of opposed elements; but unlike Burch, their possible combinations include some that are merely different as well as some that are directly opposed. Whereas Burch polemically addresses the issue of how sound should function, Bordwell and Thompson attempt to exhaust all of its possible functions. The value of both approaches lies in the construction of a grammar of sorts for film sound and of categories useful for analyzing its formal structure.

Fundamental Aesthetics of Sound in the Cinema
DAVID BORDWELL and KRISTIN THOMPSON

The Powers of Sound

Many people tend to think of sound as simply an accompaniment to the real basis of cinema, the moving images. These viewers assume that the people and things pictured on the screen just produce an appropriate noise. But in the process of film production the sound track is created separately from the images and can be manipulated independently and flexibly.

Consider some of the advantages of sound for a film. First, it engages another sense mode: our visual attention can be accompanied by an 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.) Second, sound can actively shape how we interpret the image. In one sequence of Letter from Siberia, Chris Marker demonstrates the power of sound to alter our perception of images. Three times Marker plays 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 1). The verbal differences are emphasized by the sameness of the images; the audience will interpret the same images completely differently, depending on the sound track.

The Letter from Siberia sequence also demonstrates a third advantage of sound. Film sound can direct our attention quite specifically within the image. When the commentator describes the “blood-colored buses,” we will look at the bus and not at the car. When Fred Astaire and Ginger Rogers are tapping out 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 complex 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 next shot shows the door, now open, the viewer
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 joyful spirit of socialist emulation, happy Soviet workers, among them this picturesque denizen of the Arctic reaches, apply themselves to making Yakutsk an even better place to live.

Or else:

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 to the task like slaves, the miserable Soviet workers, among them this sinister-looking Asiatic, apply themselves to the primitive labor of grading with a drag beam.

Or simply:

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. With courage and tenacity under extreme difficult conditions, Soviet workers, among them this Yakut afflict with an eye disorder, apply themselves to improving the appearance of their city, which could certainly use it.
will likely focus his or her attention on that door, the source of the off-screen sound. In an opposite way, if the next shot shows the door still closed, the viewer will likely ponder his or her interpretation of the sound. (Maybe it wasn’t a 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.

Moreover, as V. F. Perkins has pointed out, sound brings with it a new sense of the value of silence. “Only with color 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. We should, however, isolate some qualities of sound as we perceive it. These qualities are familiar to us from everyday experience.

LOUDNESS

Perceived sound results from vibrations in air; the amplitude of the vibrations produces our sense of volume. Film sound constantly manipulates volume. For example, in dozens of films a long shot of a busy street is accompanied by loud traffic noises, but when two people meet and start to speak, the loudness of the noise 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. Needless to say, loudness is also affected by perceived distance; often the louder the sound, the closer we take it to be. Some films exploit radical changes in volume for shock value, 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. Pitch is the principal way we distinguish music from other sounds in the film, but it has more complex uses. 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. 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 change. When Bernard Herrmann obtained the effects of unnatural, 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. When we call someone’s voice nasal or a certain musical tone mellow, we are referring to timbre. Again, filmmakers manipulate timbre continually. Timbre can help articulate portions of the sound track; for instance, timbre 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 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.

As the fundamental components of film sound, loudness, pitch, and timbre usually interact to define the sonic texture of a film. At the most elementary level, loudness, pitch, and timbre enable us to distinguish among all of the sounds in a film; we recognize different characters’ voices by these qualities, for example. But at a more complex level, all three interact to add considerably to our experience of the film. Both John Wayne and James Stewart speak slowly, but Wayne’s voice tends to be deeper and gruffer than Stewart’s querulous drawl. 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, and a motif is formed by the inability of Kane’s opera-singing wife to hit accurate pitch. In Citizen Kane 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 (shift in volume and timbre); Leland beginning a sen-
tence in the street cuts to Kane finishing the sentence in an auditorium, his voice magnified by loudspeakers (shift in volume, timbre, and pitch). Such examples suggest that the elementary properties of sound afford a rich set of possibilities for the filmmaker to explore.

Selection and Combination

Sound in the cinema takes three forms: speech, music, or noise (also called sound effects). Occasionally a sound may share 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 the human voice and instead hear “screaming” violins.) Nevertheless, in most cases the distinctions hold. Now that we have an idea of the role of 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 select from several shots the best image, he or she may choose what exact bit of sound from this or that source will best serve the purpose. And just as the filmmaker may link or superimpose images, so may he or she join any two sounds end to end or one “over” another (as with commentary “over” music). Though we aren’t usually as aware of the manipulation of the sound track, it demands as much selection and control as does the visual track.

Selection of the desired sound is a necessary step in the process. 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. As any amateur recordist knows, 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 achieve the desired result. Sound studios, camera blimps to absorb motor noise, directional and shielded microphones, sound engineering and editing, and libraries of stock sounds all exist so that a film’s sound track may be carefully controlled through selection. Unless a filmmaker wants to record all of the ambient noise of a scene, simply holding out a microphone while filming will rarely suffice.

Because normal perception is linked to our choices, the director’s selection of the sounds in a film can control the audience’s choices and thus guide the audience’s perception. In one scene from Jacques Tati’s Mr. Hulot’s Holiday, vacationers at a resort hotel are relaxing. 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 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 and perception. If you start to notice how such selection of sound guides our perception, you will notice that filmmakers often use sound to shift our attention.

Such examples depend not only on selection but also on the filmmaker’s combining of various sonic elements. It is the mixing of sounds in a specific pattern that constitutes the sound track as we know it. We have already seen that mixing is a careful and deliberate production procedure. What we must now notice is what functions the particular mix can have in the total film. Obviously the mix can range from very dense (e.g., a scene containing the babble of voices, the sounds of footsteps, Muzak, and plane engines at an airport) to total silence, with most films falling in between. In addition, the filmmaker may create a mix in which each sound modulates and overlaps smoothly with the others, or one that is composed of much more abrupt and startling contrasts.

The possibilities of combining sounds are 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 sounds of swords being drawn are punctuated by long pauses in which we hear only the drumming rain. Suddenly distant horses’ hoofs are heard off-screen, Kurosawa cuts to a long shot of the bandits; their horses’ hoofs become abruptly louder. (This is typical of the scene: the closer the camera is to a sound 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-scène 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. The scene climaxes after the main battle has ended: offscreen horses’ hoofs are
cut short by a new sound—the sharp crack of a bandit’s rifle shot, which falls 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 now riderless horses, and the rain all fade slowly out. The relatively dense mix of this sound track (accomplished entirely without music) gradually introduces sounds to turn our attention to new narrative elements (hoofs, battle cries) and then goes on to modulate these sounds smoothly into a harmonious whole. This whole is then punctuated by abrupt sounds of unusual volume or pitch associated with crucial narrative actions (the archery, women’s screams, the gunshots).

Dimensions of Film Sound

We have now 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, the sound relates to visual events that take place in a specific time, and this relationship gives sound a temporal dimension. And fourth, sound conveys a sense of the spatial conditions in which it occurs. These categories begin to reveal that sound in film is actually a very complex thing; let’s look at each category briefly.

Rhythm

For our purposes, sounds can be considered to be organized rhythmically when strong and weak beats form a distinct pattern and move at a distinct pace. (This definition of rhythm combines features that musicians would distinguish as “meter,” “rhythm,” and “tempo.”) But even this simple definition is complicated by the fact that the movements in the images themselves have a rhythm as well. In addition, the editing has a rhythm; a succession of short shots creates a fast rhythm, whereas shots held longer slow down the editing’s rhythm. Moreover, all three types of sound on the sound track have their own rhythmic possibilities independent of one another. The gasping voice of a character who lies dying has a slower rhythm than the voice of a racetrack announcer. Music obviously may have different rhythms in a film. Finally, sound effects also vary in rhythm (compare the plodding hoofs of a farmhorse and a cavalry company riding at full speed).

But in most cases the rhythms of editing, of movement within the image, and of sound do not exist separately. Usually sound accompanies movements and often continues over cuts. Sound may motivate figure or camera movement. Thus there exists the potential for a considerable interplay among these three types of rhythm. No one “appropriate” combination exists.

Possibly the most common tendency is for the filmmaker to match visual and sonic rhythms to each other. An obvious example is the typical dance sequence in a musical; here 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, there is no cutting at all within the dance, for the scene consists of a single long take from a long-shot distance. Another example 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. (This nondance matching of movement with music is in fact known as “Mickey Mousing.”)

The filmmaker may choose to create a disparity among the rhythms of sound, editing, and image. One way of accomplishing this is to keep the source of the sound off-screen and to show something else on-screen. Toward the end of John Ford’s She Wore a Yellow Ribbon, the aging cavalry captain, Nathan Brittiles, watches his troop 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 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 move at a rhythm matched to the sound. Instead, the scene concentrates our attention on Brittiles, standing almost motionless by his own horse. (The moderate rate of cutting lies between these two rhythms.) The contrast of fast rhythm of sound with the shots of the solitary Brittiles functions to emphasize his regret at having to stay behind for the first time in many years.

Several great directors have used music that might seem to have a rhythm inappropriate for the visuals. In Four Nights of a Dreamer Robert Bresson includes several 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 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 *Playtime*. 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 starts our expectations because it seems inappropriate to the images; in fact, it belongs with 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 rhythms to what may be its logical limit, in *La Période de République* the sea, 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 and little movement.

**Fidelity**

By fidelity we don't mean "high-fi" in the sense of the quality of recording. Here we are speaking of whether the sound is faithful to the source as we conceive it. If a film shows us a barking dog and the sound track has a barking noise, that sound is faithful to its source—the sound maintains fidelity. But if the sound of a cat meowing accompanies the picture of the barking dog, there enters a disparity between sound and image—a lack of fidelity.

Fidelity has nothing to do with what originally made the sound during filming. 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. Note, however, that fidelity is purely a matter of our expectations. In production the bark or meow might be produced electronically or by an animal-imitator. Fidelity involves conventional expectations about sources, not knowledge of where the filmmaker actually obtained the sound.

A play with fidelity most commonly functions for comic effect. Jacques Tati is one of the directors most skilful at employing various degrees of fidelity. In *Mr. Hulot's Holiday* much comedy 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 patterns in which waiters and diners pass through the door. 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, there is no recipe that will allow us to interpret every manipulation of fidelity as comic; some non-faithful 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 striking transition.

Finally, in some special cases fidelity may be manipulated by a change in volume. A sound may seem unreasonably loud or soft in relation to the 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. 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.

**Space**

Sound has a spatial dimension because it comes from a source, and that source may be characterized by the space it occupies. If the source of a sound is a character or object in the story space of the film, we call the sound die-
The voices of the characters, sounds made by objects in the story, or music coming from instruments in the story space are all diegetic sound.

On the other hand there is nondiegetic sound, which does not come from a source in the story space. Familiar examples of such sound are music coming from instruments in the story space are all diegetic sound.

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. Orson Welles speaks the nondiegetic narration in his own film "The Magnificent Ambersons" for example. Nondiegetic sound effects are also possible. In "Le Million" various characters all pursue an old coat with a winning lottery ticket in its pocket. They converge backstage at the opera and begin racing and dodging 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 as ball, this enhances the comedy of the sequence. We hear a crowd cheering and a whistle's sound; yet we don't assume that the characters present are making these sounds (so this is not a manipulation of fidelity, as with the earlier examples from "Le Million"). The nondiegetic sounds create comedy by making a sort of audiovisual pun.

What are the 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 any given moment. If we 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 dropped out of the story. We simply have a sense that those people are off-screen. And if one of those off-screen people speaks, we still assume that the sound is coming from part of the story space. Thus diegetic sound can be either on-screen or off-screen, depending on whether its source is within the frame or outside the frame.

Simple examples will illustrate this. A shot shows a character talking, and we hear the sound of his or her voice; another 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—on-screen. 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 street 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—off-screen.

Off-screen sound can suggest space extending in various directions beyond the visible action. In "American Graffiti," a film that plays heavily on the distinction between diegetic and nondiegetic music, off-screen sounds of car radios often suggest that all of the cars on the street are tuned to the same radio station. Off-screen sound may also control when we begin to formulate expectations about off-screen 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 off left, and immediately a new space comes to our attention, though we haven't seen it yet. She walks to the window and sees a gallows being prepared for an execution. Here off-screen sound initiates the discovery of fresh space.

A brilliant use of a similar device comes 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 create the situation this boldly. He shows a medium close-up of one of the men, Hatfield, who has just discovered that he is down to his last bullet; he glances off right and raises his gun. The camera pans right to a woman, Lucy, praying. During all this, orchestral music, including bugsles, plays nondiegetically. Unseen by Lucy, Hatfield's gun comes into the frame from the left, ready to shoot her to prevent her from being captured by the Indians. But before he shoots, an off-screen gunshot is heard, and Hatfield's hand and gun drop down out of the frame. Then the bugle music becomes somewhat more prominent, and Lucy's face changes as she says, "Can you hear it? Can you hear it? It's a bugle. They're blowing the charge." Only then does Ford cut to the cavalry itself racing toward the coach. Rather than focusing on the mechanics of the rescue, Ford uses off-screen sound to restrict our vision 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 which signals their rescue.

Are there other possibilities for diegetic sound? 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. A character may also remember words, snatches of music, or events as represented by sound effects. This device 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 only from 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.)
maker presents Hamlet's famous soliloquies as interior monologues. The character registers the appropriate emotion on his face, but does not move his lips while we hear his voice saying the words, "To be or not to be . . . ," and so on. Hamlet is the source of the thoughts we hear represented as speech, but the words are only in the character's mind, not in his physical space. Hence the words are simple diegetic, but internal.

To summarize: Sound may be diegetic (in the story space) or nondiegetic (outside the story space). If it is diegetic, it may be on-screen or off-screen, internal ("subjective") or external ("objective").

One characteristic of diegetic sound is the possibility of suggesting the distance of its source. Volume is one simple way to give an impression of distance. A loud sound tends to seem near; a soft one, more distant. The horses' hoofs in the Seven Samurai battle and the bugle call from Stagecoach exemplify this.

In addition to volume, timbre may suggest the texture and dimensions of the space within which a sound supposedly takes place. In the Magnificent Ambersons the conversations that take place on the baroque staircase have an echoing effect, giving the impression of huge, empty spaces around the characters. The skilful filmmaker will pay attention to the quality of the sound, taking advantage of the possibilities of variation from shot to shot.

In recent years technical developments have added the possibilities of stereo, quadraphonic, and other multichannelled systems to the filmmaker's range. This means that the sound can suggest location not only in terms of distance (volume, resonance) but also by specifying direction. In stereo versions of David Lean's Lawrence of Arabia, for example, 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 stereo sound slides from channel to channel, reinforcing the visual depiction of the planes swooping overhead. Multiple channels make it possible to delineate space precisely.

In general, the spatial relations of sounds in films are clearly diegetic or nondiegetic. But because film is such a complex art form, involving the combination of so many elements, some films blur the distinctions between diegetic and nondiegetic sound. Since we are used to placing the source of a sound easily, a film may cheat our expectations.

There 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 involved 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 gossip'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.

This example from the Magnificent Ambersons juxtaposes diegetic and nondiegetic sounds in an ambiguous way. In other films a single sound may be ambiguous because it seems to fall with equal logic into either category. This is often true in the films of Jean-Luc Godard. He narrates some of his films, but sometimes he seems also to be present in the story space just off-screen. Godard does not claim to be a character in the action, yet the characters on the screen sometimes seem to hear him. In an early scene in Two or Three Things I Know About Her Godard's voice introduces the actress Marina Vlady and describes her, then does the same with the character that Vlady plays, Juliette Janson. He speaks in a whisper, and we are not sure whether she can hear him or not. Later in the scene she gives answers to questions seemingly asked by someone off-screen. Yet we do not hear the questions themselves and don't know if Godard is asking them from his position behind the camera as director. We are never sure whether Godard is nondiegetic narrator or diegetic character; in the latter case, his role would have to be something like "director/narrator of Two or Three Things I Know About Her." This uncertainty is important for Godard, since in some of his films an uncertainty as to diegetic or nondiegetic sound sources enables him to stress the conventionality of traditional sound usage.

The distinction between diegetic and nondiegetic sound is important not as an end in itself but as a tool for understanding particular films.

Time

Sound relates temporally to filmic images in two ways: viewing time and story time. By viewing time we mean the physical length of the film—the time it takes the film to be projected. This usually differs from story time, that is, the time assumed to pass in the film's action. Events may cover a number of years in the characters' lives (story time), but most films we see in theaters
take only about two hours to watch (viewing time). Viewing time becomes an instrument of the plot's manipulation of story time.

A sound may be juxtaposed in any temporal relationship with an image. The matching of sound with image in terms of our viewing time is called synchronization. When a sound is synchronized with the image, we hear it at the same time as we see the source produce the sound on the screen. Most dialogue between characters is matched 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 (e.g., through an error in projection), the result is quite distracting. But some imaginative filmmakers have obtained good effects by using out-of-sync, or asynchronous, sound. One such example occurs in a scene in the musical by Gene Kelly and Stanley Donen, Singin' in the Rain. The story is set in the early days of sound in Hollywood; a famous 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 source is 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 sound and image are supposed to be matched, but actually occur separately.

A lengthier example of a play with our expectations about synchronization comes in Woody Allen’s film What's Up Tiger Lily? Allen has taken an Oriental spy film and dubbed a new sound track on, but the English-language dialogue is not a translation of the original; rather, 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 viewing time. But what of story time? If the sound takes place at the same time as the image in terms of the story events, it is simultaneous sound; if the sound occurs earlier or later than the story events of the image, the sound is nonsimultaneous.

Most of the time a film's sound is simultaneous, with image and sound both in the present. We are familiar with this from countless dialogue scenes, musical numbers, chase scenes, and so forth. We shall call this simultaneous diegetic sound simple diegetic.

<table>
<thead>
<tr>
<th>Temporal relation:</th>
<th>Space of source</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Displaced diegetic:</td>
</tr>
<tr>
<td>1. Sound earlier than image</td>
<td>External: sound flashback; image flashforward</td>
</tr>
<tr>
<td></td>
<td>Internal: Memories of character heard</td>
</tr>
<tr>
<td>2. Sound simultaneous with image</td>
<td>Simple diegetic:</td>
</tr>
<tr>
<td></td>
<td>External: dialogue, effects music</td>
</tr>
<tr>
<td></td>
<td>Internal: thoughts of character heard</td>
</tr>
<tr>
<td>3. Sound later than image</td>
<td>Displaced diegetic:</td>
</tr>
<tr>
<td></td>
<td>External: Sound flashforward; image flashback with sound continuing in the present; character narrates earlier events</td>
</tr>
<tr>
<td></td>
<td>Internal: Character’s vision of future heard</td>
</tr>
</tbody>
</table>

But scenes with nonsimultaneous sound are relatively familiar as well. Diegetic sound can occur in a time either earlier or later than the time of the image. In either case we shall call it displaced diegetic sound. As we saw earlier, both types of diegetic sound can have either an external or an internal source.

As these categories suggest, temporal relationships in the cinema are complex. To help distinguish them, Table 2 sums up the possible temporal and spatial relationships which may exist between image and sound.

Diegetic Sound

We have already discussed simple diegetic sound, both external and internal. This is the commonest kind of sound in films. Some concrete examples may help clarify the distinctions among the other categories.
1. **Sound earlier than image.** Displaced diegetic sound may recall an earlier scene through the repetition of sound from that scene while the images on the screen remain in the present. In a scene in Hitchcock's *Psycho* the repeated sounds are memories recalled by the central character, Marion Crane. In a previous scene Marion's boss has told her to deposit a large sum of money in the bank. Instead, she steals the money, and we see her driving away. The images show Marion in a medium close-up behind the wheel of her car. On the sound track, however, we hear an exact repetition of the lines spoken earlier by her boss; this sound is her memory of the earlier scene.

Less common but still possible is external sound that forms a sound flashback. Joseph Losey's *Accident* ends with a shot of a driveway gate. We hear a car crash, but the sound represents the crash that occurred at the beginning of the film. Since no one is remembering the scene and since the sound is from an earlier time than the image, we have an external sound flashback.

2. **Sound later than image.** Displaced diegetic sound may also occur at a later time than that of the images. Probably the most familiar use of this category is the narrator who tells a story that has occurred in the past. In Ford's *How Green Was My Valley* the man Huw narrates the story of his boyhood in Wales. Aside from a glimpse at the beginning, we do not see him as a man, only as a boy; the words of the narration are spoken by a man's voice in a period long after that of the events we see on the screen.

An internal use of sound that is later than the images is rare, but there are a few cases. Later in the scene from *Psycho* described above, Marion begins to imagine what her boss will say on Monday when he discovers the theft. While the image of her driving the car remains on the screen, we hear internal displaced diegetic sound representing a later time: a character's premonition on the sound track.

Sound may belong to a later time than the image in another way. In some cases, particularly in films of the 1960s and 1970s, the sound from the next scene begins while the images of the last one are still on the screen. This is called a sound bridge. Sound bridges create transitions, since we see one image, say, of a person's face, but hear what seems an inappropriate sound, perhaps of a band playing. Then a cut reveals a new locale and time, and we see the band which was the source of the music. Since the sound belongs to the later scene, the moment before the cut uses nonsimultaneous sound.

Examples of sound flashforwards (external displaced diegetic) are rare, perhaps nonexistent, but such a use is logically possible in the cinema.

---

**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 for us to ask if it is happening at the same time as the images, since the music has no relation to the space of the story. But occasionally the filmmaker may use a type of nondiegetic sound that does have a defined temporal relationship to the story. For example, Orson Welles's narration in *The Magnificent Ambersons* speaks of the action as having happened long ago, in a different era of American history.

**Summary**

All of these temporal categories offer us ways of making important distinctions in analyzing films. Fritz Lang's *Secret Beyond the Door*, for example, has an unusual combination of internal displaced diegetic and internal simple diegetic sound. In the first third of the film, the heroine's wedding is about to take place; her voice is heard recalling the circumstances that led to her marriage. This interior monologue is in the present tense. Over the scenes from the past, the monologue is internal displaced diegetic, but over the framing scenes of the wedding, her voice continues as internal simple diegetic. As a whole, the film depends on the contrast of two types of internal speech: the character's mind reacts to immediate situations (simple diegetic), and her mind reflects on past events (displaced diegetic). In such cases the various categories of sound help sharpen our awareness of the ways in which sound can combine with images.

---

**Note**