# SPEECH, MUSIC, SOUND

Theo van Leeuwen

# Speech, Music, Sound

# Speech, Music, Sound

THEO VAN LEEUWEN

To Tobias Icarus, the other musician in the family

# *Contents*

# Acknowledgements

# Chapter 1

# Introduction

## Integrating speech, music and 'noise'

It is the project of this book to explore the common ground between speech, music and other sounds. These three have usually been treated as separate, in theory as well as in practice. They have been talked about in different ways and with different terminologies: linguistics to talk about speech; musicology to talk about music; not much at all to talk about 'sound effects'. And they have been practised as separate disciplines too, especially in dominant modes of communication and high culture art forms. This kind of semiotic purism has not always existed. In the Middle Ages and the Renaissance the voice was still a musical instrument and music was embedded in every aspect of everyday life, just as many 'less developed' cultures had and still have songs for grinding grains, songs for harvesting crops, songs for constructing houses, songs for carrying goods, toilet training songs, puberty songs, news bulletin songs, political comment songs, and so on (cf. Merriam, 1964). But as clergical plainsong, the cries of night-watchmen, and the chanting of the ABC in schools were replaced by reading aloud, speech was divorced from music, and much flattened in the process. And as the musical sounds in our cities (church bells, the postman's horn, and so on) were replaced by mechanical noises, with music moving indoors, into the concert hall, music and 'noise', too, became separate categories: 'The string quartet and urban pandem-onium are historically contemporaneous' (Schafer, 1977: 103).

Thus music and the rest of life went their separate ways. Much effort was spent in trying to underpin the separateness of music with

watertight arguments. The most common argument was based on acoustics. Musical sound results from regular, periodic vibration, non-musical sound from irregular, non-periodic vibration: 'Music is a play of tones, that is, of fixed, clearly defined quantities. Other sounds, glissandoes, cries, noises, may occur as inserts. If they are numerous, the result is only partially musical. If they predominate, it is no longer music' (Wiora, 1965: 191–2). But many of the sounds of music (including 'classical' music) do not fit this definition, and that applies not just to percussion instruments: all conventional musical instruments have their irregularities and granularities. Another common argument was based on the effect of music. Musical sound is beautiful, it pleases us. But again, not all music is beautiful, and many of the sounds we do not normally regard as music can be and are called beautiful, for instance certain human voices, and certain of the sounds of nature.

In this century, all this has begun to reverse. Recording technology has brought music back into everyday life – through muzak, the transistor radio, the car stereo, the walkman. The boundaries between speech, music and other sound have weakened. Composers have experimented with combinations of musical instruments, singing and speaking voices, and non-musical sounds. Satie added a typewriter to the orchestra, Gershwin a taxi-horn, Antheil a propeller, Zappa a cash register, Reich car alarms. Schoenberg, Berio and others wrote Sprechgesänge ('speech songs') – and so, in a different way, do contemporary Rap artists. The Futurist composer Russolo created a noise orchestra consisting of noise makers, boxes cranked by a handle: buzzers, bursters, a thunderer, whistlers, rustlers, gurglers, a shatterer, a shriller and a snorter. And, most significantly, he redefined noise as music:

> Let us walk together through a great modern capital, with the ear more attentive than the eye, and we will vary the pleasures of our sensibilities by distinguishing among the gurglings of water, air and gas inside metallic pipes, the rumbling and rattling of engines, breathing with obvious animal spirits, the rising and falling of pistons, the stridency of mechanical saws, the loud jumping of trolleys on their rails, the snapping of whips, the whipping of flags. We will have fun imagining our orchestration: of department stores' sliding doors, the hubbub of the crowds, the different roars of railroad stations, iron foundries, textile mills, printing houses, power plants and subways (Russolo, 1986 [1913]: 3–8).

The pioneers and early theorists of the sound film saw it the same way:

> It is the business of the sound film to reveal for us our acoustic environ-
> ment, the acoustic landscape in which we live, the speech of things and the
> intimate whisperings of nature; all that has speech beyond human speech,
> and speaks to us with the vast conversational powers of life and incessantly
> influences and directs our thoughts and emotions, from the muttering of
> the sea to the din of a great city, from the roar of machinery to the gentle
> patter of autumn rain on a window-pane. The meaning of a floorboard
> creaking in a deserted room, a bullet whistling past our ear, the death-
> watch beetle ticking in old furniture and the forest spring tinkling over the
> stones. Sensitive lyrical poets always could hear these significant sounds of
> life and describe them in words. It is for the sound film to let them speak
> more directly from the screen (Balasz, 1970: 197–8).

Not only 'noise', but also the word began to be treated as music, for
instance in the 'concrete poetry', pioneered by Dada poets of the 1920s
such as Kurt Schwitters:

> priimiitittiii    tisch
> tesch
> priimiitittiii    tesch
> tusch
> priimiitittiii    tischa
> tescho
> (*and so on*)

We can now hear words as *sounds* again, whether in the 'nonsense
word' conversations of young children (cf. for example, Keenan,
1974), in the magic chants and songs of many of the world's religions,
or in the virtuoso 'scat' singing of singers like Ella Fitzgerald and the
electronic manipulation of snippets of speech in Steve Reich's 'Come
Out'. Murray Schafer makes the point with a telling example in a small
book called *When Words Sing* (reprinted in Schafer, 1986):

> I have just been listening to two tape recordings I made of Miranda, age
> seven. In the first she reads a story from her reader; in the second she
> makes up a scary story of her own. The first sounds flat and stupid. I wish
> you could hear the second.
>
>> Once there was a little old man. He wanted to get a piece of volcano rock. So he
>> got all his hiking equipment ready. Then he hiked up the mountain. It took
>> him twenty years. When he got to the top he took off the lid. Oooooahooooo!
>> Inside there was a ghost and he mo-aned and he gro-aned. The man ran away as
>> fast as he could and he said, 'I'm never going to climb volcanoes again!'

> Miranda recites her unusual story with intense emotion. Words like 'moaned' and 'groaned' are so highly inflected and attenuated they are almost chanted. The 'Oooooahoooo' is a pure glissando melody. Miranda knows that words are magic invocations and can cast spells. So she exorcizes them with music. Of course her teachers will correct all this in another year or two by muzzling her to the printed page (Schafer, 1986: 198).

This book tries to do on a *theoretical* level what many contemporary musicians, poets, film-makers, multimedia designers and so on, already do in practice (and what children have always done): *integrate* speech, music and other sound. It tries to foreground the integration of these three, rather than to talk about their specifics, and to contribute to the creation of a vocabulary for talking about this integration, and for exploring its ramifications and potentials. Above all, it tries to make you listen. Listen to the city as though it was music and to music as though it was the city, or to speech as though it was music and to music as though it was speaking to you. This listening can, in the end, only be done by you, the reader. But I hope my book will be able to give you a helping hand. It is not always going to be easy, incidentally. Not through our own fault we have, most of us, ill-educated ears. Re-educating them may take some effort.

## Some principles of semiotics

Although I have not used this term in my title, this book is about semiotics. But what is semiotics? Or rather, what do semioticians do? Three things, I think.

### Describing semiotic resources

Semioticians describe the semiotic *resources* people use in communication. Narrowing this down to sound, the semiotics of sound concerns itself with describing what you can 'say' with *sound*, and how you can interpret the things other people 'say with sound'.

Why the term 'resource'? This needs a little background. In the past the dominant semiotic understanding and explanation of communication derived from a particular conception of language which saw language as a set of rules, a *code*. Once two or more people have

mastered the same code, it was thought, they would be able to connect the same meanings to the same sounds or graphic patterns, and hence be able to understand each other. The question of how this code came about, of who made these rules, and of how and why they might be changed, was not high on the agenda. The rules were treated as simply being there.

In some contexts communication actually works like this, for example, in bureaucracies. There the use of words, the shape of documents and so on, are laid down in impersonal rules that simply must be followed to the letter and are not open to interpretation – if they do turn out to be open to interpretation that means there is some kind of 'loophole' in the system which should be plugged as soon as possible. But in other contexts, say poetry, or advertising, or conversations between children, things work differently. No hard and fast rules exist. Any bit of language you might lay your hands on could come in handy for the semiotic job at hand, whether it is grammatical or not, whether it represents a standard variety of English or not, indeed, whether it is English at all. So long as it does the job. In yet other contexts, in music for instance, either approach (or some mixture of the two) is possible. Some musicians are bureaucrats in spirit, or rather, they have to earn their living in a context where music is practised in this bureaucratic spirit, and so must play according to a definite code, a musical grammar. Anything else is frowned upon: 'That is not how Bach should be played'; 'That is not jazz' and so on. Others work in less strictly regulated contexts. They have their ears wide open and use any sound they can lay their hands on if it suits their purposes, frequently crossing generic boundaries, or even the boundaries between music and speech, or between music and 'noise'. The same applies to interpretation. There are viewers, listeners and readers who view, listen and read 'according to the book' – in educational contexts you usually have to do this if you want to get a good grade. And there are viewers, listeners and readers who use whatever resources of interpretation and intertextual connection they can lay their hands on to create their own, new interpretations and connections.

In public (in the workplace, for instance) only people with a large amount of cultural power are allowed to break or make rules. Most of us must have to follow them. In private, in the smaller groups and subcultures we live in, we may have more freedom, but our semiotic productions and interpretations are not likely to spread beyond these contexts. They will remain relatively marginal. Sometimes, however, society *needs* something new, and there is a chance for new and

adventurous modes of production and interpretation to break through. Sound (sound-*as*-sound, that is, rather than 'sound-as-music' or 'sound-as-language') belongs in this category. There is an increasing interest in it. A new discipline of 'sound design' is emerging. But the code-books have not been written, and the production and interpretation of sound have not been taught and policed in the 'bureaucratic' spirit. Sound design has remained, so far, a little bit marginal perhaps, but free. A semiotics of sound should therefore not take the form of a code book, but the form of an annotated catalogue of the sound treasures Western culture has collected over the years, together with the possible uses to which they might be put, as gleaned from past and present experience. A semiotics of sound should describe sound as a semiotic resource offering its users a rich array of semiotic choices, not as a rule book telling you what to do, or how to use sound 'correctly'.

To create a usable catalogue you need usable headings, and that means classifying. I will do a fair bit of classifying in this book, although I hope to bear in mind that 'classification is only justified if it leads to the improvement of perception, judgement and invention' (Schafer, 1986: 133). In doing my classifying I follow a 'systemic-functional' approach (cf. for example, Halliday, 1978; Martin, 1992). This means that the 'choices' which make up the 'resource' are set out as (a) *binary opposites*, (b) *ever more 'delicate' choices* (moving from left to right in a system network one moves from the broadest headings to the finest subdivisions) and (c) in terms of their *semiotic value* for the production and interpretation of sound events. Let us take these points one by one in terms of an example that will be discussed in more detail in Chapter 3.
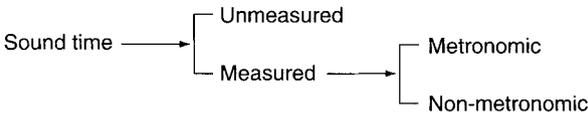


This fragment of a system network is concerned with a particular sound 'resource'. It tries to map how sound events can be structured in terms of their *timing*. The key distinction it introduces is the distinction between 'measured' and 'unmeasured' time. Measured time is time you can tap your feet to (ONE-two ONE-two and so on, or ONE-two-three ONE-two-three and so on). This is not possible with unmeasured time. The physical reaction to unmeasured time is more likely to be a slow swaying of the body. To give some examples, a long

wail is unmeasured, and so is medieval plainchant; the reciting of metrical poetry is measured, and so is dance music. The distinction is a binary distinction: sound is either measured or unmeasured. There is no in-between. Other contrasts, for instance the contrast between 'loud' and 'soft', are not binary and admit of many 'in-between' gradations – a notation for this will be introduced later. The 'metronomic' system is more 'delicate' than the 'measurement' system: 'metronomic' and 'non-metronomic' time form a subdivision of 'measured time'. 'Metronomic time' is governed by the implacable regularity of the machine, whether or not a metronome (or a drum machine or stopwatch) is actually used. It is the time of the machine, or of soldiers on the march. 'Non-metronomic time' is also measured, but it subverts the regularity of the machine. It stretches time, it anticipates or delays sounds and so on. It is the time of human speech and movement, or of Billie Holiday singing a slow blues while 'surfing on the beat'. Clearly it would be possible to go further and distinguish several kinds of 'non-metronomic' time. But these two distinctions are enough to illustrate the idea of a system network.

What is the semiotic value of these distinctions? What can be expressed by their use? To answer this question, we need to put the system in its social, cultural and historical context. The 'high art' of the Middle Ages was the music of the Church. Its favoured mode of timing music, its favoured 'choice' from our 'system' was 'unmeasured time', because unmeasured time is a particularly apt signifier for 'eternity' – it literally negates time and goes 'on and on'. When, in northern Europe, cultural dominance moved away from the Church and gradually passed into the hands of the emerging merchant class, with its interest in secular art and science, measured time became the favoured mode of timing for the new 'high art' music, that is, for what we now refer to as 'classical music'. This 'choice', too, had its semiotic value. First of all it elevated a secular mode of timing to the status of 'high art', for 'measured time' had previous been 'low art' and associated with secular life, with dance and popular song. Second, it was an apt signifier for the values of calculability and quantification which were so important in both science and capitalist enterprise. The Church saw it as desecration. 'The new school occupies itself with the measured dividing of time and pesters every composition with semibreves and minims', declared the Pope, 'We hasten therefore to banish these methods, and put them to flight, far from the House of God. Let nothing in the authoritative music be changed' (quoted in Harman and Mellers, 1962: 123).

'Non-metronomic' time became more important as a result of the influence of Afro-American music. It too was seen as a form of 'desecration' because it subverted the discipline of the clock which had been such a key value of the industrial age, and it heralded a time of changing social relations and less severe affect control. Merriam (1964: 242–3) cites some of the reactions it met with in the 1920s:

> The composer, Sir Hamilton Harty, worried that future historians 'will see that in an age which considers itself enlightened we permit groups of jazz barbarians to debase and mutilate our history of classical music and listen with patience to impudent demands to justify its filthy desecration', and a Dr Reisner added that 'Jazz is a relic of barbarism. It tends to unseat reason and set passion free.'

All this will be discussed in more detail in Chapter 3. The point here is, the 'choices' offered by semiotic resources have *semiotic value*. They carry with them a potential for semiosis, for meaning-making. This is why certain 'choices' may become mandatory, conventional or traditional in certain contexts: the kinds of meaning they allow are mandatory, conventional or traditional in those contexts. In this case there is no real choice, and this should be remembered even when, from now on, I will no longer put the word 'choice' in inverted commas. Other contexts, on the other hand, do not, or not yet, operate on the basis of strict rules, conventions or traditions, and in such contexts there is choice. It is then again the semiotic value of a given choice which makes people recognize it as an apt choice for expressing what they want to say.

It is important to *contextualize* semiotic systems, to put them in their historical and social setting. Some semiotic disciplines have neglected this, especially linguistic ones. Even now that the idea of 'context' has gained an important place in some forms of linguistic theory (for example, Halliday, 1978; Martin, 1992), it is still not always fleshed out in sufficient cultural and historical detail (the 'discourse-historical' method of Ruth Wodak and her associates is an important exception, cf. Fairclough and Wodak, 1997). Other semiotic disciplines, on the other hand, for instance art-historical iconology, are so tied up with the specificities of individual artists and schools, that the idea of a common 'language of art' is never worked out systematically. I want to combine these two approaches and profit both from the systematicity of linguistics and from the focus on social, cultural and historical detail which characterizes semiotic disciplines such as art-historical iconology.

'Describing semiotic resources', the first of the three things semioticians do, will be the main focus of this book. I will discuss six major domains, sound perspective, sound time and rhythm, the interaction of 'voices' (for instance by taking turns or speaking, singing, playing or sounding together in different ways), melody, voice quality and timbre, and modality. All are aspects of sound common to speech, music and other sounds, and in each case I will attempt to investigate how they can be used to say and do things with sound, and to interpret how people actually say and do things with them. In this my overall focus will be on the semiotics of sound as it has developed in Western culture. Yet I will often draw on sounds from 'other cultures'. This is only an apparent contradiction. Once sounds from other cultures have been written up by Western anthropologists, or (for example, in the case of 'World Music'), distributed through Western distribution chains, they have been already transformed and incorporated into Western culture and its frame of reference. My references to 'other cultures' are to be taken as references to other cultures *as they have been incorporated in Western culture and its frame of reference*. I do not pretend to be able to look at them as a member of that culture might look at them in an intracultural context.

## Explaining how semiotic resources are used

Describing semiotic resources provides the means for describing and explaining how these resources are actually used. In this book I do this, not as an end in itself, but only to try and show that my descriptions can be used in this way. Every chapter ends with an extended example in which I show how the resources I have described in that chapter are used in articulating a particular sound event, and what kinds of socially and culturally relevant ideas might arise from the analysis of its use of these resources. I have used a variety of sound events to try and demonstrate that my ideas can be used in a number of different areas of cultural studies, media studies and film studies. This includes extracts from popular songs, radio and television programmes, commercials and film soundtracks.

There is of course a degree of artificiality to the way I have done this, because in each case I analyse a sound event according to one articulatory parameter only, whereas every sound event in fact incorporates 'choices' from *all* the resources described in the book – it is the interaction between all these choices, the total 'mix', that matters in the end.

Also, when I try to formulate the semiotic values of the 'choices', I do not provide a *code*, with definite and fixed meanings, but a *meaning potential* which will be narrowed down and coloured in the given context. The importance of context cannot easily be overstated. The same sound can be used to mean one thing in one context and another in another context. Take the example of 'unmeasured time'. In a science fiction film it might be the time of a background track made up of densely shimmering electronic sounds, and come across as signifying a sinister primeval chaos. In church it might be part of a musical setting of the words 'peace on earth', sung by a choir of boy sopranos, and coming across as signifying something like 'eternal peace'. Yet both meanings derive from the same meaning potential. In both cases the idea of 'out of time', of 'eternity' is a definite part of the total meaning and effect.

## Exploring how semiotic resources can be expanded

There is yet another contribution semioticians can make: they are particularly well placed to explore how semiotic resources can be *expanded*, so as to allow more options, more tools for the production and interpretation of meaningful action. In other words, semiotics can be a tool for design.

In the past this was a by-product of semiotics. Semiotics was supposed to be 'the science of signs' and science, in turn, is supposed to be about 'what is', not about 'what could be' or 'what might be'. Still, when you systematically describe 'what is' you find gaps, you find yourself wondering why certain options are not available and why certain things cannot be done in certain semiotic modes. Which is only one step away from *unlocking semiotic doors*, from asking: Could it be done? Does it have to be impossible? And if we are going to do it, how shall we do it?

Film-makers in Soviet Russia in the 1920s were faced with the task of bringing a propagandistic message to people speaking many different languages, and they wanted to do this through film, a medium which, they thought, could be understood universally – movies were still silent at the time. In other words, they wanted to convey ideas in a medium which so far had only been used to tell stories and provide entertainment, and which therefore had not developed any resources for conveying ideas. So they asked themselves: Why can't we use metaphors in film? Why has film been so literal-

minded, while metaphors are perfectly possible in other semiotic modes? And they studied semiotic modes which could convey more abstract or generalized ideas, for instance hieroglyphics and Japanese Noh theatre (cf. Eisenstein, 1949, 1975). In times of rapid change and new communicative challenges, semiotics and design, theory and practice, can work hand in hand.

In this book I concentrate on inventorizing 'what is'. But I hope that my 'catalogue' will trigger ideas and be of some use to the people who are now pioneering the new discipline of sound design, whether in relation to the design of objects or in relation to music and the computer media, and I do believe that making theory-and-practice links of this kind, and learning to describe 'what could be' is the single most important job now facing semiotics.

## Exercises and discussion points

Write a 'system network' (see Appendix) to describe the main choices available in choosing a doorbell, knocker or other device or method for announcing your presence in front of a closed door or gate, concentrating of course on the sound, rather than the look of the devices.

What is the semiotic value of these choices? You might approach answering this question by thinking about the kinds of places (shops, office, apartments, houses and so on) or people who tend to make a particular choice, or by asking people why they chose the 'doorbell' they chose, and what they think about various other kinds of 'doorbell' which they did not choose.

Does your 'doorbell' system integrate speech, music and other sounds? If so, why? If not, why not?

Is the 'doorbell' a historically and culturally specific phenomenon? If so, how and why did it arise? How was its function fulfilled in other times and places?

Can you think of a new kind of 'doorbell'? What kind of sound would it have? Who might want to buy it and why?

# Chapter 2

# Perspective

## Perspective and social distance

Every semiotic mode can create relations between what is being presented or represented and the receiver, the reader or viewer or listener of the message. Images do it through two slightly different and complementary ways of spatial positioning, *size of frame and perspective.*

Sizes of frame such as the close shot, the medium shot, the long shot and so on create a certain *distance* between the viewer and the people, places and things represented in the picture. The relations expressed by these distances derive from our everyday experience, from the distances we keep from different kinds of people, places and things in everyday life. Edward Hall (for example, 1966: 110–20) has described this in relation to our interactions with people, but the same applies to interactions with places and things. According to Hall, we carry with us an invisible set of boundaries beyond which we allow only certain people to come. The zone of 'personal distance', the distance at which you can touch the other person, is for those who are close to us – if others enter it this will be experienced as an act of aggression. 'Social distance' is for more businesslike and formal interactions. At this distance we keep people 'at arm's length'. 'Public distance' is for larger and more formal group interactions. It is the distance we keep from people 'who are and are to remain strangers'. To all these distances correspond different fields of vision. At personal distance only head and shoulders are in sharp vision, and as it happens this corresponds to the close shot, as usually defined in the world of film and television. At social distance we see a little less than the whole figure, which roughly corresponds to the medium shot, and at public distance we see the whole figure with space around him or her, which corresponds to the long shot. In this way close shots position viewers in

a relation of *imaginary* intimacy with what is represented, while medium shots create more formal kinds of imaginary relations, and long shots portray people as though they fall outside the viewer's social orbit, either because they are strangers or because they are much lower or higher in social status. In reality this may not be the case. The people we see in long shot may be people like us. But that is not the point. The point is that viewers are addressed *as though* these people are not part of their world.

Perspective creates horizontal and vertical angles. Vertical angles can make us literally and figuratively 'look up at' or 'down on' what is represented in a picture – or make us see it from the position of an equal, at eye-level. The vertical angle is therefore connected to imaginary *power* relations, be it the power of the viewer over what is represented, or the power of what is represented over the viewer. The glamorous role models in advertisements, for instance, tend to be shown from below, so that we look up at them, while the products are shown from above, so that they seem within reach of our hands, and under our control. Horizontal angles can be frontal, confronting us directly and unavoidably with what is represented, *involving* us with what is represented, or profile, making us see it from the sidelines, as it were, in a more detached way – and there are of course many in-between possibilities.

These concepts can be used to ask what *attitudes* a given image expresses towards what it represents. Who or what is positioned close to or far away from us, and why? Who or what are we made to look up to or down on, and why? Who or what are we brought face to face with, and who or what do we see in the more detached side-on way? The answers to these questions are usually found in the context. In a Dutch junior high school geography textbook two pictures appeared side by side as part of a double page headlined 'The Third World in Our Street'. The picture on the left showed three women with headscarves, in long shot, on the other side of the street, and turned away from the viewer. In other words, these women were portrayed as strangers, as people outside the social orbit of 'us', Dutch high school students. The picture on the right showed a young couple, a white girl and a black boy, sitting at an outdoor café table, the girl's hand on the boy's arm. They were shown in a much closer shot and from a more frontal angle. A different imaginary relation was suggested here. The couple was portrayed, if not quite as 'our' friends, sitting at the same table, then at least as 'people like us', frequenting the same café. This school book therefore addresses Dutch high school students as though