

Music, Sound, and the Space of Narrative

Concepts and Terminology

Introduction

Under normal circumstances, sound is closely tied to physical space. The basic human cognition of sound is to "anchor" it in an object—a person speaks, a door creaks, a dove coos, the radio plays. Those objects necessarily reside in some physical space (the space of the real world). Narrative complicates matters, however. A storyteller speaks and in that sense is the object-source of that speech, but he or she also conjures up a fantasy world whose own "objects" can be the sources of sound, as when the storyteller attributes dialogue to a character ("Then George said, 'I don't know what you mean'").

The cinematic narrative complicates things still further because it directs most of the attention to that conjured-up fantasy world, so that the primary vehicle of narrative is no longer speech but visual representation. In film, the role of a storyteller is sometimes taken over by voice-over narration, someone unseen who speaks on the sound track, but there is always an "implied narrator" (because narrative films tell stories, there must be a storyteller "somewhere"). Therefore, the viewer-listener's cognition of sound is obliged to operate at two levels: the level of the narration (implied or spoken) and the level of sound in the fantasy or screen world.

"Background" and "source," respectively, are terms commonly used for sound in these two levels, but as we noted in chapter 1, "background" is too easily confused with the subordinate element in the foreground/background pair. (You'll see what we mean if we say that in most cases of voice-over narration background sound is foregrounded.) The terms "onscreen" and "offscreen" do not solve the problem because "screen" refers simply to the part of the film world that is within the camera's frame at any particular moment. "Offscreen," then, is whatever part of that filmic world we cannot see in the frame

but may already have seen or may imaginatively project from the part we can see (recall those chirping bird sounds in the opening minute or so of *Atonement* (2007)). Partly to solve this problem, but also partly to draw attention to the fundamental importance of visual representation in film, scholars have borrowed the term "diegetic" from literary theory to refer to the world of the narrative, the screen world or world of the film. Thus, "nondiegetic" refers to the level of narration: voice-over narration is nondiegetic—and so is underscoring.

In this chapter, we first consider the diegetic/nondiegetic pair, which is essential to the analysis and interpretation of sound and music in film but from which all kinds of ambiguities arise. Then, we will cover a number of specific traits and devices of diegetic sound related to the onscreen/offscreen pair and to sound perspective. Finally, we will discuss some specialized but commonly used devices that exploit the boundaries between diegetic and nondiegetic space.

Diegetic/Nondiegetic Music and Narrative

At its most basic level, diegetic sound refers to everything that can be heard by characters in the film. Nondiegetic sound, by contrast, cannot be heard by the characters. Consider, for example, the conversation scene from *Catch Me If You Can* (2002) that we discussed in the Introduction to Part I and have mentioned several times since. The radio on Carl Hanratty's desk produces sounds (Bing Crosby singing "Mele Kalikimaka") that Carl can hear and that Frank also hears through the telephone receiver. The melancholy underscoring near the end of the scene, on the other hand, is heard only by the audience—if we understand this music as mimicking Frank's mood, we do not assume that it is literally going through his mind at the time (in fact, that would be highly unlikely because the music continues when Frank speaks). Similarly, in the Boston Common scene from *Good Will Hunting* (1997), the dialogue and ambient sounds of birds and traffic are diegetic but the music that enters near the end is nondiegetic; and again the same in "The Shooting" from *Tizy Kolory: Bialy* (White, 1994): All the noises in the underground train station could plausibly be heard by the characters, but the music could not.

The typewriter sounds in *Atonement*, on the other hand, pose a problem. We can assume in retrospect that Briony can hear the diegetic sounds that actually first appear behind the opening titles, but what about the typewriter sounds that continue after she stops and walks away? Have they simply been transferred to the underscoring, or are they sounds of typing still mentally resounding in Briony's ears? The music's marchlike gait, which seems synchronized with her walking, also threatens to lose its nondiegetic independence and join in the action, so to speak—not to mention the stinger chords at the end, which are not so much music as sound effect.

It is by no means uncommon for ambiguities of this kind to arise, especially in connection with rhythmic mechanical sounds such as a typewriter. In

Shall We Dance (1937), Pete Peters (Fred Astaire) is sailing to America to make his debut as "Petrov," purportedly a Russian ballet dancer. To conceal his identity, Pete must refrain from openly dancing in the vernacular "jazz" style that he prefers. He therefore descends to the engine room, where the pistons of the ship's engines establish the rhythm so that the song "Slap That Bass" seems to emerge from it. Something similar occurs at the beginning of *Love Me Tonight* (1932). Here, Paris in the early morning slowly awakens in rhythm. We hear the sound of a bell chiming the hour, a lone bicyclist, then the sound of a pick ax, all laying down a basic rhythm. The hammering of cobblers, the beating of rugs, the sounds of a factory all join in rhythm, suggesting the vitality and common purpose of the city. This symphony of sound ultimately dissolves into Maurice Chevalier singing "That's the Song of Páree."

Where *Love Me Tonight* moves from sound to music, another early sound film, *Das Testament des Dr. Mabuse* (1933), essentially reverses the procedure. The film starts with very dissonant music over the title only to dissolve into the oppressive rhythm of the machine stamping out a triple meter. The opening scene, furthermore, begins *in medias res*: Hofmeister has slipped into the Forger's factory and has been forced to hide in a room (Figure 3-1). The scene plays essentially as silent film, the machine being the only noise on the sound track.

Underscoring, then, can move with ease—and speed—between diegetic and nondiegetic functions. If we are not bothered by music that lacks anchoring in the diegesis, equally we are not bothered by music that begins onscreen and then wanders offscreen or even hovers uncertainly between. This is business-as-usual in musicals: in *Meet Me in St. Louis* (1944), for instance, one number has Judy Garland singing a verse and chorus to accompaniment of a piano that we see (initially anyway), but then she sings a second chorus to the accompaniment of an invisible orchestra. This technique of passing from one to the other scholar Rick Altman has named "audio dissolve."¹ Although common in musicals, the audio dissolve occurs in dramatic films as well, as we shall see later in this chapter while discussing the lead-in to Rick's flashback in the after-hours scene from *Casablanca* (1942).

EXAMPLE FOR DIEGETIC AND NONDIEGETIC MUSIC: GLORY, BOSTON PARTY SCENE

A scene from early in the Civil War film *Glory* (1989) will illustrate both the distinction between diegetic and nondiegetic music and some of the ambiguities that can arise from them. The city (Boston) is the same one depicted in *Good Will Hunting* and the scene is probably just a few blocks away at most from the Common (because both films use location shooting, this could be literally true), but in *Glory* the time is about 130 years earlier. A party is underway to honor Massachusetts officers who fought in the recent battle of Antietam (September 1862). It is one of the few "non-military" scenes in the film, which stars Matthew Broderick, Denzel Washington, and Morgan Freeman and recounts



Figure 3-1. *Das Testament des Dr. Mabuse*. A distressed Hofmeister (Karl Meixner) tries to think amidst the din of the machinery in the Forger's factory.

the early history of the 54th Massachusetts Infantry, the first regularly formed African-American regiment in the United States Army.

The party scene begins at about 00:09:30, lasts just over 5 min, and is heavily scored—only about 30 sec are without music. Up to this point, more than half the film has had nondiegetic orchestral music: music begins with the opening titles and runs about 3½ min until the beginning of the Antietam battle scene; a second cue begins at 00:05:30, as Captain Robert Gould Shaw (Broderick) rises from the ground after the battle; and continues until he enters the field hospital (1 min and 45 sec in all).

The extended party scene that follows breaks down readily into three parts, the first being Shaw's decidedly self-conscious entry into the party, the second a conversation with the Governor and others about forming the regiment, and the third a final conversation with his friend Forbes (Cary Elwes) carried on outdoors. We will be concerned here only with the first part and its transition into the second.

The scene opens with a shot of a rose window, which is quickly revealed as a skylight providing illumination to the upper end of a staircase (top frame in

Figure 3-2). Music is heard from a piano, a pastoral piece stylistically appropriate for the era (by early 19th-century composer Franz Schubert; Figure 3-3). Our first reaction, then, is to regard the music as belonging to the physical environment, as diegetic, but the volume level is quite unrealistically high. Eventually we will learn that the music is meant to be diegetic—we see the pianist briefly in the general shot of the room (Figure 3-4), but sound levels are never quite appropriate (even when we see the piano close by in the background of segment two). The transition to this scene was achieved by a short lap dissolve (overlapping images) from an Army field hospital where, among other things, an amputation was being performed—sound participates, as the hospital sounds mingle briefly with the music. In the course of the shot, the camera pulls back, moves down and to the left to reveal Shaw (in medium shot) descending the stairs (middle frame of Figure 3-2). He is in his parents' upper-class Boston house on his way to join the large party in progress.

After a point-of-view shot (from the staircase looking down on people below and traveling across them as if Shaw is scanning what he sees below him) and a sudden, inserted close-up of a plate of food, with a woman and an older man talking (the man takes and is about to eat a small tomato), comes what we will take as an establishing shot, a general view of a large room with partygoers. Sound, on the other hand, is oddly unsettled. The piano music does provide a suitable aural counterpart to the establishing shot: its unrealistic sound levels and broad stylistic-chronological reference would work just as well as nondiegetic music. A generalized sound of party conversation is appropriate, but it is undermined by odd moments (particularly by unusually loud laughter and whispering in the close-up mentioned earlier), and the result is that we begin to suspect that the sound is subjective—specifically, it is ambient sound as filtered through Shaw's mind and emotions. This is still diegetic—unlike the generalized sense of emotion in the scenes from *Catch Me If You Can* and *Good Will Hunting*, we are hearing diegetic sounds as filtered, distorted, by Shaw's own hearing and attention.

All this is confirmed in subsequent shots: (1) Shaw in medium shot as he walks through the doorway (lower frame of Figure 3-2) (at this point, the crowd sounds begin to mingle with a wordless (and nondiegetic) boys chorus

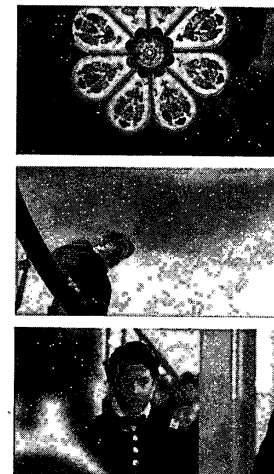


Figure 3-2. *Glory*. Boston Party
Scene: (a) Dissolve from previous
scene; (b) Shaw descending the
staircase; (c) entering the main
room of the party.



Figure 3-3. *Glory*. Boston Party Scene: Franz Schubert, *Drei Klavierstücke*, D.946, No. 2.



Figure 3-4. *Glory*. Boston Party Scene: A pianist plays on the far side of the room.

(as the piano fades out)—this music continues, mixed with bits of conversation, and gradually grows louder); (2) another extreme close-up, now of two women talking as they look at Shaw; (3) cut back to a close-up of Shaw (another soldier enters the room behind him), then (4) away to another close-up of a woman laughing, (5) again back to Shaw, (6) then a medium shot of an officer in a wheelchair entering the room, (7) back to one last close-up of Shaw before the nondiegetic music goes abruptly out and the diegetic piano music returns as Thomas (Andre Braugher) addresses him (we see Shaw in medium shot over Thomas's shoulder).

Until the greeting from Thomas, Shaw maintains his distance from the crowd, the shot/reverse-shot series emphasizes that distance (it jumps back and forth between close-ups of him and of people who seem to be at some physical distance from him), and the disjointed quality of the sound track corresponds. The uncertain diegetic status of the piano music is followed by Shaw's mental filtering of room sounds, and the intrusion of nondiegetic music over them. The crux of the scene is reached, not in conversation, but in Shaw's rapidly increasing discomfort (Figure 3-5a; music here, as in the scene from *Good Will Hunting*, contributes much of the emotional intensity). Rather than concluding and making a transition to another scene, this first part is simply cut off by Thomas's greeting (Figure 3-5b). There is a strong sense in which this moment explains everything before it as subjective: with a sudden return to the real world copes a normal mode of hearing.

The basic categories are defined easily enough: the piano, crowd, and speech are all diegetic; the wordless chorus is nondiegetic. Before Thomas's

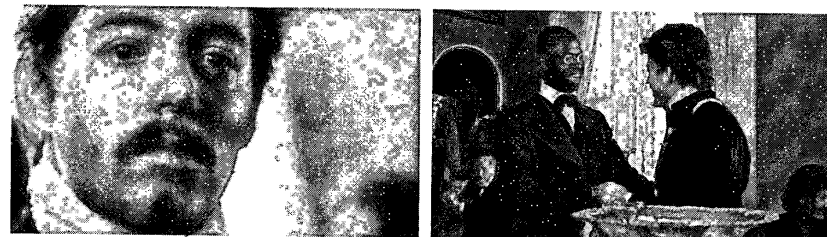


Figure 3-5. *Glory*. Boston Party Scene: (a) Shaw is deep in thought, disconnected from the party; (b) Thomas greets Shaw.

greeting, however, the diegetic sounds, including the music, were interpreted, filtered, as if they were sounds "narrated" for us by Shaw.

See Figure 3-6 for an example of a particularly subtle interplay between diegetic and nondiegetic within the continuously playing music for a dance.



Figure 3-6. *Shakespeare in Love* (1996). The de Lesseps' dance (00:27:58–00:30:30). The music is diegetic, mostly restricted to the historically accurate ensemble seen in the background—flute, lute, harp, viols and tambourine. As Will (Joseph Fiennes) integrates himself into the dance (00:28:47), the viols are subtly augmented by modern strings, with a tremolo growing ever more pronounced until Viola (Gwyneth Paltrow) says: "Master Shakespeare" (00:29:07), when the tremolo is transferred upward to the violins. The scoring retreats to the diegetic ensemble as she moves on to Lord Wessex (Colin Firth) (00:29:18). When she returns to Will (00:29:29), the scoring shifts again, now dominated by modern strings. The style of the music is also transformed. Lord Wessex drags Will away (00:29:58) and the modern instruments again recede, although not as far as before and the musical style also remains the new one, emphasized by drum fills at the ends of phrases. With the shot of the exterior of the house (00:30:30), the orchestra takes over completely.

Onscreen/Offscreen Sound and Music

The diegetic/nondiegetic pair refer to the status of sounds within (or without) the physical world depicted in the narrative, or the set of relations induced specifically by what is presented to us in the frame and the sounds that occur in the sound track.

The basic categories of these relations can be expressed by combinations of the terms. The first of these, diegetic-onscreen, is certainly the default case: we see within the frame what we expect to see in the film world (Will and Sean talking on a park bench, not a view of Sean's empty office or an insert of, say, the Voyager I spacecraft while they talk). Diegetic-offscreen is also common: a room is shown onscreen but we hear someone speaking or music playing with the correct volume and other sound qualities that would match another room connected to the one we see. Nondiegetic-onscreen, on the other hand, is much less common but is likely to be invoked when it is clear that an onscreen character imagines or remembers speech or music and the performance of that music is visualized. Nondiegetic-offscreen is the default case for voice-over narration and underscoring, but it can also apply to characters such as ghosts whose voices can be heard but who have no definable place in the physical world.

In this section, we will first stress offscreen sound and then several ways in which offscreen and onscreen sound interact.

OFFSCREEN SOUND

Without evidence to the contrary, we will take offscreen sound as simply an extension of onscreen space. For example, dialogue scenes that utilize the shot/reverse-shot syntax will often cut away at some point in the conversation to the non-speaking character. During this reaction shot, clearly, the speech that continues will be offscreen. The reaction shot is so common that we seldom pay attention to the fact that the shot itself creates offscreen sound, which has the effect of also binding offscreen and onscreen space more tightly together.

A brief but narratively important scene late in *The Big Sleep* (1946) provides a simple example (01:43:54). Philip Marlowe and Vivian Sternwood, characters played by Humphrey Bogart and Lauren Bacall, have escaped from one dangerous encounter and are heading toward another (the final scene of the film). They are shown in a car (Figure 3-7), and they discuss their situation (although Marlowe, in fact, does most of the talking). They were clearly attracted to one another early on in the film, but here they admit, for the first time, genuine romantic feelings.

The scene lasts less than 2 min and consists of twelve shots. (This scene, by the way, is the one that we graphed under Exercise 2 at the end of ch. 2.) Three of the twelve are 2-shots of Vivian and Marlowe: at beginning and end,

This overrunning of one person's image with another person's voice is a method peculiar to the talkies; it is one of the devices which help the talkies to tell a story faster than a silent film could tell it, and faster than it could be told on stage.

—Alfred Hitchcock, on the significance of offscreen sound to the reaction shot.²



Figure 3-7. *The Big Sleep*. Second car scene, Vivian and Marlowe.

as well as one in the middle. The others are alternating close-ups of the two characters. At several points one character is shown in close-up while we hear the other speaking. The effect is not only to bind offscreen and onscreen space (limited although those differences are in such a situation) but in this case also to highlight the emotional resonance in this newly forming couple. (See Figure 3-8 for another example of offscreen sound.)

When offscreen sound is localizable as an object that *could* be shown but is not—that is, a sound that suggests an object is more than simply background—we refer to it as “sound-off” (short for “sound offscreen”). Examples include doors, footsteps, telephones, and so forth. Such noises are commonly used as a means of establishing a scene, either at the beginning of the sequence itself, where an offscreen sound can motivate a cut to the location, or to introduce new characters (and so also a new direction) to the scene. Offscreen bomb sounds, for example, are used this way in *Lawrence of Arabia* (1962) (at 00:37:58). The sound of the bombs interrupts a conversation between Lawrence and Colonel Brighton; a cut to an encampment of Arabs being attacked by Turkish planes follows.

As a “noise” in the diegetic world, music can also be used in this manner. In the opening of *The Broadway Melody* (1929), for instance, music wafts from



Figure 3-8. *Trois Couleurs: Bleu* (Blue, 1993). Julie (Juliette Binoche) listens as a man, evading thugs, runs through her building knocking on doors (00:34:48-00:36:04); the scene is rendered wholly in offscreen sound.

a window of a music store before a cut shows the interior of the shop. In *The Bride of Frankenstein* (1935), the monster hears the sound of a violin and gradually finds his way to the house of its blind player; and the main-title sequence of *The Birdcage* (1996) includes (apparently) nondiegetic underscoring that is revealed as a stage performance of "We Are Family" after the camera moves gradually across the water, beach, and street into the interior of the club.

A sound-off can also be used to signal the end of a scene. In *42nd Street* (1933), the offscreen sound of a door closes to signal Julian (Warner Baxter) leaving the office, which sets up the tagline: "New York will see its first triple funeral."

The voice-off is similar to the sound-off, except that it highlights the voice. A simple voice-off will occur with a cut to a reaction shot. More characteristic, perhaps, are introductory words, like "hello," that announce the presence of a new character before we see him or her. A voice-off will sometimes involve a clear mismatch in sound scale (that is, linking sound volume and timbre to shot scale).

The film in *Emma* (1996) contains several excellent examples of the voice-off technique. At 00:13:58, Emma (Gwyneth Paltrow) and Harriet (Toni Collette) are walking along a creek, where they are presented in an extreme long shot, often even obscured by trees. Nevertheless, their dialogue is rendered with the clarity of a normal 2-shot, into which they eventually move. Another example appears shortly thereafter (00:15:16). Emma and Harriet are doing embroidery under a canopy. Again, they are first shown in extreme long shot but with the dialogue suggesting much closer proximity. The mismatch en-

courages a series of cuts that will eventually bring the image in line with the sound. At the end of this sequence, we are once again shown an extreme long shot, but this time the dialogue declines in clarity, moving in tandem with the distance as framed. A pattern of mismatched scale of image and sound continues throughout the film and seems calculated to coincide with the mismatched romantic pairings that Emma attempts to bring about.

ONSCREEN/OFFSCREEN SOUND INTERACTION: EXAMPLES

Filmmakers are very sensitive to the play between onscreen and offscreen sound. One of many striking examples we could cite occurs in *The Apartment* (1960). Bud (Jack Lemmon) has just been promoted and is settling into his office (00:43:00). As he is hanging up his overcoat, an offscreen voice offers congratulations. He finishes putting away his jacket, and responds, "Hi, fellows." Only at this point is there a cut to four men entering his office. This, then, initiates the scene proper where the men remind Bud that he is beholden to them for his promotion.

As everyone moves further into the office, the camera reframes to incorporate Bud into the group. In a sense, the camera here entertains the perspective of the men: "all for one and one for all." Nevertheless, the desk intervenes to keep Bud somewhat apart from the group (Figure 3-9a). When he is accused of not having the right "attitude," he is separated from the group with a cut for his response. There is another accusation, this time delivered completely offscreen with the camera locked on Bud through his response. A reverse shot of the four men brings a third accusation. The shot is reversed again for Bud's response. Another accusation begins offscreen before cutting back to the man to complete his line. One of the men complains of the trouble he's been having because Bud no longer lets him into the apartment. Bud's response occurs in another reverse shot.

Next comes a cut back to the full group, with the four men now clearly separated from Bud, and one of the men delivers a threat. Bud responds briefly, but this is followed by a more pointed threat. At this moment (00:44:03) Jeff (Fred MacMurray) enters, he and Bud begin to converse, and the other men leave. Bud closes the door and returns to his desk, sits down, and a relatively uncomplicated shot/reverse-shot sequence follows, but without reaction shots and offscreen sound. One exception is an odd moment with the close-up of a mirror (00:45:01): Bud delivers a line while being somewhat out of focus (Figure 3-9b) and Jeff's face appears in the broken mirror for his response (Figure 3-9c). The lack of departure from onscreen sound emphasizes the strangeness of this particular exchange, the distorted image of each man suggesting each character's relation to the mirror—or rather to Fran (Shirley MacLaine), its owner.



Figure 3-9. *The Apartment*. a
Sound and the dynamic play of
b
onscreen and offscreen space. c

In *Lost in Translation* (2003), Bob (Bill Murray) is jetlagged, having just arrived in Tokyo. He is in the hotel bar where a woman is singing. The sound track is filled with ambient chatter and on the first shot of Bob some of the chatter grows more distinct. Offscreen voices indicate that they have recognized him. As they address him, the camera pans from Bob to reveal two young American businessmen. The two continue talking on a cut back to Bob, where he briefly responds, then gets up while the other voices carry on. The overall effect of the exchange is Bob's refusal to engage the other characters, who remain defined almost entirely by their offscreen voices. The fact that the voices remain offscreen minimizes the importance of these characters, while also suggesting a narrative direction that the film will not follow. (See Figure 3-10 for an additional example with commentary.)



Figure 3-10. *Lara Croft: Tomb Raider* (2001). In this scene (00:07:34–00:09:30), Manfred Powell (Iain Glen), who has been charged with finding the Key, stands before the Council of Illuminati to report his progress. Despite a large amount of offscreen dialogue and wide shifts of shot scale, the dialogue remains at a constant level and squarely centered in the stereo field. The large space is cued, however, through the use of pronounced reverb.

Partly due to the reverberant space, the tempo of the dialogue is quite slow, appropriate for the formal setting and to the gravity of the discussion. Much of the head of the Council's dialogue is delivered offscreen, which coupled with the lack of any sound perspective has the effect of disassociating his voice from his body, allowing the words to transcend the limits of the character. (It should be noted in this respect that offscreen dialogue is not always rendered thus in the film.) The use of apparently nondiegetic singing voices at the opening of the scene and bells tolling periodically throughout reinforce the ecclesiastical setting suggested by the exterior shot of the Santa Maria Della Salute (Venice) in the establishing sequence. The interior, however, is the Painted Hall at the Old Royal Naval College (Greenwich), a secular space appropriate to a secret society. The mysterious quality of the quest for the Key is emphasized by the brooding strings and especially the thrice stated piano figure that displace the voices from the sound track later in the scene as talk turns to the Key itself. The crackling thunder interspersed with the instrumental music adds to the effect, portending that the Key will unlock something ominous and supernatural.

AMBIGUITY OF OFFSCREEN SOUND: UNDERDETERMINATION OF SOUND

Sound is often underdetermined, that is, not defined down to all its possible naturalistic details. Depending on how it is rendered, a waterfall, for instance, sounds very much like applause, which is also similar to the rustling of leaves or the crackle of fire. Crumpling paper might likewise be mistaken for fire or fallen leaves in autumn. The crack of a baseball bat can sound like a tree branch snapping; the buzz of an insect like an electric saw or a malfunctioning radio tube. Indeed, the production of sound effects often depends on misrecognitions of just this sort: the hitting of high tension wires for the sound of a laser blast, or coconuts for horses' hooves.

A scene from *King Kong* (1933) plays with precisely this ambiguity. As the ship sails into the fog surrounding Skull Island, amorphous nondiegetic music enters, the first music heard since the opening credits. Gradually, a soft crackle can be heard. Jack says, "Listen. Do you hear anything?" The sound grows louder and an offscreen voice shouts "Breakers ahead!" A moment later, Jack says, "That's not breakers—that's drums." This confusion helps to establish the sense of Skull Island as a mysterious place. (Figures 3-11 and 3-12 give additional examples from other films, along with commentary.)



Figure 3-11. *Lawrence of Arabia*. After a beating at the hands of the Turkish army, Lawrence (Peter O'Toole) has gone to Jerusalem, abandoning his Arab comrades and vowing to be nothing more than a regular man. The establishing shot shows a marching band. As Lawrence enters the military compound, there is a cut to the interior of the compound and the music begins to echo rather significantly, to the point where the echo separates into its own channel, lagging significantly behind the source. Lawrence himself is almost stooping as he walks through the courtyard, even as the echo separates further and further from the source. The sound seems to split, go out of focus, representing Lawrence's own divisions and self-doubts. The importance of the music is emphasized by its continuation throughout the sequence as Lawrence moves through the building. It only goes out when Lawrence reaches the office where General Allenby (Jack Hawkins) and Prince Feisal (Alec Guinness) are waiting.



Figure 3-12. *The Apartment*. Ambiguity of offscreen sound: Near the end of the film, Fran (Shirley MacLaine) has come to realize that she loves Bud. As she is rushing up the stairs to his apartment she hears what sounds like a gunshot coming from Bud's apartment (02:03:11). Fearing the worst, she rushes to the door. Bud opens it to reveal a bottle of bubbling champagne. Here Fran's reaction to the offscreen bang allows us to see just how deeply she feels for Bud.

POINT OF VIEW SOUND

Point of view sound is rendered from the perspective of a character in the film. The effect is generally to increase our identification with that particular character whose hearing the sound track mimics. A very common example is a conversation heard over the telephone when a shot/reverse-shot structure is not used. The filtering of the voice to render a realistic telephone sound underlines the distance between the characters and in so doing increases our focus on the character we can see.

The first part of the Boston party scene from *Glory* clearly is point of view sound, as we hear speech and music filtered through Captain Shaw's ears and mind. A more recent film, *Lost in Translation*, contains a short example that effectively illustrates point of view sound. Bob is swimming laps while a water aerobics class is also taking place. The short scene all takes place from Bob's aural perspective, with the characteristic tone of his gentle sound and music dulled.

"Imagined sound" is a special case of point of view sound. Here, the sound track presents what a character is hearing in his or her head. A good example of this occurs in *The Bourne Ultimatum* (2007). Here Jason Bourne's numerous flashbacks are presented with distorted sound (and image). Such imagined sound may be a dream, memory, or hallucination; an internal monologue; or hearing something like music in one's head. For obvious reasons, the latter case is especially common in films about musicians. Another typical use is for an individual reading a message, letter, or passage from a book silently. A po-

tentially more complex instance is when a character in a film also acts at one or more points as a voice-over narrator. Once we have associated the voice-over narration with a character in the diegesis, the sound can no longer be unambiguously nondiegetic. Instead, we hear a simple imagined sound if the character is onscreen, not shown speaking but nevertheless speaking in imagination.

Point of view sound, like the basic distinction between diegetic and nondiegetic, is open to many ambiguities. A good example to illustrate this point occurs in *The Graduate* (1967), where it is coupled with a sound advance. In this scene, Ben (Dustin Hoffman) is decked out in scuba gear that he received as a gift from his parents. As he reluctantly leaves the house to go to the pool to show off the equipment, all we hear is his heavy breathing, a point of view sound matched to the image, which also has been matted to give the impression of seeing through a scuba mask. The matched perspective of image and sound continues into the water, separating only when the image changes to an external reverse shot. After the cut, the sound continues with what at first seems to be a return to point of view, but the sound of bubbles gradually dissolves into an externalized perspective as we become aware that a decrescendo in the sound is accompanying a slow tracking back of the camera. As the camera finally comes to rest on Ben in long shot, we hear his voice speaking in normal tones over this image of him isolated in the pool. With the cut we recognize that he is speaking on the phone with Mrs. Robinson (Anne Bancroft), whom he has agreed to meet for their first rendezvous. The sound bridge here serves to tie these scenes together, his isolation in the pool being literally answered by companionship with Mrs. Robinson.

The animated feature *Madagascar* (2005) manages the linking of the sound bridge with point of view sound in a somewhat different manner. In this case, the protagonist, Alex the Lion, is tranquilized and experiences a hallucinatory production number to the strains of "Candyman" (sung by Sammy Davis, Jr.). We easily recognize that at this moment we are perceiving things from Alex's perspective; ultimately, everything goes black and silent, a sign that Alex has fallen unconscious and probably a sign that the scene has ended. Because the screen is black we recognize that the next scene will emerge out of this dark silence and a sound bridge seems a likely way to make this transition. We do in fact hear voices: a sound bridge seems to be taking shape. As a shot opens with a wipe from the bottom revealing a matted half oval, however, we understand that the voices we had heard were not a sound bridge but a representation of Alex beginning to wake up, that is, a continuation of the point of view that had led into the silence to begin with. In this way the dark, silent join between the production number and the awakening is rationalized. After Alex is shot with the tranquilizer gun again, and the sequence repeats, this time with a highly sped up version of the production number follows. The screen goes momentarily dark, but this time we hear Alex speaking at the same time that we see big

eyes—the same shape as the previous matte shot—floating in the darkness. It quickly becomes clear that these eyes belong to Alex, the continuity of shape with the previous emergence out of the dark silence cleverly allowing a seamless reversal here to the normal, external shot.

THE ACOUSMÈTRE (ACOUSTICAL BEING)

Michel Chion notes that sound film makes possible a special kind of character, one who exists in the diegetic space but is placed consistently offscreen. He calls such a character an *acousmètre*,⁴ a French neologism that means "acoustical being." Being heard but not seen, such a character is defined wholly in terms of diegetic sound. This situation is distinct from narration because the voice of the acoustical being is taken to occur in the same timeframe as the diegesis, whereas narration is necessarily after the fact, even when the voice of the narrator is also that of a character in the diegesis. The *acousmètre* is also distinct from a radio voice, which may be defined solely through sound and in that sense is "acousmatic" but lacks any expectation of being visualized: the world of radio is an acousmatic world consisting only of sound, and so no character can have other than an acoustical presence.

Neither onscreen diegetic nor offscreen nondiegetic music or sound pose any real obstacle to our attempts to make sense of what we see and hear. Onscreen diegetic testifies to the reality of the diegetic world. As typically used, offscreen nondiegetic sound (especially music) testifies to the inner psychology of the character. One of the most common functions of nondiegetic music, for instance, is to give us some sense as to how a particular piece of dialogue or incident affects a character. "Underscoring" thus operates in two senses: technically, it is music that is scored under dialogue; but figuratively, it is also music that *underscores*—that is, emphasizes—gestures and moments that are particularly psychologically fraught and revealing.

Whereas the typical character lives by the rules of psychological reality, the acoustical being lives outside them. A typical character has no awareness of and so no control of the camera and what the camera shows. The acoustical being, on the contrary, seems aware of the camera inasmuch as it has the mys-

terious ability to avert the camera's gaze, to be always just outside the frame. This ability to sense the frame so as to avoid being shown seems to endow the character with almost god-like powers. On the other hand, the acoustical being is typically represented as malevolent. The loss of its powers, through revelation of its body onscreen, that undoes and destroys the malevolent *acousmètre*. Chion points to such pathological or even psychotic characters as the child murderer in *M* (1931), the Mother in *Psycho*, and Mabuse in *The Testament of Dr. Mabuse* as the common type of acoustical being. Mabuse is a prototype of the evil genius that appears frequently in later films. Those films in the James Bond series dealing with threats of world domination or destruction often have antagonists who are revealed and gradually lose power through the process of de-acousmatization (becoming visible). Likewise, horror films often introduce their monsters as acoustical beings, and the more we see the more vulnerable they become—it is the unseen monster that is most terrifying.

Acoustical beings are not absent from musicals or comedies, where they often appear as authority figures who rule or assert control despite absence. In such cases, the process of de-acousmatization is often the process by which comedic deflation, subversion, or inversion of authority occurs, allowing the individual to escape the determinations of authority. Although represented by a giant head floating translucently in space among flames and smoke, the Wizard in *The Wizard of Oz* (1939) is in many respects an acoustical being; certainly, the revelation of the "man behind the curtain" is a particularly direct example of de-acousmatization.

The opposite procedure, or what we might call the acousmatization of a character, can also have the effect of rendering that character impotent, especially when it occurs near the end of a film. In *The Robe* (1953), for instance, the Emperor Caligula's voice desperately cries out offscreen as Gallio and Diana leave "to a better world." Similarly, at the end of *A Face in the Crowd* (1957), Marcia (Patricia Neal) walks out on Larry "Lonesome" Rhodes (Andy Griffith). As she exits the building onto the street, Lonesome calls to Marcia from a window far above. As her taxi pulls away his voice grows fainter, disappearing into the sounds of the city as "The End" fills the screen. To render the effect even more poignantly, the camera does not follow the taxi as it disappears but instead shifts to general shots of the city. The sound changes similarly so that Lonesome's voice has become just another sound of the city, and the city, like Marcia, has become indifferent to his cries.

As might be expected, actual filmic representations of God or Jesus, although typically acousmatic, are treated somewhat differently. Here, the character is understood as belonging to a different world. In *The Robe*, for instance, Christ is either not shown or seen only from a great distance, and his presence is marked by a prominent shift to nondiegetic choral music. *Cabin in the Sky* (1943), by contrast, allows the diegetic appearance of an angel, but God is

acousmatic in the extreme, sounding as "otherworldly" bell-like music that only the angel can understand. Such films as *King of Kings* or *The Greatest Story Ever Told* (1965) dealing directly with the life of Christ must necessarily treat Jesus as a diegetic character, but the representation generally follows a scenario of de-acousmatization near the beginning of the film, which is reversed at the end, with the ascension. Whereas God exists *beyond* the diegesis so that the sound or music that represents God is usually understood as a peculiar mode of the nondiegetic (as in *The Robe*) or as an intrusion in an extraordinary diegetic space (*Cabin in the Sky*), the acoustical being exists *within* the diegesis.

Summary

In this chapter, we paid particular attention to the (surprisingly complex) relations between sound in the sound track and the physical world or diegesis represented through the narrative. The distinction between diegetic and nondiegetic refers to this anchoring of sound in the filmic world (or, in the nondiegetic, the failure to be anchored in that world). Another pair, onscreen/offscreen, refers more narrowly to sound anchored in what the camera frames for us at any given moment (onscreen) or that we can reasonably assume belongs to space we can extend from what we see (offscreen).

EXERCISE 1: USING AN EVENTS TABLE

Instead of a prose description or list (modeled on our analysis checklist), you could use the following template (or your own variant of it) as a quick way to describe the sound track in a film scene (Table 3-1). The template has been

Table 3-1 Sound track elements and functions in the opening of *42nd Street*.

Sequence	Time	Music		sfx	Dia.	Comments	General comments
		ND	D				
Titles	0:00	✓				Music high on mix	Title music: "42nd Street"; new tune for cast.
Establishment	1:28	✓		✓	✓	Music levels remains high; sfx mixed; dia. added near end	Music provides continuity for montage; sfx present "life" of city in contrast to static images; dia. used more as sfx (audio collage).
Abner and Dorothy	1:59			✓	✓	Dia. pre-dominates; minimal sfx	Emphasis on dialogue; only prominent sfx is rustling paper, which underscores importance of the "contract."

ND = nondiegetic; D = diegetic; sfx = sound effects; Dia. = dialogue (speech)

filled in for the opening minutes of *42nd Street* as an example. Use column 1 to segment the sequence or scene (we did something similar by grouping shots in the analysis of the Shooting scene from *Trzy Kolory: Bialy (White)* in ch. 2). Start points for each segment can go in the second column. Use columns 3 and 4 to indicate presence or absence of that sound component on the sound track. Use the "comments on mix" field to make more nuanced observations, such as the predominance of dialogue, music sneaking in and out, use of ambient sound, and so on. Use the "general comments" field to note items of particular interest, such as whether sound is onscreen or offscreen; musical genre, tempo, or mood; point-of-view sound or music, and so forth.

EXERCISE 2: USING AN (UPDATED) ANALYSIS CHECKLIST

In chapter 2, we updated the analysis checklist to include description of the sound track in terms of tempo, texture, and other musical terms. We can do the same for the topics discussed in this chapter:

1. Background and general information
2. Synopsis of the scene
3. Shot list
4. Description of the sound track elements and their balance
5. Description and evaluation of the sound track (diegetic/nondiegetic; onscreen/offscreen; point of view sound; voice-over; audio dissolve; mickey-mousing; the *acousmètre*)
6. If needed, a summary statement about the sound track in the scene

VALUE ADDED BY MUSIC

Empathetic and Anempathetic Effects

In my book *Le Son au cinéma* I developed the idea that there are two ways for music in film to create a specific emotion in relation to the situation depicted on the screen.⁴ On one hand, music can directly express its participation in the feeling of the scene, by taking on the scene's rhythm, tone, and phrasing; obviously such music participates in cultural codes for things like sadness, happiness, and movement. In this case we can speak of *empathetic music*, from the word empathy, the ability to feel the feelings of others.

On the other hand, music can also exhibit conspicuous indifference to the situation, by progressing in a steady, undaunted, and ineluctable manner: the scene takes place against this very backdrop of "indifference." This juxtaposition of scene with indifferent music has the effect not of freezing emotion but rather of intensifying it, by inscribing it on a cosmic background. I call this second kind of music *anempathetic* (with the privative *a-*). The anempathetic impulse in the cinema produces those countless musical bits from player pianos, celestas, music boxes, and dance bands, whose studied frivolity and naiveté reinforce the individual emotion of the character and of the spectator, even as the music pretends not to notice them.

To be sure, this effect of cosmic indifference was already present in many operas, when emotional pitch was so high that it froze characters into inaction, provoking a sort of psychotic regression. Hence the famous operatic convention of madness, with the dumb little music that a character repeats while rocking back and forth. . . . But on the screen the anempathetic effect has taken on such prominence that we have reason to consider it to be intimately related to cinema's essence—its mechanical nature.

For, indeed, all films proceed in the form of an indifferent and automatic unwinding, that of the projection, which on the screen and through the loudspeakers produces simulacra of movement and life—and this unwinding must hide itself and be forgotten. What does anempathetic music do, if not to unveil this reality of cinema, its robotic face? Anempathetic music conjures up the mechanical texture of this tapestry of the emotions and senses.

Finally, there also exist cases of music that is neither empathetic nor anempathetic, which has either an abstract meaning, or a simple function of presence, a value as a signpost: at any rate, no precise emotional resonance.

The anempathetic effect is most often produced by music, but it can also occur with noise—when, for example, in a very violent scene after the death of a character some sonic process continues, like the noise of a machine, the hum of a fan, a shower running, as if nothing had happened. Examples of these can be found in Hitchcock's *Psycho* (the shower) and Antonioni's *The Passenger* (an electric fan).

INFLUENCES OF SOUND ON THE PERCEPTION OF
MOVEMENT AND PERCEPTION OF SPEED

Visual and auditory perception are of much more disparate natures than one might think. The reason we are only dimly aware of this is that these two perceptions mutually influence each other in the audiovisual contract, lending each other their respective properties by contamination and projection.⁵

For one thing, each kind of perception bears a fundamentally different relationship to motion and stasis, since sound, contrary to sight, presupposes movement from the outset. In a film image that contains movement many other things in the frame may remain fixed. But sound by its very nature necessarily implies a

Summary

In chapter 3 we covered sound in relation to the physical space of the filmic world. In this chapter the attention turned to the temporal relations of sound and image, specifically the formal effects of scene transitions that manipulate sound (sound bridge) and the close (or loose) coordination of sound and image (synchronization)—and its opposed term, the deliberate juxtaposition of sound and image (counterpoint). We also discussed synchronization and counterpoint in terms of what we might call narrative or emotional dissonance (another common meaning of “playing with” or “playing against” the film).

EXERCISE: THE COMMUTATION TEST

There is no easier way to confirm our explanation of music’s role in the sound track and its narrative functions in a scene than by seeing (hearing) what happens when we substitute other music. In fact, this kind of exercise is very closely allied in its method and even in its goals to the spotting process that is used to decide on music for a film: you can feel much more comfortable that you “got it right” by considering alternatives.

Commutation tests are easily made for any film by simply playing music on a CD player while the film runs without sound.⁵ The disadvantage is that

you lose the dialogue and effects, so that the overall character of the sound track is altered (unless you have editing software available, of course). You might try an experiment with main titles, which are often accompanied by music only. For scenes without music, on the other hand, you can judge what music does by playing both CD track and film audio simultaneously. More frequently nowadays, DVDs are released with alternate tracks, for example, *Dracula* (1931) with a new string quartet score by Philip Glass.

Suppose that we removed the quiet, slightly nervous music from the bedroom scene in *Psycho*, discussed in chapter 1, where Marion Crane packs her suitcase before running off with her employer’s money (Figure 4-12). Now, replace that music with a much more obviously ominous cue from a recent horror film. The empathy we feel for Janet Leigh’s character as she mulls over an all-too familiar human dilemma about how to balance behavior and desire would be lost—we would immediately be sharply distanced from her (“Is she about to become a victim already?”) but still wanting to communicate, to warn her about the danger that awaits her in a closet, perhaps, or in the shower we see in the background. What this substitution confirms, by negative example, is how effectively Bernard Herrmann’s music makes us empathize with Marion, despite its rather sparse and understated quality.

More radically, if we substituted a popular song sung by a woman, such as the classic 1970s-era “Get It While You Can” (Janis Joplin) or the more recent “Cornflake Girl” (Tori Amos), there would suddenly be an extra semantic layer whose meanings we would take to be emanating from Marion (because



Figure 4-12. *Psycho*. Marion hesitates for a moment about whether to take the money.

the soloist is a woman) or else from the singer as a confidante or mentor for Marion. This substitution confirms, again by negative example, that Bernard Herrmann's music emphasizes the immediate problem of desire in Marion—"Should I take the money or not?"—rather than the sexual desire that is the root motivation for her stealing the money (Figure 4-13). (Sexual desire here is best understood in the more abstract sense of wanting to establish the couple: Marion steals the money so that she and her lover, Sam, can move away and establish a life for themselves.) Alternatively, the stylistic anachronisms provoked by these songs might seem to put the singers at a distance from Marion, encouraging an interpretation of them as narrators or commentators (voice-over narration or "Greek chorus") and perhaps unsympathetic (anempathetic) commentators at that.

Finally, if we used an electronic dance track (say, early 1990s house, with volume level set lower than normal (for its style), although still prominent in the sound track), the music would pass over and ignore the very subtle shifts of emotion that occur every few seconds in this scene. What such an overbearing but neutral music confirms is that Bernard Herrmann's cue is firmly set in the classical tradition of empathetic, synchronized music. That in itself is an important observation about film style, as Herrmann is not necessarily known for closely adhering to that tradition, especially in his later film scores.

Note that the issue is not whether the music we substitute has an effect on the scene—thanks to the cognitive process that Chion called the "audiovisual contract,"⁶ any music added to the sound track will have *some* kind of effect,



Figure 4-13. *Psycho*. Marion takes the money.

leading us to draw connections among the things we see, even if those connections provoke confusion and therefore "play against" (what if we tried the opening of Beethoven's Fifth Symphony?). The issue is the narrative coherence of those connections—our ability to judge easily the appropriateness of the music to the characters, emotions, and actions of a scene and our understanding of its narrative contexts.