D., *Listening and Voice: A Phenomenology of Sound*, Athens, OH: Ohio University Press, 1976.
eau-Ponty, M., *Phenomenology of Perception*, London: Routledge and Kegan Paul, 1962.
lleton, R., *Studying Popular Music*, Milton Keynes: Open University Press, 1990.
hell, T., *Popular Music and Local Identity: Rock, Pop and Rap in Europe and Oceania*, London: Leicester University Press, 1996.
fer, R. M., *The Tuning of the World*, New York: Knopf, 1977.
effelin, B. B., *The Give and Take of Everyday Life: Language Socialization of Kaluli Children*, Cambridge: Cambridge University Press, 1990.
effelin, E. L., *The Sorrow of the Lonely and the Burning of the Dancers*, New York: St Martin's Press, 1976.
ma, A., 'Sounds oriental: the (im)possibility of theorizing Asian musical cultures', in *Dis-orienting Rhythms: The Politics of the New Asian Dance Music*, ed. S. Sharma, J. Hutnyk and A. Sharma, pp. 15–31, London: Zed Books, 1996.
herd, J., *Music as Social Text*, Cambridge: Polity, 1991.
eney, P., *The Virgin Directory of World Music*, New York: Henry Holt, 1991.
or, T., *Global Pop: World Music, World Markets*, New York: Routledge, 1997.

9 Audio-Vision and Sound

MICHEL CHION

from Sound, P. Kruth & H. Stobart, eds., 2000.

Although sound cinema has existed for several decades (since about the late twenties and early thirties), theoreticians and historians who have given the issue of sound on screen the importance it deserves are still few and far between. I can cite the remarkable articles by Rick Altman in the USA, or Claude Bailblé in France, but I should also mention my own works. Indeed, in parallel with my activities as a composer and a filmmaker, I have devoted the last 20 years or so to the study and teaching of the relationships between sound and image. In this chapter, I propose to give an overview of my personal theorization which has led me to forge a whole new vocabulary.

My approach does not involve studying the sound of films or television programs in isolation from the image. Rather, as I have demonstrated in numerous analyses and experiments, film sound cannot be studied separately from its image, and vice versa. It is their very combination which produces something entirely new and specific, in the same way as a chord or an interval in music.

Since this field of study is new and as yet 'non-coded', I have had to coin numerous expressions to refer to audiovisual effects that have been known and used for a long time, but intuitively, as 'figures' deprived of specific names. This is why I still use the term 'effects', which has lost some of its meaning today, whereas in the past it was commonly found in articles and reviews on opera, theatre or music – in short those arts which, like cinema, are based on performance.

The question is whether these effects make up, or will make up, a rhetorical

Translated from the French by Patricia Kruth and David H. Jones. The terminology used in this chapter – when already existing in English – is from *Audio-Vision: Sound on Screen* translated by Claudia Gorbman.

Although originally written for this volume, this chapter has also appeared in French in Michel Chion's book *Le Son* (Nathan, Paris) in 1998.

system consciously known to the public, or whether they will remain 'effects'. But this is no different from musical effects, such as harmonic progressions or diminished seventh chords, which have long been received by audiences who could feel them without being able to identify and understand them. This is still true for most music listeners today.

I should stress that I work within a descriptive logic where it is never a case of all or nothing, and where exceptions do not disprove the rule.

Audio-vision, added value, illusion of redundancy

I have named ***audio-vision*** the perceptive process by which sound in cinema, television, and video modifies and influences the perception of what is seen. Indeed, audiovisual combination does not work as an addition of similar or opposed components but as a mixture in which sound is rarely taken into account. (In a similar way, to continue the musical comparison suggested above, listeners who have no formal musical training will not be able to distinguish between the overall emotion created by the melody, and the chords that 'accompany' it. They may therefore attribute solely to the melodic line – and in the case of audio-vision, to the image – the emotion or meaning that actually derives from the combination and association of all the musical components.)

Very frequently, when sound adds meaning to the image, the meaning seems to emanate from the image itself. This is what I call ***added value*** (i.e. value added by sound to the image). Added value is at the basis of most 'audiovisual effects', and may be defined as a sort of simultaneous 'Kuleshov effect' between sound and image. (In cinema this effect is named after the Russian director who, in an experiment, juxtaposed the same shot of an actor with a neutral expression and shots of other subjects like a baby, a bowl of soup, a dead body, and the like. Each time the actor's face seemed to register the appropriate emotion.) This value – be it sensorial, informative, semantic, narrative, structural, or expressive – which a sound heard in a scene leads us to project onto the image, can create the impression that we view what in fact we 'audio-view'. Added value is a widely used effect, most of the time experienced unconsciously. To become aware of it and analyze its mechanism, it is necessary to separate out the audiovisual mix by observing the sound and the image of a given sequence independently. Only then do we appreciate how, in different ways, sound never ceases to influence what we see.

Added value is partly bilateral – the image likewise influences our perception of sound. Yet, because of the conscious focusing of the spectator of a film or a television program towards the screen and what is visible, it is ultimately onto the image that the overall product of the mutual influences between sound and image is most often reprojected. On the other hand, in a cultural situation such as a concert – what I call a situation of 'visu-audition' – where conscious attention is projected by cultural tradition onto listening, added value functions mainly the other way round. For example, if we see a player make a vigorous gesture we will 'hear' a more powerful sound.

To come back to cinema, is it appropriate to speak of *audiovisual* effects? These effects do indeed have an audiovisual *cause*, but the result of the combination does not consist in perceptions of sounds and images as such, but rather in perceptions of space, matter, volume, meaning, expression, and organization of space and time. This is why I prefer to speak of ***audiovisiogenic effects***, i.e. effects generated by sounds and images. The peculiarity of these effects, as mentioned, is not to be detectable as such, but rather for the most part to create the illusion that sound only duplicates what the image would already say 'by itself'. The relationship between the audio and the visual is therefore based on a fundamental misperception: the belief in a ***redundancy*** between sound and image.

The most banal and seemingly least questionable example of redundancy one might think of is a dialogue in a film; but it is precisely not a redundancy. Indeed, sound, as a general rule, cannot be inferred from the image, and neither can a spoken text be deduced from what is seen, or just barely, except for a deaf person trained to lip-read (and besides, only in the original language with the actors facing the audience!). Symmetrically, the characters' faces, the way they are dressed and move, or the setting where they evolve can only rarely be deduced from the sound alone. Audiovisual redundancy is therefore impossible.

Sound/image: a skewed symmetry

My theory of audio-vision in the cinema contrasts with the lazily symmetrical model prevalent in a number of film courses, which describes sound on the one hand, and image on the other. On the contrary, audio-vision rests on a dissym-

metrical model of description – a model in which sound and image are not two complementary and well-balanced elements. I have demonstrated how the (visible) frame of the image is also the frame in relation to which sound locates itself in space, and onto which sounds project their effects.

If sounds are easily projected by the spectator onto the film image, it is because the image is circumscribed by a frame that can be located in space, whereas sound lacks a frame. The visual frame is therefore the support of a double projection on the 'audio-spectator's' part – the projection of images (since he or she reprojects onto one image the previous ones in the film), and that of sounds.

Cinema indeed depends on the principle of *a visual frame for images*, an all but unique frame that pre-exists their random, sweeping succession. At the same time, it is the frame that allows one to speak of 'the image' in the singular, since images never extend beyond the frame. On the other hand, it can be said that *there is no sound frame for sounds*. Possibly, sounds are only framed by the image which grounds them (through spatial magnetization, see below), anchors and binds them – or not – to an object defined in space; conversely, if sounds are not incorporated into the image, they are made to exist on another invisible stage or in a contiguous (offscreen) space. Furthermore, contrary to the image that is enclosed within a frame, film sounds can be layered on top of one another without any limit of quantity or complexity, and they are free from all laws of realism. Film music, voice-over narration, dialogues or realistic atmospheric noises can all be mixed.

The absence of a sound frame is one of the main reasons which, for a long time, has led me to assert that *there is no soundtrack*. By this I mean that the different sounds which are present in a film (words, noises, diverse musics and sounds) and contribute to its meaning, its shape and its effects do not by themselves, by the sheer virtue of their all being sound elements, make up a comprehensive entity that is interdependent and homogeneous. In other words, in the cinema the relations of meaning, contrast, concordance or divergence that words, noises and musical elements are likely to entertain with one another are much weaker, even non-existent, in comparison with the relations each of the sound elements, on its own, has with a given visual or narrative element present simultaneously in the image. I refer you to my previous books on sound for detailed demonstrations of this assertion which allows few exceptions (see Further reading).

How audiovisiogenic effects originate

Audiovisual relationships are largely cultural and historical but, in everyday life as well as in the audiovisual arts, they rely also on relatively little-known universal psycho-physiological phenomena. (This is probably due to the increasing specialization of scholars, which has led them to pay less attention to the connections between the senses and focus their study on one of them.) First among these phenomena is the effect of ***synchresis***.

'Synchresis' – a Lewis Carroll-style word I have forged (from synchronism and synthesis) – is the name I give to a spontaneous and reflex psycho-physiological phenomenon that depends on our nervous and muscular connections. It consists in perceiving as one and the same phenomenon – which manifests itself both on the visual and sound levels – the 'concomitance' of a precise sound event with a precise visual event on the sole and only condition that they happen simultaneously.

Through this phenomenon, which literally cannot be controlled, we are led instantaneously to establish a tight link of interdependence between sounds and images that are often quite unrelated in reality, and to assign them a common origin even if they are of completely different natures and sources. Synchresis therefore allows for the use of almost any sound effect for the footsteps of a character on the screen, in total freedom of expression.

Synchresis also permits effects based on contradiction and discrepancy (like a disproportion between voice and body in cartoons, or a gender inversion in certain comic or fantastical stories), and without it the 'audio' would purely and simply break away from the 'visual'. In brief, without synchresis, sound would have to mimic reality and its range of possibilities of expression would be much smaller. (It should be specified that film sound bears only a very remote resemblance to the sound in real situations.)

The word 'synchresis' is possibly ambiguous; it is not really a synthesis in the sense that no difference is 'transcended' or resolved; the image remains the image, and the sound remains the sound; what they have come to represent exists beyond them, like a projected shadow. If there is such a thing as an ***audio-image*** – an expression I occasionally use – then, *it is not the image which is on the screen.* It is a mental image, like the space created in a mise-en-scène through cutting and editing.

The second psycho-physiological condition, which is universal (i.e.

non-cultural) and permits audiovisual relations, is what I call ***spatial magnetization***, i.e. magnetization of sound by the image. This is the process whereby when we visually locate a sound source (a human being, an animal, a machine, an object, etc.) in a certain place in space, and when for diverse reasons (the sound is electrically amplified, it bounces off the walls . . .) the associated sound comes mainly from another direction, we can still 'hear' the sound come from what we see as its source. Consequently, during the projection of a film on an airplane, the sound of the actors' voices seems to come from the screen whereas we hear the sound through earphones. An important counter-example is when sound really originates from different sources in space (for instance coming out of one loudspeaker, and then out of the other in the case of Dolby cinema) and therefore, for psycho-physiological reasons, our attention is reminded of its real acoustic location. Spatial magnetization has made possible classical talking pictures in which we accept that in mono sound the characters' voices do not really move about, especially when we see the characters walk across the screen. In the same way, sounds which are located 'offscreen' are only mentally so, in the minds of spectators who project onto the sound of a scene movements they have witnessed with their eyes. (This is a case of 'added value' in which value is added by the image to sound.)

Spatial magnetization works all the better when sounds are synchronized with images, and in many cases it implies synchresis. In the case of auditoriums equipped with Dolby, and so with multiple tracks, reflex spatial magnetization may be strengthened or on the contrary weakened by the *real* sources of the sound that is broadcast, depending on the positioning of the loudspeakers (i.e. whether they are more or less distant from one another and located outside the axis of the screen) and the position of the spectators themselves in the auditorium.

Why conscious hearing attention is hierarchical

The analysis of audiovisual relations must also take into account the fact that a human being's conscious hearing attention is not directed indifferently towards all types of sound. It is structured and hierarchical, and in particular ***voice-centered***.

I call ***voice-centering*** the process by which, in a sound environment, the voice

attracts and centers our attention, in the same way as the human face in the image of a film. Voice-centering can be obliterated or toned down by specific procedures: this is what happens, for example, in the films of Jacques Tati. The director introduces fluctuations in the sound level and the intelligibility of the text, while he also carefully establishes that the dialogues are not essential to the action proper, and at the same time, of course, puts the characters at a distance with his camera; these are all devices destined to 'prevent' our attention from focusing on the voices.

This does not mean that in the classical voice-centered cinema other sounds, noises and music, are 'unimportant'. On the contrary their role is as important, only at a less conscious level, just like the 'inner parts' (those of the tenor and alto which sound neither above nor below) in a string quartet or mixed four-voice choir. It is only when these parts are missing or different that one can feel that 'something has changed', even though the melody to which one consciously pays attention remains the same.

In sound cinema, voice is also the main, if not the exclusive, vehicle for the text. I have therefore suggested the term ***audio(logo)visual*** – instead of 'audiovisual' – to highlight the fact that most of the time in cinema the presence of language is central. It is a determining and privileged component, whether as a written text (intertitles in silent films, titles and subtitles in talking pictures, etc.) or as an oral text (dialogues, interior monologues, voice-over, etc.), and in these different forms language can determine, regulate and justify the overall structure of a film. By using the term audio(logo)visual one can avoid reducing cinema to a mere question of 'sounds' and 'images'.

Words, indeed, are not only the center of conscious attention but also frequently the key to audiovisual structuring; in some cases they even completely guide and organize the other elements around them. *Le Roman d'un tricheur* (1936) by Sacha Guitry uses this effect in a particularly conspicuous way. This film, which was very much admired by Orson Welles and the French New Wave, is indeed 'told' from beginning to end by a voice-over narrator-protagonist who 'comments' on the pictures, and even interprets the different voices of the characters on the screen. In the case of classical films with dialogues – which we call 'voice-centered' – this is more insidious and implicit. The whole film is then conceived and structured so as to justify and help the hearing of dialogues and to treat them as action, while at the same time the perception of the dialogue as such is obliterated. The – willing – spectator of classical voice-

ıtered films does not realize that he is really listening to a flow of dialogues ɔund which everything is organized; he is convinced that he is witnessing a nplex action whose dialogues make up only what he considers to be an nost negligible part. This is what happens in the films of Hitchcock.

Other films, on the contrary, can be called ***logo-decentered.*** They correspond the apparently paradoxical cases where dialogues are abundant and import- t but instead of being 'dissimulated' or 'absorbed' by the mise-en-scène, their undance is perceived as such because the other filmic elements do not courage us to listen to them. Logo-decentered films range from the works of llini, where multilingual dialogues abound, to those of Tarkovsky, where the aracters chatter away as they are confronted with their lack of power in the ːe of life and nature's mysteries. Also belonging to this category are cases ıere the audiovisual style renders speech relative and treats it like one noise ıong others. As already mentioned, this is the case in the films of Tati.

Effects of rendering and matter

ːt me start with the notion of ***rendering***, which I defined in my book *Audio-ːsion* (Figures 1 and 5). We can speak of rendering when 'the film spectator cognizes sounds to be truthful, effective, and fitting not so much if they *repro-ɹce* what would be heard in the same situation in reality, but if they *render* (i.e. nvey, express) the feelings associated with the situation'. The use of sound as means of rendering (and not of reproduction) is facilitated because of the most endless range of causes with which it can be associated. In other words, und is easily justifiable, or, if you prefer, the spectator is very tolerant of the ct that a sound does not resemble what one would hear in reality; as I have ɜmonstrated in other writings (e.g. *Le Promeneur écoutant*), there is no rigid w linking a sound to its cause(s).

Examples of 'rendering', i.e. of a sound which does not express another sound ɹt rather a sense of speed or strength, would include the sound effects that ınctuate action scenes (like the hissing of various swords, which conveys agil- y in kung fu movies), the noises of falling bodies that impart the violence ıflicted upon the characters (whereas the same fall in real life might not make ıy noise) or the sound of blows in boxing films. Other examples include ɔunds meant to give an impression of matter or non-materiality, of fragility or

FIGURE 1. Jean-Jacques Annaud's *L'Ours* (*The Bear*, 1988). The sound of the animal's footsteps, created in a studio, are not aimed at reproducing the often faint noises made by a bear walking on the soles of its feet on the ground, but at *rendering* its massive power.

resistance, of hollowness or plenitude, of weight or lightness, of having been used or being brand new, of luxury or misery. This is why sounds are generated – rather than to reproduce the real sound of a given object or character. *A rendering is always the rendering of something.* Besides, we must not forget that the issue of rendering is part of an audiovisual context. The effect of rendering is thus projected onto the image and falsely perceived as directly expressed by it (hence an illusion of redundancy).

Another means of audiovisual expression, dealing in this case with perception of matter, is the variable use of ***materializing sound indices (m.s.i.s).*** This expression refers to any aspect of a sound which reflects with more or less precision the material nature of its source and the concrete history of its production. M.s.i.s will reveal the nature of the source – which may be solid (woody, metallic), ethereal, gooey, grainy, liquid – its material consistency, the accidents occurring while it progresses, and so on. A sound may contain a certain number of materializing sound indices, or ultimately none at all. These m.s.i.s often consist of *unevennesses*, or slight or more pronounced irregularities, which reveal the material conditions of the sound source. In a specific listening situation such as a film or a piece of 'musique concrète', a voice, footsteps or a note may contain a variable number of materializing indices. These include throat clearing and breathing noises in the sound of a voice; crunching, squeaking or hissing noises in footsteps; slight accidents in the attack, the resonance or the tuning of a musical sequence caused by an out-of-tune piano or instruments that do not attack together. In the sound conception of a film, m.s.i.s are variously apportioned, particularly through the production of sound effects that either totally erase them (which creates an abstract, dematerialized universe) or on the contrary accentuate them (which foregrounds matter and bodies) with every possible combination in between. They are all important means of rendering in the cinema.

Materializing indices do not only concern noises but also play a part in the sound of dialogues. Voices in films may be more or less 'materialized' by details such as slight oral clicks, breathing noises between sentences and words, coughing or hoarse voices, or on the contrary be more or less dematerialized. Voice-over narration, especially, is often intentionally purified of its m.s.i.s so as not to attract attention to the physical body emitting the sound.

Scenographic effects

Other audiovisiogenic effects contribute to what I call ***audiovisual scenography.*** By this I mean everything that, in a combination of sounds and images, has to do with the construction of a narrative scene. This is done especially through ways of entering and exiting the sound frame (characters and vehicles entering and exiting the visual frame, and announced or followed by sound), through the contrast or identity between 'extension of sound space' (see below) and visual framing, through the comparison of the sizes of the screen characters and, acoustically, through the proximity or distance of their voices and more generally of the sounds they produce. It can be stressed right away that only rarely does 'sound perspective' reproduce and strictly redouble 'visual perspective', or if it happens it is in an approximate and tentative manner.

Here are two examples of audiovisual scenography borrowed from well-known films. In Ridley Scott's *Blade Runner* (1982), the setting and the characters are often shown in close shots while their surroundings are acoustically portrayed or suggested by sounds that evoke wide open spaces. This results in a sort of complementarity and compensation between vision in close-up and hearing in long shot. In Fellini's *Satyricon* (1969), on the contrary, several scenes combine visual scenography based on an empty space and the decentering of characters (often shown in long shot and at the bottom of a gigantic cinemascope screen) with fantasy sound scenography: the voices of the characters in question, which are not at the same distance as the bodies who 'emit' them, have an invading intimacy and speak into our ears as if in a dream.

Extension of sound space is one of the effects concerned with construction of space through the combination of sounds and images (Figure 2). By this I mean the concrete, more or less wide open space that the sonic environment defines beyond the screen, which constructs the geographical, human and natural framework from which the image on the screen is extracted. Let us consider the case where the setting of the action is confined to the inside of an apartment from which the camera does not escape. Extension of sound space will be *restricted* if the sound heard are only those made within the apartment; it will be *wider* if one can hear noises coming from the landing and the adjoining apartments offscreen, even wider if street noises are included, and wider still if distant noises such as foghorns or train whistles are perceived. All these choices are left to the director and the sound editor and depend on the scene and what

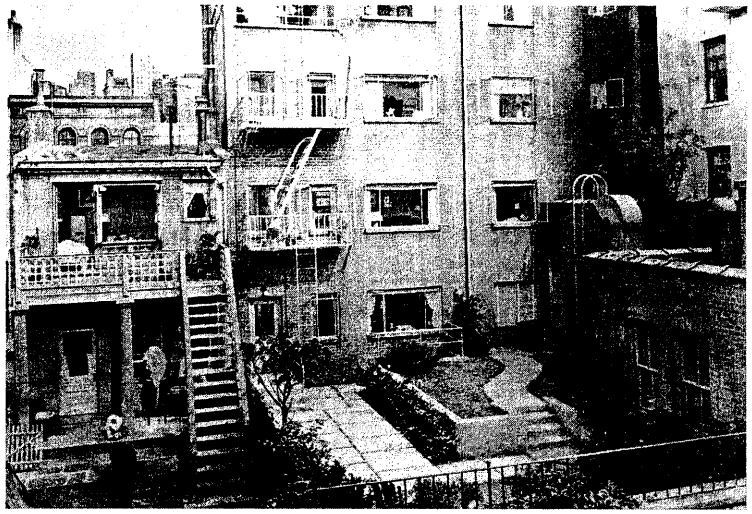

FIGURE 2. The courtyard set of Alfred Hitchcock's *Rear Window* (1954). Through sound, in this famous film, Hitchcock freely plays with *extension*. He focuses our attention onto the courtyard, the street, other apartments, or on the contrary makes us forget all that is not the inside of James Stewart's character's apartment.

eds to be expressed. All the above possibilities will be accepted as 'natural' the viewer. Extensions of an interior set will for instance make us feel 'natu- '(with the wind blowing outside the house in John Huston's *Key Largo* [1948]) 'solitude' (through a dog barking in the distance), or on the contrary crowds d promiscuity (with the noises of a Vietnamese street in the love scenes of an-Jacques Annaud's *The Lover* [1992]); they will also direct the attention of e spectator or of a character, and create an effect of meaning or contrast.

In this chapter, I can mention only briefly the contributions of Dolby to diovisual scenography, and I would particularly like to stress the notion of **perfield** which I developed in *Audio-Vision*. 'I call "superfield" the space eated, in multitrack films, by ambient natural sounds, city noises, music, and l sorts of rustlings that surround the visual space and that can issue from udspeakers outside the physical boundaries of the screen.' (This is the case for stance in *Blade Runner*.) By virtue of its acoustic precision and relative stab- ty, this ensemble of sounds takes on a kind of quasi-autonomous existence in

FIGURE 3. François Périer and Giulietta Masina in Federico Fellini's *Le notti di Cabiria* (1957). The *suspension* of natural sounds in this idyllic scene gives us a premonition that the couple's happiness is just an illusion.

relation to the visual field, in that it does not depend moment by moment on what we see onscreen, but neither does it acquire the autonomy and pregnancy of sound relations amongst themselves which would justify speaking of a *soundtrack*.

In a fictional scene in which our audiovisual habits call for (urban or natural) ambient noises, **suspension** is the dramatic audiovisiogenic effect which consists of interrupting these noises or even eliminating them from the start, while the causes of these sounds are still present in the action and even in the image. This often results in a feeling of mystery or threat, and sometimes in a sort of poetic suspension where the world loses some of its reality. For example, at the end of Fellini's *The Nights of Cabiria* (1957) during the romantic walk through the scenery of an enchanted wood, no bird song can be heard and the atmosphere is one of the impending doom. We soon learn that the man who has taken Cabiria to the edge of the cliff wants to kill her (Figure 3).

Effects concerning time and phrasing

After space, I now turn to time. The question of time is often neglected in the study of cinema and I have taken a particular interest in the way in which sound – by definition a temporal element – contributes to its construction, in general through added value.

One can call ***audiovisual phrasing*** everything that in a film sequence concerns the construction of time and rhythm through devices including phrasing, punctuations and pauses, freeze frames, anticipation and release. Sound is an important means of audiovisual phrasing because in editing one can easily incorporate sudden punctuating sounds (a honking horn, the cry of an animal) or sounds evolving in time (a passing car, gusts of wind) which help to cut up, enliven and build time. One can also speak of phrasing because sound is an important means of *temporalization.*

Temporalization is an audiovisiogenic effect – a case of added value – in which sound either influences and 'contaminates' the duration which images already have, or endows images with a duration they do not have in themselves e.g. the completely static shots in Chris Marker's *La Jetée* (1962) or shots of an empty scene or motionless characters). Sound can, in particular, impose a linear and chronological succession to a sequence of images whose relationship does not presuppose temporal succession; this is what I have called ***linearization*** of images through sound. For example, at the beginning of *Citizen Kane* (Orson Welles, 1941), the first shots of Xanadu are *linearized* by Bernard Herrmann's music, and a similar effect is achieved at the end of *L'Eclisse* (Michelangelo Antonioni, 1962) by the music of Giovanni Fusco. Sound can also 'vectorize' individual shots, i.e. orient them in time by imparting a sense of expectation, of progression, of a movement forward or an imminent action that they do not convey by themselves; this is what I have called ***vectorization.*** A good example is the sound crescendo, from pianissimo to fortissimo, created by Bergman over some of the shots in the silent prologue sequence of *Persona* (1966). I refer you to my analysis of this sequence in *Audio-Vision: Sound on Screen.* Also, some of the static shots showing no movement in the sequence of the crop-duster attack in Hitchcock's *North by Northwest* (1959) are *vectorized* by the crescendo or decrescendo of the airplane's engine (Figure 4).

Audiovisual phrasing in the cinema also relies on the existence of what I have called 'points of synchronization'. A ***point of synchronization***, or ***synch point***,

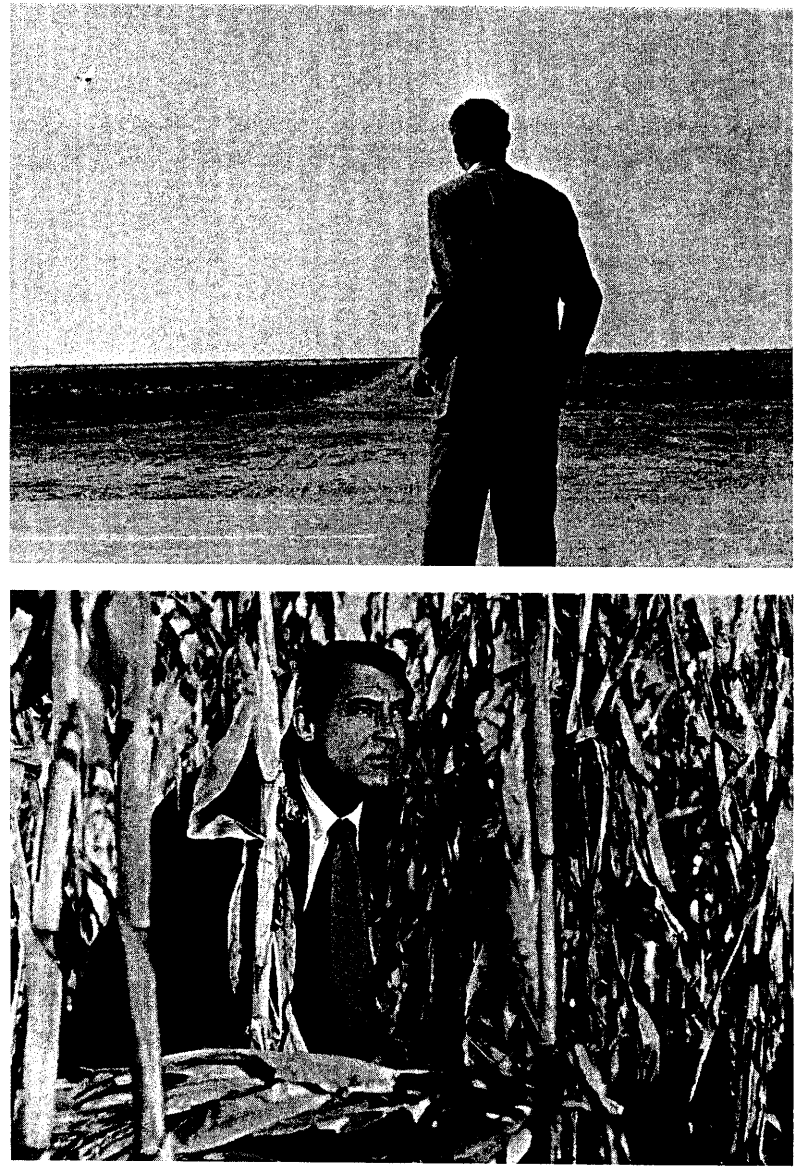

FIGURE 4. Two photograms of a scene from Alfred Hitchcock's *North by Northwest* (1959). In this famous scene, the plane attacking Cary Grant is not always visible, but the sound of the plane, which keeps evolving, coming and going, creates a sense of time and suspense even in the most static shots.

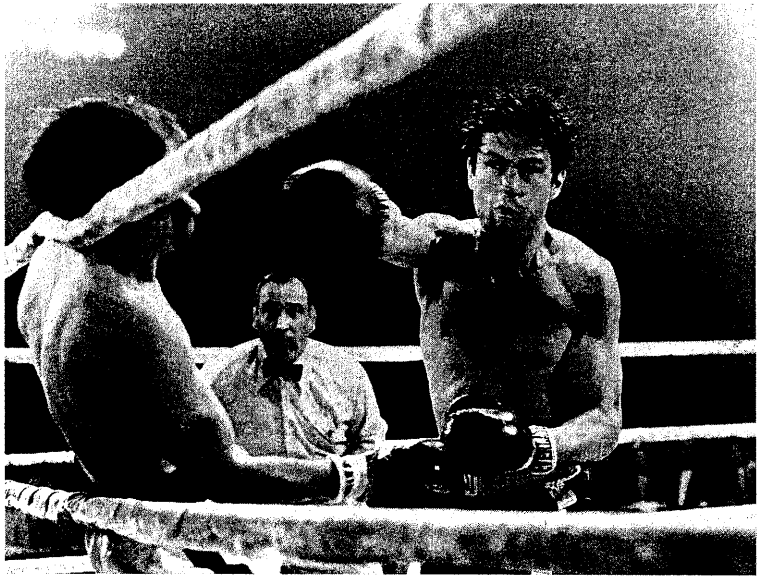

FIGURE 5. Martin Scorsese's *Raging Bull* (1980). The sound of the blows given and received by Robert De Niro (recreated in a studio by the sound editor Frank Warner) is used to *render* physical violence, but also to create *points of synchronization* and build the rhythm of the sequence.

; a salient moment in an audiovisual sequence when a sound event and a visual vent meet in synchrony. In other words, it is a point where the effect of synch- esis is particularly prominent. The frequency and distribution of synch points ithin a sequence contribute to its phrasing and rhythm, and they also provide ieaning and emphasis. A point of synchronization can occur between an nage and a sound *within a shot*, between a visual cut (followed by another shot) nd a sound cut, as well as between a visual cut and a line from the dialogue igure 5).

To obtain a synch point, synchronism is necessary but not sufficient. This ieans that a scene of filmed dialogue with much lip synch does not necessarily ontain points of synchronization. Synch points correspond to particularly sali- nt and meaningful moments linked to varied criteria, such as the impact of a reak in perception (when sound and image are cut simultaneously), the pres- nce of an emphasis which is both visual (a close-up) and acoustic (a particularly

close or powerful sound), a synchronous detail that has a dramatic or emotional value. The context is also meaningful: therefore the first synchronous meeting between a sound and an image after a long absence of synchronism (for example, after a series of long takes of someone listening to a character offscreen) becomes a point of synchronization. The latter synch point may often also have been prepared and arranged as a meeting of temporal vanishing lines (see below). One can for example speak of a synch point when the camera shows the face of an actor at the very moment he says 'Yes' after his hands have been shown for 30 seconds while he was speaking off camera; the 'reunion' between the movement of the lips and the voice is a typical point of synchronization.

A very simple experiment may be performed by starting a piece of music taken at random (from a record) together with a sequence of a film (from a video cassette with the original sound cut out), thereby creating an aleatory audiovisual superposition. This highlights the way in which the spectator, who is 'eager for synchronism', is on the lookout for the slightest synch points, even to the absurd, and uses any pretext to make them up. The experiment confirms the need for a form of scansion and punctuation in an audiovisual sequence, and the tendency for the spectator to find meaning in any concomitance, whether intentional or fortuitous.

One can speak of ***temporal vanishing lines*** when a certain number of sound and/or visual elements are superimposed and arranged in a way leading to the anticipation that they will cross, meet or collide within a more or less predictable time span. This anticipation is then either realized or 'disappointed' and this crossing may occur sooner or later than expected. Let us take the example of a character (filmed by a static camera) walking towards us at a certain pace while a musical theme with its melodic line can be heard on the soundtrack. We have here two temporal vanishing lines: one based on the anticipation of the 'collision' between the character and the camera, and the other on the conclusion – the 'cadence' – of the musical phrase. This intuitive feeling of anticipation, which leads spectators to project themselves forward in time, does not require any technical knowledge, on their part, of the language of music or film.

Two simultaneous temporal vanishing lines can be found within a single image, for example when a character is walking towards a certain point while the camera is moving diagonally or perpendicularly in relation to him or her. The same is true of course within a sequence of sounds when for example a sentence in a dialogue (in which the object comes after the verb) and a musical

phrase (with its cadential flow) heard simultaneously build two temporal vanishing lines.

Audio-division, phantom audio-vision and audiovisual dissonance

I have so far dealt with the cases where sound and image are mutually compounded and result in what is generally experienced as an 'effect of the image'. I should now point out that audiovisual relationships are also based on shortcomings where sound brings out what is missing in the image, or the image what is missing in sound, as in the above-mentioned example of 'suspension'. I would therefore be tempted to make a play on words and speak of ***audio-division.***

The term audio-division does not describe audio(logo)visual relationships as complementary and self-contained recreations of an imaginary natural entity; rather, it suggests a concomitance which, as well as generating audiovisionic effects of association, of added value pertaining to rendering or audiovisual phrasing and scenography, also develops new shortcomings, 'phantom' effects (see below) and diverse divisions. In other words, even if sound is 'realistic', it does not answer the question raised by the image; sound divides the image, and vice versa.

For instance, in an audiovisual sequence a ***phantom sound*** is a sound suggested by the image but not heard, while other sounds associated to the scene are audible. If, in a Fellini film, one can hear the characters constantly talking while they are walking at a brisk pace without hearing their footsteps, the sound of the footsteps becomes a *phantom sound* in the context of the other sounds heard. However, we hear the sound mentally, we imagine it, it does not seem to be missing but rather contributes to a feeling of lightless and fluidity. We find another example in a scene at the beach in *Mr Hulot's Holiday* (1952) by Tati: the sound of the sea, which is visible in the background, cannot be heard whereas one can make out the shouts of the bathers. The distinction we can make between a *phantom sound* and *suspension* is that *suspension* specifically concerns sounds whose absence evokes something similar to a lull before a storm. On the other hand, a ***phantom image*** is a precise image, suggested by a sound but not visible. For example, the sound of the sea accompanying the close-up of a character's face creates a phantom image of the sea.

Finally, ***audiovisual dissonance*** is an effect of contradiction between sound and image at a precise moment in a story, or between a realistic sonic environment and the setting with which it is associated. Examples can be found in Godard's *First Name Carmen* (1983), when cries of seagulls and sounds of waves are heard over shots of the Austerlitz Bridge by night, or in the gender inversion between the hero and the heroine's voices in Patrick Schulmann's fantastical comedy *Rendez-moi ma peau* (1980) or also in the contrast between a harsh and booming voice and a tiny body as in Tex Avery's cartoon *The Cat That Hated People.* When contradiction concerns size, the effect does not seem to be of dissonance, but rather of monstrosity.

The term 'dissonance' seems to me more appropriate here than the often misused term 'counterpoint' (which in music concerns a superposition of lines). It can be noted that the effect of audiovisual dissonance is almost always limited to pre-coded rhetorical cases such as gender opposition, contrast between voice and body, city versus nature in Godard's films, nature versus culture in Paolo and Vittorio Taviani's *Padre Padrone* (1977) or past versus science fiction with the *Blue Danube* waltz in Stanley Kubrick's *2001: A Space Odyssey* (1968). Furthermore, audiovisual dissonance is difficult to obtain because of the lack of demand on the spectator's part for a sonic 'verisimilitude' of the image, and also because of the power of the process of synchresis that compounds sounds with images.

Conclusion

In this chapter I have highlighted a selection of the audiovisual effects I have studied and I have provided a basis for their description. Many questions remain: 'How did these effects come about? Can they be assimilated to a "code"? Will the new multitrack technique now widespread in the cinema (what is called Dolby) transform them altogether?' These are three questions, among many others, to which I will provide a very brief and general answer.

The genesis and history of these different effects is a fascinating field of study which I have just barely opened up in my own work. It seems to me that one should take two precautions: first, cinema must be resituated within the history of theatre, music, ballet, pantomime, radio and opera, from which it has borrowed extensively. A history of sound in the cinema separate from the history

of sound in audiovisual arts in general is as absurd as telling the history of France without reference to that of the countries with which it was involved.

Second, I do not think that these effects can be approached as a 'code', in the exact sense that there exist codes of visual editing with a fixed meaning (e.g. the use of a shot/reverse shot). It seems to me that it is necessary to have a theory of the audiovisual effect, and of the filmic effect in general – a notion that needs to be rehabilitated. The contributions of scholars such as Christian Metz have very significantly advanced this cause.

Third, I do not expect a complete 'revolution' to take place overnight because of a new technique. Film language, like an individual or a species, is constructed layer by layer and through partial reconfigurations of an overall structure. What is called sound cinema has developed from a structure that was partially dictated by the requirements of silent cinema – a structure it has always retained, as is very well shown by David Bordwell and Kristin Thompson in their study of classical Hollywood cinema. In the same way, the different types of sound cinema in existence today accumulate effects and practices linked to different stages in the evolution of cinema in general. A truly historical approach is necessary to study this evolution, and an exciting task awaits historians who wish to pursue such research.

FURTHER READING

Altman, R. (ed.), *Sound Theory Sound Practice*, New York, London: Routledge, American Film Institute, 1992.

Bailblé, C., 'Pour une nouvelle approche de l'enseignement de la technique du cinéma: programmation de l'écoute', four-part article in *Cahiers du Cinéma*: nos. **292** (September 1978), 52–9; **293** (October 1978), 4–12; **297** (February 1979), 45–54; **299** (April 1979), 16–27.

Bordwell, D., Staiger, J. and Thompson, K., *The Classical Hollywood Cinema*, London: Routledge and Kegan Paul, 1985.

Chion, M., *Guide des objets sonores*, Paris: INA/Buchet-Chastel, 1982.

Chion, M., *La Voix au cinéma*, Paris: Cahiers du Cinéma, 1982. [*The Voice in Cinema* (translated by Claudia Gorbman),] New York: Columbia University Press, 1999.

Chion, M., *Le Son au cinéma*, Paris: Cahiers du Cinéma, 1985.

Chion, M., *La Toile trouée ou La Parole au cinéma*, Paris: Cahiers du Cinéma, 1988.

Chion, M., *L'Audio-vision: Son et image au cinéma*, Paris: Nathan, 1990. [*Audio-Vision: Sound on Screen* (translated by Claudia Gorbman, foreword by Walter Murch), New York: Columbia University Press, 1994.]

Chion, M., *Le Promeneur écoutant*, Paris: Plume, 1993.

Chion, M., *La Musique au cinéma*, Paris: Fayard, 1995.

Chion, M., *Le Son*, Paris: Nathan, 1998.

Schaeffer, P., *Traité des objets musicaux*, Paris: Seuil, 1967.