NINE

SOUND IN THE CINEMA

Most films create a strong impression that the people and things pictured simply produce an appropriate noise. But as we have seen in Chapter 1, in the process of film production the sound track is built up separately from the images, and it can be manipulated independently. This makes sound as flexible and wide-ranging as other film techniques.

Yet sound is perhaps the hardest of all techniques to study. We are accustomed to ignoring many of the sounds in our environment. Our primary information about the layout of our surroundings comes from sight, and so in ordinary life sound is often simply a background for our visual attention. Similarly, we speak of "watching" a film and of being movie "viewers" or "spectators"—all terms which suggest that the sound track is a secondary factor. We are strongly inclined to think of sound as simply an accompaniment to the real basis of cinema, the moving images.

Moreover, we cannot stop the film and freeze an instant of sound, as we can study a frame to examine mise-en-scene and cinematography. Nor can we lay out the sound track for our inspection as easily as we can examine the editing of a string of shots. In film, the sounds and the patterns they form are elusive. This elusiveness accounts for part of the power of this technique: Sound can achieve very strong effects and yet remain quite unnoticeable. To study sound, we must learn to listen to films.

Fortunately, many viewers are becoming more sensitive to the way movies sound. The film industry has heavily promoted improvements in sound recording, mixing, and reproduction. These developments include Dolby noise-reduction processes and four- and six-track theater sound. *Star Wars*,

with its stereo and "surround" channels, demonstrated that audiences appreciated better sound tracks. During the early 1990s digital sound mixing and reproduction became routine attractions of big-budget pictures.

Although theatrical reproduction remains the weak link in the sound chain, many exhibitors are upgrading their equipment and paying more attention to the auditorium's acoustic design. George Lucas's THX theater-certification program promotes high-quality sound-design systems. Audiences who have become used to digital compact-disc music have also embraced stereo-encoded videocassettes and surround-encoded laserdiscs. Not since the first "talkies" of the late 1920s have filmgoers been so aware of cinema sound, and filmmakers now compete in creating rich and engaging sound tracks.

THE POWERS OF SOUND

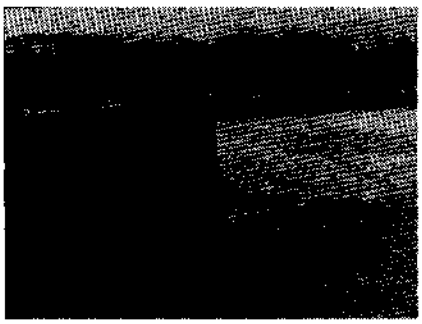
Whether noticed or not, sound is a powerful film technique for several reasons. For one thing, it engages a distinct sense mode. Our visual attention is accompanied by aural attention. Even before recorded sound was introduced in 1926, the "silent" cinema recognized this by its use of accompaniment by orchestra, organ, or piano. At a minimum, the music filled in the silence and gave the spectator a more complete perceptual experience. More significantly, the engagement of hearing opens the possibility of what the Soviet director Sergei Eisenstein called "synchronization of senses"—making a single rhythm or expressive quality unify both image and sound.

Secondly, sound can actively shape how we perceive and interpret the image. In one sequence of *Letter from Siberia,* Chris Marker demonstrates the power of sound to alter our understanding of images. Three times Marker shows the same footage—a shot of a bus passing a car on a city street, three shots of workers paving a street. But each time the footage is accompanied by a completely different sound track. Compare the three versions tabulated alongside the sequence (Table 9.1). The verbal differences are emphasized by the identical images; the audience will construe the same images differently, depending on the sound track.

The *Letter from Siberia* sequence demonstrates a third advantage of sound as well. Film sound can direct our attention quite specifically within the image. When the commentator describes the "blood-colored buses," we are likely to look at the bus and not at the car. When Fred Astaire and Ginger Rogers are executing an intricate step, chances are that we watch their bodies and not the silent nightclub spectators looking on. In such ways, sound can guide us through the images, "pointing" to things to watch.

This possibility becomes even more fertile when you consider that the sound cue for some visual element may *anticipate* that element and relay our attention to it. Suppose we have a close-up of a man in a room and we hear the creaking of a door opening. If the second shot shows the door, now open, the viewer's attention will probably shift to that door, the source of the offscreen sound. But if the second shot shows the door still closed, the viewer will likely ponder his or her interpretation of the sound. (Maybe it wasn't a

Table 9.1 ***LETTER FROM SIBERIA* FOOTAGE**

IMAGES	FIRST COMMENTARY	SECOND COMMENTARY	THIRD COMMENTARY
 **Fig. 9.1**	Yakutsk, capital of the Yakutsk Autonomous Soviet Socialist Republic, is a modern city in which comfortable buses made available to the population share the streets with powerful Zyms, the pride of the Soviet automobile industry. In the	Yakutsk is a dark city with an evil reputation. The population is crammed into blood-colored buses while the members of the privileged caste brazenly display the luxury of their Zyms, a costly and uncomfortable car at best. Bending	In Yakutsk, where modern houses are gradually replacing the dark older sections, a bus, less crowded than its London or New York equivalent at rush hour, passes a Zym, an excellent car reserved for public utilities departments on account of its scarcity.
 **Fig. 9.2**	joyful spirit of socialist emulation, happy Soviet workers, among them this picturesque denizen	to the task like slaves, the miserable Soviet workers, among them this sinister-looking Asiatic,	With courage and tenacity under extremely difficult conditions, Soviet workers, among them this Yakut
 **Fig. 9.3**	of the Arctic reaches, apply themselves	apply themselves to the primitive labor	afflicted with an eye disorder, apply themselves to
 **Fig. 9.4**	to making Yakutsk an even better place to live. Or else:	of grading with a drag beam. Or simply:	improving the appearance of their city, which could certainly use it.

door, after all?) Thus the sound track can clarify image events, contradict them, or render them ambiguous. In all cases, the sound track can enter into an active relation with the image track.

This example of the door opening suggests a fourth advantage of sound: It cues us to form expectations. If we hear a door creaking, we anticipate that someone has entered a room and that we will see that person in the next shot. But, if the film draws upon conventions of the horror genre, the camera might stay on the man, staring fearfully. We would then be in suspense as to the appearance of the monster offscreen. Horror and mystery films often utilize the power of sound from an unseen source to engage the audience's interest, but all types of films can take advantage of this aspect of sound. During the town meeting in *Jaws,* the characters hear a grating sound and turn to look offscreen; a cut reveals Quint's hand scraping on a blackboard—creating a dramatic introduction to this character. We shall see as well several cases in which the use of sound can creatively cheat or redirect the viewer's expectations.

In addition, as V. F. Perkins has pointed out, sound brings with it a new sense of the value of silence. "Only with colour as an available resource can we regard the use of black-and-white photography as the result of a conscious artistic decision. Only in the sound film can a director use silence for dramatic effect." In the context of sound, silence takes on a new expressive function.

A final advantage: Sound bristles with as many creative possibilities as editing. Through editing, one may join shots of any two spaces to create a meaningful relation. Similarly, the filmmaker can mix any sonic phenomena into a whole. With the introduction of sound cinema, the infinity of visual possibilities was joined by the infinity of acoustic events.

FUNDAMENTALS OF FILM SOUND

■ ACOUSTIC PROPERTIES

To pursue in detail the acoustic processes that produce sound would take us on a long detour. (See Notes and Queries for reading on the subject.) We should, however, isolate certain qualities of sound as we perceive it. These qualities are familiar to us from everyday experience.

Loudness. The sound we hear results from vibrations in the air. The amplitude, or breadth, of the vibrations produces our sense of loudness, or volume. Film sound constantly manipulates volume. For example, in many films a long shot of a busy street is accompanied by loud traffic noises, but when two people meet and start to speak, the volume of the traffic drops. Or a dialogue between a soft-spoken character and a blustery one is characterized as much by the difference in volume as by the substance of the talk.

Loudness is also related to perceived distance; often the louder the sound, the closer we take it to be. This sort of assumption seems to be at

work in the street-traffic example already mentioned: The couple's dialogue, being louder, is sensed as in the acoustic "foreground," while the traffic noise sinks to the background. In addition, a film may startle the viewer by exploiting abrupt and extreme shifts in volume (usually called changes in *dynamics*), as when a quiet scene is interrupted by a very loud noise.

Pitch. The frequency of sound vibrations governs pitch, or the perceived "highness" or "lowness" of the sound. Certain instruments, such as the tuning fork, can produce pure tones, but most sounds, in life and on film, are "complex tones," batches of different frequencies. Nevertheless pitch plays a useful role in picking out distinct sounds in a film sound track. It helps us distinguish music and speech from other sounds. Pitch also serves to distinguish among objects. Low-pitched sounds, such as thumps, can evoke hollow objects, while higher-pitched sounds (like those of jingle bells) suggest smoother or harder surfaces and more dense objects.

Pitch can also serve more specific purposes in a film. When a young boy tries to speak in a man's deep voice and fails (as in *How Green Was My Valley*), the joke is based primarily on pitch. Marlene Dietrich's vocal delivery often depends upon a long upward-gliding intonation which makes a statement sound like a question. In the coronation scene of *Ivan the Terrible,* Part I, a court singer with a deep bass voice begins a song of praise to Ivan, and each phrase rises dramatically in pitch—which Eisenstein emphasizes in the editing, with successively closer shots of the singer coinciding with each vocal change. When Bernard Herrmann obtained the effects of shrill, birdlike shrieking in Hitchcock's *Psycho,* even many musicians could not recognize the source: violins played at extraordinarily high pitch.

Timbre. The harmonic components of a sound give it a certain "color" or tone quality—what musicians call timbre. Timbre is actually a less fundamental acoustic parameter than amplitude or frequency, but it is indispensible in describing the texture or "feel" of a sound. When we call someone's voice nasal or a certain musical tone mellow, we are referring to timbre. In everyday life, the recognition of a familiar sound is largely a matter of various aspects of timbre.

Filmmakers manipulate timbre continually. Timbre can help articulate portions of the sound track, as when it differentiates musical instruments from one another. Timbre also comes forward on certain occasions, as in the clichéd use of oleaginous saxophone tones behind seduction scenes. More subtly, in the opening sequence of Rouben Mamoulian's *Love Me Tonight,* people starting the day on a street pass a musical rhythm from object to object—a broom, a carpet beater—and the humor of the number springs in part from the very different timbres of the objects. In preparing the sound track for Peter Weir's *Witness,* the editors drew upon sounds recorded twenty or more years before, so that the less modern timbre of the older recordings would express the rustic seclusion of the Amish community.

As fundamental components loudness, pitch, and timbre interact to define the overall sonic texture of a film. At the most elementary level, these three acoustic factors enable us to distinguish the various sounds in a film. For example, these qualities enable us to recognize different characters'

voices. At a more complex level, all three components of film sound interact to add considerably to our experience of the film. For instance, both John Wayne and James Stewart speak slowly, but Wayne's voice tends to be deeper and gruffer than Stewart's querulous drawl. This difference works to great advantage in *The Man Who Shot Liberty Valance*, where their characters are sharply contrasted. In *The Wizard of Oz* the disparity between the public image of the Wizard and the old charlatan who rigs it up is marked by the booming bass of the effigy and the old man's higher, softer, more quavering voice.

Citizen Kane offers a wide range of sound manipulations. Echo chambers alter timbre and volume. A motif is formed by the inability of Kane's wife Susan to sing pitches accurately. Moreover, in *Citizen Kane* the plot's shifts between times and places are covered by continuing a sound "thread" and varying the basic acoustics. A shot of Kane applauding dissolves to a shot of a crowd applauding (a shift in volume and timbre). Leland beginning a sentence in the street cuts to Kane finishing the sentence in an auditorium, his voice magnified by loudspeakers (a shift in volume, timbre, and pitch).

Recent noise-reduction techniques, multitrack reproduction, and digital sound have yielded wider ranges of frequency and volume, as well as crisper timbres.

■ SELECTION, ALTERATION, AND COMBINATION

Sound in the cinema is of three types: *speech*, *music*, and *noise* (also called *sound effects*). Occasionally a sound may cross categories—is a scream speech or noise? is electronic music also noise?—and filmmakers have freely exploited these ambiguities. In *Psycho*, when a woman screams, we expect to hear a human voice and instead hear "screaming" violins. Nevertheless, in most cases the distinctions hold. Now that we have an idea of some basic acoustic properties, we must consider how speech, music, and noise are selected and combined for specific functions within films.

The creation of the sound track resembles the editing of the image track. Just as the filmmaker may pick the best image from several shots, he or she may choose what exact bit of sound will best serve the purpose. Just as footage from disparate sources may be blended into a single visual track, so too sound that was not recorded during filming may be added freely. Moreover, just as a shot may be rephotographed on the optical printer or tinted in color or jigsawed into a composite image, so too may a bit of sound be processed to change its acoustic qualities. And just as the filmmaker may link or superimpose images, so may he or she join any two sounds end to end or place one "over" another (as with commentary "over" music). Though we are not usually as aware of sonic manipulations, the sound track demands as much choice and control as does the visual track.

Animated films depend on careful sound planning, since drawings and puppets cannot make sounds themselves. Studio-made animated cartoons typically record music, dialogue, and sound effects *before* the images are filmed, so that the figures may be synchronized with the sound frame by frame. For many years Carl Stalling created frantically paced jumbles of

familiar tunes, weird noises, and distinctive voices for the adventures of Bugs Bunny and Daffy Duck.

Experimental films, especially those using abstract form, also frequently build their images around a preexisting sound track. Some filmmakers have even argued that abstract cinema is a sort of "visual music" and have tried to create a synthesis of the two media.

As with other film techniques, sound guides the viewer's attention. Normally, this means clarifying and simplifying the sound track so that important material stands out. Dialogue, the transmitter of story information, is usually recorded and reproduced for maximum clarity. Important lines should not have to compete with music or background noise. Sound effects are usually less important. They supply an overall sense of a realistic environment and are seldom noticed; yet if they were missing, the silence would be distracting. Music is also subordinate to dialogue, entering during pauses in dialogue or effects.

Dialogue does not always rank highest in importance. Sound effects are usually central to action sequences, while music can dominate dance scenes, transitional sequences, or emotion-laden moments without dialogue. And some filmmakers have shifted the weight conventionally assigned to each type of sound. Charles Chaplin's *City Lights* and *Modern Times* eliminate dialogue, letting sound effects and music come to the fore. The films of Jacques Tati and Jean-Marie Straub retain dialogue but still place great emphasis on sound effects. Later in this chapter we will consider how music and noise in Robert Bresson's *A Man Escaped* fill out a sparse dialogue track by evoking offscreen space and creating thematic associations.

In creating a sound track, then, the filmmaker must select sounds that will fulfill a particular function. In order to do this, usually the filmmaker will provide a clearer, simpler sound world than that of everyday life. Normally, our perception filters out irrelevant stimuli and retains what is most useful at a particular moment. As you read this, you are attending to words on the page and (to various degrees) ignoring certain stimuli that reach your ears. But if you close your eyes and listen attentively to the sounds around you, you will become aware of many previously unnoticed sounds—distant voices, the wind, footsteps, a radio playing. Any amateur recordist knows that if you set up a microphone and tape recorder in a "quiet" environment, all of those normally unnoticed sounds suddenly become obtrusive. The microphone is unselective; like the camera lens, it does not automatically filter out what is distracting. Sound studios, camera blimps to absorb motor noise, directional and shielded microphones, sound engineering and editing, and libraries of stock sounds all allow the filmmaker to choose exactly what the sound track requires. Unless a filmmaker actually wants the ambient noise of a scene, simply holding out a microphone while filming will rarely be selective enough.

By selecting certain sounds, the filmmaker guides our perception of the image and the action. In one scene from Jacques Tati's *Mr. Hulot's Holiday*, vacationers at a resort hotel are relaxing (Fig. 9.5). In the foreground guests quietly play cards; in the depth of the shot, Mr. Hulot is frantically playing Ping-Pong. Early in the scene, the guests in the foreground are murmuring

Fig. 9.5

quietly, but Hulot's Ping-Pong game is louder; the sound cues us to watch Hulot. Later in the scene, however, the same Ping-Pong game makes *no* sound at all, and our attention is drawn to the muttering card players in the foreground. The presence and absence of the sound of the Ping-Pong ball guides our expectations. If you start to notice how such selection of sound shapes our perception, you will also notice that filmmakers often use sound quite unrealistically, in order to shift our attention to what is narratively or visually important.

Our scene from *Mr. Hulot's Holiday* also points up the importance of how a chosen sound may have its acoustic qualities transformed for a particular purpose. Thanks to a manipulation of volume and timbre, the Ping-Pong game gains in vividness and clarity. Similarly, a character speaking will usually sound about as loud in long shot as in close-up, even though this is a flagrant violation of realism.

At the limit, wholly new sounds may be made of old ones. The noises emitted by the demonically possessed girl in *The Exorcist* blended screams, animal thrashings, and English spoken backward. To create the roar of a Tyrannosaurus Rex for *Jurassic Park*, sound engineers fused a tiger's roar, a baby elephant's trumpeting at midrange frequencies, and an alligator's growl for the lower tones.

Nowadays, film sound is normally reprocessed to yield exactly the qualities desired. A "dry recording" of the sound in a fairly nonreflective space will be manipulated electronically to yield the desired effect. For instance, the voice of someone on the telephone is typically treated with filters to make it more tinny and muffled. (In Hollywood parlance, this is called "futzing" the sound.) The almost nonstop rock-and-roll music of *American Graffiti* used two recordings of the music: a "dry" one for moments when the music was to dominate the scene and had to be of high sonic quality, and a more ambient one for background noise. The latter was derived from a tape recorder simply playing the tune in a backyard.

Guiding the viewer's attention, then, depends on selecting and reworking particular sounds. It also depends on *combining* them. It is useful to think of the sound track not as a set of discrete sound units but as an ongoing stream of auditory information. Each sonic event takes its place in a specific pattern. This pattern involves linking events in time as well as "layering" them at any given moment.

We can easily see how the sound track offers a stream of auditory information by considering a scene cut according to classical continuity principles. When filmmakers edit conversations in shot/reverse shot, they often utilize a *dialogue overlap* to smooth down the visual change of shot. In a dialogue overlap the filmmaker continues a line of dialogue across a cut. During a conversation in John McTiernan's *Hunt for Red October*, we get the following shots and dialogue:

1. (ms) Over political officer's shoulder, favoring Captain Ramius (Fig. 9.6).
   Officer: "Captain Tupalev's boat."
   Ramius: "You know Tupalev?"
   Officer: "I know he descends . . ."

Fig. 9.6

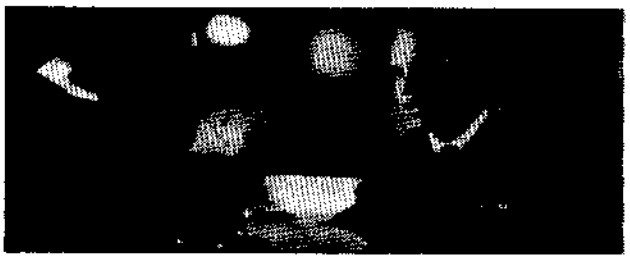


Fig. 9.7

Fig. 9.8

2. (ms) Reverse angle over Ramius's shoulder, favoring officer (Fig. 9.7).
   Officer (continuing): ". . . from aristocracy, and that he was your student. It's rumored he has a special . . ."
3. (mcu) Reverse angle on Ramius (Fig. 9.8).
   Officer (continuing): ". . . place in his heart for you."
   Ramius: "There's little room in Tupalev's heart for anyone but Tupalev."

Here the officer's chatter provides an auditory continuity that distracts from the shot changes. Moreover, by cutting to the listener before a sentence is finished, the sound and editing concentrate our attention on Ramius's response. As a Hollywood editor puts it: "The minute a telling word or a question is posed . . . I go for a reaction to see . . . how they are trying to formulate the answer in their face or dialogue." The principle of dialogue overlap can be used with noise as well. In the *Red October* scene just mentioned, sounds of a spoon clinking in a tea cup and of papers being riffled also carry over certain cuts, providing a continuous stream of sonic information.

This stream can involve more than simply linking one line of dialogue or bit of noise to another. We have already seen that in production, combining sounds is usually done after shooting, in the mixing process. The mixer can precisely control the volume, duration, and tone quality of each sound. In

modern filmmaking, a dozen or more separate tracks may be mixed in layers at any moment. The mix can be quite dense, as when an airport scene combines the babble of several distinct voices, footsteps, luggage trolleys, Muzak, and plane engines. Or the mix can be very sparse, with an occasional sound emerging against a background of total silence. Most cases will fall somewhere in between these extremes. In our *Red October* scene, distant throbbing engines and slight brushings of fabric form a muted background to the dialogue exchange.

The filmmaker may create a mix in which each sound blends smoothly with the others. This is commonly the case when music and effects are mixed with speech. In classical Hollywood cinema of the 1930s, the musical score may become prominent in moments in which there is no dialogue, and then it is likely to fade unnoticeably down just as the characters begin to talk. (In studio parlance, this is called "sneaking in" and "sneaking out.") Alternatively, the acoustic stream may contain much more abrupt contrasts. Contemporary Hollywood films often exploit the dynamic range of Dolby technology to fill chase sequences with startling shifts between low, rumbling engines and whining sirens or squealing tires.

The ways in which sounds may combine to create an ongoing stream of information is well illustrated by the final battle sequence of Akira Kurosawa's *Seven Samurai.* In a heavy rain, marauding bandits charge into a village defended by the villagers and the samurai. The torrent and wind form a constant background noise throughout the scene. Before the battle, the conversation of the waiting men, footsteps, and the sound of swords being drawn are punctuated by long pauses in which we hear only the drumming rain. Suddenly distant horses' hooves are heard offscreen. This draws our attention from the defenders to the attackers. Then Kurosawa cuts to a long shot of the bandits; their horses' hooves become abruptly louder. (The scene employs conventional "sound perspective": the closer the camera is to a source, the louder the sound.) When the bandits burst into the village, yet another sound element appears—the bandits' harsh battle cries, which increase steadily in volume as they approach.

The battle begins. The muddy, storm-swept mise-en-scene and rhythmic cutting gain impact from the way in which the incessant rain and splashing are explosively interrupted by brief noises—the screams of the wounded, the splintering of a fence one bandit crashes through, the whinnies of horses, the twang of one samurai's bowstring, the gurgle of a speared bandit, the screams of women when the bandit chieftain breaks into their hiding place. The sudden intrusion of certain sounds marks abrupt developments in the battle. Such frequent surprises heighten our tension, since the narration frequently shifts us from one line of action to another.

The scene climaxes after the main battle has ended. Offscreen horses' hooves are cut short by a new sound—the sharp crack of a bandit's rifle shot, which fells one samurai. A long pause, in which we hear only the driving rain, emphasizes the moment. The samurai furiously throws his sword in the direction of the shot and falls dead into the mud. Another samurai races toward the bandit chieftain, who has the rifle; another shot cracks out and he falls back, wounded; another pause, in which only the relentless rain is

heard. The wounded samurai kills the chieftain. The other samurai gather. At the scene's end, the sobs of a young samurai, the distant whinnies and hoofbeats of riderless horses, and the rain all fade slowly out.

The relatively dense mix of this sound track gradually introduces sounds which turn our attention to new narrative elements (hooves, battle cries) and then modulates these sounds into a harmonious stream. This stream is then punctuated by abrupt sounds of unusual volume or pitch associated with crucial narrative actions (the archery, women's screams, the gunshots). Overall, the combination of sounds enhances the unrestricted, objective narration of this sequence, which shows us what happens in various parts of the village rather than confining us to the experience of a single participant.

The choice and combination of sonic materials can also create patterns which run through the film as a whole. This is most readily seen by examining how the filmmaker uses a musical score. Sometimes the filmmaker will select preexisting pieces of music to accompany the images, as Bruce Conner does in using portions of Respighi's *Pines of Rome* as the sound track for *A Movie.* In other cases the music will be composed for the film, and here the filmmaker and the composer make several choices.

The rhythm, melody, harmony, and instrumentation of the music can strongly affect the viewer's emotional reactions. In addition, a melody or musical phrase can be associated with a particular character, setting, situation, or idea. *Local Hero,* a film about a confused young executive who leaves Texas to close a business deal in a remote Scottish village, uses two major musical themes. A rockabilly tune is heard in the urban Southwest, while a slower, more poignantly folkish melody is associated with the seaside village. In the final scenes, after the young man has returned to Houston, he recalls Scotland with affection, and the film plays the two themes simultaneously. By contrast, a single musical theme can change its quality when associated with different situations. In *Raising Arizona,* the hapless hero has a terrifying dream in which he envisions a homicidal biker pursuing him, and the accompanying music is appropriately ominous. But at the film's end, the hero dreams of raising dozens of children, and now the same melody, reorchestrated and played at a lilting tempo, conveys peace and comfort.

By reordering and varying musical motifs, the filmmaker can subtly compare scenes, trace patterns of development, and suggest implicit meanings. A convenient example is Georges Delerue's score for François Truffaut's *Jules and Jim.* Overall, the film's music reflects the Paris of 1912–1933, during which the action takes place; many of the melodies resemble works by Claude Debussy and Erik Satie, two of the most prominent French composers of that era. Virtually the entire score consists of melodies in ¾ meter, many of them in waltz time, and all the main themes are in keys related to A major. These rhythmic and harmonic decisions help unify the film.

More specifically, musical themes are associated with particular aspects of the narrative. For instance, Catherine's constant search for happiness and freedom outside conventional boundaries is conveyed by her singing the "Tourbillon" ("whirlwind") song, which says that life is a constant changing of romantic partners. Settings are also evoked in musical terms. One tune is heard every time the characters are in a café. As the years go by, the tune

changes from a mechanical player-piano rendition to a jazzier version played by a black pianist.

The characters' relations become more strained and complicated over time, and the score reflects this in its development of major motifs. A lyrical melody is first heard when Jules, Jim, and Catherine visit the countryside and bicycle to the beach. This "idyll" tune will recur at many points when the characters reunite, but as the years pass it will become slower in tempo, more sombre in instrumentation, and shifted from a major to a minor mode. Another motif that reappears in different guises is a "dangerous love" theme associated with Jim and Catherine. This grave, shimmering waltz is first heard when he visits her apartment and watches her pour a bottle of vitriol down the sink. (The acid, she says, is "for lying eyes.") Thereafter, this harmonically unstable theme, which resembles one of Satie's *Gymnopédies* for piano, is used to underscore Jim and Catherine's vertiginous love affair. At times it accompanies scenes of passion, but at other times it accompanies their growing disillusionment and despair.

The most varied theme is a mysterious phrase first heard on the flute when Jules and Jim encounter a striking ancient statue. Later they meet Catherine and discover that she has the statue's face; a repetition of the musical motif confirms the comparison. Throughout the film, this brief motif is associated with the enigmatic side of Catherine. In the film's later scenes, this motif is developed in an intriguing way. The bass line (played on harpsichord or strings) that softly accompanied the woodwind tune now comes to the fore, creating a relentless, often harsh pulsation. This "menace" waltz underscores Catherine's fling with Albert and accompanies her final vengeance on Jim: driving her car, with him as passenger, into the river.

Once musical motifs have been selected, they can be combined to evoke associations. During Jim and Catherine's first intimate talk after the war, the bassline-dominated version of the "enigma" waltz is followed by the love theme, as if the latter could drown out the menacing side of Catherine's character. The love theme accompanies long tracking shots of Jim and Catherine strolling through the woods. But at the scene's end, as Jim bids Catherine farewell, the original woodwind version of her theme recalls her mystery and the risk he is running by falling in love with her. Similarly, when Jim and Catherine lie in bed, facing the end of their affair, the voice-over narrator says: "It was as if they were already dead" as the "dangerous love" theme plays. This sequence associates death with their romance and foreshadows their fate at the film's end.

A similar sort of blending can be found in the film's final scene. Catherine and Jim have drowned, and Jules is overseeing the cremation of their bodies. As shots of the coffins dissolve into detailed shots of the cremation process, the "enigma" motif segues into its sinister variant, the "menace" motif. But as Jules leaves the cemetery and the narrator comments that Catherine had wanted her ashes to be cast to the winds, the string instruments glide into a sweeping version of the "whirlwind" waltz. The film's musical score thus concludes by recalling the three sides of Catherine that attracted the men to her: her mystery, her menace, and her vivacious openness to experience. In such ways, a musical score can create, develop, and associate motifs that enter into the film's overall formal system.

DIMENSIONS OF FILM SOUND

We have seen what sounds consist of and how the filmmaker can take advantage of the widely different kinds of sounds available. In addition, the way in which the sounds relate to other film elements gives them several other dimensions. First, because sound occupies a duration, it has a *rhythm.* Second, sound can relate to its perceived source with greater or lesser *fidelity.* Third, sound conveys a sense of the *spatial* conditions in which it occurs. And fourth, the sound relates to visual events that take place in a specific time, and this relationship gives sound a *temporal* dimension. These categories reveal that sound in film offers a great many creative possibilities to the filmmaker.

■ RHYTHM

Rhythm is one of the most complex features of sound. We have already considered it briefly in relation to mise-en-scene (p. 196) and editing (p. 278). As we suggested, rhythm involves, minimally, a *beat* or pulse, a pace or *tempo,* and a pattern of *accents,* or stronger and weaker beats. In the realm of sound, all of these features are naturally most recognizable in film music, since there beat, tempo, and accent are basic compositional features. In our examples from *Jules and Jim* (p. 325), the motifs can be characterized as having a 3/4 metrical pulse, putting an accent on the first beat, and displaying variable tempo—sometimes slow, sometimes fast.

Speech also has rhythm. People can be identified by "voice prints" which show not only characteristic frequencies and amplitudes but also distinct patterns of pacing and syllabic stress. In fictional films, speech rhythm is a matter for the performer's control, but the sound editor can also manipulate it at the dubbing phase.

Sound effects have distinct rhythmic qualities as well. The plodding hooves of a farmhorse differ from a cavalry company riding at full speed. The reverberating tone of a gong may offer a slowly decaying accent, while a sudden sneeze provides a brief one. In a gangster film, a machine gun's fire creates a regular, rapid beat, while the sporadic reports of pistols may come at irregular intervals.

Any consideration of the rhythmic uses of sound is complicated by the fact that the movements in the images have a rhythm as well, distinguished by the same principles of beat, speed, and accent. In addition, the editing has a rhythm. As we have seen, a succession of short shots helps create a rapid tempo, whereas shots held longer tend to slow down the rhythm.

In most cases the rhythms of editing, of movement within the image, and of sound all cooperate. Possibly the most common tendency is for the filmmaker to match visual and sonic rhythms to each other. In a dance sequence in a musical, the figures move about at a rhythm determined by the music. But variation is always possible. In the "Waltz in Swing Time" number in *Swing Time,* the dancing of Astaire and Rogers moves quickly in time to the music. But no fast cutting accompanies this scene. Indeed, the scene consists of a single long take from a long-shot distance.

Another prototype of close coordination between screen movement and sound comes in the animated films of Walt Disney in the 1930s. Mickey Mouse and the other Disney characters often move in exact synchronization with the music, even when they are not dancing. (As we have seen, such exactness was possible because the sound track was recorded in advance of the drawing of the cels.) Such matching of nondance movement with music in fact came to be known as "Mickey Mousing."

The filmmaker may also choose to create a disparity among the rhythms of sound, editing, and image. One of the most common options is to edit dialogue scenes in ways that "cut against" natural speech rhythms. In our specimen of dialogue overlap from McTiernan's *Hunt for Red October* (Figs. 9.6–9.8), the editing does not coincide with accented beats, cadences, or pauses in the officer's speech. By cutting "against" the rhythm of his lines, the editing smoothes over the changes of shot and emphasizes the words and facial expressions of Captain Ramius. If the filmmaker wants to emphasize the speaker and the speech, the cuts usually come at pauses or natural stopping points in the line. McTiernan uses this sort of rhythmic cutting at other points in the film.

The filmmaker may contrast the rhythm of sound and picture in more noticeable ways. For instance, if the source of sound is primarily offscreen, the filmmaker can utilize the behavior of onscreen figures to create an expressive counter-rhythm. Toward the end of John Ford's *She Wore a Yellow Ribbon*, the aging cavalry captain, Nathan Brittles, watches his troops ride out of the fort just after he has retired. He regrets leaving the service and desires to go with the patrol. The sound of the scene consists of two elements: the cheerful title song sung by the departing riders, and the quick hoofbeats of their horses. Yet only a few of the shots show the horses and singers, who ride at a rhythm matched to the sound. Instead, the scene concentrates our attention on Brittles, standing almost motionless by his own horse. The contrast of brisk musical rhythm and the static images of the solitary Brittles functions expressively, to emphasize his regret at having to stay behind for the first time in many years.

At times, accompanying music might even seem rhythmically inappropriate to the images. At intervals in *Four Nights of a Dreamer* Robert Bresson presents shots of a large, floating nightclub cruising the Seine. The boat's movement is slow and smooth, yet the sound track consists of lively calypso music. (Not until a later scene do we discover that the music comes from a band aboard the boat.) The strange combination of fast music with the slow passage of the boat creates a languorous, mysterious effect.

Jacques Tati does something similar in *Play Time*. In a scene outside a Parisian hotel, tourists climb aboard a bus to go to a nightclub. As they file slowly up the steps, raucous, jazzy music begins. The music again startles us because it seems inappropriate to the images. In fact, it primarily accompanies action in the next scene, in which some carpenters awkwardly carrying a large plate-glass window seem to be dancing to the music. By starting the fast music over an earlier scene of slower visual rhythm, Tati creates a comic effect and prepares for a transition to a new locale.

In Chris Marker's *La Jetée* the contrast between image and sound rhythms dominates the entire film. *La Jetée* is made up almost entirely of still

shots; except for one tiny gesture, all movement within the images is eliminated. Yet the film utilizes voice-over narration, music, and sound effects of a generally rapid, constantly accented rhythm. Despite the absence of movement, the film does not seem "uncinematic," partly because it offers a dynamic interplay of audio-visual rhythms.

These examples suggest some of the ways in which rhythms may be combined. But of course most films also vary their rhythms from one point to another. A change of rhythm may function to shift our expectations. In the famous battle on the ice in *Alexander Nevsky*, Sergei Eisenstein develops the sound from slow tempos to fast and back to slow. The first twelve shots of the scene show the Russian army anticipating the attack of the German knights. The shots are of moderate length, and they contain very little movement. The music is comparably slow, consisting of short, distinctly separated chords. Then, as the German army rides into sight over the horizon, both the visual movement and the tempo of the music increase quickly, and the battle begins. At the end of the battle Eisenstein creates another contrast with a long passage of slow, lamenting music, majestic tracking shots, and little figure movement.

■ FIDELITY

By fidelity we do not mean the quality of recording. In our sense, fidelity refers to the extent to which the sound is faithful to the source as we conceive it. If a film shows us a barking dog and we hear a barking noise, that sound is faithful to its source; the sound maintains fidelity. But if the picture of the barking dog is accompanied by the sound of a cat meowing, there enters a disparity between sound and image—a lack of fidelity.

From our standpoint, fidelity has nothing to do with what originally made the sound in production. As we have seen, the filmmaker may manipulate sound independently of image. Accompanying the image of a dog with the meow is no more difficult than accompanying the image with a bark. If the viewer takes the sound to be coming from its source in the diegetic world of the film, then it is faithful, regardless of its actual source in production.

Fidelity is thus purely a matter of expectation. Even if our dog emits a bark on screen, perhaps in production the bark came from a different dog or was electronically synthesized. We do not know what a laser gun "really" sounds like, but we accept the whang they make in *Return of the Jedi* as plausible. (In production, their sound was made by hammering guy wires that anchored a radio tower.)

When we do become aware that a sound is unfaithful to its source, that awareness is usually used for comic effect. In Jacques Tati's *Mr. Hulot's Holiday* much humor arises from the opening and closing of a dining-room door. Instead of simply recording a real door, Tati inserts a twanging sound like a plucked cello string each time the door swings. Aside from being amusing in itself, this sound functions to emphasize the rhythmic patterns created by waiters and diners passing through the door. Because many of the jokes in *Mr. Hulot's Holiday* and other Tati films are based on quirkily unfaithful noises, his films are good specimens for the study of sound.

Another master of comically unfaithful sound is René Clair. In several

scenes of *Le Million* sound effects occur that are not faithful to their sources. When the hero's friend drops a plate, we hear not shattering crockery but the clash of cymbals. Later, during a chase scene, when characters collide, the impact is portrayed by a heavy bass drum beat. Similar manipulations of fidelity commonly occur in animated cartoons.

But as with low- or high-angle framings, we have no recipe that will allow us to interpret every manipulation of fidelity as comic. Some nonfaithful sounds have serious functions. In Hitchcock's *The Thirty-Nine Steps* a landlady discovers a corpse in an apartment. A shot of her screaming face is accompanied by a train whistle; then the scene shifts to an actual train. Though the whistle is not a faithful sound for an image of a screaming person, it provides a dramatic transition.

In some cases fidelity may be manipulated by a change in volume. A sound may seem unreasonably loud or soft in relation to other sounds in the film. Curtis Bernhardt's *Possessed* alters volume in ways that are not faithful to the sources. The central character is gradually falling deeper into mental illness. In one scene she is alone, very distraught, in her room on a rainy night, and the narration restricts us to her range of knowledge. But sound devices enable the narration to achieve subjective depth as well. We begin to hear things as she does; the ticking of the clock and dripping of raindrops gradually magnify in volume. Here the shift in fidelity functions to suggest a psychological state, a movement from the character's heightened perception into sheer hallucination.

■ SPACE

Sound has a spatial dimension because it comes from a *source.* Our beliefs about that source have a powerful effect on how we understand the sound.

Recall that for purposes of analyzing narrative form, we described events taking place in the story world as *diegetic* (p. 92). For this reason, *diegetic sound* is sound which has a source in the story world. The words spoken by the characters, sounds made by objects in the story, and music represented as coming from instruments in the story space are all diegetic sound.

Diegetic sound is often hard to notice as such. It may seem to come naturally from the world of the film. But as we saw in the sequence of the Ping-Pong game in *Mr. Hulot's Holiday,* the filmmaker may manipulate diegetic sound in ways that are not at all realistic.

On the other hand there is *nondiegetic sound,* which is represented as coming from a source outside the story world. Music added to enhance the film's action is the most common type of nondiegetic sound. When Roger Thornhill is climbing Mount Rushmore in *North by Northwest* and tense music comes up, we do not expect to see an orchestra perched on the side of the mountain. Viewers understand that the "movie music" is a convention and does not issue from the world of the story. The same holds true for the so-called omniscient narrator, the disembodied voice that gives us information but does not belong to any of the characters in the film. An example is *The Magnificent Ambersons,* in which the director Orson Welles speaks the nondiegetic narration.

Nondiegetic sound effects are also possible. In *Le Million* various char-

acters all pursue an old coat with a winning lottery ticket in the pocket. The chase converges backstage at the opera, where the characters race and dodge around one another, tossing the coat to their accomplices. But instead of putting in the sounds coming from the actual space of the chase, Clair fades in the sounds of a football game. Because the maneuvers of the chase do look like a football game, with the coat serving as a ball, this enhances the comedy of the sequence. Although we hear a crowd cheering and a referee's whistle, we do not assume that the characters present are making these sounds. (Thus this is not a manipulation of fidelity, as with the earlier cases from *Le Million*.) The nondiegetic sounds create comedy by making a sort of audio-visual pun.

Entire films may be made with completely nondiegetic sound tracks. Conner's *A Movie*, Kenneth Anger's *Scorpio Rising*, and Derek Jarman's *War Requiem* use only nondiegetic music. Similarly, many compilation documentaries include no diegetic sound; instead, omniscient voice-over commentary and orchestral music guide our response to the images.

As with fidelity, the distinction between diegetic and nondiegetic sound does not depend on the real source of the sound in the filmmaking process. Rather, it depends on our understanding of the conventions of film viewing. We know that certain sounds are represented as coming from the story world, while others are represented as coming from outside the space of the story events. Such viewing conventions are so common that we usually do not have to think about which type of sound we are hearing at any moment. At many points in this chapter, however, we will find that the film's narration deliberately blurs boundaries between different spatial categories. Such a play with convention can be used to puzzle or surprise the audience, to create humor or ambiguity, or to achieve other purposes.

Let us survey some possibilities of diegetic sound. We know that the space of the narrative action is not limited to what we can see on the screen at a given moment. If we already know that several people are present in a room, we can see a shot that shows only one person without assuming that the other people have left. We simply have a sense that those people are offscreen. And if one of those offscreen people speaks, we still assume that the sound is coming from part of the story space. Thus diegetic sound can be either *onscreen* or *offscreen*, depending on whether its source is within the frame or outside the frame.

A shot shows a character talking, and we hear the sound of his or her voice. Another shot shows a door closing, and we hear a slam. A person plays a fiddle, and we hear its notes. In each case the source of the sound is in the story—diegetic—and visible within the frame—onscreen. But the shot may show only a person listening to a voice without the speaker being seen; another shot might show a character running down a corridor and the sound of an unseen door slamming; lastly, an audience is shown listening while the sound of a fiddle is heard. In all of these instances, the sounds come from within the story—again diegetic—but are now in a space outside the frame—offscreen.

At first this may seem a trivial distinction, but we know from our study of framing in Chapter 7 how powerful offscreen space can be. Offscreen sound can suggest space extending beyond the visible action. In *American*

Graffiti, a film that plays heavily on the distinction between diegetic and nondiegetic music, offscreen sounds of car radios often suggest that all of the cars on a street are tuned to the same radio station.

Offscreen sound may also control our expectations about offscreen space. In *His Girl Friday* Hildy goes into the press room to write her final story. As she chats with the other reporters, a loud clunk comes from an unseen source. Hildy glances offscreen left, and immediately a new space comes to our attention. She walks to the window and sees a gallows being prepared for an execution. Here offscreen sound initiates the discovery of a new area of action.

Offscreen sound can make the film's narration less restricted. In John Ford's *Stagecoach*, the stagecoach is desperately fleeing from a band of Indians. The ammunition is running out, and all seems lost until a troop of cavalry suddenly arrives. Yet Ford does not present the situation this baldly. He shows a medium close-up of one of the passengers, Hatfield, who has just discovered that he is down to his last bullet (Fig. 9.9). He glances off right and raises his gun (Fig. 9.10). The camera pans right to a woman, Lucy, praying. During all this, orchestral music, including bugles, plays nondiegetically. Unseen by Lucy, the gun comes into the frame from the left as Hatfield prepares to shoot her to prevent her from being captured by the Indians (Fig. 9.11). But before he shoots, an offscreen gunshot is heard, and Hatfield's hand and gun drop down out of the frame (Fig. 9.12). Then the bugle music becomes somewhat more prominent. Lucy's expression changes as she says, "Can you hear it? Can you hear it? It's a bugle. They're blowing the charge" (Fig. 9.13). Only then does Ford cut to the cavalry itself racing toward the coach.

Fig. 9.9

Fig. 9.10

Rather than showing the cavalry riding to the rescue, the film's narration uses offscreen sound to restrict our awareness to the initial despair of the passengers and their growing hope as they hear the distant sound. The sound of the bugle also emerges imperceptibly out of the nondiegetic music. Only Lucy's line tells us that this is a diegetic sound that signals their rescue, at which point the narration becomes far less restricted.

Diegetic sound harbors other possibilities. Often a filmmaker uses sound to represent what a character is thinking. We hear the character's voice speaking his or her thoughts even though that character's lips do not move; presumably other characters cannot hear these thoughts. Here the narration

Fig. 9.11

Fig. 9.12

Fig. 9.13

uses sound to achieve subjectivity, giving us information about the mental state of the character. Such spoken thoughts are comparable to mental images on the visual track. A character may also remember words, snatches of music, or events as represented by sound effects. In this case the technique is comparable to a visual flashback.

The use of sound to enter a character's mind is so common that we need to distinguish between *internal* and *external* diegetic sound. External diegetic sound is that which we as spectators take to have a physical source in the scene. Internal diegetic sound is that which comes from "inside" the mind of a character; it is subjective. (Nondiegetic and internal diegetic sounds are often called *sound over* because they do not come from the real space of the scene.) In the Laurence Olivier version of *Hamlet,* for example, the filmmaker presents Hamlet's famous soliloquies as interior monologues. Hamlet is the source of the thoughts we hear represented as speech but the words are only in his mind, not in his objective surroundings.

Fig. 9.14

A more complex use of internal diegetic sound occurs in Wim Wenders's *Wings of Desire.* Dozens of people are reading in a large public library. As the camera tracks along past them, we hear their thoughts as a throbbing murmur of many voices in many languages (Fig. 9.14). Incidentally, this sequence also constitutes an interesting exception to the general rule that one character cannot hear another's internal diegetic sound. The film's premise is that Berlin is patrolled by invisible angels who can tune in to humans' thoughts. This is a good example of how the conventions of a genre (here, the fantasy film) and the film's specific narrative context can modify a traditional device.

To summarize: Sound may be diegetic (in the story space) or nondiegetic (outside the story space). If it is diegetic, it may be onscreen or offscreen, and internal ("subjective") or external ("objective").

One characteristic of diegetic sound is the possibility of suggesting the *sound perspective.* This is a sense of spatial distance and location analogous to the cues for visual depth and volume which we get with visual perspective. "I like to think," remarks sound designer Walter Murch, "that I not only record a sound but the space between me and the sound: The subject that generates the sound is merely what causes the surrounding space to resonate."

Sound perspective can be suggested by volume. A loud sound tends to seem near; a soft one, more distant. The horses' hooves in the *Seven Samurai* battle and the bugle call from *Stagecoach* exemplify how rising volume suggests closer distance. Sound perspective is also created by timbre. The combination of directly registered sounds and sounds reflected from the environment creates a timbre specific to a given distance. Such timbre effects are most noticeable with echoes. In *The Magnificent Ambersons* the conversations that take place on the baroque staircase have a distinct echo, giving the impression of huge, empty spaces around the characters.

Multichannel recording and reproduction tremendously increase the filmmaker's ability to suggest sound perspective. In most 35mm theaters equipped with multitrack sound systems, three speakers are located behind the screen. The center speaker transmits most of the onscreen dialogue, as well as the most important effects and music. The left and right speakers are stereophonic, carrying not only important dialogue but also sound effects,

music, and minor dialogue. These channels can suggest a region of sound within the frame or just offscreen. Surround channels principally carry minor sound effects, and they are divided among several speakers arranged along the sides and in the back of the theater.

By using stereophonic and surround tracks, a film can more strongly imply a sound's distance and placement. In farcical comedies like *The Naked Gun* and *Hot Shots*, stereophonic sound can suggest collisions and falls outside the frame. Without the greater localization offered by the stereophonic channels, we might scan the frame for sources of the sounds.

In addition, stereo reproduction can specify a moving sound's direction. In David Lean's *Lawrence of Arabia*, for instance, the approach of planes to bomb a camp is first suggested through a rumble occurring only on the right side of the screen. Lawrence and an officer look off right, and their dialogue identifies the source of the sound. Then, when the scene shifts to the besieged camp itself, the sound slides from channel to channel, suggesting the planes swooping overhead.

With stereophonic and surround channels, a remarkably convincing three-dimensional sound environment may be created within the theater. Sound sources can alter in position as the camera pans or tracks through a locale. The *Star Wars* series uses multiple-channel sound to suggest space vehicles whizzing not only across the screen but also above and behind the spectators.

Like other techniques, sound localization in the theater need not be used for realistic purposes. *Apocalypse Now* divides its six-track sound among three channels in the rear of the theater and three in the front. In the film's first sequence, the protagonist Ben Willard is seen lying on his bed. Shots of his feverish face are superimposed on shots of U.S. helicopters dropping napalm on the Vietnamese jungle. The sound oscillates between internal and external status, as Willard's mind turns the whoosh of a ceiling fan into the whir of helicopter blades. These subjective sounds issue from both the front and the back of the theater.

Abruptly, a POV shot tracking toward the window suggests that Willard has gotten to his feet and is walking. As the camera moves, the noises fade from all rear speakers and become concentrated in the front ones at screen left, right, and center. Then, as Willard's hand opens the venetian blinds to reveal his vision of the street outside, the sound fades out of the left and right front speakers and is heard only from the center channel. Our attention has been narrowed: As we leave Willard's mind, the sound steers us back to the outside world, which is rendered as unrealistically monophonic. In addition, the disparity in acoustic dimensions suggests that the protagonist's wraparound memory of jungle destruction is richer than the pallid environment of Saigon.

In most films, the sources of the sounds are clearly diegetic or nondiegetic. But some films blur the distinction between diegetic and nondiegetic sound, as we saw in the cavalry-rescue scene of *Stagecoach*. Since we are used to identifying a sound's source easily, a film may try to cheat our expectations.

At the beginning of Mel Brooks's *Blazing Saddles,* we hear what we think is nondiegetic musical accompaniment for a cowboy's ride across the prairie—until he rides past Count Basie and his orchestra. This joke depends on a reversal of our expectations about the convention of nondiegetic music. A more elaborate example is the 1986 musical version of *Little Shop of Horrors.* There a trio of female singers strolls through many scenes, providing musical commentary on the action without any of the characters noticing them. (To complicate matters, the three singers also appear in minor diegetic roles and only then do they interact with the main characters.)

More complex is a moment in *The Magnificent Ambersons* when Welles creates an unusual interplay between the diegetic and nondiegetic sounds. A prologue to the film outlines the background of the Amberson family and the birth of the son, George. We see a group of townswomen gossiping about the marriage of Isabel Amberson, and one predicts that she will have "the worst spoiled lot of children this town will ever see." This scene has presented diegetic dialogue. After the last line, the nondiegetic narrator resumes his description of the family history. Over a shot of the empty street, he says: "The prophetess proved to be mistaken in a single detail merely; Wilbur and Isabel did not have *children.* They had only one." But at this point, still over the shot of the street, we hear the gossiper's voice again: "Only one! But I'd like to know if he isn't spoiled enough for a whole carload." After her line, a pony cart comes up the street, and we see George for the first time. In this exchange the woman seems to reply to the narrator, even though we must assume that she cannot hear what he says. (After all, she is a character in the story and he is not.) Here Welles playfully departs from conventional usage to emphasize the arrival of the story's main character and the hostility of the townspeople to him.

This passage from *The Magnificent Ambersons* juxtaposes diegetic and nondiegetic sounds in an ambiguous way. In other films a single sound may be ambiguous because it could fall into either category. In the *Apocalypse Now* sequence already mentioned, the throbbings of the ceiling fan and the helicopter blades are clearly diegetic, but Coppola accompanies these with Jim Morrison's song, "The End." This might be taken as either a subjective part of Willard's fantasy or as nondiegetic, an external commentary on the action in the manner of normal "movie music."

A more disturbing uncertainty about whether a sound is diegetic or not often crops up in the films of Jean-Luc Godard. He narrates some of his films in nondiegetic voice-over, but in other films, such as *Two or Three Things I Know About Her,* he seems also to be in the story space, whispering questions or comments whose sound perspective makes them seem close to the camera. Godard does not claim to be a character in the action, yet the characters on the screen sometimes behave as though they hear him. This uncertainty as to diegetic or nondiegetic sound sources enables Godard to stress the conventionality of traditional sound usage.

The distinction between diegetic and nondiegetic sound is important for understanding particular films, as we shall see in some cases discussed at the end of the chapter.

■ TIME

Sound also permits the filmmaker to represent time in various ways. This is because the time represented on the sound track may or may not be the same as that represented in the image.

The most straightforward audio-visual relations involve sound-image synchronization. The matching of sound with image in projection creates *synchronous sound.* When a sound is synchronized with the image, we hear it at the same time as we see the source produce the sound. Dialogue between characters is normally synchronized so that the lips of the actors move at the same time that we hear the appropriate words.

When the sound does go out of synchronization during a viewing (often through an error in projection or lab work), the result is quite distracting. But some filmmakers have obtained imaginative effects by putting out-of-sync, or *asynchronous,* sound into the film itself. One such effect occurs in a scene in the musical by Gene Kelly and Stanley Donen, *Singin' in the Rain.* In the early days of Hollywood sound filming, a pair of silent screen actors have just made their first talking picture, *The Dueling Cavalier.* Their film company previews the film for an audience at a theater. In the earliest days of "talkies," sound was often recorded on a phonograph record to be played along with the film; hence the chances of the sound's getting out of synchronization with the picture were much greater than they are today. This is what happens in the preview of *The Dueling Cavalier.* As the film is projected, it slows down momentarily, but the record keeps running. From this point all the sounds come several seconds before their sources are seen in the image. A line of dialogue begins, *then* the actor's lips move. A woman's voice is heard when a man moves his lips and vice versa. The humor of this disastrous preview in *Singin' in the Rain* depends on our realization that the synchronization of sound and image is an illusion produced by mechanical means.

A lengthier play with our expectations about synchronization comes in Woody Allen's *What's Up Tiger Lily?* Allen has taken an Asian spy film and dubbed a new sound track on, but the English-language dialogue is not a translation of the original. Instead, it creates a new story in comic juxtaposition with the original images. Much of the humor results from our constant awareness that the words are not perfectly synchronized with the actors' lips. Allen has turned the usual problems of the dubbing of foreign films into the basis of his comedy.

Synchronization relates to screen duration, or *viewing* time. As we have seen in Chapter 4, narrative films can also present *plot* and *story* time. To recall the distinction: Story time consists of the order, duration, and frequency of all the events pertinent to the narrative, whether they are shown to us or not. Plot time consists of these temporal qualities (order, duration, and frequency) of the events actually represented in the film. Plot time shows us selected story events but only refers to others. Thus it usually covers a shorter span than the complete story does.

Story and plot time can be manipulated by sound in two principal ways. If the sound takes place at the same time as the image in terms of the story events, it is *simultaneous sound.* This is by far the most common usage. When characters speak onscreen, the words we hear are occurring at the same moment in the plot's action as in story time.

But it is possible for the sound we hear to occur earlier or later in the story than the events which we see in the image. In this manipulation of story order, the sound becomes *nonsimultaneous.* The most common example of this is the sonic flashback. For instance, we might see a character onscreen in the present but hear another character's voice from an earlier scene. By means of nonsimultaneous sound, the film can give us information about story events without showing them to us. And nonsimultaneous sound may, like simultaneous sound, have either an external or an internal source—that is, a source in the "objective" world of the film or the "subjective" realms of the character's mind.

As these categories suggest, temporal relationships in the cinema are complex. To help distinguish them, Table 9.2 sums up the possible temporal and spatial relationships that image and sound can display.

Diegetic sound. Because the first and third of these possibilities are comparatively uncommon, we start by commenting on the second, most normal option.

2. *Sound simultaneous in story with image.* This is by far the most common temporal relation which sound has in fiction films. Noise, music, or speech that comes from the space of the story almost invariably occurs at the same time as the image. Like any other sort of diegetic sound, simultaneous sound can be either external (objective) or internal (subjective).

1. *Sound earlier in story than image.* Here the sound comes from an earlier point in the story than the action currently visible onscreen. A clear example occurs at the end of Joseph Losey's *Accident.* Over a shot of a driveway gate, we hear a car crash. The sound represents the crash that occurred at the *beginning* of the film. Now if there were cues that the sound was internal—that is, that a character was recalling it—it would not strictly

Table 9.2 TEMPORAL RELATIONS OF SOUND IN CINEMA

	SPACE OF SOURCE	
TIME	DIEGETIC (STORY SPACE)	NONDIEGETIC (NONSTORY SPACE)
1. Nonsimultaneous: sound from *earlier* in story than image	Sound flashback; image flashforward; sound bridge	Sound marked as past put over images (e.g., sound of a John Kennedy speech put over images of United States today)
2. Sound *simultaneous* in story with image	*External:* dialogue, effects, music *Internal:* thoughts of character heard	Sound marked as simultaneous with images put over images (e.g., a narrator describing events in the present tense)
3. Nonsimultaneous: sound from *later* in story than image	Sound flashforward; image flashback with sound continuing in the present; character narrates earlier events; sound bridge	Sound marked as later put over images (e.g., reminiscing narrator of *The Magnificent Ambersons*)

be coming from the past, since the memory of the sound would be occurring in the present. But here no character is remembering the scene, so we have a fairly pure case of a "sonic flashback." In this film, an unrestricted narration makes an ironic final comment on the action. Similarly, early in Ron Howard's *The Paper*, we see images of two teenagers inspecting bloody corpses in a car while the sound of gunfire provides a flashback to the murders.

Sound may belong to an earlier time than the image in another way. The sound from one scene may linger briefly while the image is already presenting the next scene. This is called a *sound bridge*. Sound bridges of this sort may create smooth transitions by setting up expectations that are quickly confirmed. One scene of Jonathan Demme's *The Silence of the Lambs* ends with the heroine talking on the phone, identifying a location as the "Your Self Storage facility. . . ." Her over-the-phone description continues on the sound track (". . . right outside central Baltimore") while the image track presents a medium shot of the Your Self Storage facility sign, introducing the next scene.

Sound bridges can also make our expectations more uncertain. In Tim Hunter's *The River's Edge*, three high-school boys are standing outside school, and one of them confesses to having killed his girlfriend. When his pals scoff, he says, "They don't believe me." There is a cut to the dead girl lying in the grass by the river, while on the sound track we hear one of his friends respond to him by calling it a crazy story that no one will believe. For an instant we cannot be sure whether a new scene is starting or we are seeing a cutaway to the corpse, which could be followed by a shot returning to the three boys at school. But the shot dwells on the dead girl, and after a pause we hear, with a different sound ambience: "If you brought us. . . ." Then there is a cut to a shot of the three youths walking through the woods to the river, as the same character continues, ". . . all the way out here for nothing. . . ." The friend's remark about the crazy story belongs to an earlier time than the shot of the corpse, and it is used as an unsettling sound bridge to the new scene.

3. *Sound later in story than image.* Nonsimultaneous diegetic sound may also occur at a later time than that depicted by the images. Here we are to take the images as occurring in the past and the sound as occurring in the present or future. A simple prototype occurs in many trial dramas. The testimony of a witness in the present is heard on the sound track, while the image presents a flashback to an earlier event. The same effect occurs when the film employs a reminiscing narrator, as in John Ford's *How Green Was My Valley*. Aside from a glimpse at the beginning, we do not see the protagonist Huw as a man, only as a boy, but his narration accompanies the bulk of the plot, which is set in the distant past. Huw's voice on the sound track creates a strong nostalgia for the past and constantly reminds us of the pathetic decline that the characters will eventually suffer.

Since the late 1960s, it has become somewhat common for the sound from the next scene to begin while the images of the last one are still on the screen. Like the instances mentioned above, this transitional device is called a *sound bridge*. In Wim Wenders's *American Friend*, a nighttime shot of a little boy riding in the back seat of a car is accompanied by a harsh clacking.

There is a cut to a railroad station, where the timetable board flips through its metal cards listing times and destinations. Since the sound over the shot of the boy comes from the space of the later scene, this portion is nonsimultaneous.

If the sound bridge is not immediately identifiable, it can surprise or disorient the audience, as in this *American Friend* transition. A more recognizable sonic lead-in can create more clear-cut expectations about what we will see in the next scene. Federico Fellini's *8½* takes place in a town famous for its health spa and natural springs, and several scenes have shown an outdoor orchestra playing to entertain the tourists and guests. Midway through the film a scene ends with the closing of a window on a steam bath. Near the end of the shot, we hear an orchestral version of the song "Blue Moon." There is a cut to an orchestra playing the tune in the center of the town's shopping area. Even before the new scene has established the exact locale of the action, we can reasonably expect that the musical bridge is bringing us back to the public life of the spa.

In principle, one could also have a sound flashforward. The filmmaker could, say, use the sounds that belong with scene 5 to accompany the images in scene 2. In practice, such a technique is very rare. In Godard's *Band of Outsiders,* the sound of a tiger's roar is heard as sound "over," not as sound "off," several scenes before we see the tiger. A more ambiguous case can be found in Godard's *Contempt.* A husband and wife quarrel, and the scene ends with her swimming out to sea while he sits quietly on a rock formation. On the sound track we hear her voice, closely miked, reciting a letter in which she tells him she has driven back to Rome with another man. Since the husband has not yet received the letter, and perhaps the wife has not yet written it, the letter and its recitation presumably come from a later point in the story. Here the sound flashforward sets up strong expectations which a later scene confirms: We see the wife and the husband's rival stopping for gas on the road. In fact we never see a scene in which the husband receives the letter.

Nondiegetic sound. Most nondiegetic sound has no relevant temporal relationship to the story. When "mood" music comes up over a tense scene, it would be irrelevant to ask if it is happening at the same time as the images, since the music has no place in the world of the action. But occasionally the filmmaker may use a type of nondiegetic sound that does have a defined temporal relationship to the story. Welles's narration in *The Magnificent Ambersons,* for instance, speaks of the action as having happened in a long-vanished era of American history.

■ SUMMARY

As we watch a film, we do not mentally slot each sound into each of these spatial and temporal categories. But our categories do assist us in analysis. They offer us ways of noticing important aspects of films—especially films that play with our expectations about sounds. By becoming aware of the rich range of possibilities, we are less likely to take a film's sound track for granted and are more likely to notice unusual sound manipulations.

One such manipulation comes early in Alain Resnais's *Providence*. As the film begins, we see a mysterious house and men hunting down a wounded old man. Suddenly we are in a courtroom, where a witness is being interrogated. These abrupt transitions give us little time to form expectations. Apparently a prosecutor is questioning a young man accused of the mercy killing of the old man during the hunt. The young man justifies his act by saying that the man was not only dying but turning into an animal. (We had seen the man's hairy face and clawlike hands, so now we begin to see the links between the scenes.) The prosecutor pauses, astonished: "Are you suggesting some kind of actual metamorphosis?" He pauses again, and a man's voice whispers, "A werewolf." The prosecutor then asks, "A werewolf, perhaps?"

The whispered words startle us, for we cannot immediately account for them. Are they whispered by an unseen character offscreen? Are they perhaps even nondiegetic, coming from outside the story world? Only much later in the film do we find out whose voice whispered these words, and why. The whole opening of *Providence* provides an excellent extended case of how disorienting an ambiguity of sound sources can be when the filmmaker departs from conventional usage.

In the *Providence* sequence, we are aware of the ambiguity immediately, and it points our expectations forward, arousing curiosity as to how the whisperer can be identified. The filmmaker can also use sound to create a retrospective awareness of how we have *mis*interpreted something earlier. This occurs in Francis Ford Coppola's *The Conversation*, a film that is virtually a textbook on the manipulation of sound and image.

The plot centers on Harry Caul, a sound engineer specializing in surveillance. Harry is hired by a mysterious corporate executive to tape a conversation between a young man and woman in a noisy park. Harry cleans up the garbled tape, but when he goes to turn over the copy to his client, he suspects foul play and refuses to relinquish it.

Now Harry obsessively replays, refilters, and remixes all his tapes of the conversation. Flashback images of the couple—perhaps in his memory, perhaps not—accompany his reworking of the tape. Finally Harry arrives at a good dub, and we hear the man say, "He'd kill us if he could."

The overall situation is quite mysterious. Harry does not know who the young couple are (is the woman his client's wife or daughter?). Nevertheless Harry suspects that they are in danger from the executive. Harry's studio is ransacked, the tape is stolen, and he later finds that the executive has it. Now more than ever Harry feels that he is involved in a murder plot. After a highly ambiguous series of events, including Harry's bugging of a hotel room during which a killing takes place, Harry learns that the situation is not as he had thought.

Without giving away the revelation of the mystery, we can say that in *The Conversation* the narration misleads us by suggesting that certain sounds are objective when at the film's end we are inclined to treat them as subjective, or at least ambiguous. The film's surprise, and its lingering mysteries, rely on unsignaled shifts between external and internal diegetic sound.

The two films just discussed point to a second justification for our set of categories. These categories seem to correspond to tacit assumptions and

inferences which spectators actually make. The films and our reactions to them suggest that we quickly learn to distinguish between internal and external, diegetic and nondiegetic, simultaneous and nonsimultaneous sound. We are surprised or amused when films violate these categories; we are puzzled or misled when a sound source shifts from one category to another. If these categories did not tally with our assumptions, such films as *Providence* and *The Conversation* would not have the power to undermine our expectations, to create suspense or surprise or ambiguity. Our taxonomy, then, is offered both as a tool for film analysis and as a systematic outline of viewers' ordinary intuitions.

FUNCTIONS OF FILM SOUND: *A MAN ESCAPED*

Robert Bresson's *A Man Escaped (Un Condamné à mort c'est échappé)* illustrates how a variety of sound techniques can function throughout an entire film. The story takes place in France in 1943. Fontaine, a Resistance fighter arrested by the Germans, has been put in prison and condemned to die. But while awaiting his execution, he works at an escape plan, loosening the boards of his cell door and making ropes. Just as he is ready to put his plan in action, a boy named Jost is put into his cell. Deciding to trust that Jost is not a spy, Fontaine reveals his plan to him, and they are both able to escape.

Throughout the film, sound has many important functions. As in all of his films, Bresson emphasizes the sound track, rightly believing that sound may be just as "cinematic" as images. At certain points in *A Man Escaped,* Bresson even lets his sound technique dominate the image; throughout the film, we are compelled to *listen.* Indeed, Bresson is one of a handful of directors who create a complete interplay between sound and image.

A key factor in guiding our perception of the action is the commentary spoken over by Fontaine himself. The voice-over is nonsimultaneous, since it occurs at a time later than the images. But it could be either internal or external sound, since we never learn whether he is simply thinking back over these events or telling them to someone.

Fontaine's narration has several functions. First, the commentary helps clarify the action. Certain temporal cues suggest how long Fontaine spends in prison. As we see him working at his escape plan, his voice-over tells us, "One month of patient work and my door opened." At other points he gives us additional indications of time. His commentary is particularly important during the final escape scene, where hours of action occupy only 15 minutes of viewing time and the narration is narrowly limited to what Fontaine could know. Fontaine's voice calmly tells us of his and Jost's patient, cautious progress toward freedom.

We receive other vital information through the commentary. Sometimes the narration simply states facts: that the pin Fontaine obtains came from the women's wing of the prison or that certain prison officials' quarters were at various places in the building. More strikingly, Fontaine often tells what his thoughts had been. After being beaten and put in his first cell, he wipes the