1

The Material Heterogeneity of Recorded Sound

Rick Altman

When we understood cinema as a text, we borrowed our terminology and our methodology from previously established textual domains. An understanding of cinema as event requires new terms and models for a new type of multi-dimensional analysis. It seems especially appropriate to begin this retooling process with the development of a new vocabulary for sound analysis, for sound itself is particularly event-oriented itself. Whereas image analysis has given us many terms and techniques fully consonant with a textual approach to cinema, sound's heterogeneity has much to offer to an event-oriented aesthetic. In order to reap sound's harvest, however, we must take a new approach to sound, replacing the idealist models offered by musical analysis with increased sensitivity to sound's three-dimensional materiality.

Current approaches to film sound systematically borrow a musical model. The most influential introductory film textbook of the last decade defines the acoustic properties of sound as *loudness*, *pitch*, and *timbre*.[1] This definition is based on the apparent assumption that all film sounds have the nature of musical notes, that is, they are single phenomena, produced instantaneously, emitted from a point source, and perceived in an immediate and direct fashion. With a definition like this one, we can explain many aspects of film sound, such as contrasts or confluences in volume, frequency, and tone.

In fact, since the terminology is borrowed from the realm of music, we find that with these terms we can handle almost any of the types of analysis typically practiced on a musical score. We note Hitchcock's suspenseful diminuendo from a loud slam to muffled scratching, the harmony of Orson Welles' bass and Joseph Cotten's tenor, the melodic gifts of Cary Grant and Katharine Hepburn, the awkward timbre of Zasu Pitts and Jerry Lewis, or the varied instrumentation of the "Symphony of Sounds" with which Rouben Mamoulian opens *Love Me Tonight*. It we could notate all film sounds according to the musical criteria of loudness, pitch, and timbre, then these three criteria would suffice for the analysis of film sound.

Yet this is precisely what we cannot do. While all film sounds have loudness, pitch, and timbre, not a single sound in cinema can be adequately described with this musical terminology. In fact, not even musical sounds can be fully described with musical terminology. More appropriate for describing musical scores than individual performances, musical terminology pays little attention to the details of any particular performance, concentrating instead on the common factors joining all performances of the same score. If I attend three concerts of Mozart's "Little Night Music," one in a well-upholstered salon, another in a large concert hall, and a third in a city park, I am in one sense hearing the "same" music three times, that is, music that is represented by a single, identical score. Yet how different are the sounds that reach my ears during the three concerts!

Musical notation assumes that each sound is single, discrete, uniform, and unidimensional. Stressing the formal concerns of music's internal, self-referential aspect, musical notation diverts attention from sound's discursive dimensions, concealing the fact that sound is in reality multiple, complex, heterogeneous, and three-dimensional. As a concept, middle C exists independently of space and time, in the abstract notion of a sound of approximately 262 cycles per second. As a reality, however, no two versions of middle C are identical, because of the different temporal and spatial circumstances in which they originate and are heard. The middle C located on the first line below the G clef may be only a concept, but the sound that we hear with our ears—whether on the street or in a movie theater—is a heterogeneous event that carries its own temporal and spatial dimensions and constitutes a full-fledged narrative. When we listen to recorded sound we are therefore always listening to a particular account of a specific event.

In order to respect the discursive complexity that is characteristic of all sound events, we can no longer continue to depend on a fundamentally conceptual terminology that remains insensitive to sound's phenomenality. Instead we must have a terminology capable both of respecting sound's heterogeneous nature and of figuring the narrative component built into the very process of recording and reproducing sound. This article proposes such a terminology, based on a schematic but systematic review of the physical phenomenon that we call sound.

Sound Events: The Production of Sound

What is sound? What happens when a sound is made? While this is hardly a technical treatise, it will nevertheless be useful to recall the

manner in which sounds are produced. Three elements are required for the production of any sound. First, there must be *vibration*, such as that of the vocal cords or a violin string. Second, the vibration must take place in a *medium* whose molecules can be set in motion, such as air, water, or a railroad rail (sound cannot be transmitted through a vacuum). Third, the transmitting medium must absorb and transmit the original vibrations in the form of *changes in pressure*. When a violin string is plucked or bowed, the molecules of the surrounding medium are compressed, with the pressure passed on from one molecule to the next. When the string reaches the end of its travel, maximum *compression* is achieved. As the string starts back, the molecules rush back to fill up the void left by the departing string. When the string reaches the end of its travel in the opposite direction, maximum *rarefaction* occurs. In order to create a specific, recognizable note this process must be repeated in rapid succession hundreds or even thousands of times a second. For example, the G string on a violin causes the surrounding air to go through 196 *compression/rarefaction cycles* per second, commonly expressed as a frequency of 196 cps or 196 Hz. In other words, what we call the musical note G below middle C is in fact a series of rapid changes in pressure.

Even taking the three-dimensional nature of sound events into account, however, this description vastly oversimplifies the situation. Whereas an electronic tone generator is capable of producing pure tones, all musical instruments produce notes that combine a *fundamental* frequency (such as the violin's 196 Hz G string) with a series of partials: *harmonics* (tones whose frequency is a whole number multiple of the fundamental) and *overtones* (tones whose frequency is related to the fundamental according to a more complex formula). Depending on the instrument and the way it is played, the combination of harmonics and overtones can vary tremendously. When played in such a way as to emphasize the upper harmonics, for example, the violin sounds harsh and strident, while a mellow tone results from stressing instead the lower harmonics. If the oboe, trumpet, flute, and cello sound so recognizably different, it is primarily because they produce radically different combinations of partials.

While few people are trained to hear harmonics and overtones, most listeners will rapidly recognize their absence, as when music is played through the telephone or over an old record player with limited frequency response. While the loss of these partials reduces our pleasure in listening to music, it may have an even more radical effect on other sounds. Spoken language becomes far more difficult to understand, voices and familiar sound effects may become harder to differentiate, even our ability to judge the distance and direction of a sound source may be impaired. In other words, the composite nature of sounds is hardly limited to music.

In fact, most of the events that we think of as a single sound are not singular at all. The musical model of tone generators and violin strings is extremely misleading. If a violin note could be produced by a violin string alone, then Stradivarius would never have become a household name. Every violin note is a complex event combining the vibrations of a string, a wooden case, and the air trapped inside that case. Each of these three contributes to the overall tone of the note played.

For what we call a sound is typically made up not only of multiple frequencies, but actually has multiple different fundamentals produced over a period of time. Think of the following familiar sounds: a refrigerator, snoring, a lawnmower, the wind, a squeaky door. We think of each as a single sound, but none is actually single in the way that an A-440 produced by a tuning fork is unitary. Each of these sounds constitutes an event taking place in time, involving multiple separate sounds organized in a familiar, recognizable fashion. Given the importance of rhythmic and melodic elements for our recognition of each of these sounds, it would be more appropriate to compare them to musical phrases than to individual notes.

Yet even individual notes have a temporal dimension. Returning for a moment to our violin string, consider the difference between plucking and bowing the string. In one case the sound starts suddenly, reaching its full volume extremely rapidly; in the other case the violinist seems to be sneaking up on the note, teasing the molecules into moving rather than suddenly shoving them. Whether violent or peaceful, this initiation of the sound event is termed the *attack*. It is followed by the *sustain*. How long is the note hold? How long does it stay at full volume? Finally, the sound fades away. This stage is called the *decay*, implying not only a temporal measure but also a qualitative one. Compare, for example, the decay of a plucked string that is simply allowed to spend its own energy and the decay of a plucked string instantaneously dampened by a finger.

As parts of the *sound envelope*, the stages of attack, sustain, and decay apply equally to any sound event. Contrast, for example, the smooth attack of Orson Welles' opening voice-over in *The Magnificent Ambersons* to the sharp attack of Georgie Minafer's dialogue. How essential to the *soundscape* of *The Wizard of Oz* is the gulf separating Margaret Hamilton's staccato attack and nearly instantaneous decay from Judy Garland's ability to ease in and out of speech! Anyone who has ever tried to edit dialogue will understand just how important the elements of the sound envelope are for the establishment of auditory realism. Even when the initial or final words of a sentence are perfectly comprehensible, they create an uneasy feeling whenever part of their attack or decay has been cut off in the editing process.

The production of sound is thus a material event, taking place in space and time, and involving the disruption of surrounding matter. This doesn't

mean that we have to be molecular physicists or sound engineers to understand sound, but it does suggest a very precise basis for our description of sound events. It is no longer sufficient to analyze a musical score or a written text to understand the effects of a particular performance event. Recognizing the extent to which sound sets matter in motion—albeit invisibly—we readily see the importance of developing a vocabulary and a methodology appropriate to the complex materiality of sound. Instead of describing just a sound's loudness, pitch, and timbre, we stress the extent to which every sound event includes multiple sounds, each with its particular fundamental and array of partials, each with its characteristic sound envelope, each possessing its own rhythm within the sound event's overall temporal range.

The Sound Narrative: The Story of a Sound Event

In order to understand sound as it is produced, we need to recognize the material heterogeneity of sound events. Sound production is only part of the story, however, for sound, like the proverbial tree falling in the forest, must be heard in order to take on its narrative and social significance. By offering itself up to be heard, every sound event loses its autonomy, surrendering the power and meaning of its own structure to the various contexts in which it might be heard, to the varying narratives that it might construct. Beginning as the vibration that induces molecular movement, sound is not actualized until it reaches the ear of the hearer, which translates molecular movement into the sensation of sound. Just as the sound event necessarily introduces a temporal dimension into the production of every sound, so the process of perception always guarantees sound's spatial nature.

When we speak of language, we implicitly agree to disregard certain aspects of linguistic discourse as somehow sub-linguistic. Fred Astaire and Ginger Rogers may make something of the difference between ee-ther and eye-ther, but no normal user of the English language shows such a concern. Regional accents and personal idiosyncracies produce recognizable differences, but these are not taken to be differences in language. Whether it's ee-ther or eye-ther, it's still the English word "either." Our understanding of sound works in a very similar fashion. We know that our neighbor's lawn mower sounds very different when it's mowing on the near or the far side of the house (and *vive la différence!*), yet that difference does not change our nomenclature. Whether the sound comes from the near side or the far side of the neighbor's house, it is still the sound of the neighbor's lawn mower. The sounds are different, but the *name* of the sounds is not (Metz). Systematically, the name of a sound refers to the production of sound and not to its consumption, to the object making the sound rather than the person perceiving it.

Yet the hearing process necessarily involves important variables that often outweigh the sound itself in importance. It doesn't take children long to learn that the word "Boo!" does not by itself produce surprise. When a child jumps out from behind a chair at the other end of the room and shouts "Boo!" the reaction is likely to be mild indeed. When my ten-year-old suddenly emerges from beneath my desk, on the other hand, she can be assured of a good return on her "Boo!", however quietly it may be spoken. To be sure, the sharp attack of the letter "b" contributes to the effect. (If perchance you are not convinced of this, try to scare someone with the pastoral attack of a "Moo!") Still more important, though, is the proximity effect obtained by a good surprise. The effect is dissipated if the booer holds her hand in front of her mouth or looks away from the booee. The reason for this is very simple: the surprise is created largely by the sudden arrival of a zone of sound pressure on the ear. Anything that diminishes the sharpness of this experience (standing too far away, whispering instead of shouting, facing away from the booee, or uttering the "Boo!" before emerging from the hiding place) spoils the effect. Having learned to distinguish between various versions of the "same" sound, our ears tell us how to react not on the basis of the sound event alone, but also according to our perceived relation to that sound event.

How does a sound event contribute to hearing? And what are we actually perceiving when we hear? In the previous section, I explained the molecular basis for sound's characteristic compression/rarefaction cycle. Vibration creates pressure, which is communicated through a medium. At the other end of sound's path, the human ear collects that pressure and transforms its mechanical energy into electrical impulses that the brain understands as sound. Sensitive to frequency (pitch), amplitude (loudness), and many other factors, the human ear is a marvelously sensitive organ capable of very minute distinctions. The ear hears not only a sound's fundamental frequencies, but its harmonics and overtones as well, thus facilitating the distinction between male and female voices or French horns and saxophones. Through the ear's ability to sense not only pressure but the rate of changes in pressure as well, we are able to measure even minute differences in the sound envelope, and thus to distinguish between individual voice patterns.

The ear must do far more than this, however, for until now I have assumed that sound arrives directly to the ear, in a single pencil of pressure. This is precisely not the case. Imagine an actress standing in the center of a stage in a large auditorium, 150 feet wide and 200 feet deep. Since sound travels at about 1130 feet per second at 70 degrees Fahrenheit, the actress's voice takes approximately one-eighth of a second to reach a member of the audi-

But what happens to the sound that goes straight to the spectator on the edge of the auditorium? Certainly it doesn't die there; it must eventually reach the ears of other audience members as well. Radiating out like a cone from the actress's mouth, the sound pressure soon fills up the entire auditorium, bouncing off the walls, the floor, and the ceiling, and bending around audience members, chairs, and posts until it is finally completely absorbed.[2] The notion that sound travels in a straight line from sound event to hearing ear is thus radically incomplete. In addition to *direct sound*, there is also a great deal of *reflected sound* or *reverberation*, produced by the sound that reaches the hearing ear only after bouncing off one or more surfaces. In a large room, the delay between the arrival of direct sound and the arrival of the last reflections can be quite long. When the full effect of three-dimensional reverberation is considered, delays of multiple seconds may easily be encountered.

Contrary to popular assumptions, even apparently instantaneous sounds thus have a considerable temporal dimension. Our notation systems for sound reinforce a received notion that separately produced sounds are also perceived separately. As they are printed, Hamlet's words "To be or not to be" provide a blueprint for sounds that are clearly separate and sequential. As they are perceived, however, the direct sound of one word is often heard before the reflected sound of the previous word ceases. Musical notation systematically distinguishes between *melody* (sequential sounds) and *harmony* (simultaneous sounds). Yet the sounds notated as sequential are heard as overlapping, thus confounding the received distinction. The reflected sounds of the first beat of the measure continue to be heard as the direct sound of later beats reaches our ears.

Such a distinction might easily appear purely academic and theoretical. Our ears know, however, that this is not just a question of splitting hairs. Who has not been in a large auditorium, cafeteria, or gymnasium and had trouble making out the speaker's words. The master of ceremonies may be saying "The winning numbers are seventeen, forty-three, fourteen, and seventy-two," but what we actually hear is more like "The win—— num—— seven—— four—— four—— seven——." Because we know what to expect in this context, we easily complete the opening words, but strain as we might, there is no understanding the all-important numbers, for the reflected sound of the first part of each word is bouncing all around the cafeteria—off chairs, tables, floor, walls, and ceiling—long enough to obscure the direct sound of the second half of each number. We are all aware of the difficulty of understanding a telephone message with the competition of a nearby conversation; our ears know that speakers in large halls often provide their own competition.

The subsistence of reflected sound does more than block understanding, however. Our ears are marvelously tuned instruments, extraordinarily

Imagine that there are two actors on our stage, one facing the audience, the other facing backstage. The lights are low; the audience cannot always be sure of seeing which actor's lips are moving. Yet we never have any doubt whatsoever about who is speaking. Our ears tell us.

The first actor, facing the audience, sends a strong ray of sound directly to each spectator (as well as an infinite number of rays that reach individual spectators as reflected rather than direct sound). The words pronounced by actor two, away from the audience, are prevented by the actor's own head from reaching spectators directly. In order to be heard at all, these sounds must rebound off the set or backdrop, thus taking up to three times as long to reach spectators. Fortunately, these spectators are also auditors. Their ears rapidly process this data and easily distinguish between the words that are being spoken directly to them and the words that have to bounce around the theater before arriving. This ability to measure the *ratio of direct to reflected sound* provides one of our most important capacities: the ability to distinguish between sounds that are being spoken to us and those that are meant for others. Imagine John Wayne walking down a line of new recruits standing at attention. The script might read "Johnson, straighten up. Jackson, button that top button. Jones, get that chin down. Altman, where'd you learn to tie a tie?" and so forth. The name at the beginning of each sentence, apparently spoken to gain each recruit's attention in turn, is actually quite redundant, for as Wayne moves down the line there is a change in the ratio of direct to reflected sound heard by each recruit. I need no course in acoustics to know when it's my turn, when I am the one being addressed.

In other words, the fact that a "single" sound reaches our ears over a period of time permits us to reconstitute certain facts about the circumstances surrounding the production of that sound. What our ears are doing is a form of narrative analysis. They are analyzing the narrative produced by sound pressure, in all of its complexity, in order to ascertain how, by whom, and under what conditions that sound pressure was produced. To be sure, some people have ears that are better trained in this process of narrative analysis than others, but we have *all* developed over the years a great deal of expertise in this area. We use the delay between visual information and the first arrival of direct sound to determine the distance of the sound source. The difference in the characteristics of sound arriving at our two ears permits us to locate the sound source laterally. The ratio of reflected to direct sound helps us to decide whether the speaker is facing us or not. Combined with other information, this ratio also helps us recognize the size of the room in which the words are spoken. By noting how long the reflected sound lasts, we refine our conclusions about the originating space.

In fact, we regularly draw still other conclusions from the other aspects

of sound's itinerary. During the course of its picaresque journey from production to perception, sound not only takes many specific courses, it sets in motion a particular medium and is reflected off particular surfaces. Imagine the difference that would be made by staging the preceding example of two onstage actors in two different theaters, one a plush Broadway theater with a velvet backdrop and the other a high school gymnasium with a concrete wall at the back of the stage. Just like a tennis ball thrown toward the back of the stage, the actor's words will in one case be muffled by the backdrop while in the other case they will shoot off of it at nearly their original velocity and volume. While few people are aware of the theory underlying such differences, our ears are surprisingly attentive to them. They seem to know that certain surfaces reflect different frequencies better than others, that some surfaces absorb more sound and dampen specific frequencies more than others, that some environments will continue to reflect sound almost indefinitely, while others will restrict reflected sound to a minimum. In this way, even before we look out the window, our ears tell us that it has snowed during the night. They help us distinguish between the recording of a junior high concert made in the gymnasium and the next day's recording of the "same" concert made in an upholstered, acoustically treated auditorium. They help us to get the "feel" of every room we enter, without ever touching any of the room's surfaces.

The fact that we come equipped with two functioning ears each makes still more information available to us. Because all sounds that are not exactly equidistant from both ears arrive at our ears one after the other, and under slightly different conditions, our ears are able to localize sound laterally as well as in terms of distance. Especially when aided by a radar-like rotation of the head, our own personal sonar gives us varied information about our soundscape.

Our ears are so good at decoding sound that it would be a shame to deprive our terminology of our ears' expertise. Without entering the specialized worlds of acoustics, audio engineering, and otology, we must nevertheless find ways of respecting not only sound's material heterogeneity, but also the cleverness of our ears in analyzing the auditory narratives that it constitutes. Constantly delayed, dampened, reinforced, overlapped and recombined, sound provides us with much of the information we need to understand its origins and its itineraries—but the existing terminology clearly does not.

The Sound Record: Recording the Story of a Sound Event

Every sound initiates an event. Every hearing concretizes the story of that event. Or rather, it concretizes a particular story among the many that could be told about that event. When the baseball broke the window, I was outside, more than a little worried; I heard the sound of the break directly, with little reflected sound, since there are no walls and ceiling outdoors to keep the reverberation going. My father was sitting in his favorite chair, right next to the broken window; he was subjected not only to the direct sound of the impact but also to a roomful of reflected sound. My mother was ironing in the back room; she thought something had broken, but the muffled reflected sound that reached her didn't specify whether it was a window, a vase, a car headlight, or something else still. Doing her homework in the second floor back bedroom, my sister hardly knew anything had happened. At least until she heard my father bellow.

All four of us heard the "same" sound, yet all four of us heard a "different" sound. Or, to put it in a more useful fashion, each of us heard a different narrative of the same event. Sound's existence as both event and narrative immensely complicates—and enriches—our understanding. Usually discussed as the most transparent of classical narratives, sound is in fact a *Rashomon* phenomenon, existing only in the separate stories of various perceivers of the original event. Potentially important apropos of any sound and its perception, this fact takes on special significance in all media that make use of recorded sound. For what the record contains is not the sound event as such but a record of a particular hearing, a specific version of the story of the sound event. Every recording is thus signed, as it were, with the mark of the particular circumstances in which it was heard. A recording of the shattering window made next to my father's easy chair will be signed in a different way from a recording of the "same" event made next to my sister's desk. Every recording carries the elements of this *spatial signature*, carried in the audible signs of each hearing's particularities. Even when those signs are contradictory or have been tampered with, even when they seem not to match the visual data provided with the sound record, they still carry information that is narrative and spatial in nature.

The situation is immensely complicated by the fact that sound records never convey exactly the same information that a given auditor would experience. Far from arresting and innocently capturing a particular narra-tive, the recording process simply extends and complicates that narrative. Just as the upholstery of a particular soundscape has an impact on the sound narrative, so the way in which sound is collected and entered into memory becomes part and parcel of the overall sound phenomenon.

Even in the simplest of sound collection systems, decisions regarding the location of the microphone carry enormous importance, especially when the sound is to accompany a related image. Should sound collection take place in the same room as the sound to be recorded? At what distance? Under what acoustic conditions? Or should sound collection be in a remote location, thus reducing volume, dampening certain frequencies, and in-

creasing the ratio of reflected to direct sound? This approach will certainly convince auditors that they are not located in the same sound space as the speaker. In fact, if the reverb level is high enough and the image slightly out of focus, the sound may even appear to have been collected in a time frame different from its production.

The process of editing further complicates the question of microphone location. Should the microphone location be changed every time the camera is moved and the shot changes? Or should sound logic remain entirely independent of image logic? To what extent is consistency of sound collection needed? Must sound collection decisions be subordinated to narrative concerns? Under what conditions may the volume and spatial characteristics of synchronized sound be modified during the editing process? Are there special volume and reverberation requirements for sound effects recorded separately from dialogue? All these and many other questions are implied by the simple necessity of choosing a microphone location.

Since the very beginnings of sound cinema, filmmakers have been convinced that intelligibility is one of the most important requisites of recording speech. Indeed, nowhere else are the stakes of microphone location so clear. Imagine that we are recording a sentence spoken by a woman to the man she is facing. While she is speaking a child walks silently past, catching the woman's attention and causing her to turn away from her interlocutor. Now, in order to maximize the intelligibility of the woman's words we might legitimately decide to "pan" the microphone with her, so that she is always talking directly into the mike, maximizing direct sound and thus intelligibility. Note, however, that this decision robs the sound track of its spatial characteristics. Instead of telling us that the woman turned away from her initial position, the sound track implies that she continued to face in the same direction.

If, instead, we choose to retain the initial microphone position throughout, the sound track will exhibit a faithful spatial signature, but it will almost certainly reduce our ability to understand the final parts of the woman's sentence. We will realize that the woman has turned her head while talking, but, like the man to whom she speaks, we may miss some of her words. Recording choices, as we easily see from this example, govern our perception of particular sound events. Far from simply recording a specific story of a specific sound event, the sound engineer actually has the power to create, deform, or reformulate that event. In the example just illustrated, the sound engineer must choose to allow either deformation of the dialogue or mistaken perception on the part of the auditor.

Nor is microphone location the only variable available to the sound engineer. The microphone itself makes many choices regarding the type, amount, and source of sound that will be collected. It is perhaps useful, in an image-oriented world, to think of the microphone as a "sound-camera," a collection device for sound that shares many of the characteristics of familiar image-collection devices. Just as cameras may have wide-angle or telephoto lenses, changing the angle of image collection and thus the apparent distance of the object filmed, so microphones vary from omnidirectional to narrowly focused, thus changing both the angle of sound collection and the apparent distance of the sound source. In addition, the change in the ratio of direct to reflected sound that accompanies a change in microphone may also affect perception of room size and other characteristics.

Microphones also vary in their sensitivity to specific sound frequencies. The familiar carbon microphone in our telephones has an extremely limited frequency response. Sound heard over the telephone thus always sounds dull and lifeless. Close-miking with a telephone mike (or stripping the sound of appropriate frequencies in postproduction) thus gives the impression that all sounds presented are being heard through a telephone. Since no microphone is equally sensitive to all frequencies, the choice of a microphone fairly assures that some sounds will be boosted, while others will be dampened.

Many other microphone characteristics may come into play as well. It is often assumed that every microphone produces a faithful sound record. Actually, *no* microphone produces an entirely faithful sound record. Not only does every microphone have its own particular *directional characteristics* (omnidirectional, bidirectional, cardioid, shotgun and so on), but every microphone also has its own particular frequency response, sound configuration, and power requirements. In addition, many microphones produce unwanted sounds of various types (hum, pop, hiss, buzz, crackle and so on) in a wide variety of situations (loud sound signal, wind pressure, close sound source, vibration and so on).

Recorded sound thus always carries some record of the recording process, superimposed on the sound event itself. Added to the story of sound production we always find the traces of sound recording as well, including information on the location, type, orientation, and movement of the sound collection devices, not to mention the many variables intervening between collection and recording of sound (amplification, filtering, equalization, noise reduction, and so forth). Indeed, the recording system itself provides one of the most important determinants of sound characteristics; as such it not only provides a record of sound, it also participates in the overall sound narrative. Think for example of the differing frequency responses of 78 rpm records and digital compact disks. It is so difficult to compare musical performances recorded on these two radically different technologies that the masterworks of Toscanini and Furtwangler seem diminished

without the wonders of digital remastering (which is none other than an attempt to restore the frequencies to which pre-war disk recording was not sensitive).

To record is thus to recall to mind, as the dictionary would have it, but like most mnemonic devices, sound recordings must heighten some aspects of the original phenomenon at the expense of others. So-called recordings are thus always representations, interpretations, partial narratives that must nevertheless serve as our only access to the sounds of the past.

Sound Reproduction: Playing the Record of the Story of a Sound Event

But how can we gain access to those sounds? A recording, as we all know, is not a sound. Without some sort of playback device, a recording can only sit silently on the shelf. And as long as it sits on the shelf, it has only one space: the space of the recording of the original sound event. My record of Oistrakh and Rostropovich playing the Brahms Double Concerto with George Szell and the Cleveland Philharmonic Orchestra was recorded in Severance Hall. Once I put the record on my stereo and set the needle down, however, the Concerto becomes very Double indeed. Not only do I hear the fabulous acoustics of the Cleveland Orchestra's home concert hall, but at the same time I have to put up with the less than ideal acoustics of my own living room. Every sound I hear is thus double, marked both by the specific circumstances of recording *and* by the particularities of the reproduction situation.

The late twenties were a time of particularly intense reflection on this problem. Throughout the twenties, movie theaters had grown increasingly large and ornate. While the desire to accommodate growing numbers of higher class patrons was an important factor in the rise of the picture palace, theater acoustics played a part as well. Silent films depended heavily on music chosen from familiar nineteenth-century sources. Now, the nineteenth century had little use for chamber music or small baroque organs, preferring instead large orchestras and enormous choirs, along with the long-lasting reverberations and high indirect-to-direct sound ratios characteristic of spacious concert halls or churches. The music that silent cinema inherited from late romantic composers was thus expected to sound as if it were being played in large, enclosed spaces. Ornate picture palaces, with their multiple levels, private boxes, rows of fluted columns, and endless plaster moldings, were thus a perfect environment for the sounds of silent cinema.

When synchronized dialogue came to cinema, however, a new set of requirements was rapidly imposed. The words had to be comprehensible; to that end the amount and duration of reflected sound had to be kept extremely low. Studios thus rushed to create acoustically treated sets, open-ended studios, and other devices designed to limit reverberation and maximize intelligibility. The biggest stumbling block of all turned out, however, to have nothing to do with film production. It was film exhibition that caused all the problems. Built to maximize reverberation, those drafty barns called picture palaces made it nearly impossible to understand the spoken word. When played over the loudspeakers of a huge, hard-surfaced Roxy with three second decay time, even intimate scenes recorded to give the impression of small, private spaces sounded as if they were set in cavernous public halls. Carefully recorded speech turned into the same auditory mush that had become the trademark of romantic church organs and mighty theater Wurlitzers alike. Only careful redesign and costly acoustic treatment were able to solve this problem.

Even with the practical problem solved, however, the theoretical difficulty remains. Which acoustics am I listening to? The Hollywood sound stage or the Rialto? Severance Hall or my living room? For that matter, which sound am I listening to? The original sound event or its loudspeaker reproduction? In order to understand sound—cinema sound in particular—we must recognize both the narrative and the represented nature of sound as it reaches our ears in the movie theater. The sound system plays the record of the story of an event. At every point in that chain, new variables enter, new elements of uncertainty. Sound heads, amplifiers, leads, loudspeakers, and theater acoustics all force new auditory data on the audience, just as the recording process itself had earlier introduced an implicit viewpoint.

Hearing Events: Hearing the Record of the Story of a Sound Event

Just as sound events remain only hypothetical sound sources until they are actualized by a hearer, so the playing of a sound record takes on meaning only in the presence of an audience. Yet the process of hearing a recording differs significantly from listening to a live sound event. This should come as no surprise to anyone who has contemplated the difference between a photograph and the scene that it represents. When we look directly at a scene we gain a sense of depth from our binocular vision, by rotating our head, or by moving to the left or right. If we want to know what's underneath a chair we have but to lean down. In order to get a clearer view of a specific object, we need only adjust the focus of our eyes. Yet all of this is to no avail when we view a photograph. No amount of rotating, moving, leaning, or adjusting will deliver information that the photograph lacks. We may have two eyes, but we might as well all be Cyclops when it comes to sensing distance in a photograph, for here the concept of distance is encoded through size, masking, and detail rather

than sensed by the parallax implicit in binocular vision. Without requiring any special education, we have all learned to use our eyes radically differently when we view three-dimensional space and when we view a two-dimensional representation of that space.

A similar situation obtains with the sound representations that we call recordings. When attempting to locate a crying child we normally call heavily on our binaural hearing system to provide cues regarding lateral location. When we listen to a recording of a crying child, no such localization is possible. However much we might rotate our heads or change positions, we remain unable to make use of the directional information that was present when the sound was produced, but which is no longer available in the recording (unless it is in stereo, and even then the location of microphones and speakers plays just as important a role as the location of the original sound source). For listening to the sound pouring out of a loudspeaker is like hearing a lawn mower through an open window: wherever the lawn mower may actually be, it always appears to be located on the side of the house where the open window is.

When we listen for a crying child, we are marvelously effective at cutting out extraneous sounds and concentrating on the cries that we recognize as those of our own child. Dubbed the *cocktail party effect* by Colin Cherry, the process of selective auditory attention is far more difficult when we are listening to recorded material. Whereas live sound provides an extraordinary number of variables, each permitting and promoting selective attention, recorded sound folds most of those variables into a single, undifferentiated source. In a live situation, we easily differentiate among the various sound sources surrounding us, but with recorded sound no such clear distinctions are possible.

Live sound situations reveal the actual relationship between the sound producer and perceiver, while recordings suggest only an apparent relationship. If I sit in an auditorium and listen with my eyes closed to a series of speeches, I remain constantly aware of the speakers' location. I know what direction they are facing, how loud they are speaking, and what tones of voice they are using. When I listen to a recording of the same meeting, I can no longer locate the speakers. Nor can I be sure of their original body positions, volume, or tones. Depending on the type, location, and movement of the microphone(s) used in the recording process, the recorded sound substitutes an apparent sound event for the original phenomenon. Revealing its mandate to *represent* sound events rather than to *reproduce* them, recorded sound creates an illusion of presence while constituting a new version of the sound events that actually transpired.

What happens in the course of a hearing event is thus not the expected detective activity wherein the hearer searches the recorded sound track for clues permitting reconstitution of the original sound event. Instead, we follow the trail that has been laid for us all the way to an apparent sound event having all the aural guarantees of reality but only partial correspondence to the original sound event. Indeed, it is the partial nature of the relationship that makes hearing events so fascinating. If there were no connection between the apparent sound event and the original sound source, recorded sound would not have its extraordinary capacity for ideological impact. It is precisely because recorded sound seems to reproduce an original phenomenon that recordings attract and hold audiences so readily. Between the illusion of reproduction and the reality of representation lies the discursive power of recorded sound.

We hear recordings with the same ears we use for live sound. We reach conclusions about the evidence provided by recordings in the same way that we interrogate and evaluate live sound. We constitute apparent sound events just as we directly perceive live sound events. Yet recordings systematically fail to justify our confidence in them. Most listeners have learned to concentrate on the aspects of sound events that are most faithfully rendered by recordings and to pay little attention to the aspects introduced or transformed by the recording process. A proper theory of sound will accept no such selective deafness. It will pay special attention to those very points where confusion is possible, recognizing in such moments of imprecision, indecision, or incoherence the very place where sound seizes the opportunity to take an active role in the definition and exploitation of culture.

Sound Terminology: Talking about Hearing the Record of the Story of a Sound Event

Often called "distortions," on the theory that sound recording is a science of reproduction rather than an art of representation, the variables introduced by sound's material heterogeneity, along with the system constituted to record (that is, represent) it, lie at the very heart of film sound. Though they may constitute distortions for the sound engineer, the marks of the sound narrative and the recording process that appear as part of any sound record constitute the very text of the sound analyst, the fundamental signs of the sound semiotician, the basic facts of the sound historian.

Central to the interpretation of film sound is the fact that multiple moments and operations must be carried simultaneously by the same final sound track. The characteristics of sound production, sound recording, sound reproduction, and audience perception are all superimposed in a single experience. When we hear any particular film sound, how do we know to whom to attribute it? Which part of the sound chain has produced, selected, highlighted, or masked it? Does a decrease in the ratio of direct to indirect sound mean that the character has turned away, an obstacle has

been introduced, the microphone has been moved, the sound engineer has fiddled with the dials, or the spectator has shifted her position?

To study film sound is to take seriously the multiplicity of possible determinants of any given audience perception. As a complex representation of a complex sound event, cinema sound offers sound designers infinite possibilities for creation and obfuscation. As such, it also offers theoreticians and critics of cinema sound fascinating opportunities to recognize and analyze the techniques, conventions, codes, and ideological investments of the sound chain. This work is only beginning. It will move more quickly if we adopt a vocabulary that reflects the material, heterogeneous nature of sound presented here.

Notes

1. The Material Heterogeneity of Recorded Sound

1. Bordwell 1990. Not surprisingly, technical manuals aimed at the production of sound rather than at aesthetic analysis of sound present a far broader terminology. The best and most complete of these manuals is by Stanley R. Alten.

2. As a general rule, sound waves will be reflected by (that is, bounce off) obstacles having dimensions greater than the wavelength of the sound. Since audible sounds have wavelengths varying from about one and one-half inches (the C above the piano, 8372 Hz) to 70 feet (the C below the piano, 16 Hz), with the fundamentals of most sounds having wavelengths between one and eight feet, most acoustic situations will produce a combination of reflected and refracted (that is, bent) sound.

2. Sound Space

1. "Reproducing Sound from Separate Film," *JSMPE* 16 (Feb. 1931), p. 152. Just three months later a patent taken out by W. Bouwa, for "Apparatus and Method for Localization of Sound on Screen," was reported on in *JSMPE* 16.5 (May 1931), 643–44.

2. On the particular outlook of sound technicians during this period, see Altman 1992.

3. It is difficult, however, to reconcile Maxfield's total mastery over everything related to reverberation with an obvious oversight in his 1938 reprinting of the microphone placement chart from the 1931 article (p. 73). Whereas before 1931 only omnidirectional microphones were available in Hollywood, by 1938 the ribbon bidirectional mike and the cardioid mike were widely used, especially where dialogue had to be recorded with lenses of large focal length, such as those charted in Maxfield's graph. Now, it is generally recognized that the cardioid (directional) microphone collects far less reflected sound than an omnidirectional mike at the same distance, thus permitting placement of the cardioid mike at 1.7 times the distance appropriate for an omnidirectional mike. Is Maxfield's blind spot perhaps a sign of fidelity to the parent Bell/WE/ERPI complex? Whereas Olson's directional mikes were developed for RCA in the early thirties, Western Electric did not have a [illegible]