

AUDIO-VISION

• • •

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 Ministère de la Culture in the preparation of the translation.

• • •



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

Copyright © 1994 Columbia University Press
All rights reserved

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.

PN1995.7.C4714 1994

791.43'024—dc20 93-23982

CIP

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

c 10 9 8 7 6 5 4 3 2 1

p 10 9 8 7

CONTENTS

Foreword by Walter Murch vii • Preface xxv

PART ONE • THE AUDIOVISUAL CONTRACT	1
1 PROJECTIONS OF SOUND ON IMAGE	3
2 THE THREE LISTENING MODES	25
3 LINES AND POINTS: HORIZONTAL AND VERTICAL PERSPECTIVES ON AUDIOVISUAL RELATIONS	35
4 THE AUDIOVISUAL SCENE	66
5 THE REAL AND THE RENDERED	95
6 PHANTOM AUDIO-VISION	123
PART TWO • BEYOND SOUNDS AND IMAGES	139
7 SOUND FILM—WORTHY OF THE NAME	141
8 TELEVISION, VIDEO ART, MUSIC VIDEO	157
9 TOWARD AN AUDIOLOGOVISUAL POETICS	169
10 INTRODUCTION TO AUDIOVISUAL ANALYSIS	185

Notes 215 • Glossary 221 • Bibliography 225 • Index 229

ONE

**PROJECTIONS OF
SOUND ON IMAGE**

. . .

The house lights go down and the movie begins. Brutal and enigmatic images appear on the screen: a film projector running, a closeup of the film going through it, terrifying glimpses of animal sacrifices, a nail being driven through a hand. Then, in more “normal” time, a mortuary. Here we see a young boy we take at first to be a corpse like the others, but who turns out to be alive—he moves, he reads a book, he reaches toward the screen surface, and under his hand there seems to form the face of a beautiful woman.

What we have seen so far is the prologue sequence of Bergman’s *Persona*, a film that has been analyzed in books and

university courses by the likes of Raymond Bellour, David Bordwell, Marilyn Johns Blackwell. And the film might go on this way.

Stop! Let us rewind Bergman's film to the beginning and simply *cut out the sound*, try to forget what we've seen before, and watch the film afresh. Now we see something quite different.

First, the shot of the nail impaling the hand: played silent, it turns out to have consisted of three separate shots where we had seen one, because they had been linked by sound. What's more, the nailed hand in silence is abstract, whereas with sound, it is terrifying, real. As for the shots in the mortuary, without the sound of dripping water that connected them together we discover in them a series of stills, parts of isolated human bodies, out of space and time. And the boy's right hand, without the vibrating tone that accompanies and structures its exploring gestures, no longer "forms" the face, but just wanders aimlessly. The entire sequence has lost its rhythm and unity. Could Bergman be an overrated director? Did the sound merely conceal the images' emptiness?

Next let us consider a well-known sequence in Tati's *Monsieur Hulot's Holiday*, where subtle gags on a small bathing beach make us laugh. The vacationers are so amusing in their uptightness, their lack of fun, their anxiety! This time, let's cut out the visuals. Surprise: like the flipside of the image, another film appears that we now "see" with only our ears; there are shouts of children having fun, voices that resonate in an outdoor space, a whole world of play and vitality. It was all there in the sound, and at the same time it wasn't.

Now if we give Bergman back his sounds and Tati his images, everything returns to normal. The nailed hand makes you sick to look at, the boy shapes his faces, the summer vacationers seem quaint and droll, and sounds we didn't especially hear when there was only sound emerge from the image like dialogue balloons in comics.

Only now we have read and heard in a different way.

Is the notion of cinema as the art of the image just an illusion? Of course: how, ultimately, can it be anything else? This book is about precisely this phenomenon of *audiovisual illusion*, an illusion located first and foremost in the heart of the most important of relations between sound and image, as illustrated above with Bergman: what we shall call *added value*.

By *added value* I mean the expressive and informative value with which a sound enriches a given image so as to create the definite impression, in the immediate or remembered experience one has of it, that this information or expression "naturally" comes from what is seen, and is already contained in the image itself. Added value is what gives the (eminently incorrect) impression that sound is unnecessary, that sound merely duplicates a meaning which in reality it brings about, either all on its own or by discrepancies between it and the image.

The phenomenon of added value is especially at work in the case of sound/image synchronism, via the principle of *synchresis* (see chapter 3), the forging of an immediate and necessary relationship between something one sees and something one hears. Most falls, blows, and explosions on the screen, simulated to some extent or created from the impact of nonresistant materials, only take on consistency and materiality through sound. But first, at the most basic level, added value is that of text, or language, on image.

Why speak of language so early on? Because the cinema is a vococentric or, more precisely, a verbocentric phenomenon.

VALUE ADDED BY TEXT¹

In stating that sound in the cinema is primarily vococentric, I mean that it almost always privileges the voice, highlighting and setting the latter off from other sounds. During filming it is the voice that is collected in sound recording—which therefore is

almost always voice recording—and it is the voice that is isolated in the sound mix like a solo instrument—for which the other sounds (music and noise) are merely the accompaniment. By the same token, the historical development of synch sound recording technology, for example, the invention of new kinds of microphones and sound systems, has concentrated essentially on speech since of course we are not talking about the voice of shouts and moans, but the voice as medium of verbal expression. And in voice recording what is sought is not so much acoustical fidelity to original timbre, as the guarantee of effortless intelligibility of the words spoken. Thus what we mean by vococentrism is almost always verbocentrism.

Sound in film is voco- and verbocentric, above all, because human beings in their habitual behavior are as well. When in any given sound environment you hear voices, those voices capture and focus your attention before any other sound (wind blowing, music, traffic). Only afterward, if you know very well who is speaking and what they're talking about, might you turn your attention from the voices to the rest of the sounds you hear. So if these voices speak in an accessible language, you will first seek the meaning of the words, moving on to interpret the other sounds only when your interest in meaning has been satisfied.

Text Structures Vision

An eloquent example that I often draw on in my classes to demonstrate value added by text is a TV broadcast from 1984, a transmission of an air show in England, anchored from a French studio for French audiences by our own Léon Zitrone². Visibly thrown by these images coming to him on the wire with no explanation and in no special order, the valiant anchor nevertheless does his job as well as he can. At a certain point, he affirms, "Here

are three small airplanes," as we see an image with, yes, three little airplanes against a blue sky, and the outrageous redundancy never fails to provoke laughter.

Zitrone could just as well have said, "The weather is magnificent today," and that's what we would have seen in the image, where there are in fact no clouds. Or: "The first two planes are ahead of the third," and then everyone would have seen *that*. Or else: "Where did the fourth plane go?"—and the fourth airplane's absence, this plane hopping out of Zitrone's hat by the sheer power of the Word, would have jumped to our eyes. In short, the anchor could have made fifty other "redundant" comments; but their redundancy is illusory, since in each case these statements would have guided and structured our vision so that we would have seen them "naturally" in the image.

The weakness of Chris Marker's famous demonstration in his documentary *Letter from Siberia*—already critiqued by Pascal Bonitzer in another context³—where Marker dubs voiceovers of different political persuasions (Stalinist, anti-Stalinist, etc.) over the same sequence of innocuous images, is that through his exaggerated examples he leads us to believe that the issue is solely one of political ideology, and that otherwise there exists some neutral way of speaking. The added value that words bring to the image goes far beyond the simple situation of a political opinion slapped onto images; added value engages the very structuring of vision—by rigorously framing it. In any case, the evanescent film image does not give us much time to look, unlike a painting on a wall or a photograph in a book that we can explore at our own pace and more easily detach from their captions or their commentary.

Thus if the film or TV image seems to "speak" for itself, it is actually a ventriloquist's speech. When the shot of the three small airplanes in a blue sky declares "three small airplanes," it is a puppet animated by the anchorman's voice.

VALUE ADDED BY MUSIC

Empathetic and Anempathetic Effects

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

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

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

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

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

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

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

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

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

displacement or agitation, however minimal. Sound does have means to suggest stasis, but only in limited cases. One could say that "fixed sound" is that which entails no variations whatever as it is heard. This characteristic is only found in certain sounds of artificial origin: a telephone dial tone, or the hum of a speaker. Torrents and waterfalls can produce a rumbling close to white noise too, but it is rare not to hear at least some trace of irregularity and motion. The effect of a fixed sound can also be created by taking a variation or evolution and infinitely repeating it in a loop. As the trace of a movement or a trajectory, sound thus has its own temporal dynamic.

Difference in Speed of Perception

Sound perception and visual perception have their own average pace by their very nature; basically, the ear analyzes, processes, and synthesizes faster than the eye. Take a rapid visual movement—a hand gesture—and compare it to an abrupt sound trajectory of the same duration. The fast visual movement will not form a distinct figure, its trajectory will not enter the memory in a precise picture. In the same length of time the sound trajectory will succeed in outlining a clear and definite form, individuated, recognizable, distinguishable from others.

This is not a matter of attention. We might watch the shot of visual movement ten times attentively (say, a character making a complicated arm gesture), and still not be able to discern its line clearly. Listen ten times to the rapid sound sequence, and your perception of it will be confirmed with more and more precision.

There are several reasons for this. First, for hearing individuals, sound is the vehicle of language, and a spoken sentence makes the ear work very quickly; by comparison, reading with the eyes is notably slower, except in specific cases of special train-

ing, as for deaf people. The eye perceives more slowly because it has more to do all at once; it must explore in space as well as follow along in time. The ear isolates a detail of its auditory field and it follows this point or line in time. (If the sound at hand is a familiar piece of music, however, the listener's auditory attention strays more easily from the temporal thread to explore spatially.) So, overall, in a first contact with an audiovisual message, the eye is more spatially adept, and the ear more temporally adept.

Sound for "Spotting" Visual Movements and for Sleight-of-Hand

In the course of audio-viewing a sound film, the spectator does not note these different speeds of cognition as such, because added value intervenes. Why, for example, don't the myriad rapid visual movements in kung fu or special effects movies create a confusing impression? The answer is that they are "spotted" by rapid auditory punctuation, in the form of whistles, shouts, bangs, and tinkling that mark certain moments and leave a strong audiovisual memory.

Silent films already had a certain predilection for rapid montages of events. But in its montage sequences the silent cinema was careful to simplify the image to the maximum; that is, it limited exploratory perception in space so as to facilitate perception in time. This meant a highly stylized visual mode analogous to rough sketches. Eisenstein's *The General Line* provides an excellent example with its closeups in the cream separator sequence.

If the sound cinema often has complex and fleeting movements issuing from the heart of a frame teeming with characters and other visual details, this is because the sound superimposed onto the image is capable of directing our attention to a particular visual trajectory. Sound even raises the possibility of sleight-of-

hand effects: sometimes it succeeds in making us see in the image a rapid movement that isn't even there.

We find an eloquent example in the work of sound designer Ben Burtt on the *Star Wars* saga. Burtt had devised, as a sound effect for an automatic door opening (think of the hexagonal or diamond-shaped automatic doors of sci-fi films), a dynamic and convincing pneumatic "shhh" sound. So convincing, in fact, that, in making *The Empire Strikes Back*, when director Irving Kershner needed a door-closing effect he sometimes simply took a static shot of the closed door and followed it with a shot of the door open. As a result of sound editing, with Ben Burtt's "pssst," spectators who have nothing before their eyes besides a straight cut nevertheless think they see the door slide open. Added value is working full steam here, in accordance with a phenomenon specific to sound film that we might call faster-than-the-eye.

Deaf people raised on sign language apparently develop a special ability to read and structure rapid visual phenomena. This raises the question whether the deaf mobilize the same regions at the center of the brain as hearing people do for sound—one of the many phenomena that lead us to question received wisdom about distinctions between the categories of sound and image.

The Ear's Temporal Threshold

Further, we need to correct the formulation that hearing occurs in continuity. The ear in fact listens in brief slices, and what it perceives and remembers *already* consists in short syntheses of two or three seconds of the sound as it evolves. However, within these two or three seconds, which are perceived as a gestalt, the ear, or rather the ear-brain system, has minutely and seriously done its

investigation such that its overall report of the event, delivered periodically, is crammed with the precise and specific data that have been gathered.

This results in a paradox: we don't hear sounds, in the sense of recognizing them, until shortly after we have perceived them. Clap your hands sharply and listen to the resulting sound. Hearing—namely the synthesized apprehension of a small fragment of the auditory event, consigned to memory—will *follow* the event very closely, it will not be totally simultaneous with it.

INFLUENCE OF SOUND ON THE PERCEPTION OF TIME IN THE IMAGE

Three Aspects of Temporalization

One of the most important effects of added value relates to the *perception of time in the image*, upon which sound can exert considerable influence. An extreme example, as we have seen, is found in the prologue sequence of *Persona*, where atemporal static shots are inscribed into a time continuum via the sounds of dripping water and footsteps. Sound temporalizes images in three ways.

The first is temporal animation of the image. To varying degrees, sound renders the perception of time in the image as exact, detailed, immediate, concrete—or vague, fluctuating, broad.

Second, sound endows shots with temporal linearization. In the silent cinema, shots do not always indicate temporal succession, wherein what happens in shot B would necessarily follow what is shown in shot A. But synchronous sound does impose a sense of succession.

Third, sound *vectorizes* or dramatizes shots, orienting them toward a future, a goal, and creation of a feeling of imminence