In *Audio-Vision,* the French composer-filmmaker-critic Michel Chion presents a reassessment of the audiovisual media since sound's revolutionary debut in 1927 and sheds light on the mutual influences of sound and image in audiovisual perception.

Chion expands on the arguments from his influential trilogy on sound in cinema—*La Voix au cinema, Le Son au cinema,* and *La Toile trouee*—while providing an overview of the functions and aesthetics of sound in film and television. He considers the effects of evolving audiovisual technologies such as widescreen, multitrack sound, and Dolby stereo on audio-vision, influences of sound on the perception of space and time, and contemporary forms of audio-vision embodied in music videos, video art, and commercial television. His final chapter presents a model for audiovisual analysis of film.

Walter Murch, who contributes the foreword, has been honored by both the British and American Motion Picture Academies for his sound design and picture editing. He is especially well-known for his work on *The Godfather, The Conversation,* and *Apocalypse Now.*

"Michel Chion is the leading French cinema scholar to study the sound track. . . . I know of no writer in any language to have published as much in this *area,* and of such uniformly high quality, a, he."

ALAN W|LUAMS

RUTGERS UNIVERSITY

MICHEL CHION is an experimental composer, a director of short films, and a critic for *Cahiers du cinema.* He has published books on screenwriting, Jacques Tati, David Lynch, and Charlie Chaplin, in addition to his four books on film sound.

CLAUDIA GORBMAN is a Professor in the Liberal Studies Program at the University of Washington, Tacoma.

Jacket illustration: *Eratorhmad* by David Lynch, 1976.

Jacket design: John Costa

Printed in **U.S.A.**

COLUMBIA UNIVERSITY **PRESS**
NEW YORK

ISBN 0-231
9 780231

AUDIO-VISION

SOUND ON SCREEN

MICHEL CHION

FOREWORD BY

WALTER MURCH

EDITED AND TRANSLATED BY

CLAUDIA GORBMAN

A U D I O - V I S I O N

. . .

SOUND ON SCREEN

Michel Chion

. . .

edited and

translated by

Claudia Gorbman

with a foreword by

Walter Murch

COLUMBIA UNIVERSITY PRESS • NEW YORK

Columbia University Press wishes to express its appreciation of assistance given by the government of France through Le Ministere de la Culture in the preparation of the translation.

Columbia University Press
New York Chichester, West Sussex
L'Audio-Vision © *1990 Editions Nathan, Paris*

Library of Congress Cataloging-in-Publication Data
Chion, Michel
[Audio-vision, French]
Audio-vision: sound on screen/Michel Chion; edited and translated by Claudia Gorbman; with a foreword by Walter Murch.
p. cm
Includes bibliographical references and index.
ISBN 0-231-07898-6
ISBN 0-231-07899-4 (pbk.)
1. Sound motion pictures. 2. Motion pictures—Sound effects. 3. Motion pictures—Aesthetics. I. Gorbman, Claudia. II. Murch, Walter, 1943- . III. Title.
PN1995J.C4714 1994
791.43'024r-4c20 93-23982
CIP

Casebound editions of Columbia University Press books are printed on permanent and durable acid-free paper

c 10 987654321
p 10 987654321

CONTENTS

FOUR

THE AUDIOVISUAL SCENE

* * *

Is THERE AN AUDITORY SCENE?

"The Image" = The Frame

Why in the cinema do we speak of "the image" in the singular/when a film has thousands of them (only several hundred if it's shots we're counting, but these too are ceaselessly changing)? The reason is that even if there were millions, there would still be only one container for them, the frame. What "the image" designates in the cinema is not content but container: the frame.

The frame can start out black and empty for a few seconds (Ophuls's *Le Plaisir,* Preminger's *Laura)* or even for several minutes (Duras's *L'Homme Atlantique).* But it nevertheless remains perceivable and present for the spectator as the visible, rectangular, delimited place of the projection. The frame thus affirms itself as a preexisting container, which was there before the images came on and which can remain after the images disappear (end credits reaffirm this role in a certain way).[1]

What is specific to film is that it has just *one place for images*—as opposed to video installations, slide shows, Sound and Light shows, and other multimedia genres, which can have several. This fact, and no other, accounts for why we speak of the image in the singular.

Let us recall that in the first years of the cinematograph people sought to soften the hard borders of the frame, through irising, masking, or haloing, similar to such effects in photography. But these techniques were abandoned little by little, and, aside from the rare experiment with changing frame dimensions within a single film (Max Ophuls in *Lola Montes),* the principle of the full-frame image came to dominate in 99 percent of movies. Similarly, the occasional experiment with multiscreen cinema—Abel Gance's *Napoleon,* Michael Wadleigh's *Woodstock,* or even Paul Morissey's *Forty Deuce*—have not spawned many descendants, and as exceptions they prove the rule of the classical frame.

There Is No Auditory Container for Sounds

What is the corresponding case for sound? The exact opposite. For sound there is neither frame nor preexisting container. We can pile up as many sounds on the soundtrack as we wish without reaching a limit. Further, these sounds can be situated at different narrative levels, such as conventional background music (nondiegetic) and synch dialogue (diegetic)—while visual elements can hardly ever be located at more than one of these levels

at once. So there is no auditory container for film sounds, nothing analogous to this visual container of the images that is the frame.

What do sounds do when put together with a film image? They dispose themselves in relation to the frame and its content. Some are embraced as synchronous and onscreen, others wander at the surface and on the edges as offscreen. And still others position themselves clearly outside the diegesis, in an imaginary orchestra pit (nondiegetic music), or on a sort of balcony, the place of voiceovers. In short, we classify sounds in relation to what we see in the image, and this classification is constantly subject to revision, depending on changes in what we see. Thus we can define most cinema as "a place of image's, plus sounds," with sound being "that which seeks its place."[2] This relation differs from that of television, as we will see later on.

If we can speak of an audiovisual scene, it is because the scenic space has boundaries, it is structured by the edges of the visual frame. Film sound is that which is contained or not contained *in an image;* there is no place of the sounds, no auditory scene already preexisting in the soundtrack—and therefore, properly speaking, *there is no soundtrack.*

But Jean-Marie Straub's and Daniele Huillet's highly idiosyncratic 1969 film *Othon* (which acts out a Roman tragedy by Corneille on modern-day Roman locations) demonstrates what a sound scene or an auditory container-of-sounds might be in a monaural film. We'd have to agree that the sounds are the actors' voices declaiming their lines, and that the container would be the urban hum of distant traffic in which the voices and lines are heard. Actors in *Othon* often give long monologues offscreen, and yet such voices are not perceived as the traditional offscreen voice entirely determined by the image. Their voices seem to be "in the same place" as voices of actors we do see, a space defined by the background noise. A related effect can be felt in another film of the same year, Jacques Rivette's *La Religieuse.* Here, the reverb around voices, which results from direct sound (as with Straub and Huillet), has a similar role of enveloping and homogenizing the voices, inscribing them in a space like the medium of city traffic noise does in *Othon.* The price each film pays is a relative loss of intelligibility. Generally speaking, certain effects of the "spatial signature," as Rick Altman calls it, can provide the framework for an auditory scene.[3]

At least all this holds true until the arrival of Dolby, which now creates a space with fluid borders, a sort of superscreen enveloping the screen—the superfield, which I expand upon in a later chapter. But the superfield does not altogether upset the structure we have described, even if it has set it trembling on its base.

How THE IMAGE "MAGNETIZES" SOUND IN SPACE

What does a sound typically lead us to ask about space? Not "Where is it?"—for the sound "is" in the air we breathe or, if you will, as a perception it's in our head—but rather, "Where does it come from?" The problem of localizing a sound therefore most often translates as the problem of locating its source.

Traditional monaural film presents a strange sensory experience in this regard. The point from which sounds physically issue is often not the same as the point on the screen where these sounds are supposed to be coming from, but the spectator nevertheless does perceive the sounds as coming from these "sources" on the screen. In the case of footsteps, for example, if the character is walking across the screen, the sound of the footsteps seems to follow his image, even though in the real space of the movie theater, they continue to issue from the same stationary loudspeaker. If the character is offscreen, we perceive the footsteps as if they are outside the field of vision—an "outside" that's more mental than physical.

Moreover, if under particular screening conditions the loudspeaker is not located behind the screen, but placed somewhere else in the auditorium or in an outdoor setting (e.g., at the drive-in), or if the soundtrack resonates in our head by means of earphones (watching a movie on an airplane), these sounds will be perceived no less as coming from the screen, in spite of the evidence of our own senses.

This means that in the cinema there is *spatial magnetization* of sound by image. When we perceive a sound as being offscreen or located at screen right this is a psychological phenomenon, at least if a monaural projection is involved.

During the first years of multitrack sound, attempts at real spatialization were made—that is, really locating the sound on the left side of the screen if its source was shown there. The problem with these efforts is precisely that they ran into this psychological phenomenon of spatialization. Mental spatialization had been a blessing for the sound film, since it allowed movies to function for well over forty years without problems. We only need imagine the mess if sounds had to issue from the points where their sources on the screen were shown: one would have to install veritable beehives of speakers behind and around the screen. Not to mention, of course, the headaches of sound matching that would have resulted.

In using Dolby today filmmakers have learned the lesson from these first efforts in realistic spatialization and their "in-the-wing effects" (see p. 83). Today's multitrack mixes very often strike a compromise between psychological localization and real localization.

Note that sound coming from another point than the screen is "magnetizable" only if the sound itself maintains a basic spatial stability. If it constantly moves back and forth among loudspeakers, the image will have a harder time absorbing it, and the sound

takes on a centrifugal force of its own that resists visual "attraction."

Even in the classic case of a single loudspeaker, there is one real sonic dimension that the sound cinema capitalized on in its infancy, and neglected later: depth, the sensation of distance from the source. The ear detects depth from such indices as a reduced harmonic spectrum, softened attacks and transitions, a different blend of direct sound and reflected sound, and the presence of reverberation. The factor of depth has figured importantly in experiments with sound perspective in some films.[4] Let us note, however, that sound perspective was not so much a true depth, necessarily situating the sound source to the rear of the spatial plane of the screen, as a *distance* interpreted by the spectator in various different directions, depending on what she or he saw on the screen and could infer about the place of the source. In other words, a distant sound can be interpreted as being distantly to the left, far to the right, far behind the spectator, far to the rear of the screen; in other words, always localized in space depending on mental factors.

Thus to mental localization, determined more by what we see than by what we hear (or rather by the relationship between the two), we may oppose the absolute spatialization made possible by multitrack film sound.

THE ACOUSMATIC

Acousmatic, a word of Greek origin discovered by Jerome Peignot and theorized by Pierre Schaeffer, describes "sounds one hears without seeing their originating cause."[5] Radio, phonograph, and telephone, all which transmit sounds without showing their emitter, are acousmatic media by definition. The term *acousmatic music* has also been coined; composer Francis Bayle, for example, uses

it to designate concert music that is made for a recorded medium, intentionally eliminating the possibility of seeing the sounds' initial causes.

What can we call the opposite of acousmatic sound? Schaeffer proposed "direct," but since this word lends itself to so much ambiguity, we shall coin the term *visualized* sound—i.e., accompanied by the sight of its source or cause.

In a film an acousmatic situation can develop along two different scenarios: either a sound is visualized first, and subsequently acousmatized, or it is acousmatic to start with, and is visualized only afterward. The first case associates a sound with a precise image from the outset. This image can then reappear with greater or lesser distinctness in the spectator's mind each time the sound is heard acousmatically. It will be an "embodied" sound, identified with an image, demythologized, classified.[6]

The second case, common to moody mystery films, keeps the sound's cause a secret, before revealing all. The acousmatic sound maintains suspense, constituting a dramatic technique in itself. A theatrical analogy to this treatment of sound might be to announce and then to delay a stage entrance; think of Tartuffe, who finally enters during the third act of Moliere's play. The cinema gives us the famous example of M; for as long as possible the film conceals the physical appearance of the child-murderer, even though we hear his voice and his maniacal whistling from the very beginning. Lang preserves the mystery of the character as long as he can, before "de-acousmatizing" him.[7]

A sound or voice that remains acousmatic creates a mystery of the nature of its source, its properties and its powers, given that causal listening cannot supply complete information about the sound's nature and the events taking place.

If s fairly common in films to see evil, awe-inspiring, or otherwise powerful characters introduced through sound before they are subsequently thrown out to the pasture of visibility, de-acousmatized. Odile Larere has discussed the example of Visconti's *Conversation Piece,* where the intruders who disturb the lovely universe of the hero, the old professor played by Burt Lancaster, systematically make their entrance on the soundtrack before being visible.[8]

The opposition between visualized and acousmatic provides a basis for the fundamental audiovisual notion of offscreen space.

THE QUESTION OF OFFSCREEN SPACE

Onscreen, Offscreen, Nondiegetic

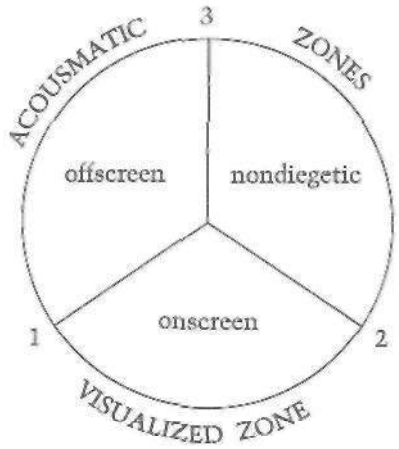
The question of offscreen sound has long dominated an entire field of thinking and theorizing about film sound, and it occupies a central place in my first two books on sound as well. Although we can see now that it seems to have been privileged at the expense of other avenues of investigation, it has yet to lose its importance as a central problem—even if the recent evolution of film sound, involving mainly multitrack sound and the "superfield" it establishes, has modified some of its basic traits.

In the narrow sense *offscreen sound* in film is sound that is acousmatic, relative to what is shown in the shot: sound whose source is invisible, whether temporarily or not. We call *onscreen sound* that whose source appears in the image, and belongs to the reality represented therein.

Third, to designate sound whose supposed source is not only absent from the image but is also external to the story world, I shall use the term *nondiegetic?* This is the widespread case of voiceover commentary and narration and, of course, musical underscoring.

Do Exceptions Disprove the Rule?

In *Le Son au cinema* I presented onscreen, offscreen, and nondiegetic as three zones of a circle, wherein each communicates with the other two:



1. Onscreen/offscreen border
2. Onscreen/nondiegetic border
3. Offscreen/nondiegetic border

But in recent years, the distinction onscreen-offscreen-nondiegetic, which arises from very basic considerations, has often been denounced as obsolete and reductive. Critics have problematized it with increasing fervor, because of the exceptions and special cases it doesn't seem to account for. For example, where should we situate sounds (usually voices) that come from electrical devices located in the action and that the image suggests or directly shows: telephone receivers, radios, public-address speakers? And what to do with a character who speaks with her back to us, so we don't actually see her speak? Is her voice acousmatic (offscreen)? And what can we say about the so-called internal voice of a character who can be seen in the image—the voices of his conscience, of his memory, of his imaginings and fantasies?[10]

What about Amy Heckerling's *Look Who's Talking,* where an adult voice accompanies the facial expressions of a baby, and articulates the baby's thoughts and feelings when the baby obviously doesn't have the physical and intellectual ability to do so? The voice is definitely connected to the present of the action, but it is not visualizable; so it seems unconcerned with these distinctions, being tied to the image via the loosest of synchronization. And finally how should we classify general background sounds such as birdsongs and wind, heard with natural exteriors? It seems rather ridiculous to characterize them as offscreen, on the basis that we don't "see" the little birds chirping or the wind blowing.

These exceptions, though distressing, do not by any means cancel out the validity or interest of a basic distinction between onscreen, offscreen, and nondiegetic sound, or of the basic division between acousmatic and visualized.

A Topological and Spatial Perspective

Anyone who brings up such exceptions in order to claim the categories useless or trivial is throwing out the baby with the bathwater. Why reject a valuable distinction simply because it isn't absolute? It is a mistake to see things in a binary, all-or-nothing logic. These distinctions only have meaning from a geographical, topological, and spatial perspective, analogous to zones among which one finds many shadings, degrees, and ambiguities. Of course we must continue to refine and fill in our typology of film sound. We must add new categories—not claiming thereby to exhaust all possibilities, but at least to enlarge the scope, to recognize, define, and develop new areas. -

Ambient Sound (Territory-Sound)

Let us call *ambient sound* sound that envelops a scene and inhabits its space, without raising the question of the identification or visual embodiment of its source: birds singing, churchbells ringing. We might also call them *territory sounds,* because they serve to identify a particular locale through their pervasive and continuous presence.

Internal Sound

Internal sound is sound which, although situated in the present action, corresponds to the physical and mental interior of a character. These include physiological sounds of breathing, moans or heartbeats, all of which could be named *objective-internal* sounds. Also in this category of internal sounds are mental voices, memories, and so on, which I call *subjective-internal* sounds.

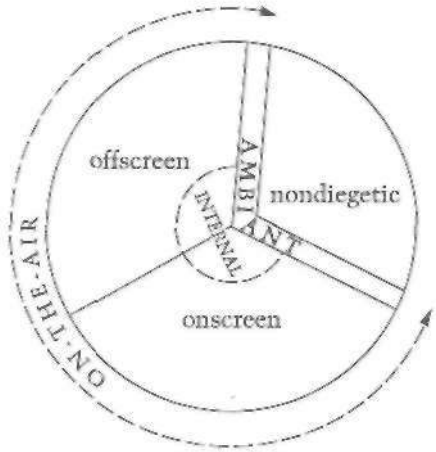
Bruce Willis's voice in *Look Who's Talking* gives us an interesting case of an internal voice, partly externalized through gesture. The film establishes it as not being heard by the other characters. In the voice of the adult that the baby will become, it tells us what the baby might be thinking, even as this voice is associated with the gestures in a way that is faithful to codes of realism regarding the baby's physical abilities.

"On-the-Air" Sound

I shall refer to sounds in a scene that are supposedly transmitted electronically as *on-the-air*—transmitted by radio, telephone, amplification, and so on—sounds that consequently are not subject to "natural" mechanical laws of sound propagation. In fact, to an ever greater degree, these sounds from television sets, clock radios, and intercoms are taking on a unique status in the films they appear in. Sometimes we hear them in sound closeup—clear and sharp, as if the film's loudspeaker were directly plugged into the radio, telephone, or phonograph depicted on the screen. At the other extreme thay can be identified in the setting by acoustical traits to produce an effect of distancing, reverb, and the particular tone color of the speakers or whatever their onscreen source is. Between these two cases lie infinite degrees of variation. On-the-air sounds, usually situated in the scene's real time, enjoy the freedom of crossing boundaries of cinematic space.

A particular case of on-the-air sound is that of recorded or broadcast music. Depending on the particular weight given by such factors as mixing, levels, use of filters, and conditions of music recording—i.e., whether the emphasis is on the sound's *initial source* (the real instruments that play, the voice that sings) or on the *terminal source* (the speaker present in the narrative whose material presence is felt through use of filters, static, and reverb), the sound of on-the-air music can transcend or blur the zones of onscreen, offscreen, and nondiegetic. It can also be read, to greater or lesser degrees, as screen music or pit music. Road movies 6uch as Barry Levinson's *Rain Man* constantly play with this oscillation. As early as 1975 George Lucas's *American Graffiti*—with the help of its sound designer Walter Murch—explored the entire gamut of possibilities between these two poles. The film was based on the simple setup of placing its characters in their cars for much of the action, all listening to a single rock-and-roll station.

The same problem exists for *dialogue* presented in the diegesis as recorded: does it refer to the time of its production or to the time at which we are hearing it? Imagine a scene in a film where a man is listening to a taped interview. If the sound being listened to has technical qualities of directness and presence, it refers back to the circumstances of its original state. If it has aural qualities that highlight its "recordedness," and if there is emphasis on the acoustic properties of the place where it is being listened to in the diegesis, we tend to focus on the moment where the recording is being heard. In *The Passenger* there is a sequence where Jack Nicholson listens to the recording of a conversation he had with a man he met by chance. Antonioni shuttles from one position to the other and in this way leads into a flashback. The interview Nicholson is listening to becomes real, becomes the scene of the interview itself.



So our tripartite circle becomes more complicated, but also richer. Through the very exceptions we introduce, it continues to illustrate the dimensions and oppositions involved:

- the opposition between acousmatic and visualized,
- the opposition between objective and subjective or real and imagined,
- the differences between past, present, and future.

It is important to think of the circle as consisting of interlocking sectors. In fact, this would probably be expressed much better by a topological model in three dimensions. We also return to the question of the *source*, which conditions such distinctions. First, the idea of sound source must be relativized and "unpacked," since "the" sound source is usually a multifaceted phenomenon. And second, the making and even the conception of a film and its screenplay are likely to emphasize one of these facets to a greater or lesser degree.

Place of the Sound, Place of the Source

Spatially speaking, a sound and its source are two different entities. In a film the emphasis may fall on one or the other, and the onscreen-offscreen question will pose itself differently, according to which thing—the sound or its cause—the spectator reads as being "in" the image or "outside" it. For sound and cause, though quite distinct, are almost always confused. But surely this confusion is inscribed also at the very heart of our experience itself, like an unsettling knot of problems.

For example, the sound of a shoe's heel striking the floor of a reverberant room has a very particular source. But as sound, as an agglomerate of many reflections on different surfaces, it can fill as big a volume as the room in which it resonates. In fact, no matter how precisely a sound's source can be identified, the sound in itself is by definition a phenomenon that tends to spread out, like a gas, into whatever available space there is.

In the case of ambient sounds, which are often the product of multiple specific and local sources (a brook, bird songs), what is important is the space inhabited and defined by the sound, more than its multisource origin. The same goes for films of musical performances. Depending on choices in the editing and technical directing of sound and image, the emphasis can fall either on the specific material source of the sound (the instrument, the singer) or on the sound as it fills the auditory space, considered independently from the source.

The more reverberant the sound, the more it tends to express the space that contains it. The deader it is, the more it tends to refer to its material source. The voice represents a special case. In a film, when the voice is heard in sound closeup without reverb, it is likely to be at once the voice the spectator internalizes as his or her own and the voice that takes total possession of the diegetic

space. It is both completely internal and invading the entire universe. This is what I have called the *I-Voice.*[n] Of course the voice owes this special status to the fact that it is the original, definitive sound that both fills us and comes from us.

In the play of onscreen and offscreen space, background music also stands out as a type of exception that proves the overall rule.

THE EXCEPTION OF MUSIC

I have given the name *pit music* to music that accompanies the image from a nondiegetic position, outside the space and time of the action. The term refers to the'classical opera's orchestra pit. I shall refer as *screen music,* on the other hand, to music arising from a source located directly or indirectly in the space and time of the action, even if this source is a radio or an offscreen musician.

These ideas were developed in chapters on music in *Le Son au cinema.* They correspond to a distinction that has long been noted, with a variety of names. Some say nondiegetic for the first and diegetic for the second, or commentative and actual, or objective and subjective. For music I prefer to rely on terms that simply designate the place where each (supposedly) comes from. A music cue inscribed in the action can of course be just as "commentative" as a nondiegetic music cue, as in Siodmak's *Abschied,* where the protagonists' neighbor is a pianist whose music accompanies and punctuates their emotional states. *Rear Window* conclusively demonstrates this as well.

Once this distinction is established it is relatively simple to describe ambiguous or mixed cases. Consider the case of screen music framed by a pit music cue with ampler orchestration: someone plays a piano in the action, to the accompaniment of the pit orchestra. This occurs in many musicals; an example that comes to mind is in Raoul Walsh's *The King and Four Queens.* In

another kind of case music begins as screen music and continues as pit music by separating from the action. Or, inversely, a grand pit music cue can narrow into screen music being played by an instrument onscreen, for example in older movies, when opening credit music segues into the start of the action.

Not to mention the numerous cases in current films where music established as on-the-air freely circulates between the two levels. In *Taxi Driver* Bernard Herrmann's main theme, heard as pit music throughout much of the film, crops up as the music on a phonograph to which the pimp (Harvey Keitel) and his young hooker (Jodie Foster) dance.

Music as Spatiotemporal Turntable

All music in a film, especially pit music, can function like the spatiotemporal equivalent of a railroad switch. This is to say that music enjoys the status of being a little freer of barriers of time and space than the other sound and visual elements. The latter are obliged to remain clearly defined in their relation to the diegetic space and to a linear and chronological notion of time.

Another way to put it is that music is cinema's *passe-muraille,* capable of instantly communicating with the other elements of the action.[12] For example, it can accompany from the nondiegetic realm a character who is onscreen. Music can swing over from pit to screen at a moment's notice, without in the least throwing into question the integrity of the diegesis, as a voiceover intervening in the action would. No other auditory element can claim this privilege. Out of time and out of space, music communicates with all times and all spaces of a film, even as it leaves them to their separate and distinct existences.

Music can aid characters in crossing great distances and long stretches of time almost instantaneously. This use of music is fair-

ly frequent, ever since the beginning of sound. In King Vidor's *Hallelujah* protagonist Zeke moves through several locales during the singing of one spiritual, "Going Home": a boat on the Mississippi, the roof of a train, a prairie. We can recognize here the embryonic structure of the music video, which, governed by musical form (its only constraint being to include points of synchronization here and there to solder the music and image together) allows the image to wander at will through time and space. In the music video there really no longer exists an audiovisual scene anchored in coherent time and space.

In Vidor's film music gives the characters winged feet; it functions to contract both space and time. In general, however, we can say that music makes space and time pliable, subject to contraction *or* distention. In suspense scenes, it is music that makes us accept the convention of a frozen moment, eternalized by editing.

And in the long confrontations in Sergio Leone's films, where characters do little but pose like statues staring at each other, Ennio Morricone's music is crucial in creating the sense of temporal immobilization. True, Leone also tried to stretch time without the help of music. Notably, at the opening of *Once Upon a Time in the West,* he made do with the occasional creaking of a weather vane or a noria. But there, the plot situation—a long period of waiting and inaction—was chosen to justify the immobility of the characters. At any rate, Leone developed this sort of epic immobility with reference to opera and by generally using music overtly on the soundtrack.

RELATIVE OFFSCREEN SPACE AND ABSOLUTE OFFSCREEN SPACE

The term *offscreen sound* is deceptive; it might lead us to think that the sound itself has some intrinsic quality. We only have to close

our eyes at a film or look away from the screen to register the obvious: without vision, offscreen sounds are just as present—at least as well-defined acoustically speaking—as onscreen sounds. Nothing allows us to tell the two apart. Acousmatized and reduced to an ensemble of sounds that certainly constitute a soundtrack worthy of the name, the film completely changes. I already cited the example of certain scenes of Mr *Hulot's Holiday:* listen to the sounds without the image and they reveal a different character.

Thus sound's "offscreenness," in monaural cinema, is entirely a product of the combination of the visual and aural. It is really a *relation* of»what one hears to what one sees, and exists only in this relation; consequently it requires the simultaneous presence of both elements.

Without the image, the sound of numerous great films of the past is meaningless. In particular, the magical voices that fascinated us would atrophy or become prosaic. The voices of Norman's mother in *Psycho,* Dr. Mabuse in *The Testament of Dr. Mabuse,* or Marguerite Duras in *L'Homme Atlantique* would no longer be extraordinary if they ceased to interact with a screen where they encountered the void of their presence.

Multitrack Cinema's "In-the-Wings Effect" and "Offscreen Trash"

Characteristic of real spatialization and of early multitrack sound film experiments, and generally avoided since then, the "in-the-wings effect" is produced whenever a sound linked to a cause likely to appear onscreen, or which has just exited, lingers in one of the offscreen loudspeakers to one side. Examples are the footsteps of a character approaching or leaving, the engine of a car that has just gone offscreen or that is about to appear, or the voice of a protagonist just out of view.

At these times we have the feeling, which is disconcerting to our normal sense of spectatorship, that we're being encouraged to believe that the audiovisual space is literally being extended into the theater beyond the borders of the screen, and that, over the exit sign or above the door to the restrooms, the characters or cars are *there,* preparing their entrance or completing their exit.

Sometimes this in-the-wings effect cannot be attributed to the direction and mixing of the film, but is simply created by an aberrant placement of speakers in the theater. Sometimes it is indeed due to an attempt by the sound engineers or the director to exploit the effect of *absolute offscreen space,* an effect made possible by multitrack.

Slowly, this practice has been dropped. Sounds of entrances and exits are now rendered with greater discretion and subtlety, or they are opportunely drowned in the sound mix (numerous ambient sounds, music) so as to avoid the sense of the nearby offstage wings.

Certainly, the in-the-wings effect created a nagging problem by violating the conventions of continuity editing and making sound matching problematic. But maybe it could have gained more permanent admittance into film practice had it been systematized along with some partial adjustments in editing conventions—just as the superfield of the multitrack cinema was able to strike a compromise with traditional editing. So perhaps it was a mistake to have given it up so quickly.

The *offscreen trash* is a particular case of passive offscreen space (see below) that results from multitrack sound. It is created when the loudspeakers outside the visual field "collect" noises—whistles, thuds, explosions, crashes—which are the product of a catastrophe or a fall at the center of the image. Action and stunt movies often draw on this effect. Sometimes poetic, sometimes intentionally comic, the "offscreen trash" momentarily gives an

almost physical existence to objects at the very moment they are dying. A modern action movie like John McTiernan's *Die Hard,* a veritable feast of glass-breaking and deflagration taking place in a tower where a man fights terrorists, is filled with such effects.

Active and Passive Offscreen Sound

I shall give the name *active offscreen sound* to acousmatic sound that raises questions—What is this? What is happening?—whose answer lies offscreen and which incite the look to go there and find out. Such sound creates a curiosity that propels the film forward, and it engages the spectator's anticipation: "I'd like to see his face when the other character says that to him." The sounds in active offscreen space necessarily issue from objects that could be identified by sight. Active offscreen sound is used frequently in traditional sound-image editing, bringing objects and characters into a scene by means of sound, then showing them. Films like *Psycho* are based entirely on the curiosity aroused by active offscreen sound: this mother we keep hearing, what does she look like?

Passive offscreen sound, on the other hand, is sound which creates an atmosphere that envelops and stabilizes the image, without in any way inspiring us to look elsewhere or to anticipate seeing its source. Passive offscreen space does not contribute to the dynamics of editing and scene construction—rather the opposite, since it provides the ear a *stable place* (the general mix of a city's sounds), which permits the editing to move around even more freely in space, to include more close shots, and so on, without disorienting the spectator in space. The principal sounds in passive offscreen space are *territory sounds* and *elements of auditory setting.*

Dolby multitrack has naturally favored the development of passive offscreen space over active. Why? The answer may he in the fact that active offscreen space mobilizes identifiable, single

sources—a human body, an object—and in multitrack sound, real (no longer mental) localization of offscreen sound poses the problem of the too realistic "in-the-wings" effect I have already mentioned. If you want to avoid this effect, it is hardly advisable to employ insistent offscreen sounds that pose enigmas and demand to be de-acousmatized, for, logically speaking, this sound should be situated outside the field of the screen. The entrance of Roy Batty, the antagonist in *Blade Runner,* would have been done by the sound of his voice or his footsteps if the film had been recorded in mono. In the actual film this character is almost always present in the image at the same time as his voice. It is as if we were in a perpetual present. In the traditional monaural cinema, on the other hand, offscreen sound demands its resolution from the center of the image, from the very heart of the image, and thus can be called active.

But as early as 1954 *Rear Window* included much passive offscreen sound: city noise, apartment courtyard sounds, and radio, which, full of reverb, cued the ear into the contextual setting of the scene without raising questions or calling for the visualization of their sources.[13]

EXTENSION

Recall the fixed images, like photographs, in Bergman's *Persona:* shots of the park, a hospital wall, and a pile of dirty snow. Over these shots we heard churchbells and no human sound; these created the impression of a small slumbering village.

Let us take away Bergman's sounds and replace them with, say, the sound of the ocean. We see the same pile of snow, the same grillwork, but offscreen space takes on a salty sea smell. If we now remove the ocean sound and instead dub in a crowd of voices and footsteps, the offscreen space becomes a busy street. Nothing prevents us from taking these same images and beginning with a nearby sound (e.g., footsteps in the snow), then bringing in other sounds that suggest a larger space—car sirens—and so on: someone walks by, the siren passes and fades into the distance, faraway churchbells begin to ring. On one static long take we can thus infinitely dilate the offscreen space imagined and evoked by the soundtrack. And we can shrink it just as easily, in which case we will retain the memory of the vast space evoked at the beginning.

Extension of the the sound environment is our designation for the degree of openness and breadth of the concrete space suggested by sounds, beyond the borders of the visual field, and also within the visual field around the characters.

We can speak of *null extension* when the sonic universe has shrunk to the sounds heard by one single character, possibly including any inner voices he or she hears. At the other end of the spectrum we might call *vast extension* the arrangement wherein, for example, for a scene taking place in a room, we not only hear the sounds in the room (including those offscreen) but also sounds out in the hallway, traffic in the street nearby, a siren farther away, and so on. Ambient extension has no absolute limit except those of the universe—if, of course, sounds could ever be found that were capable of maximally dilating the perception of space surrounding the action.

Obviously what is interesting in the cinema is not only the extension that remains the same throughout a scene and even throughout a film but also contrasts and variations in extension from one scene to another, or even within one and the same scene. Sound designer Walter Murch alludes to variation in extension (without using this term) in describing his work on Coppola's *The Conversation* and *Apocalypse Now.*[u]

Dolby stereo, having dramatically increased the possibilities of layering sounds and deploying them in wide concentric spaces,

encourages experimentation with extension. Almost forty years ago *Rear Window*—where everything is seen from a flat in a Greenwich Village courtyard apartment house—made magnificent use of variations in extension. Sometimes it lets us hear the big city thrumming outside this courtyard that the film never leaves. At other times the soundtrack eliminates the larger cityscape entirely, so as to reconcentrate the spectator on the apartment itself, which then becomes for our couple, Grace Kelly and James Stewart, a theater stage cut off from its surroundings. At the very end of the film, the extension becomes extremely narrow, focussing on a single point, like a lone spotlight pursuing a character on a stage—the footsteps' of the killer in the stairway, which Stewart can hear approaching ...

The final scene of *Children of a Lesser God* achieves a similar tightening of spatial extension. As the two estranged lovers reunite in the cool night air, we perceive more and more faintly the noise of a disco dance going on nearby; then it fades down entirely.

Although variations in extension can also consist of sudden contrasts between one scene and the next, they are generally executed in such a way as not to be noticed as a technical manipulation. The occasions on which they *are* made obvious usually contribute toward some emotional effect. This is not like reframings, for example, which are tolerated as technical and coded.

Some films adopt a single fixed strategy for spatial extension and maintain it throughout. In Lang's M extension is generally quite limited. All we hear during a conversation scene is what the characters onscreen are saying; almost never do we hear ambient sounds outside the frame. On the other hand, certain modern films adopt a consistently vast extension: think for example of *Blade Runner,* where rumblings of the city behind characters in the frame constantly remind the viewer of the presence of a huge spatial context.

In fact one of the trickiest things in Dolby stereo films is narrowing extension down to one sound or one point in space, since this necessitates silencing several loudspeakers. The final effect of *Rear Window,* for example, would be very difficult in multitrack sound.

Varying extension to the point of absolute silence is of course used for achieving effects of subjective sound. The suppression of ambient sounds can create the sense that we are entering into the mind of a character absorbed by her or his personal story. A good example occurs in the scene in Bob Fosse's *All That Jazz* where the protagonist has a heart attack.

POINT OF AUDITION

Spatial and Subjective Point of Audition

The notion of a point of audition is a particularly tricky and ambiguous one. Several scholars (Francois Jost in particular) have approached this subject, and I myself devoted a chapter to it in Le *Son au cinema*—where, to tell the truth, I raised more questions than I provided answers. It might be useful to return to it here with greater precision.

Let us first note that critics have come up with the concept of a point of audition based on the model of point of view. Here begins the problem, since cinematic point of view can refer to two different things, not always related:

1. The place from which I the spectator see; from what spatial location the scene is presented—from above, from below, from the ceiling, from inside a refrigerator. This is the strictly *spatial* designation of the term.

2. Which character in the story is (apparently) seeing what I see. This is the *subjective* designation.

In most shots of a modern-day film the camera's point of view is not that of a specific character. Which does not mean that it is necessarily arbitrary: it tends to obey certain laws and constraints. For example, the camera will rarely be located where the eye of a normal human character couldn't be (on the ceiling, in a closet, etc.). Or it only shoots along certain privileged axes, excluding others (e.g., Bergman's *After the Rehearsal,* which takes place on a theater stage, excludes the fourth side, which is the auditorium, the theater seats).

The notion of point of view in this first spatial sense rests on the possibility of inferring fairly precisely the position of an "eye" based on the image's composition and perspective.

Let us recall too that point of view in the subjective sense may be a pure effect of editing. If I cut from a shot of a character looking out the window to a shot of an exterior scene, it is highly likely that the second shot will be perceived as the character's point of view, as long as the information in shot B doesn't contradict anything in shot A.

Now, by comparison, let us examine the notion of a point of audition. This too can have two meanings, not necessarily related:

1. A spatial sense: from where do I hear, from what point in the space represented on the screen or on the soundtrack?
2. A subjective sense: which character, at a given moment of the story, is (apparently) hearing what I hear?

In the first definition, we should start by noting that the specific nature of aural perception prevents us, in most cases, from inferring a point of audition in space based on one or more sounds. This is because of the omnidirectional nature of sound (which, unlike light, travels in many directions) and also of listening (which picks up sounds in the round), as well as of phenomena involving sound reflection.

Consider a violinist playing in the center of a large round room, her audience grouped in various places against the wall. Most of the listeners, even those standing at diametrically opposite points of the room, will hear roughly the same sound, with slight differences in reverberation. These differences, related to the acoustics of the space, are not sufficient to locate specific points of audition. Every *view* of the violinist, on the other hand, can immediately situate the point from which she is being looked at.

So it is not often possible to speak of a point of audition in the sense of a precise position in space, but rather of a place of audition, or even a zone of audition.

In the second, subjective sense of point of audition, we find the same phenomenon as that which operates for vision. It is the *visual* representation of a character in closeup that, in simultaneous association with the hearing of sound, identifies this sound as being heard by the character shown.[15]

The classic example of audiovisual counterpoint cited in Eisenstein's manifesto—the image of a man on lookout duty, and the creaking of a character's boots offscreen—is of the type that is commonplace today. The question is not what characteristics of distance, color, and reverberation *in terms of sound* allow us to infer that the sound is heard by character X. For it is the image that always creates the point of audition, which in this case is worthy of the term *point.*

A special case of point of audition is one defined by sounds that "don't carry," supposedly of such a nature that one must be right up close in order to hear them. Upon hearing these sounds or indices of proximity (e.g., breathing in a voice), the spectator can locate the point of audition as that of a character in the scene—provided of course that the image, the editing and the acting all confirm the spectator's hunch. Phone conversations are the most common example. When the spectator hears the voice of the

unseen person speaking clearly in sound closeup, with its characteristic filtering, she or he can identify the point of audition as being that of the character seen receiving the call. Unless, of course, we are in a situation of on-the-air, which unhooks the sound from its point of departure or arrival and accordingly renders the notion of point of audition no longer pertinent.

Frontal Voice, Back Voice

In some special cases it is nevertheless possible to attribute a direction to what is heard. A sound's high frequencies actually travel in a more directional manner than the low; and when someone speaks to us with his back turned we perceive fewer of the voice's high harmonics and find the voice less present. We can therefore speak of an audible difference between the *frontal voice* and the *back voice.*

In certain films shot in direct sound we can hear variations in a voice's color, due to the fact that a character turns away now and then from the microphone, which is generally above his head. These fluctuations in tone color help to give a particular kind of life to direct sound, and they also function as "materializing indices" (see chapter 5).

Note, however, that, first, there is no law against simulating or reconstituting such variations during postsynchronization, by moving the actor or the mike. (See for example the postsynch of *L'Homme blesseby* Patrice Chereau.) Second, conversely, the mike during shooting can be arranged so as to follow the actor constantly "in front," particularly when the actor wears a lavaliere mike.

If the cinema usually employs the frontal voice, with the most treble allowed by the equipment, it is for a reason: high frequencies are crucial for intelligibility.

However, when the spectator hears a back voice, he or she cannot automatically infer the shot's point of audition from this: for one thing because in most cases this is a momentary effect, not stable and pronounced enough. For another, the spectator does not associate the point of audition with any mental representation of a microphone.

Blind Spots in Theorizing the Mike/Ear

This important question of the "scotomization" of the role of the microphone applies not only to the voice but also more generally to all sounds in a film.[16] And not only to the cinema but equally to most radiophonic, musical, and audiovisual creations that rely on sound recording. The camera, though excluded from the visual field, is nonetheless an active character in films, a character the spectator is aware of; but the mike must remain excluded not only from the visual and auditory field (microphone noises, etc.) but also from the spectator's very *mental representation.* It remains excluded, of course, because everything in movies, including films shot in direct sound, has been designed to this end. This naturalist perspective remains attached to sound, but it is a perspective from which the image—60s and 70s theories on the "transparency" of mise-en-scene notwithstanding—has long been liberated. The naturalist conception of sound continues to infuse real experience and critical discourse so completely that it has remained unnoticed by those who have referred to it and critiqued this same transparency on the level of the image.

We might locate the reasons for such a difference of status between image and sound, in different technical, aesthetic, physiological, and ideological problems, by asking which ones serve as alibis or coverups for which others. We must, for example, explore the implications of the fact that the ears are not disposed

directionally like the eyes. Or the technical possibility, unexplored in the image but utilized in soundtracks since the coming of sound, of *mixing* sounds recorded simultaneously by several mikes placed at different locations: what becomes of the mike/ear then?

But perhaps this is beside the point. After all, the camera has little to do with our eyes (to begin with, it is monocular), which hasn't prevented it from becoming the agent of the look. So the problem lies rather in ways of thinking. To disengage sound thinking and its technical and aesthetic applications from its naturalist rut might well take many years. A concern which lies at the heart of our project.

FIVE

THE REAL AND THE RENDERED

THE ILLUSION OF UNITY

A common perspective to which we made reference in the preceding chapter, which might be called naturalist, postulates that sounds and images start out in "natural harmony." Proponents of this approach seem surprised net to find it working in the cinema; they attribute the lack of this natural audiovisual harmony to technical falsifications in the filmmaking process. If people would only use the sounds recorded during shooting, without trying to improve on them, the argument goes, this unity could be found.

Such is of course rarely the case in reality. Even with so-called