with arranging the audio processes in certain preferred orders, combining signals, and routing the signals throughout the console. Issues included in configuration include signal routing (e.g., which microphone input goes to what channel of a multitrack recorder), *buses, auxiliary sends*, and *pan pots*.

Auditory streaming an audio psychoacoustics term that describes the segmentation of sound into a variety of parts such as speech. A part of auditory scene analysis.

Azimuth see *head alignment*.

B chain the part of a theater sound system following the *A chain* from the top of the volume control, through speaker/room equalization of at least one-third-octave band resolution, electronic crossovers where used, power amplifiers, and

loudspeakers and their acoustical environment, both local to the loudspeakers and global throughout the room. The B chain encompasses every factor having an aural effect on the listener after the processes occurring in the A chain.

Background noise usually refers to the acoustical noise present on a location or set without the presence of the production, although more generally it applies to any type of unwanted noise.

Backgrounds synonym for ambience.

Bandpass filter an electrical filter designed to pass only a certain range of frequencies while suppressing signals at frequencies outside the range. Bandpass filters are often adjustable as to frequency and may be composed of a combination of a *highpass filter* and a *lowpass filter*. An example of the use of a bandpass filter is in limiting the frequency range of a speech recording to a narrow band to simulate speech heard over a telephone.

Bandstop filter an electrical filter designed to pass only frequencies lying outside a defined range; both lower and higher frequencies are passed, and frequencies lying within the range of the filter are suppressed. A bandstop filter is usually used to suppress unwanted noise that lies in only one frequency region. It is similar to a *notch filter*, but has a broader frequency range of suppression than a notch filter.

Bandwidth the frequency range usually stated from low to high frequency, over which a system has a stated uniform *frequency response*. Usually, if left unstated, the bandwidth is the frequency range over which the output does not fall more than 3 dB from its midrange value, thus, a specification such as "bandwidth from 30 Hz to 20 kHz" probably means ± 3 dB from 30 Hz to 20 kHz. Clearly, it is far better when both the frequency range and the response tolerances over the range are given.

Barney a generally soft motion-picture camera cover meant to reduce the acoustic noise of the camera for sync sound shooting.

Beat in music, the beat is the underlying meter or rhythm. In sound in general, however, it has an additional meaning. Whenever tones of two or more frequencies are passed through a device exhibiting nonlinear *distortion*, new tones will be created at new frequencies corresponding to sums and differences of the form $f_2 + f_1, f_2 - f_1, 2 \times (f_2 - f_1), f_2 - (2 \times f_1)$, etc., out to higher orders. These new tones are called beat notes. Often the term is applied to the first-order difference tone at $f_2 - f_1$, because it may be well separated in frequency from f_2 and f_1 and thus be audible because it is not subject to much *frequency masking*. The term also applies to, for example, multiple piano strings used for one note that are not perfectly tuned to one another; the resulting amplitude modulation leads to an audible "wobbling" in time.

Bias to employ bias is to add an inaudible DC or AC signal to a desired audio signal to overcome nonlinearities of amplification or the medium. In the case of amplifiers, bias is, for example, the DC idle current that is present in the circuit in the absence of a signal. Here the bias serves to place the particular stage of the audio amplifier at an operating point at which it can most accommodate the range of signals expected.

Ultrasonic AC bias is used in analog tape recording to linearize the tape medium, which would otherwise be highly distorted. Bias in this case is a high-frequency sine-wave signal supplied by a bias oscillator that is added to the audio signal, usually after the last stage of amplification and before application to the record head. The bias frequency is usually from 100 kHz upward in professional machines. The level of ultrasonic bias is important in analog tape recording because it has an impact on many important parameters, including medium- and high-audio-frequency sensitivity, headroom versus frequency, and noise of various kinds. Choice of a bias operating point is made by considering the characteristics of the tape medium, tape speed, and the record head gap in use. Usually the manufacturer of a tape or film machine will supply a procedure for best results, but a common example would be to operate at a certain amount of overbias, such as 2 dB, at 10 kHz. This means finding the bias operating point having the highest sensitivity by adjusting the bias level up and down while measuring the effect on the level of a 10-kHz tone and then adding more bias until the desired degree of overbias is achieved.

Bilateral a term applied to analog variable-area optical soundtracks to describe a soundtrack that is symmetrical and a mirror image about its centerline. The advantage of a bilateral soundtrack is that its use helps to cancel out variations in output level and the accompanying distortion occurring due to nonuniform illumination across the soundtrack area; if one side is moving into an area of less illumination, then it is likely that the other side is moving into an area of greater illumination, the effects of which cancel. To accommodate stereophonic information, two bilateral soundtracks are placed side by side, and the whole track is called the "dual-bilateral stereo variable area."

BKSTS see *British Kinematograph Sound and Television Society*.

Black-track print a motion-picture prerelease print format with picture elements generally complete, but printed black in the optical soundtrack area. The black-track procedure is often used for early answer prints before an optical soundtrack is available.

Blimp a more or less solid, continuous, camera cover meant to reduce the acoustic noise of a motion-picture camera for sync sound shooting. See also *barney*.

Bloop

1. In optical sound recording, the term bloop refers to making the soundtrack either transparent or opaque, as called for on a negative or positive recording to prevent hearing a transient noise, such as a click or a pop, particularly at splices.

2. Generally obsolete: in production sound recording, the term bloop is a shortened form of bloop slate. The bloop slate consists of a pushbutton and light connected to an electronic oscillator supplied in portable tape machines to indicate synchronization points between picture and sound. It was most often used in documentary film production, in which the use of the traditional clapperboard slate may be intrusive to the subject. The bloop oscillator is at a different frequency and has a different timbre compared to the reference oscillator so that they may be easily distinguished by listening.

Board synonym for *console*.

Boom a mechanical device for holding a microphone in the air and manipulating it in several dimensions by swinging the boom arm and rotating the microphone on the end of the boom. Booms are generally floor-standing, fairly heavy devices. Booms are preferred in fixed-set situations, but their size and weight make them more problematic for productions shot on location, where the *fishpole* is perhaps more often used.

Boom op see *boom operator*.

Boom operator the user of a microphone boom, whose other duties often include operation of *fishpoles*, planting of hidden microphones on the set, and placement of radio microphones and their transmitters on actors. The boom operator often has a sophisticated job to do in balancing among the actors to get the best recording of each of them. They also learn the script through rehearsal in order to anticipate moves on the part of the actors; to accommodate camera moves, dialog overlaps, and other things; to make a coherent-sounding soundtrack production sound recording.

British Kinematograph Sound and Television Society (BKSTS) a British-based professional group with interests in film, video, sound, and television.

Bulk eraser see *degausser*.

Bus an electrical interconnection among many points, so called because it can be viewed as a bus line, with stops (connections) along its way. Signal buses include main buses, often called mixdown buses; buses intended to send signals to a multitrack tape recorder, called multitrack buses; and buses to send signals from input channels to outboard devices, called auxiliary buses. Auxiliary buses may be named for their purpose: reverb send (signals sent to a reverberation process) and cue send (signals sent to performers to cue them).

Buzz track a special type of recorded *optical sound* test film used in the alignment of the area scanned by analog optical sound heads (playback devices). It consists of two recordings made outside the usual soundtrack area, with their maximum peaks just touching the outer edges of the area to be scanned, each at a different frequency. The optical sound head is adjusted correctly for lateral position when essentially no sound is heard, whereas adjusting to one side results in hearing one frequency and to the other side in hearing a different frequency. This is one adjustment that sometimes must be made directly on release prints of classic movies rather than on test film, as they may have shrunk.

Cable person the third person on a *production sound* crew responsible for cable handling and maintenance of sound equipment.

Calibration tape or film a magnetic tape or film prepared under laboratory conditions with recordings having prescribed characteristics. In audio, these characteristics include the absolute level of reference *fluxivity* sections at a standard frequency, such as 1 kHz, and the relative levels of various frequencies according to the standard in use. Specialized test tapes or films also are available recorded with low *flutter* so that flutter due to transports can be measured. In video, the prescribed characteristics include reference levels for white and black and reference level and phase for color, among many others.

Cans slang for *headphones*.

Capacitor microphone see *electrostatic microphone*.

Capstan the rotating cylindrical shaft against which tape is pressed by a *pinch roller* to impart the correct linear speed to tape or film. The average thickness of the tape must be accommodated in the design to produce the correct linear speed. Typical professional tape speeds are 30, 15, and $7\frac{1}{2}$ inches/sec, whereas film speeds are $22\frac{1}{2}$ (70mm film), 18 (35mm film), and 7.2 (16mm film) inches/sec.

Channel an audio term specifying a given signal path. When audio is physically recorded on film or tape, the term applied to the physical representation on the medium is *track*. Thus, the input of a tape machine may be labeled in channels, whereas the recording it makes on tape is made on tracks. In console terminology, the term applies to both input channels and output channels. Input channels usually represent various single sounds coming into a console for mixing, and output channels represent various combined sounds; for example, all sounds destined for the left loudspeaker represent the L output channel.

Cinema Audio Society (CAS) a U.S.-based group of recording professionals with interests in film and television sound.

Clapper see *clapperboard*.

Clapperboard the traditional device used to synchronize sound and picture starts by banging a board down on top of another within view of the camera so that a reference mark is made in both sound and picture.

Click track a soundtrack, usually prerecorded on one track of a multitrack medium, to guide musicians and others in making recordings in synchronization with the picture and other soundtracks, consisting of clicks at the correct intervals to correspond to the beat. Use of the click track for analog recording requires using *Sel-Sync* (playback off the record head of the track assigned to provide the clicks) to ensure synchronization.

Cochlea the inner ear, pronounced coke-lee-ah. The organ that converts mechanical sound energy, delivered by way of the outer and middle ear, to electrical impulses for detection by the brain. The fundamental mechanism is a spatially dependent spectrum analyzer, with the input end of the membrane stretched in the cochlea responding to high frequencies, and the far end to low frequencies, with the membrane covered in hair cells and nerves, which make the transduction from mechanical to electrical energy.

Compact disc audio (*Compact Disc*, Philips) the popular standardized audio disc format using optical recording to encode two channels of 16-bit linear pulse code modulated audio information sampled at 44.1 kHz, of up to approximately 72 min in length.

Compression

1. a part of the process of sound propagation wherein the density of particles in the medium is momentarily increased: compression and rarefaction, the two parts of the process, raise, and lower the sound pressure level from ambient respectively.

2. a device that controls the volume range of source material by measuring its level and controlling the reproduction range. A compressor may be feed-forward wherein the signal controls the level downstream from the point of measurement, or feedback, wherein the output of the level controlled stage is fed back to control the level. Typical controls include level threshold (where the action starts), compression ratio (such as 2:1 wherein for each 2 dB change into the compressor 1 dB in change is found at the output), attack and release time controls, and "make up" gain that adjusts for the fact that the effect of a compressor is typically to reduce the level.

Compression driver an electroacoustic transducer designed to be coupled to the throat of a *horn*. The operating principle of most compression drivers is similar to a conventional loudspeaker, with a coil of wire, called the "voice coil," suspended in a strong magnetic field and supplied with audio current from a power amplifier, often by way of a crossover network. Application of electrical current to the voice coil produces corresponding motion due to the interaction of the current and the magnetic field. The voice coil is attached to a suspension, to keep it mechanically centered in the magnetic field, and to a diaphragm, which moves air in and out of the throat. Because of the differing path length from the outer diameter compared with the center of the diaphragm down the throat, which would cause high-frequency cancellation as the sound waves from different parts of the diaphragm arrived at the throat at slightly different times, it is customary to insert a "phasing plug" into the throat, with multiple paths machined through the plug designed to time the sound waves so that they arrive at the throat simultaneously.

Compressor a special kind of audio amplifier arranged so that equal level changes in the input result in smaller level changes at the output, often used to help fit a wide *dynamic range* program into a narrower dynamic range channel. Compressors have many potential adjustments, and any one hardware or software compressor may contain a range of controls such as *threshold* (the level above which to begin compression), *slope* (the amount of compression such as 2 dB in for 1 dB out), *gain make-up* (a level control used to compensate for the fact that the compressor is, on the whole, reducing the level above threshold), *attack time* (a control to set how fast the compression function operates: if set too fast small transients that are not very loud take over the compression, if set too long the gain control function will be audible, as when the program gets loud one will hear the compressor turn the level down), *release time* (a control to set how long the compressor takes to restore the normal gain), and others.

Condenser microphone synonym for *electrostatic microphone*.

Console a piece of audio equipment designed to amplify, combine, and otherwise process multiple inputs. See *audio mixer*.

Contact microphone a specialized microphone designed to pick up vibration directly from a solid body. Contact microphones have been used to record items as diverse as a violin and a bridge structure stimulated by traffic.

Container a specialized dynamic range controller designed to set an upper limit on recorded level with minimum audible side effects for use in preparing *print masters*, used especially for transfer to analog optical soundtracks, respecting the headroom limitations of the variable area soundtrack versus frequency.

Cross talk undesired audio signals coming from adjacent sources or tracks. The sources of cross talk include inductive and capacitive coupling among, for example, the various channels of a tape head.

Crystal sync the term used generally to describe a method for synchronizing tape recordings with film cameras. A separate recording derived from an accurate quartz-crystal-based oscillator is made as a reference on conventional non-perforated tape simultaneous with audio recording. Crystal sync relies on accurate motor speed control of the camera and an accurate frequency of the reference oscillator in the recorder. Analog tapes played with synchronization recordings are subsequently *resolved*; that is, the speed of playback is controlled to precisely match the original conditions, and then the tape is copied to a digital audio workstation or to perforated film, which will be in sync with the original picture.

Cue tone any system in which a tone is used to cause an action. An example is a slide-change system based on tones on one track of a tape recorder, whereas other tracks may contain program material intended to be heard.

Cue track the track of a multitrack recorder assigned to cue tones, or a track with incidental information, such as editing information.

DC (direct current) the power derived from a battery or a specialized generator, as opposed to *AC*. DC refers to power delivered at 0 Hz.

DDL see *digital delay line*.

Decibel (dB) literally, one-tenth of a bel. The use of the term decibel means that logarithmic scaling of the amplitude of a quantity divided by a reference amplitude has been used. Such scaling is useful because the range of amplitudes encountered in sound is extremely large and because hearing generally judges the relative loudness of two sounds by the ratio of their intensities, which is logarithmic behavior. Differing factors are used when applying decibels to various quantities so that the number of decibels remains constant: A 3-dB increase is always 3 dB, although is represents twice as much power but only 1.414 times as much voltage. For reference, 3 dB is twice as much power, 6 dB is twice as much voltage, and 10 dB is twice as loud. Thus,

approximately nine times the sound power is required to make a sound twice as loud. Because the use of the term decibel implies a ratio, the reference quantity must be stated. Some typical reference frequencies are as follows:

dB SPL	Referred to threshold of hearing at 1 kHz
dBm	Reference 1 mW, usually in 600 ohms
dBV	Reference 1 V
dBu	Reference 0.7746 V
dBFS	Reference full scale

Deemphasis rolloff of a frequency range of an audio signal, usually corresponding to a preemphasis applied earlier in the audio chain. The object of the preemphasis/deemphasis loop is generally to trade off *headroom* against *signal-to-noise ratio*, both versus *frequency*. That is, a strong high-frequency boost applied to audio before transmission of FM radio signals is counteracted by an equal and opposite high-frequency rolloff in reception, with the advantage of reduced high-frequency noise but the disadvantage of equally reduced high-frequency headroom.

Degauss to demagnetize. When done deliberately in a degausser, this means first applying a strong AC magnetic field that can continuously reverse the state of all magnetic domains in the object being demagnetized and then decreasing the field strength in an orderly way to zero. This process leaves the state of the various magnetic domains utterly random and thus demagnetized.

Desk (British usage) synonym for *console*.

Diffraction the property of sound that permits it to be heard around corners; sound encountering an edge will reradiate into the space beyond the edge.

Diffusion the property of acoustical barriers to provide usually desirable scattering in reflected sound. An example of a lack of diffusion is the flutter echoes arising between two flat, parallel, hard walls. Good diffusion promotes smooth-sounding reverberation, without discrete audible events such as echoes.

Digital a signal that has been converted to numerical representation by means of *quantizing* the amplitude domain of the signal (putting amplitudes into bins, each bin denoted by a number). This means that for each amplitude a number is derived for storage or transmission. This process must be done frequently enough to adequately capture the signal through *sampling*. The advantage of digital is that to the extent that the numbers are incorruptible by means of error-protecting coding, the signal will be recovered within the limits of the original quantizing process, despite the number of generations or transmission paths encountered by the signal.

Digital audio the use of *digital* techniques to record, store, and transmit audio. Audio, once digitized, may also be processed in a huge variety of ways in the digital domain, some of which are impossible or impractical to do in the analog domain. Digital reverberators, for example, which construct models of acoustic spaces mathematically through a variety of tactics, may produce a very convincing sonic illusion of space.

Digital delay line a method of producing a time delay of an audio or video signal, by analog-to-digital conversion at the input of the device, delaying by means of digital memory, and subsequent conversion from digital-to-analog representation at the output. Delay lines have many uses in both audio and video signal processing. In audio, the output of the delay is added together with the original signal to produce a range of effects from "thickening" a voice to discrete echoes. In video, one-line delay lines are used for many purposes, including color decoding of composite video signals.

Digital recording any method of recording, be it on tape, disk, or film, that uses digital *sampling* and *quantization*.

DIN the German standards organization.

Dip filter an electrical circuit having one, usually tunable, frequency that is suppressed, while passing audio at both lower and higher frequencies. A dip filter is most useful for removing tonal noise from recordings, such as hum or certain whistles. Engineers are more likely to call this a *notch filter*.

Direct sound that sound that arrives by way of the shortest path from source to receiver within line of sight.

Directional microphone a microphone exhibiting any *microphone polar pattern* other than omnidirectional.

Directivity the factor describing a preference for one direction of sound propagation over another; a highly directional sound source is said to have high directivity or high Q.

Distortion any undesired alteration of the signal that is related to the signal (this definition excludes noise). Distortion is separable into two types, linear and nonlinear. Linear distortions include *frequency response* and phase errors and are characterized by reversibility by equal and opposite signal processing. Nonlinear distortions involve the generation of new frequency components not present in the original and are thus not generally reversible.

The simplest nonlinear distortion is harmonic distortion. With harmonic distortion, a sine wave, passed through a distorting device, appears at the output with added frequency components at harmonic intervals above, and potentially below, the original input frequency. Thus, if a 1-kHz sine-wave input suffers harmonic distortion, there will be an output of at least one of the following: 2 kHz, 3 kHz, 4 kHz, etc. Total harmonic distortion is often used in the measurement of audio equipment and is the root mean square sum[2] of the amplitude of all of the harmonics in the output of a device, often expressed as a percentage of the amplitude of the sine wave that appears at the output.

All other distortions are called "intermodulation distortions" because they involve the input of more than one sine wave (or sometimes square waves combined with sine waves) into a device under test and the examination of the output for the amplitude of energy at new frequencies not present in the input. One type is called "SMPTE intermodulation distortion" and involves the combination of a low-frequency sine wave such as 60 Hz and a high-frequency sine wave such as 8 kHz. The output is examined for the level of distortion components at

---

[2]The square root of the sum of the squares.

$f_2 + f_1$ and $f_2 - f_1$, which is basically a measure of how much amplitude or frequency modulation the high-frequency tone undergoes during a cycle of the low-frequency tone. Another intermodulation distortion test involves using two high-frequency sine waves and generally examining the output for the distortion product at the frequency $f_2 - f_1$. This is called "difference-tone intermodulation distortion."

DM&E (dialog, music, and effects) master that is intended to be used with the three stated elements summed together in the proportions 1:1:1 but kept separate in case the primary language or balance must be changed. A three-track DM&E is useful only for monaural work because each source track is only mono, but the idea may be extended to stereo through the use of multiple channels for each of the elements.

Dolby a set of technologies introduced by Dolby Laboratories with applications in professional and consumer audio.

For motion-picture-based program material, the use of the Dolby 4:2:4 amplitude-and-phase coding matrix has made practical widespread use of four-channel stereo, with left, center, and right screen channels and a surround channel. In this system, four channels of information are encoded into the two available tracks on the medium, such as the two tracks of an analog optical soundtrack, or the two tracks on a Hi-Fi VCR. Suitable decoders in the theater or home can then decode the signal into four channels for application to appropriate loudspeakers. This process uses the trade name on the film program material of Dolby Stereo or Dolby's double-D logo and is decoded with Dolby Pro Logic or Pro Logic II decoders.

Dolby is also active in the area of low-bitrate *digital audio* coders, particularly AC-3. When recorded on release prints as SR-D, AC-3 code is used at 320 kbps for 5.1 channels, recorded as blocks of data in the area between the perforations on the soundtrack side of the print. This area is not used by any other process, so it permits recording along with an analog track and either one or both of the other digital sound for film formats, DTS and SDDS. AC-3 carries the name Dolby Digital in the consumer marketplace on the DVD and elsewhere and may use up to 448 kbps data rate for 5.1 channels.

The use of the term Dolby may be confusing in some contexts, because more than one of the systems described earlier may be used together for one application. For conventional Dolby stereo analog optical sound on film for example, Dolby A- or SR-type noise reduction is used, and the 4:2:4 matrix is also used.

Doppler effect the effect on frequency caused by the source or the receiver of sound moving, exemplified by the pitch drop of a train whistle as it passes an observer. The Doppler effect is put to good advantage during postproduction of motion pictures by using a digital pitch shifter to simulate objects in motion.

Double perf abbreviation for double perforation, used to describe 16mm film perforated along both edges, suitable only for center track recording, which was more rare than edge track recording on single-perf film.

Double system a system wherein the picture and sound elements are on separate strands of film or tape, or in a digital audio workstation, designed to be synchronized during reproduction by mechanical or electronic means. The separate pieces of film to be synchronized must be unambiguously marked with edit sync marks so that they may be threaded up in sync. Usually double system is used until nearly the end of postproduction, when married answer prints become available, with sound and picture on one piece of film.

Dropout a momentary loss of signal, usually applied to audio or video signals recovered from a medium or transmission path. The signal loss may not be complete but may result only in a change in level.

Drum (aka impedance drum, inertial roller) a cylindrical rotating element in the path of a film or tape transport having substantial mass. The tape or film wraps around the element, whose inertia helps to reduce speed variations and thus to improve *wow and flutter* performance. In the Davis loop drive design used on many film sound transports, the tape path through the head assembly is immediately preceded and followed by such drums.

DTS (digital theater systems) method of recording sound on a CD-ROM format disk and following a custom time-code track on release prints, placed between the analog soundtrack area and the picture. This space is used for no other purpose, so prints may contain an analog track and either one or both of the other two competing digital sound for film formats, Dolby SR-D and SDDS.

Dub see *dubbing*.

Dubber a playback-equipped only motion-picture film transport designed to play various soundtrack formats as required.

Dubbing
1. The process of rerecording soundtracks. This occurs when the *elements* or *units* cut by sound editors are dubbed into *premixes*. Subsequently, a final mix is prepared by dubbing the various premixes to make the final mix *stems*. Then the stems are used in preparation of the various *print masters* needed for all the release media. Dubbing uses a wide variety of processes, including equalization, filtering, dynamic range control such as expansion and limiting, noise-reduction devices, reverberation, pitch shifting, panning, and many others, to produce a desired mix from the individual elements. Subsequent processing includes limiting to the capabilities and desires of various media: A digital theatrical release has both a greater dynamic range capability and a larger desired range than, say, a mix for broadcast television.
2. Any copying operation, particularly applied to videotape.

Dubbing mixer (British usage) synonym for rerecording mixer.

Dummy synonym for *dubber*.

Dynamic microphone a transducer that changes acoustical energy into electrical energy by the motion of a conductor in the field of a magnet. A typical construction involves a diaphragm exposed to a sound field and allowed to move. Attached to the diaphragm is a voice coil, which in turn is suspended in the field of a magnet. The motion of the diaphragm moves the voice coil in and out of the field, thus

cutting the lines of magnetic force with a conductor. Dynamic microphones have a reputation for high reliability and ruggedness and may be built using most of the common *microphone polar patterns*. A special case of the dynamic microphone is the *ribbon microphone*, using the same principle for electroacoustical conversion, but with a different construction.

Dynamic range the range stated in decibels from the background noise level to the overload level of a medium. Dynamic range may be stated in terms of audio-band-wide noise to midrange overload or in a variety of other manners. A complete description would include a graph of the headroom and of the noise versus frequency.

Early reflections sound arriving at a receiver by way of one, two, or a few reflections off the environment. Distinguished from reverberation because the individual reflections can still be discriminated, at least by instruments. What is generally considered the psychoacoustically useful part of early reflections for live sound events usually lies between 0 and 50–80 msec of the direct sound, whereas later-arriving discrete reflections may be heard as echoes. On the other hand, for reproduction systems, early reflections can alter timbre, location, and spaciousness of a sound event and may be regarded in this role as detrimental to reproduction of a soundtrack.

EBU European Broadcasting Union, a group of European broadcasters who issue standards, among other things.

Echo a discrete repetition of a source signal, distinguished by its separation from reverberation by its greater level. To be defined as an echo, the repetition must be late enough (>50 to 80 msec) to not be integrated into the source sound by the ear and high enough in level to be distinct from reverberation. A special case of echo is the "flutter echo," occurring when two parallel surfaces reflect acoustical energy back and forth between them. The result is a characteristic patterned reflection.

Echo unit an older term generally used to describe what is thought of today as a reverberator. *Echo* means specifically what is heard as a discrete repetition of a sound, and that is not generally what is meant when the term echo unit is used. On the other hand, a *digital delay line* can be used to produce a discrete echo.

Eigentones synonym for *room modes*.

Eight track see *track formats for tape recording*.

Electret microphone an *electrostatic microphone* type in which the insulator is made of a plastic material that can be permanently charged electrostatically so that the microphone needs no externally applied polarization voltage. Electret microphones are usually less expensive than microphones using air as an insulator. The principle is used for everything from very high volume cheap microphones for answering machines to recording microphones.

Electroacoustic response the *frequency response* of a complete sound system, including loudspeakers and the effects of room acoustics measured with microphones. The electroacoustic response is usually measured with either a *real-time analyzer*

(RTA) or a time-domain-based analyzer, such as time-domain spectrometry, maximum-length sequence, or Fast Fourier Transform analyzers. For use during original design, the utility of the time-domain-based analyzers cannot be denied, but in day-to-day use as an alignment tool for sound systems, the use of the RTA predominates. The electroacoustic response for both kinds of analyzers is made much more reliable and repeatable by spatial averaging, and the output of an RTA is made more repeatable when used with computed temporal averaging.

Electrostatic microphone a transducer capable of changing acoustic energy to electrical energy by means of measurement of the small changes in electrical capacitance that result as one or more diaphragms move with respect to a fixed back plate. Both the diaphragm(s) and the back plate are made electrically conductive and are separated by air. In one type, the capacitance is measured by applying a bias voltage, usually between 48 and 200 VDC, between the diaphragm(s) and the back plate, and measuring the AC voltage resulting from motion between the elements. Alternately, in the rf (radio frequency) electrostatic microphone (not to be confused with the radio microphone, which radiates the output of a microphone on radio waves for pickup at a distance), the varying capacitance is used to frequency modulate a radio-frequency oscillator, and a frequency-modulation detector is used to demodulate the audio, all within the microphone body. Because of its lack of a bias voltage, the rf radio microphone is claimed to have high reliability in the field. In a third type, called *electret microphones*, the back plate is permanently polarized with an electrical charge during manufacture, and electronics convert the changing capacitance with diaphragm motion into a voltage. Electrostatic microphones may be built having many different polar patterns: omnidirectional, bidirectional, cardioid, hypercardioid, and superdirectional microphone configurations are all possible. See *microphone polar pattern*.

Element(s) a single strand of film cut by a sound editor to synchronize with the picture for use in *rerecording*. The original elements are broken down by discipline into dialog, music, and sound effects. Within each category there is further breakdown, particularly with sound effects, into, for example, *ambience, Foley*, and principal effects, called simply *cut effects*. The breakdown is done so that the process of premixing the elements can be accomplished readily. Many elements contain a large amount of fill leader used as spacing between sounds just to maintain synchronization with the picture.

Envelop to surround or immerse oneself in a sound field, as in, "the surround sound of the jungle in the tiger scene in *Apocalypse Now* envelops the listener."

Equal-loudness contours a set of curves plotting amplitude versus frequency characterizing the sensitivity of human listeners across the audible frequency range at various levels. The lowest curve is the minimum audible sound field versus frequency, and it shows that human hearing is most sensitive to midrange frequencies centered around 2–4 kHz, with less sensitivity at lower and higher frequencies. The curves also

show that at no sound pressure level does human hearing have a flat *frequency response*. Although the best known of these curves are named after the original researchers who found the effect, Fletcher–Munson, more recent work is more widely accepted and has been standardized as an ISO standard.

Equalizer a device to produce *equalization*. The device may produce fixed equalization, as in compensating for deliberate response variations (such as RIAA equalization used on phonograph records) or it may be adjustable. Among adjustable types, some are designed as *program equalizers* for users to adjust by listening, and others are designed to be set using the results of instrumented measurements, such as *one-third-octave equalization*.

Equalization generally, equalization means to adjust the *frequency response* of audio by means of either fixed or adjustable devices.

1. Fixed equalization consisting of standardized frequency response alterations (bass boost or cut; treble boost or cut) is usually used to correct an audio system for overall flat response from input to output, where it may be impractical or undesirable to have flat response on the medium itself. Examples include phonograph records, all analog tape and film audio formats, and the compact disc when using the optional "emphasis." Equalization may be applied both before and after a medium: In the first case it is called "record equalization" or *preemphasis*, and in the second "playback equalization" or *deemphasis*. The purpose of the process is to tailor the capabilities of the medium to the needs of the signal. Thus, for example, if there is little energy in expected signals at high frequencies, it is useful to boost highs on record and to cut them correspondingly in playback. The result is a reduction in noise of the medium because it encounters only the playback cut. The downside of the process occurs if an occasional signal violates the assumptions under which the equalization was developed: If strong high frequencies should occur along with treble-boosted preemphasis, then distortion would result.

2. Program equalization is any deliberate frequency response alteration, such as the use of tone controls or equalizers to improve sound.

European Broadcasting Union a standards-making body of European broadcasters, concerned especially with the interchange of audio and video programs.

Exciter lamp the light bulb that, together with optics, provides the illumination for playing optical soundtracks. Today the traditional incandescent exciter lamp has been largely replaced by arrays of red-light-emitting diodes.

Expansion the opposite of compression; any device or process that increases the volume range of source material. An "expander" is an amplifier whose gain is deliberately controlled by the signal, effectively turning the volume down in low-level passages and turning the volume up in high-level passages. A special case of the expander is a downward expander that expands the signal only below a threshold, usually adjustable. The purpose of a downward expander is to

reduce noise by adjusting the threshold level to be set between the level of the desired program material and the level of the noise. Then only passages in which the only content is the noise will be subject to downward expansion and a useful noise reduction occurs.

Fader a level control, also known as a "pot" (for potentiometer), a gain control, or a volume control.

Feedback, acoustic (aka howl round) the familiar sound system squeal occurring when an open microphone output is amplified and applied to a loudspeaker with sufficient gain. It seems to always precede a line from a character who is speaking through a sound reinforcement system in movies—one of the silliest movie clichés.

Filtering removing a part of the frequency spectrum, usually performed to reduce the effect of undesired sound lying within the rejected frequency range in favor of the desired sound lying within the passband of the filter. Categorized by highpass, lowpass, bandpass, bandstop, or notch filters, depending on shape and what range is rejected. Note that a *highpass filter* is one that "passes" highs; it rejects lows. The term "high filter" usually means, in contrast, that it is the highs that are being filtered, so a "high filter" is really a lowpass filter.

Final see *final mix*.

Final mix the outcome of mixing together the various component parts of a sound mix. In a preceding stage, individual sound *elements*, cut by sound editors, were rerecorded to form *premixes*, such as dialog, music, ambience effects, Foley effects, and principal on-screen sound effects. In final mixing, these are combined to produce the final mix *stems*, which are the number of tracks and strands of film or tape needed to completely represent the mix, while still keeping various kinds of sound separate so that they can be replaced for specific purposes, such as for foreign-language mixes.

The final mix stems usually remain broken down into at least separate dialog and combined music and effects component parts. In addition, stereo mixing calls for separate tracks for at least left, center, right, and surround, with two separate surrounds left and right and an extra low-frequency-only enhancement channel gaining acceptance. Stereophonically related program material must lie on one piece of film because otherwise the small time variations on different pieces of analog film would cause potential cancellations.

The final mix thus represents everything that is intended to be heard, mixed in proper proportion, but separated enough so that it is easy to prepare alternative versions. The final mix stems are mixed once again to the various *print masters* to represent the sound just as it will appear on the final motion-picture prints or video releases.

Fisher boom a brand name microphone *boom*.

Fishpole a light-weight, hand-held microphone pole widely used to get a microphone close to actors but out of the shot being filmed.

Fletcher–Munson curves the first measured *equal-loudness contours*, now considered obsolete, replaced by international standards.

Foley a kind of sound effect that is made in a sound studio while watching a picture. The most well known Foley effect is footsteps, but many other sounds are also recorded in this manner. The use of Foley effects started with the need for such kinds of sounds in pictures that were dubbed into foreign languages and lacked any of the other sound picked up on the set during *production sound recording*. The need grew as the field became more specialized and production sound mixers concentrated more on recording dialog; in these cases, it has been found expeditious to add many low-level kinds of sounds with the Foley process to make the sound seem more real.

Free-field microphone a measurement microphone calibrated for use with sound fields in which one principal direction of the field is predominant, such as in an anechoic chamber. Usually such microphones are used with the diaphragm perpendicular to the incident sound field.

Frequency for a sinusoidal wave, the number of occurrences of one complete wave per second. The unit for frequency is hertz (Hz), equal to one cycle per second. For a complex wave consisting of many sinusoids, typically the lowest occurring is the fundamental frequency, which, along with its harmonic overtones (also called "partials," such as second, third, fourth, . . .), together form the waveform. Frequency is related to wavelength and the speed of sound through the mathematical relation $f = c/\lambda$, where f is the frequency in hertz, c is the speed of sound in the medium (1130 feet per second nominally in air), and λ is the wavelength in the medium.

Frequency analysis See *spectrum analysis*.

Frequency masking see *masking*, *frequency*.

Frequency response the amplitude output with respect to frequency characteristic of a system under test. One method of measurement is to drive the system with a sine wave that sweeps from one end of the desired frequency range to the other while measuring the amplitude of the output. The resulting curve of amplitude versus frequency, usually plotted in decibels of amplitude versus logarithmic frequency, is a "frequency-response curve."

Full coat 35mm perforated film coated with magnetic oxide across at least the full area between the perforations, and usually fully across the width. It may be recorded in track formats accommodating three, four, or six tracks. (Nothing prevents recording just one track, but these recordings are generally made on *stripe coat* film.)

Futz to deliberately distort for a desired effect.

Galvanometer recording one of the methods of producing a variable-area optical image to represent sound. The galvanometer is a voltage-sensitive coil of wire placed in a magnetic field and connected to a mirror. Current through the coil moves the coil and, thus, the mirror, which, with a light source and appropriate lenses and masks, produces the varying area needed to expose the sound negative.

Graphic equalizer an audio *equalizer* having a multiplicity of controls arranged so that in use they draw out a visual representation of the *frequency-response* curve that they produce.

Group delay A measure of the difference in time that it takes a signal of various frequencies to pass through a system compared to a reference frequency. The time for a signal at the reference frequency to transit from input to output is latency. Group delay is the deviation from the latency as a function of frequency. See Blauert and Laws, "Group delay distortions in electroacoustical systems," *J. Acoust. Soc. Am.,* 63(5), May 1978.

Harmonic distortion when a pure sine wave is applied to a system, the output of the system may contain the original sine wave accompanied by sine waves at harmonic frequencies of $2 \times f$, $3 \times f$, etc. Harmonic distortion is caused by a lack of perfect replication of the original sine wave.

Head alignment a term applied to film and tape recording of analog or digital and to audio, video, or other signals meaning the correct positioning of the head with respect to the tape or film. Important considerations include the need for continuous intimate contact between the tape or film and the head, or else severe spacing loss will occur. Head alignment terms include *azimuth, footprint, gap, height, rotation*, and *zenith*.
  Azimuth the angle formed between the reference edge of the film or tape and the recording. It is 90 ° in analog audio recording, but other angles are used for other purposes, particularly on helical-scan analog and digital videotape, to permit closer spacing of the tracks.
  Footprint the area of the tape or film that is in contact with the head. It is typical to check that the footprint produces an area that is centered on the.

Headphones small electroacoustic transducers meant to be worn by individual users and classified according to whether they are worn with contact around the outer ear, sit on the outer ear, or are inserted in the ear canal.

Headroom the amount of signal-handling capacity above a reference level before reaching a specified distortion. For example, "the headroom of magnetic film at 90 feet per minute is +17 dB for 3 percent total harmonic distortion above a reference level of 185 nW/m."

Headset see *headphones*.

Hertz the unit of frequency in cycles per second. kHz is thousands of Hz; MHz is millions of Hz.

Hi-Fi tracks audio recording recorded to and replayed from the tape by heads on the video-scanning drum of VHS, S-VHS, Betacam, and other video recorders. Kept separate from the video by the use of a different carrier frequency (rather like tuning into a different station on a radio), a differing azimuth (see *head alignment*) from the video, and other tricks. Better quality than longitudinal recording because the relative tape-to-head speed is much higher (the heads are rotating). Uses a specialized companding noise-reduction system that can sometimes be tricked by video, with the result of hearing a video-based buzz coming and going with the audio.

High filter a source of unending confusion, this is a term used in consumer audio for a filter that rolls off high frequencies.

The corresponding professional term is "lowpass filter" because the standard used by engineers is to define what is passed rather than what is rejected in the name of the filter.

Highpass filter an electrical filter that passes frequencies above a specified and usually adjustable one while rejecting frequencies below the cutoff frequency. Highpass filters are most often used to reduce the effects of background noise, which often has predominant low-frequency content, while minimally disturbing the desired signal, such as voice. For example, a highpass filter set to 80 Hz, with a steep rolloff below cutoff, can be very useful in reducing the impact of traffic noise, while minimally affecting a voice recording. In consumer and semiprofessional applications, this may be called a "low filter."

Horn an acoustical amplifying device known since antiquity (the megaphone is an example, or even cupping one's hands to shout) whose principal purpose is to provide better coupling between an acoustical source and the air. The basic configuration of a horn is to control the expansion of sound waves from a small source at the throat of the horn to a larger end radiating to the air, called the "mouth." Horns come in many different shapes, as progress over 75 years has dictated ever-higher levels of performance, particularly with respect to uniformity of output coverage over differing output angles. Early exponential horns, while still in use, were known by the 1930s to be not very uniform with frequency, collapsing their output to smaller angles while moving upward in frequency. The 1940s brought "multicell horns," with dividers placed in the mouth of the horn to better "steer" high frequencies across the desired output angles. These horns, although an improvement over the simplest ones, suffered from nonuniform coverage because different sources from the different cells interfered with one another at high frequencies. A 1980s development of horns is the constant directivity type, which is nontraditional in shape, but better performing than earlier types for most purposes. The drivers coupled to the throat of a horn are called *compression drivers*.

Howl round synonym for acoustic feedback. See *feedback, acoustic*.

Hz hertz.

IEEE see *Institute of Electrical and Electronic Engineers*.

Infrasonic frequency range the region below the audible frequency range, from extremely low frequencies associated with slow changes in air pressure up to 20 Hz. Very high levels of infrasonic acoustic energy, especially around 12 Hz, can lead to discomfort and nausea.

Institute of Electrical and Electronic Engineers a very large association of engineers, with standards making and honors and awards among their activities.

Intensity among acousticians, used precisely—the complex characteristics of sound propagation, quantifying both amplitude and direction of sound. Slang-used interchangeably with amplitude, but applied usually to sound pressure. Street slang—loudness.

International Electrotechnical Commission (IEC) an international standards-making body located in Geneva, Switzerland, with activity in audio, electroacoustics, video, television, and photography, among others. It is probably best known to the lay public as the organization that standardized the categorization of Compact Cassette tape types in four types, numbered I, II, III, and IV (see *compact cassette*), and film speeds. Its standards are generally available within each country from the local standards-making body, such as ANSI in the United States and B.S. in England.

International Standards Organization (ISO) an international standards-making body located in Geneva, Switzerland, with activity in acoustics and cinematography, among others. It is probably best known in audio as the organization that standardized the electroacoustic response curve of motion-picture theaters (the X-curve). Its standards are generally available within each country from the local standards-making body, such as ANSI in the United States and B.S. in England.

Intermodulation distortion see *distortion*.

Inverse square law the property of sound waves to spread out from a point source as a function of distance leads to a decreasing intensity of sound with distance. The "law," or rule, that the fall-off follows is a "square-law," that is, for twice the distance the amplitude is reduced by four times.

Iron oxide one of several possible active ingredients in a magnetic coating for audiotape and videotape or film.

kHz kilohertz (see *hertz*).

Layback the process of copying audio from a sound-only master "back" to a video master tape. The processes of *laydown* and layback together permit sound postproduction to be done on media especially well suited to audio.

Laydown the process of copying audio from the soundtracks of an original videotape onto the postproduction audio recording medium.

Layover any audio copying stage in postproduction, especially in television postproduction, also called dubbing or x-copying.

Limiter an electronic device that acts to keep the amplitude of signals from exceeding a usually adjustable maximum amount. The limiter "turns down" the signal by the amount by which it exceeds the maximum as required so that the output contains no higher level than the maximum.

Limiting the process of applying a *limiter*.

Logarithmic scaled along an axis on which equal intervals are related by powers of 10: 1, 10, 100, 1000, 10,000, 100,000, etc., lie equally spaced.

Logs (production sound) see *production sound logs*.

Longitudinal
1. Generally, any recording that occurs along the length of a piece of tape, used with video recording to describe recordings made with other than the rotating head assembly.
2. An audio recording made along the edge of a piece of videotape by conventional direct analog magnetic recording

methods. Limited in quality by the nature of the tape (optimized for video recording) and by the low tape speed.

Loudness subjective strength of a sound field, a complex function of sound pressure level, spectrum, duration, and other factors.

Loudspeaker an electroacoustic transducer that takes an electrical input and converts it to sound meant to be heard by a number of people.

Low filter see *highpass filter*.

Lowpass filter an electrical filter that passes signals at frequencies below a cutoff frequency while attenuating frequencies above cutoff. The term is used by electrical engineers and by professional audio personnel, but note that the term used for the same thing in the consumer marketplace is *high filter*.

LT RT (literally, left total right total) two-track recording format that embeds information content from four channels—left, center, right, and surround—through the use of an amplitude-phase matrix. Also used to describe a *print master* recorded in this format.

Magnetic pertaining to the property of certain materials observed in antiquity to attract one another and one of the fundamental forces of nature. Magnetism may occur naturally or may be induced by means of electromagnetism arising from, say, a coil of wire energized from an electrical source.

Married print any motion-picture print that contains an integral soundtrack.

Masking, frequency the property of hearing that makes one frequency cause another to be inaudible to human listeners. Masking is more effective close to the masker tone or noise than further away in frequency, is more effective toward high frequencies than toward low frequencies, and is more effective at higher levels than at lower ones.

Masking, temporal the property of hearing that causes one sound to audibly suppress another, despite their not being precisely simultaneous in time. Divided into premasking, in which the sound covered up occurs very slightly before the masking sound, and postmasking, in which the sound covered up occurs after the masking sound. A use by filmmakers of premasking is in close cutting of dialog. An editor will cut a plosive "p" with a butt splice to make it produce premasking, which will tend to hide the edit.

Mic (also spelled mike) slang for microphone.

Microphone an electroacoustic transducer responsive to sound pressure, velocity, or intensity, producing an electrical output.

Microphone polar patterns the directional character of a microphone traced out by varying the angle between a free sound field and the microphone body.

    Bidirectional microphone a microphone directivity pattern that responds to sound energy from two sides, in a figure-8-shaped pattern. Bidirectional microphones are most likely to be ribbon or condenser types.

    Cardioid microphone a microphone directivity pattern that responds to sound energy in a generally heart-shaped pattern, with a null point located to the back of the microphone. Useful because by aiming the microphone at a source, the relative pickup of the source compared with noise and reverberation in the room is enhanced over omnidirectional microphones. In addition to pointing the sensitive side of the microphone at the desired source of sound, another useful property of the cardioid is that one can "point" the null point at the rear of the microphone at an offending noise source.

    Hypercardioid microphone similar to the cardioid microphone but having a narrower front angle of acceptance, two symmetrical nulls rather than one straight to the rear, and a small "lobe" showing sensitivity to sounds from the rear of the microphone.

    Interference tube (shotgun or rifle) microphone a microphone with a strong frontal sensitivity and with all other directions deemphasized. It has the highest ratio of direct-to-reverberant sound energy of all normally used microphones, but suffers from changing directivity with frequency when used in any practical length. Also, most models show strong off-axis frequency-response coloration. Still, probably the most often used production sound microphone.

    Omnidirectional microphone a microphone directivity pattern that responds to sound energy nearly equally well from all directions.

MIDI (musical instrument device interface) a popular interconnection bus for synthesizers, drum machines, and other musical instruments capable of electronic control. May be interfaced to SMPTE time code to synchronize the action of the MIDI bus with film or video.

Mix used both as a noun and as a verb to describe the activity required to produce a complete soundtrack from the various sources required. Mixing activities may be further broken down into premixing, final mixing, and print mastering. In premixing, the original cut sound *units* or *elements* are hung on *dubbers*, which run many strands of film in synchronization. Each of the outputs of the dubbers forms an input to the *console*. The output of the console is recorded on a new piece of film, called a *premix*. Sound editors and sound designers usually break down the premixes first into dialog, music, and sound effects and often further into the type of sound effects such as *Foley, ambience*, and conventional cut effects. Following premixing, the various premixes are hung on dubbers and the console is used to produce a *final mix*. The final mix may still be recorded on many tracks, and even across multiple strands of film, forming the final mix *stems*. Then, *print masters* may be prepared for each of the different release formats. Keeping the dialog separate from the other parts of the sound in a separate stem makes foreign-language dialog mixing simpler because the dialog stem may simply be replaced by one prepared in the new language and then new print masters made for the foreign-language dub.

Mixdown synonym for *mix*.

Monaural having only one audio channel.

Monophonic synonym for *monaural*.

Multitrack pertaining to audio formats using more channels than the release format used for flexibility in mixing.

Mute silent, especially a control to turn off an audio signal.

Mylar a tough plastic material used as the backing for magnetic films and tapes.

NC curves a set of curves used to rate the background noise of spaces in terms of human perception of noise. The NC curves permit a higher measured sound pressure level of low-frequency noise than of high-frequency noise because human hearing is less sensitive to low-frequency noise than to high. Some typical ratings are as follows:

Recording studios	NC 10
Dubbing stages	NC 20
High-quality theaters	NC 25
Maximum permissible in theaters	NC 30

In particular, the NC-10 rating for studios is lower than historical ratings given for these spaces. This is because of the advent of improved recording methods, with their wider dynamic range.

Neopilottone see *pilot tone*.

Noise technically a sound is said to be noise-like when the signal cannot be distinguished as having tonal properties. Examples include a waterfall, most jet-engine noise, and most air-conditioning background noise. In more common usage, noise is any undesired signal.

Noise filter usually a *lowpass filter*, sometimes made dynamic by responding to the high-frequency content of program material. The idea is that if there is high-frequency program material it will mask high-frequency noise, so that there should be no filtering, but in the absence of high-frequency program, the filter should be applied to attenuate noise that often takes on the audible character of high-frequency hiss.

Notch filter an electrical filter arranged to produce generally sharp attenuation in a fairly narrow frequency band, passing frequencies both lower and higher than it without attenuation. Notch filters are particularly useful in attenuating tonal noise, such as hum or whistles, while having minimum audible effect on the program material.

Omnidirectional see *microphone polar patterns*.

One-third-octave equalization adjustment of the *frequency response* of a sound system, which usually includes loudspeakers operating in a room, to produce a standardized or desired octave-to-octave spectral balance. Equalization typically is performed by means of a set of boost or cut *equalizers* located on center frequencies uniformly spaced at intervals of approximately one-third of an octave over the frequency range of interest, such as from 25 Hz to 20 kHz. The frequency centers are usually taken from the ISO series 25, 31.5, 40, 50, 63, 80, 100, 125, 250, etc., rationalizing the strictly calculated center frequencies to nearby round numbers. The one-third-octave spacing was chosen by the marketplace as close enough to the critical bands of human hearing (which are closer to one-fifth octave above 200 Hz) to do productive work without

excessive cost and is the most popular spacing in use. For aligning sound systems in real rooms, the use of multiple microphones for spatial averaging and computed temporal averaging is dictated for reliability of the results.

Optical sound recording the process of recording sound photographically by means of an optical camera recording a sound negative, using a variable area to represent the analog soundtrack and up to three different digital sound formats, and subsequent development of the negative, printing, and print developing. The result is an economical means of supplying large quantities of release prints with sound, which can be printed in the same printer as the picture negative. Thus, its costs are much lower than magnetic sound recording on release prints and its use predominates.

Overdub the process of using a *multitrack* recorder for recording different parts at different times, playing back already recorded sounds from the *record head* as a synchronization reference. In this way, a prerecorded click track, for instance, can be played into headphones on musicians who then play their parts for recording on other tracks. Synchronization in time is ensured by playing from the same head that is recording on separate tracks. See also *Sel-Sync*.

Oxide magnetically responsive material forming the basis of magnetic recording. It comes in a variety of types with different recording properties, especially suited to the needs of audio tape or videotape recording at a variety of densities of recording.

Parametric equalizer an audio *equalizer* with added controls that include such factors as center frequency of equalization and Q (relative sharpness of the *frequency response* curves) along with the boost/cut level controls available on other equalizers.

Peak meter a level meter that responds quickly to incoming signals. Peak meters are generally used to avoid audible distortion; they are not particularly useful for determining *loudness*, for which even the primitive *VU meter* is superior. A typical response time for a peak meter used for analog recording to reach full indication is 10 msec, whereas a digital one must be much faster to catch content that would be over the top of the coding range, that is, above 0 dB FS.

Perceptual coding a method of bit rate reduction that relies upon human perception for its models of what to save and what to throw away. Frequency and temporal masking by humans is exploited.

Phase pure sine waves can be described by only their amplitude and frequency. Most useful *signals*, on the other hand, are constituted of a multiplicity of sine waves simultaneously that change in time. For periodic waveforms, those with repeating waves, an additional element is needed beyond amplitude and frequency to completely describe the signal, and that element is phase. Phase is the relative starting point of the various component sine waves in a composite waveform. Thus, when a 1-kHz sine wave is mixed with a 3-kHz one, and the 1-kHz sine wave crosses 0 amplitude going in the positive direction at a point in time, but the 3-kHz sine wave

is at its crest, the 3-kHz sine wave is said to have a phase relative to the 1-kHz reference wave of $+90°$.

Phasing the process of making certain that the various paths in a sound system have identical polarity; that is, if presented with a positive-going increase at the input of the system, a positive-going signal will appear at all accessible points within the system. If sound systems are not phased correctly, undesirable cancellations of sound may occur.

Phon a unit of subjective loudness. By plotting a point at a frequency and sound pressure level on an equal-loudness contour graph, one can determine where such a point lies with respect to the equal-loudness contours. The loudness in phons for such a tone is the nearest loudness contour that can be drawn through the point. The reference for phons, which are given in decibels, is a sinusoidal tone at 1 kHz and 40 dB SPL.

Pilot tone a synchronization method for $\frac{1}{4}$-inch magnetic tape (see *crystal sync*). Using pilot tone, a precise 60-Hz tone (in the United States) is recorded to correspond to 24 fps. The recording is made with two small gaps out of phase with one another, and in the same *azimuth* as the main recording, so that the net sum of the two at a full-track playback head is 0, but a head having matching gaps, and wired with the correct polarity, can play the tone back.

Pink noise a random signal having equal energy in equal logarithmic units of frequency. Thus pink noise has equal energy in each one-third-octave band, for example, making it useful as a source in testing sound systems with a *real-time spectrum analyzer*.

Pinna the external ear parts. Sound fields approaching the ear from different angles encounter various reflections and shadowing from the head and outer ear. The resulting frequency-response changes can be detected as location and movement information.

Pitch the sensation of a tonal sound ascribed as lower or higher; affected mostly by the frequency of the sound, but also by *sound-pressure level* and *timbre*.

Post slang for *postproduction*.

Postproduction the phase of making motion-picture and television programs that occurs after principal photography or recording has been accomplished; includes picture and sound editing, special effects production, and sound mixing.

Postproduction sound the phase of making motion-picture and television programs that occurs after shooting of the production sound, including editing, mixing, and print mastering. The outcome of the process is a set of *print masters*, which represent the soundtrack formats needed to produce all the various release formats desired.

Pot (potentiometer) see *fader*.

Precedence effect (aka Law of the First Wavefront or Haas effect) localization in the direction of the first-arriving sound despite later reflections. It can be overcome if the later reflections are of sufficient amplitude.

Predub synonym for *premix*.

Preemphasis the process of applying *equalization* before a medium, combined with *deemphasis* to form a complete system.

Prelay applied in television applications both to a room and to the equipment used to rerecord sound effects from a sound library to an editorial medium, such as a digital audio workstation.

Premix one of the stages of preparation of soundtracks for film and television programs during which individual edited *units* of like kind are mixed together to form one complete component part of a final mix. Examples of premixes include *dialog, Foley*, and *ambience* premixes. Also called predub.

Presence (atmos and room tone) background noise recorded during *production sound* recording meant to match the production soundtrack so that it may be intercut with the track and provide a continuous-sounding background.

Pressure microphone a microphone whose output is proportional to varying air pressure caused by a sound field. Pressure microphones generally consist of a diaphragm enclosing a sealed chamber. It is the changes between the externally applied varying sound pressure and the fixed internal air pressure that the microphone converts into electrical energy. The term is used specifically with measurement microphones to indicate microphone types that are calibrated in such a way that the high-frequency response rises on axis because of the buildup of pressure in front of a barrier, which becomes increasingly true at higher frequencies in the case of microphone-size objects. Pressure-calibrated measurement microphones generally measure sound fields having a strong predominant single-direction component with the best flatness when used at grazing incidence, with the sound field passing across the diaphragm rather than perpendicular to it.

Principal photography that part of a motion-picture production in which the main actors are photographed and sound is recorded.

Print master a complete soundtrack containing all sound elements, such as dialog, music, and sound effects, recorded in a format compatible as to number of available tracks with the needs of mass production of prints or video discs or tapes.

Print-through a defect of analog magnetic tape recording wherein a recording made and wound onto a reel affects the adjacent layers, causing noticeable pre- and post-echoes. Print-through is exacerbated by time and increasing temperature.

Production sound sound recorded during *principal photography*.

Production sound logs written notes in prescribed form recording the content of each production sound tape. Usually included are scene and take information, printed take information through the convention of circling the printed take numbers, and other production and technical information needed to locate information and play back the tape correctly.

Production sound mixer the person in charge of recording on the set, who generally controls the microphone technique, operates the *audio mixer*, often in portable form, and operates the recorder.

Program equalization *equalization* deliberately added by a mixer to make the program better balanced in octave-to-octave response for the intended purpose, as opposed to "fixed equalization," which is used around media to compensate for deficiencies.

Propagation the process by which sound is distributed through a medium, by molecules nearer the source progressively imparting energy to ones farther away through mechanical vibration.

Punch-in/punch-out to begin and end recording, especially while rolling to record just one section of tape or film.

Quantized a signal once sampled is binned according to its amplitude. This means comparing the sampled amplitude to a staircase of numbered steps, and assigning the number of the closest step as the representative digital code.

Quantizing in digital representation of a signal, the process of "binning" or measuring the amplitude of each sample to represent it as a set of numbers. Then, the numbers are combined with additional numbers used to provide protection against errors and to tailor the stream of resulting numbers for a specific purpose, such as for a medium before recording or transmission. Some issues relevant to this process are the basic method, the number of bins, the regularity of the bins, and the deliberate addition of a small amount of noise, called "dither," which acts to smooth the transitions across the bin boundaries.

The most widely used method for audio today is 16-bit linear pulse code modulation, used for professional audio recording and the compact disc medium. This method uses a straight, uniformly stepped ramp to compare the amplitude of the signal against, producing one 16-bit binary number each sample period for each channel. Thus, the compact disk, for example, uses 16 bits/sample $\times$ 44,100 samples/sec $\times$ 2 channels = 1,411,200 bps, not counting error overhead, etc.

Radio frequency any frequency that may be used for electromagnetic modulation of a carrier by a signal, extending from near audio (for transmission underwater) to many megahertz.

Radio microphone a microphone system using a radio transmission link. Usually the transmission link will be by means of frequency modulation of a high-frequency carrier with noise-reduction companding for professional quality. Optionally, "diversity reception," that is, reception at more than one antenna with the receiver constantly examining the inputs and switching to the strongest one, helps to reduce the influence of multipath. Multipath results in distortion caused by a drop in signal level as the main path is interfered with by other radio frequency paths, which may cause complete signal cancellation at some positions of the transmitter and receiver.

Real-time spectrum analyzer a device for measuring the amplitude versus frequency response characteristic of a device or system under test that measures all parts of the desired frequency range simultaneously. Real-time analyzers generally use a bank of frequency-selective filters and a means of measuring the output levels from the bank of filters that is displayed as a graph of amplitude versus frequency.

Recording in audio, any means to store a signal for use at a later time is said to produce a recording. Short recording into temporary memory may also be called "sampling," but this is a secondary meaning. The means may be analog or digital, and the storage medium may vary among tape formats or magnetic or optical discs.

Recording galvanometer see *galvanometer*.

Reflection sound encountering a barrier may be reflected off of it. A flat barrier will reflect sound with a "specular reflection," just like a billiard ball encountering a bumper at the edge of a pool table. A shaped barrier can either concentrate or diffuse sound, depending on its shape. One example of a deliberately shaped barrier to capture sound energy is a whispering gallery, which has two points of focus, many feet apart, which conduct sound between them very well.

Rerecording the process of sound postproduction in which sound elements are combined into premixes, then again into final mix stems, and then finally into print masters. Because these processes do not involve live recording typically, the process became known as rerecording to distinguish it from original recording.

Reverberation acoustic reflections from the environment occurring so frequently at such close spacing in time that the individual reflections making up the ensemble can no longer be distinguished. Reverberation time is the amount of time for an interrupted steady-state sound field to decay away to 60 dB below its original sound pressure level.

Reverberator a device that by mechanical or by computer-calculated means supplies the effect of *reverberation*.

Ribbon microphone a microphone type that converts acoustical energy into electrical energy by means of a fine, light-weight conductive metal ribbon suspended in the field of a magnet. The ribbon moves in the magnetic field corresponding to the sound energy, and a voltage is obtained across the ends of the ribbon.

Rifle microphone a slang name for an interference-tube microphone having a club-shaped *microphone polar pattern*.

rms root mean square. A mathematical method of computing AC waveforms in terms of their equivalency with DC ones. For a sine wave, the rms value is 0.707 times the peak amplitude of the sine wave (zero to one peak). The resulting value is equal to the heating value of a DC voltage of that value. The line voltage 120 VAC rms has a peak value of 120/0.707 = 170 VAC peak, and that equals 340 VAC peak to peak. The average voltage of that same waveform, made by averaging all the values over the length of the sine wave, is 0.636 times the peak value, which equals 108 VAC average. Most voltmeters are calibrated in rms volts on a sine wave, although many actually measure average values and then scale them by 120/108 = 1.111 times to produce a reading, making them read the rms value on a sine wave. Potential confusion comes about when other kinds of waveforms are measured with unknown meters, like noise signals. With these, the use of a true rms meter is dictated. Note that sound pressure level and voltage are both normally specified as rms values.

Room modes see *standing waves*.

Room tone see *presence*.

Room synthesizer a specialized digital audio reverberator especially designed to simulate room acoustics. Room synthesizers are often used in postproduction to add spaciousness to "dry" recordings. The system of adding reverberation in postproduction is often preferable to recording the right amount in the first place on many sound elements because it is difficult to judge the effects of reverberation on only a single element of a complex soundtrack. Also, it is difficult if not impossible to reduce the amount of recorded reverberation and simple to add it, leading to the method of making generally dry recordings and then adding reverberation during mixing.

RTA see *real-time spectrum analyzer*.

Sampler a device that can store audio by means of *sampling* a waveform, digitizing it, and storing it away using a variety of means. A term prominent among musicians, not to be confused with definition 1 of *sampling*.

Sampling
1. The process of capturing the amplitude of a signal at regular intervals quickly enough that the desired audio signal is completely represented. The sampling theorem states that this process must occur at least somewhat more frequently than twice each cycle of the audio to be represented, so popular sampling frequencies have become 32 kHz (some broadcast links, especially in Europe), 44.1 kHz (consumer audio, including compact disc), and 48 kHz (professional audio). Because frequencies higher than one-half the sampling frequency would be misrepresented as aliases, that is, as frequencies other than themselves, much like wagon wheels when photographed at 24 fps may appear to run backward, antialiasing filters are used to strip off any ultrasonic components of the audio signal before sampling to prevent this effect. The sampling process also produces an aperture effect as the input frequency approaches one-half the sampling frequency, and this effect is compensated by adjustment of the frequency response with an equalizer.
2. Colloquial: the action done by a *sampler*.

SDDS Sony Dynamic Digital Sound, a system for recording sound on motion-picture release prints of up to 7.1 channels.

Shotgun microphone see *microphone polar pattern*.

Signal the desired information content of a program. Used as a term to distinguish the desired program from the undesired added effects of noise, or distortion.

Signal-to-noise ratio the amount that noise of a medium or transmission path lies below the reference level in the medium. For example, "the signal-to-noise ratio is 65 dB re: 185 nW/m." This is an incomplete specification because many conditions affect the measurement of signal to noise: Is the measurement weighted for human perception or measured flat? What kind of measuring voltmeter was used? A better specification is "the signal-to-noise ratio is 65 dB re: 185 nW/m, measured with flat weighting and an rms meter."

Single system a method of recording both picture and sound on one piece of film. The sound recording may be by either photographic or magnetic means, used to describe the type of sound camera often used for, for example, newsreel shooting, for which simplicity of setup is paramount in use.

SMPTE see *Society of Motion Picture and Television Engineers*.

Society of Motion Picture and Television Engineers (SMPTE) the U.S. group having standards-making activity in film and television. Certain of its standards are adopted as American National Standards.

Sone an alternate unit of subjective loudness. In addition to accounting for equal-loudness contours such as the phon, the sone goes one step further and scales the logarithmic measures so that twice the number of sones represents twice as loud. 40 dB SPL at 1 kHz = 1 phon = 1 sone; 50 dB SPL = 10 phons = 2 sones. Regrettably it is not as widely used as it might be, but it is specified for certain items, such as bathroom fans.

Sony Dynamic Digital Sound a method of recording digital audio in the 5.1- or 7.1-channel formats (7.1 adds left center and right center front channels to the 5.1-channel standard) in the area outside the perforations on motion-picture release prints. It may appear simultaneous on a print with an analog optical soundtrack and with either one or both of the competing digital audio formats, Dolby SR-D and DTS.

Sound motion of particles in a medium (air, water, steel) about a mean position.

Sound camera a motion-picture camera that is either inherently made quiet or is equipped with a noise-reducing blimp or barney so as to make the recording of sound possible without excessive camera acoustical noise interference.

Sound design the process of selecting sounds, editing them, and mixing them together to form a desired artistic effect, usually used as a term when sound is accompanying a picture. Sometimes today it is used to recognize a sound artist who makes complex sound effects for insertion in larger-scale productions.

Sound drum a mechanical cylinder having mass located in the path of film or tape transports used to impart greater stability of speed to the film or tape.

Sound-effects editing in traditional motion-picture production, the process of choosing sounds, then cutting sound elements or units into synchronization with the picture, and delivering them with the cue sheet to the dubbing stage for premixing. In television production, it is more likely that sound effects editing involves choosing sounds and then rerecording them to multitrack tape or to a digital audio workstation.

Sound head in production and postproduction, the term applies to the general class of tape or film heads, including erase, record, or playback heads. In cinema usage, it means the whole assembly used to recover sound from the film, including mechanical devices for guiding the film and stabilizing its speed, light sources needed for photographic

soundtracks, and a pickup device, such as a solar cell for photographic sound or a magnetic playback head for magnetic soundtracks.

Sound-level meter an instrument for measuring sound pressure level. Sound pressure is referenced to a small pressure variation near the threshold of hearing, 20 μN/m² rms, called 0 dB SPL.

Sound log a record kept by a production sound recording corresponding to each reel of tape used, noting scene and take information and including notes regarding any wild sound (not shot with the picture) recording. On sound logs it is conventional to circle the number of takes that the director wants printed so that only those takes will be transferred for dailies, saving time in transfer. However, note that even outtakes that are never seen at dailies can prove to be a fertile ground later for dialog editing.

Sound mixer an ambiguous term equally applicable to a sound console and to a production sound recordist.

Sound pressure the primary factor used to assess the amplitude of sound waves. Although sound particle velocity, intensity, and power can all be measured and indicate the strength of a sound field, sound pressure is usually preferred because it is the simplest to measure. Also, human hearing is basically a pressure receiver, because it is sound pressure that moves the eardrum.

Sound pressure level measured sound pressure, usually given in decibels re: a reference pressure corresponding approximately to the threshold of hearing of young adults. The dynamic range of human hearing at frequencies near 1 kHz is from the minimum audible sound around 0 dB SPL to the threshold of a sensation of tickle in the ear, approximately 120 dB. Common alternative reference levels used for microphone sensitivity specifications are 1 microbar = 74 dB SPL and 1 pascal = 94 dB SPL.

Soundtrack often used to describe the physical sound placement on a motion-picture release print containing sound. Also used to describe the recording contained on a motion-picture release print and sometimes to describe the musical score of a motion picture available in another medium.

Spectrum derived from the usage of the term in optics. Used in audio to describe a representation of the range from infrasonic through ultrasonic of a signal.

Spectrum analysis a method of representing a signal as a function of level versus frequency.

Speech mode the idea that the brain separately distinguishes speech from other sounds and processes it in a different part of the brain reserved for speech. The presence of a speech mode helps explain otherwise disparate observations about human hearing.

Speed of sound 1130 feet per second in air at 72°F. For the range −22 to +86°F, the speed is given by $c = 1052 + 1.106F$, where c is in ft/sec and F is in degrees Fahrenheit.

Standing waves a term in room acoustics used to describe the effects of sound interacting with multiple enclosure surfaces in ways that tend to strengthen or diminish the sound field at various frequencies, measured at a point in the space. Standing waves occur because sound waves cover a range of wavelengths that include wavelengths wherein integral ratios of waves fit precisely between the boundaries. At these frequencies, mutual reflection of the waves back and forth between two surfaces or among more surfaces produces points in space at which the sound pressure varies greatly (at the maxima) and where it varies very little (at the minima). Standing waves generally lead to a lack of uniformity of sound fields in enclosures, being worse in the bass frequencies and in small rooms.

Stems the constituent parts of a final mix. The various channels may be used to represent direction, such as left, center, right (corresponding to the picture), and surround and to represent the various kinds of sounds, such as dialog, music, and sound effects. The advantage of keeping the final mix as a multitrack representation is that it makes the preparation of subsequent special versions, such as foreign language release, simpler because all that needs to be done is to substitute only a new language element for the dialog, and then new print masters may be made as needed in the new language.

Stereophonic pertaining to any of a number of sound systems that can represent sound spatially, starting with conventional two-channel home stereo, up to multichannel sound systems often used to accompany pictures in large-scale presentations. For conventional film sound, what is usually meant today is 5.1 channel sound, with two surround channels and a dedicated low-frequency channel added to the front channels. It contrasts with single-channel monaural sound, whose only spatial capacity is the depth dimension along one line radiating away from the listener.

Stereophony synonym for stereophonic.

Sticks colloquial for clapperboard.

Streams See *auditory streaming*.

Stripe coat the alternative to full coat. Magnetic film that is coated with oxide in a stripe meant for recording a single channel. A "balance stripe" is applied near the opposite edge to produce better winding characteristics. Stripe coat is a term that typically applies only to such 35mm magnetic film stock.

Subdub synonym for *premix*.

Surround sound any of a series of stereophonic sound systems that uses at least left, center, and right screen channels and one or more surround-sound channels. One popular variation uses L, C, R, left surround, right surround, and a dedicated extra low-frequency channel. This system is called 5.1-channel sound.

SVA (stereo variable area) generic name applied to the Dolby Stereo format wherein two different bilateral sound records, side by side, carry LT RT information that can be decoded in a theater to in L, C, R, and S channels.

Sync sound sound recorded in conjunction with a picture, either single system or double system. In the case of double-system sync sound, a means must be provided to keep the picture and sound media in sync, such as *crystal sync*.

Temporal masking see *masking, temporal*.

Test tape a carefully prepared tape containing test signals for the alignment of playback on audio tape and videotape recorders. Most adjustments on tape machines start with making playback fit a set of standards so that when the record side is adjusted, it is done under a known set of playback conditions.

THX a set of technologies licensed by Lucasfilm, first developed for the cinema and subsequently for the home. In the theater, this means a specific electronic crossover, choice of power amplifiers and loudspeakers, local acoustical environment around the loudspeakers, global acoustical environment of the theater, and tuning of the theater sound system in accordance with standard SMPTE 202. In the home, both electronic- and loudspeaker-design tactics are used to better translate the program material prepared in a dubbing stage into the home environment.

Timbre the sensation about a tonal sound that cannot be described by loudness or pitch. Related to the strength of the various harmonics and the fundamental and how they change across time.

Time code (SMPTE time code) a method of encoding position into hours, minutes, seconds, and frames, along with other information, so that unsprocketed media, such as videotape, open reel tape, etc., can be run in frame-accurate synchronization by use of an electronic synchronizer. Time code occupies the space of one audio channel typically but may be recorded to a dedicated time-code channel or embedded with the video by means of a vertical interval time code placed in the video lines above the active picture area.

Tone a sound is said to be "tonal" when the constituent sine waves of which it is made can be distinguished. Tonal sounds usually can be assigned a pitch. Examples include an organ pipe, sine-wave oscillator, and whistle. Also used to describe the reference sine wave used for level and response adjustment, as in "send me the 1-K tone."

Track the physical area or other space on a medium such as film, tape, disc, etc., within which the record of the sound, picture, or data are recorded or printed.

Transfer the process of recording from one piece of tape or film to another, applied widely in audio and video.

Tweeter a high-frequency loudspeaker, combined with a *woofer* in a complete loudspeaker system for full-frequency range coverage. In theater usage it is more common to use the term horns to apply to the combination of a *compression driver* and *horn* that emits high frequencies.

Ultrasonic frequency range the region above the audible frequency range, above 20 kHz. The region is of interest to dogs, bats, and burglar alarms but not to people unless the amplitude is extremely high.

Unblooped pertaining to a motion-picture soundtrack negative wherein the discontinuities due to splices have not been minimized by making opaque the area near the splice with "blooping" tape or ink.

Unit synonym for *element*.

Unmodulated track an optical soundtrack containing no deliberate signal. For *variable-area soundtracks*, unmodulated track is made narrow when there is no signal present to reduce the noise caused by the optical grain of the film.

Volume range the range in decibels of a program, from the loudest passage to the softest. The volume range should fit within the *dynamic range*.

VU meter a meter intended to read the level of dynamic program material in a standardized fashion, used by audio mixers to improve the interchangeability of programs and to avoid audible distortion. The VU meter standards were originally designed so that a meter movement could be used directly without electronics to read the program level and, thus, VU meters are relatively slow, requiring about 300 msec to rise to 0 VU from rest. This makes the VU meter a better indicator of *loudness* than a *peak meter*. Although rigorously standardized in ANSI C16.5, many meters marked as VU meters do not meet the standard.

Walla speech babble recorded in sync with picture to indicate groups, usually of extras. No recognizable speech is intended.

Wave a disturbance in a medium that progresses through *propagation*. Transversal waves are perpendicular to the direction of travel, and longitudinal waves are in the direction of travel. Water waves are typically transversal; sound waves are longitudinal, that is, they oscillate to and fro in a way that expands and contracts along radial lines away from the source.

Waveform variation in amplitude of a quantity versus time, such as sound pressure versus time, voltage versus time, etc.

Wavelength for a sinusoidal pressure wave, the distance from peak to peak of the wave. The wavelengths corresponding to the audible frequency range cover from more than 50 ft to less than 1 inch.

White noise a random signal containing all frequencies containing equal energy for equal increments of frequency. In sound, reproduced white noise sounds quite "bright" because it contains as much energy between the high audio frequencies of 10 and 20 kHz as between 0 and 10 kHz. See also *pink noise*.

Wild sound sound recorded without a camera rolling. The term is used for sound gathered independent of a picture, such as a close-up sound of an effect that is going to be photographed, for use by sound editors in postproduction who must match a sound to what is seen on a picture.

Woofer a low-frequency *loudspeaker*. It is generally combined with a *tweeter* (in consumer terminology) or with a *compression driver* and *horn* (in typical theatrical usage) to cover the full frequency range.

Worldize to rerecord sound through a channel that includes a loudspeaker and a microphone, generally in an acoustical space. The purpose of worldizing is to make the sound seem as though it came from the space in question, rather than being a direct electrical recording. See also *futz*.

X-copy a copy made as closely as possible to a 1:1 match. In analog this applies to "equal flux" copying, wherein a copy is made at the same flux levels as the master.

X-curve one particular *electroacoustic response* curve having a prescribed high-frequency rolloff. Its use in motion-picture dubbing theaters, screening rooms, and exhibition theaters helps greatly to standardized what is heard by the audience. It is specified in SMPTE Standard 202 and ISO 2969.

Zenith see *head alignment*.

This page intentionally left blank

Note: Page numbers followed by *f* indicates figures and *t* indicates tables.

Tomlinson Holman is a Professor at the University of Southern California School of Cinematic Arts and at the Viterbi School of Engineering, Chief Scientist at Audyssey Laboratories, and is the founder of TMH Corporation. He was Chief Engineer of Advent Corporation and founded Apt Corporation, manufacturer of the Apt/Holman Preamplifier. Tom was subsequently Corporate Technical Director for Lucasfilm Ltd., where he pioneered the THX Sound System, Home THX, and introduced innovations into postproduction that became widely used. He named the most widely used multichannel sound system "5.1", and in recent years has been working on a 10.2-channel one.

As a consultant, Tom worked on many postproduction and other technical facilities, and as a recording engineer he made a number of successful recordings, including recordings with the Handel and Haydn Society of Boston, for which he served on the Board of Governors. His recording as engineer for Thomas Dunn's conducting of Handel's *Messiah* was reviewed in the Washington Post: "the recorded sound is superb."

Tom holds twenty-three U.S. and foreign patents and is a Fellow of Audio Engineering Society, the British Kinematograph Sound and Television Society, the IEEE, the Society of Motion Picture and Television Engineers, and the World Technology Network. He was awarded a Technical Achievement Award by the Academy of Motion Picture Arts and Sciences in March, 2002. He received a Career Achievement Award and an Honorary Membership from the Cinema Audio Society, an Honorary Membership from the Motion Picture Sound Editors, an Honorary Membership in the Society of Television Engineers, and a Lifetime Achievement Award from the Custom Electronic Design and Installation Association. In October 2006, he received the Silver Medal of the Audio Engineering Society "for more than three decades of work in the advancement of audio electronics, multichannel sound techniques, and motion picture sound production and reproduction systems." In January 2007, he received the IEEE Masaru Ibuka Consumer Electronics Award, "for engineering contributions and innovative developments in audio and cinema multichannel playback systems."

Tom Holman is experienced in standards making, having been chairman of IEC TC-100 Audio, Video, and Audio-Visual Systems for six years, chairman of the Audio Recording and Reproduction Technology of the SMPTE committee, and co-chaired the Audio Engineering Society Task Force on High Capacity Audio with John Eargle. He was one of five members who wrote the original AES 17 Digital Audio Measurement standard, and contributed to AES TD1001 on Surround Sound.

He is the author of three books, *Sound for Film and Television*, *Surround Sound Up and Running,* and *Sound for Digital Video,* all published by Focal Press. Various editions of these books have been translated into Japanese, Korean, and Chinese, and a local edition of *Sound for Digital Video* is available in India. He has written 29 refereed journal articles (17 sole author), and 140 magazine articles. He was founding editor of *Surround Professional* magazine, is an associate editor of the *Journal of the Audio Engineering Society*, and is a member of the editorial board of *Motion Imaging*, the journal of the Society of Motion Picture and Television Engineers.

This page intentionally left blank

The associated web site contains extensively researched information about microphones, tape and film formats, analog recording and editing, methods of synchronization, and more.

The URL is: http://www.focalpress/Holman/Sound FilmTV/. The password is *frequency*.

This page intentionally left blank

Instructions for accompanying DVD

The DVD inserted in the back of this book can be played on stereo or surround-sound systems. Most demonstrations work well on either type. However, a few are certainly best in 5.1. If you lack the ability to play 5.1 discrete channels, then the 2-channel mixdown may be played from the 2-channel output of a DVD player to a receiver equipped for surround sound decoding, such as Pro Logic II. If only a stereo system is available, most demonstrations will work with such a system, but obviously surround sound will not be available.

Track No.	Description	Explanation
Section 1: Setup and Test		
1	–20 dB FS 440 Hz sine wave on 5 main channels in phase, and –30 dB FS 40 Hz sine wave on LFE	Signal present check, level adjustment of consoles—set to the console reference, such as –20 dB FS, 0 VU, etc. (2-channel mixdown is complicated and level matching with 5.1 cannot be assured.)
2	Channel ID by voice, and beeps on LFE	Channel identification
3	Mid-range or LF band-limited Pink Noise at –20 dB FS rotating among the channels	Acoustical level set. See Table 13.4 for the correct level setting for your space.
Section 2: Acoustics		
4	Direct sound level vs. distance	Demonstration of dramatic fall off of direct sound (inverse square law).
5	100 Hz, 1 kHz, 10 kHz sine wave tones alternating	Shows wide frequency range of hearing, this demonstration being 100:1 range. (More extreme frequencies are not used albeit in the audible spectrum due to limitations of sound reproduction systems.)
6	Boinks covering the 10 octaves of the audible frequency range	Boinks are shaped tone bursts that stay within a narrow frequency range but challenge the peak capability of sound systems—a sort of "all your eggs in one basket," that is, Full Scale at specific frequencies, one at a time. EXERCISE CAUTION. Start at a low level on this one and advance the volume as it can break sound systems. Most systems cannot play this track at full reference level.
7	400 Hz tone with a variety of waveforms: sine, square, triangular, synthesized.	Demonstrates how the waveform of a sound produces a variety of spectra. Only the pure sine wave contains only one tone; the others contain harmonics as well as the fundamental.
8	Radiation pattern of the voice shown by recording simultaneously at different angles.	Demonstrates that recording is an art: best placement of the microphone is determined by how the source radiates sound in various directions.
9	Doppler shift.	Often used by filmmakers to portray sound in motion, along with its characteristic pitch drop as a source comes towards then away from the point of view.
10	One reflection added to direct sound.	This 10 ms reflection at the same level as the direct sound "colors" the timbre of the sound. Where have we heard this before?
11	Wide dynamic range test: 80 dB range	Noise played at 10 dB intervals over an 80 dB range, a loudness range of approximately 256 times from the softest to loudest. Once again EXERCISE CAUTION. The extremely high level of the end could break sound systems, and could damage hearing if you were unusually close to a loudspeaker.

Track No.	Description	Explanation
Section 3: Psychoacoustics		
12	Loudness: the effect of playback at various levels and frequencies	Demonstrates this effect. It is not your sound system but your hearing that does this, although it helps to have an extended frequency range and flat frequency response (equalized) sound system.
13	Loudness vs. time	The longer a sound is around the louder it gets. This sequence of 1, 2, 5, 10, 20, 50, and 100 ms of the same sound source demonstrates this effect: the peak level is constant, but you hear it as louder as it becomes longer.
14	Frequency masking	A loud sound covers up a soft one, especially nearby in frequency and at higher levels.
15	Temporal masking	A loud sound can cover up a softer one that comes before it. The beeps before the noise change in level and time relative to the noise, but are present in all the segments.
16	Timbre I: the harmonic structure of instruments, their spectrum, helps to identify them.	The level, pitch, and duration of the samples are roughly constant, but we distinguish them by their timbre, which is most associated with their spectrum and how it evolves in time.
17	Timbre II: a piano note played backwards in time demonstrates a different timbre from the conventional head-to-tail presentation.	This demonstrates that timbre is not only due to level, pitch, and duration, but also to the evolution of the spectrum as time runs along. A piano note contains a hard starting transient, the hammer strike, which registers as a much different timbre backwards.
Section 4: Audio		
18	Clipping distortion: feeding a line level signal into a microphone level input	All too common since the connectors are interchangeable.
19	Aliasing distortion	If a proper anti-aliasing filter, taking out all frequencies below one-half the sampling frequency, is not present, this type of distortion may be heard in a digital audio system. Programming by David Ranada.
20	Quantizing distortion	If correct dither is not added, low level signals will be distorted. Programming by David Ranada.
21	Analog vs. digital overload	Although most recording today is digital, this track illustrates the difference of over-recorded analog vs. digital.
Section 5: Microphones		
22	Microphone polar patterns illustrated	Walking around an array of microphones and listening to one at a time shows their polar pattern.
23	Polar patterns II.	The microphones are the same as in track 18, but the different types have been intercut at various angles to demonstrate the differences in another way. Shorter than track 19.
24	Microphone proximity effect.	Demonstrates bass boost when a directional microphone is used close to a source.
25	Pressure (omni) vs. Pressure Gradient (in this case, a cardioid) on very low-frequency response	Demonstrates that a pure pressure mic, an omni, goes lower than a similar pressure-gradient one. Seems to contradict track 20, but proximity effect only works down to the mics cutoff frequency.
26	Same as track 25, but mic types are intercut	Shows more readily the differing extreme low frequency response.
27	Microphone wind susceptibility	By blowing constantly across the two main types, the superior wind susceptibility of the pressure type is illustrated.
28	Making a cardioid	By adding together a pure pressure mic, an omni; and a pure pressure-gradient mic, a bi-directional one; a cardioid is formed when the two are added in 1:1 proportion. Other patterns such as subcardioid, hypercardioid, and supercardioid can also be made by using varying proportions of the two main types.

Track No.	Description	Explanation
The following seven tracks illustrate the sound of various microphones recorded simultaneously. The distance is adjusted to maintain a more-or-less constant direct-to-reverberation ratio, except in the case of the lavaliere where that is impossible. The words spoken are form a phonetically balanced word list.		
29	Omni measurement microphone	A reference, with known frequency response, etc. B&K Model 2230 sound level meter with its supplied pressure mic. Used at 2 ft. from the talker.
30	Omni recording mic with an hf rise	Schoeps MK2S. The hf rise is useful at a distance from a source to overcome air losses. Same distance as track 29.
31	Bi-directional pressure-gradient mic at 1.7 × the distance of the omni	Neumann U89 set to bidirectional.
32	Cardioid at 1.7 × the distance of the omni	Schoeps MK4
33	Supercardioid at 2.0 × the distance of the omni	Schoeps MK41
34	Shotgun at 3.0 × the distance of the omni	Sennheiser 815
35	Lavaliere clipped to the front of a shirt 8" below the mouth	Tram; each of the distances above was at the one corresponding to the same direct-to-reverb ratio except for this one, which is dryer sounding due to its close placement.
Section 6: Editing		
36	A slow crossfade is less noticeable than an abrupt cut.	
37	Delaying a sound cut from that made at a picture edit to one just before the first sound of the incoming shot enters may hide the edit.	
38	Use presence from the noisier of two shots to get the two to match.	
39	A loud off-camera noise can be cut out, but it leaves a hole that is best filled with presence, also called atmos (for atmosphere, in UK usage) or room tone.	
40	A sound that would cut off abruptly at a picture cut should be extended so that the cut doesn't draw so much attention.	
41	Presence (also known as room tone, fill, or atmosphere (UK usage) is sound used to cut into production sound tracks in order to make it sound seamless. Ambience also known as backgrounds (B.G.s colloquially) is additional sound cut into separate tracks used to set the production into a space and is often multichannel.	
42	The "shape" of a fade, its relative gain versus time of the outgoing and incoming sound, matters.	When two different sounds appear on either side of a crossfade an equal power fade is usually best, whereas if a sound is being crossfaded into itself identically (equal in all ways, not time displaced), the equal gain is probably best. This is demonstrated by crossfading presence (where equal power is best) and a tone into itself (where equal gain is best).

Track No.	Description	Explanation
Section 7: Mixing		
43	Level set	Level is the primary adjustment among the ingredients of a mix.
44	Equalization	This demonstration operates in four bands, each up and then down, one at a time.
45	Equalized pink noise	As an exercise, see if you can match this with a 4-band parametric equalizer.
46	Filtering	High-pass, low-pass, band pass filters are demonstrated.
47	Compression	
48	Limiting	
49	De-essing	
50	Noise reduction	
51	Equalizing lavalieres	
Section 8: Spatial Sound		
52	Reverberation. From mono through frontal stereo to multichannel, reverberation sets the stage for sound space.	
53	Universal ambience: equalized, uncorrelated noise	This is often useful for interior ambiences and may certainly be equalized again by the user for their purpose.
54	Rain storm	Demonstrates both envelopment and imaging. Built into multichannel from 2-channel library tracks.
55	Busy Beverly Hills restaurant interior ambience.	Recorded with Schoeps double-MS setup and decoded to L, R, LS, RS. Original recording by Martin Krieger.
56	Farmer's Market, mid-city Los Angeles exterior ambience.	Recorded with Schoeps double-MS setup and decoded to L, R, LS, RS. Original recording by Martin Krieger.
57	"Tamales!" A tamale vendor in the Boyle Heights neighborhood of Los Angeles is heard passing the microphone and fading into the distance. It is remarkable that you can hear her, separating "figure from ground," very far below the noise level of the environment. How many "Tamales!" can you count?	Recorded with Schoeps MS set up by Martin Krieger. Post production includes Dolby Pro Logic II decoding and decorrelation to produce two surround channels.

Disc contents © 2010 Tomlinson Holman.