Sound-to-Picture

the role of sound in the audio-visual semiosis of non-fiction film

Odysseas Constantinou
BA (Cardiff) MA (Cardiff)

Thesis submitted for the degree
Doctor of Philosophy
April 2007

Centre for Language and Communication Research
Cardiff University

UMI Number: U495109

UMI

Dissertation Publishing

ProQuest

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

ABSTRACT

This thesis develops an approach to analysing sound-to-picture and audio-visual semiosis in non-fiction film. Its broad aim is to contribute to the development of a critical audio-visual literacy that provides the 'listener-viewer' with a conceptual framework for engaging with – particularly non-fiction – audio-visual texts. Its central research questions are: (1) What is the semiotic function of sound in audio-visual semiosis?; and (2) what functions does sound-to-picture perform in audio-visual non-fiction film. The thesis builds on existing work in multimodal discourse analysis that incorporates *sound* as a site of semiosis too important to ignore (Van Leeuwen, 1999; Thibault, 2000; Iedema, 2001; Baldry and Thibault, 2005). It also extends work on non-fiction film (esp. Nichols, 1991) by integrating its observations on documentary film and TV news with the systemic-functional and social semiotic notions of context of situation and context of culture.

The study focuses on the audio-visual constructions of agency and subject position across fictional and non-fiction contexts in order to generate an understanding of what is peculiar about non-fiction film sound. Four texts are analysed – two fiction, two non-fiction – according to a framework comprising analytic resource categories derived from the traditional and technical resources of sound-to-picture.

To delineate non-fiction sound-to-picture, an attempt is made at developing a model of constraints that derive from context. Non-fiction is formulated as 'context of situation' and 'context of culture' in order to establish a framework for analysing the film and TV sound track that illuminates the active constraints on sound-to-picture.

vi

Acknowledgments

This work would not have been possible without the help and support of certain people:

My life-friend, Taina Niemelä, who provided sunshine throughout. Her support during the last few months of the writing cannot possibly be repaid (unless, of course, she chooses to pursue doctoral research). *Suurkiitokset.*

My parents, who offered shelter and food to sustain the final moments of writing.

My supervisor, Adam Jaworski, who has helped me more than he believes.

Theo van Leeuwen, who introduced me to semiotic theory, and made the task of theorising sound-to-picture seem possible.

My reader, Lisa El Refaie, for being so honest.

Hywel, for always being 'on', and Shanti, for her confidence in me.

Aila and Jorma Niemelä, who provided shelter, hospitality and snow on many occasions.

And, finally, Lucy Robinson and Andrew Wilson-Dickson for their generosity in providing accommodation that made the task of writing up as stress-free as possible – as well as romantic.

Preface

Several years ago, I worked as an assistant dubbing editor. My assigned task was to select and edit sound effects for television programmes and films. Coming from a background in electronic music production, I approached this task with a vast repertoire of creative resources. However, I would come to be frustrated with what I perceived as the boundaries around my practice. These were limits on meaning making that, I felt, were reified, regulated and maintained by my immediate superior, and through the trade discourses that permeated our professional lives. As I learned, there are ways of manipulating and editing sound that are simply not done in the practice of sound-to-picture. This is especially true for *television* programmes and films, where a 'conveyor-belt' ethos governs decisions over the available scope for creative and 'experimental' uses of sound.

Further frustrations were to stem from certain contradictions. While even the suggestion of a particular creative manipulation of sound, one that *might* draw attention to itself, would be dismissed out of hand as surplus to the narrative or expository aims of a film, the addition of sounds where there were none previously was unproblematic, even in non-fiction. Adding sound effects to fiction film is a fundamental of sound-to-picture practice, but I was surprised when I was asked to scour the effects library for sounds to augment a non-fiction film. In one instance, I was called on to reconstruct the sound track of an interview with a vicar that took place outside a church in a rural setting. As my superior argued, there weren't enough bird sounds on the track. I countered that this was *non-fiction* – surely the practices we applied to fiction film would not apply in this case? After all, the rural setting was obvious – and there *were* birds both visible and audible. Apprentices don't often win arguments.

All this happened more than a decade ago, and I was, quite obviously, naïve and idealistic. Then, I was unable to harness the critical view of sound-to-picture practice that had begun to manifest in my professional approach to a constructive end. Instead I would become dispirited with the weight of the constraints that bore down upon my practice. I simply decided one day that I'd had enough.

Where then I saw a problem, now I would see an opportunity. Following an introduction to social semiotics, I realised that not only could I legitimately set out

my concerns and grievances about sound-to-picture by systematically examining its practices, but that there was also a great deal of potential for change. By investigating how certain audio-visual texts made meaning, I began to see how changes in practices might come about. They certainly would not, of course, be instigated by quitting the practice altogether (though my frustration at the time was understandable). Rather, change can be brought about by firstly understanding the ways in which semiotic practices are regulated. Only a critical analysis of the constraints on sound-to-picture – that is to ask: what are they and how do such constraints arise? – can establish a platform from which to explain how to initiate semiotic change.

My approach to sound-to-picture practice nowadays is far different. Not only have the semiotic resources expanded considerably over the last decade, but I choose to occupy a position from which I am able to both fulfil the requirements of the industry *and* to maintain independence from them. I am increasingly interested in using my understanding of the semiotics of sound-to-picture, and my knowledge of its subtlety, in creating sound tracks for television and web advertising that, while appearing to fulfil their immediate function, also work reflexively. Subversion of the rhetorical objectives of certain genres is perhaps the most extreme outcome of this approach. This is the semiotics of sound-to-picture as *praxis*.

This thesis is a systematic attempt to understand what is peculiar about non-fiction sound-to-picture. It attempts to lay the foundations of an understanding of sound-to-picture as a political practice. My hope is that it may contribute to the development of a critical audio-visual literacy, and a multimedia literacy more generally, from which practitioners, analysts and listener-viewers may benefit.

Chapter 1: Sound-to-picture in non-fiction film
An Introduction

1.0 Introduction

This thesis aims to generate understanding in two related areas. Firstly, it is concerned to explicate the functioning of the sound track in the broader, audio-visual functioning of film texts – phrased here as *sound-to-picture*. Secondly, it aims to understand the functioning of sound-to-picture in the particular context of non-fiction filmmaking.

Given the contemporary practical and theoretical concerns of multimedia and hypertext (e.g. Lemke, 2002; Martinec and Van Leeuwen, 2006) to define the core issue of this research as audio-visual semiosis is to take one or two steps backwards. Films and television programmes appear somewhat mundane when contrasted with the more extravagant multimedia texts that flood the contemporary world. But in another sense, sound and image constitute the hardcore of multimodal discourse, as hearing and sight are, what both John Cage and Bertrand Russell have called, the 'public senses': those most engaged by the mass media (Sterne, 2003; Kahn, 1988; Russell, 1921). In this way, 'audio-visuality' becomes a crucial site for semiotic investigation.

Establishing a foundation for investigation into sound-to-picture is an urgent task. There are two reasons for this. Firstly, there is a dearth of research on the semiotic functioning of sound in the field of multimodal discourse analysis (MDA). Moreover, there are very few active researchers in that field who engage with sound and audio-visual semiosis. While the semiotic potential of sound has received an increase in interest in social semiotic research in the past few years (Van Leeuwen, 1999; Kress and van Leeuwen, 2001; Iedema, 2001; Thibault, 2000; O'Halloran, 2004; Baldry and Thibault, 2005), very few have undertaken focussed research committed to understanding how sound and image are integrated to make meaning. Moreover, I am not aware of any attempts that aim to explicate the role of sound in the audio-visual relationship – that is, to *isolate* the specific contribution that sound is making in a text by means of describing how its technical resources are deployed and analysing the meanings they are mobilised to make in an audio-visual context.

The second reason concerns the need to develop an audio-visual 'literacy', as part of a broader project of multimedia literacy (cf. Lemke, 2003). Our engagement with the modes and media that pervade contemporary social life is constrained by what we understand of their functioning. While many appear to have migrated from paper page to web page in their pursuit of the daily news, when it comes to 'watching' the news, TV still appears to dominate over web-based alternatives. Yet, increasingly, TV news – and documentary film – deploy sound and image to rhetorical and communicative ends in ways that escape the notice of all but the most informed and critically engaged listener-viewer. This is not surprising: sound's character as a medium – that is, its transient nature – entails that critical engagement is enormously difficult; understanding how sound operates in the context of the moving image involves a level of technical knowledge unavailable to the average listener-viewer. As Iedema (2001: 202) puts it:

> If we turn our attention to images or sound, we often have no other resources for dealing with them than intuition and commonsense. But if we cannot deconstruct editorial and camera strategies, or visuals and sounds, a whole universe of meanings escapes critical notice. To mount a penetrating critique of [audio-visual texts], we need, in addition to a critical linguistic literacy, a critical visual-audial literacy enabled by tools as made available by social semiotics.

But there is a more profound reason behind the need to develop such audio-visual literacy, and it serves as the critical backbone to this project. The media that dominate in a culture are intimately related to the knowledges and beliefs that constitute that culture (Postman, 1985; Luhmann, 2000). Postman (1985) is concerned specifically with the medium of television when he argues that an increasing dependence on 'the visual', and a diminishing reliance on the printed word, is shaping our knowledge systems and, by extension, our systems of belief. To lend support for this view, Postman looks farther back into media history. As he argues, the predominance of the photograph in American newspapers by the late-nineteenth century led to a situation in which 'seeing, not reading, became the basis for believing' (1985: 76)[1]. Although Postman does not attend to the sound track, his argument is still pertinent here. Without an audio-visual literacy, we, as listener-

2

viewers, cannot hope to determine the ways in which sound and image contribute to our knowledge of the world.

Furthermore, there are very real consequences to these epistemological concerns of audio-visual semiosis. As Luhmann (2000: 1) argues, '[w]hatever we know about our society, or indeed about the world in which we live, we know through the mass media.' If our knowledge about the world is constrained by 'the media', and the particular modes and media that are deployed (sound and image) afford certain meanings to be made at the expense of others, then the actions we take in response to the mediated discourse on, for example, issues of war, poverty and climate change, are dependent to a large extent on semiosic practices that we are unable to engage with analytically or critically.

In recent years this has been made abundantly clear in the imperialistic actions in Iraq and the surrounding Middle East, and the lack of major resistance from the Western world. It also is apparent in the lack of direct action with regard to climate change, particularly in the United States, but also increasingly in the UK. These actions – including the action of not taking action – could not be conducted without 'the media' and its constituent semiotic modes playing a central role in the generation of knowledge, and therefore belief and action (cf. Hodge and Kress, 1988: 147). One other example is particularly relevant here. In 2004, Howard Dean, a US presidential candidate, made the international news with what was considered an inappropriately enthusiastic vocal performance at a Democratic party rally. A professional news clip was circulated among the news press containing footage of the event, and the candidate's response. Sure enough, the candidate's response was extremely loud, his voice even distorting the audio track, and was certainly 'inappropriately' louder (and wilder) than his audience. It was a damning indictment of his character: Dean appeared 'intoxicated with power'. However, soon after the initial story circulated, an audience recording of the same event emerged. The experience of viewing this newly available clip was like viewing an entirely different event: the candidate's voice could hardly be heard over the extremely vocal audience and he apparently had to strain his voice to be heard. This explained why he appeared far 'wilder' than the audience in the professional clip which diminished the audience's collective voice.

3

The difference between the two clips was due to the techniques and technology deployed to record the event. In the first, professionally recorded, clip, we heard the candidate's voice as it was captured by his noise-cancelling, directional microphone. A directional microphone, as far as possible, is deployed to isolate the sound or voice of the intended subject from surrounding noise, considered extraneous (in this case, the candidate). In this case, a 'feed' was taken ostensibly directly from this microphone. On the later-released video clip, however, the sound we heard was recorded using a comparatively conventional microphone that also captured the audience's vocal presence, and was therefore more faithful to the actual audience experience. But by the time the second video clip began circulating, beliefs about the candidate's personality had already been formed. Even if such beliefs could be challenged by the follow-up story, it was not considered sensational enough to gain widespread international coverage. Several news networks subsequently made official apologies for showing the clip on such heavy rotation.

What arises from this discussion is the need for a literacy that incorporates knowledge of the technical construction of audio-visual texts, including the technologies that are often deployed in such construction. The present work aims to contribute to this project of developing a critical audio-visual literacy.

This thesis also represents, in part, the search for a way of critically engaging with the *non-fiction* film sound-track. The focus on non-fiction film stems from my conviction that documentary film and TV news most directly shape our view of the world. Whereas our engagement with fiction film involves a metaphorical relationship to the historical world (Nichols, 1991: 28) the sounds and images of non-fiction film 'partake of the same order of reality as that to which they refer'. This indexicality is at the heart of the rhetorical power of non-fiction film. In engaging with documentary film, to quote Nichols,

> [w]e are less engaged by fictional characters and their destiny than by social actors and destiny itself (or social praxis). We prepare ourselves not to comprehend a story but to grasp an argument. We do so in relation to sounds and images that retain a distinct bond to the world we all share. (1991: 5)

4

Understanding how sound-to-picture plays a role in documentary film and TV news, then, helps us to generate an understanding of how our knowledge of the historical ('actual') world is constructed (cf. Postman, 1985).

As Noël Carroll (1996: 286) has argued, however, the representational techniques of fiction and non-fiction are *transgeneric*. In his words,

> [t]he distinction between nonfiction film and fiction film cannot be grounded in differences of formal technique, because, when it comes to technique, fiction and non-fiction can and do imitate each other.

On that basis, it makes little sense to attempt an explication of non-fiction sound-to-picture that is guided by the view that understanding non-fiction sound is a matter of inventorising its manifest techniques of representation. Certainly, such description may constitute a part of the project. But without attempting to delineate non-fiction *as context*, and without constructing a framework with which to integrate technical analysis of sound-to-picture with a contextual model, that description will be of limited value.

The kind of critical audio-visual literacy we must aim to develop, therefore, cannot limit itself to an understanding of how a particular sound-image configuration has been accomplished technically and semiotically. Rather, in consideration of the difference between fiction and non-fiction film, an audio-visual literacy should also incorporate a formulation of the constraints of non-fiction, as context, on audio-visual semiosis. Once these components are in place, a framework for analysing sound-to-picture in non-fiction film can be constructed, enabling the development of a more comprehensive critical audio-visual literacy.

1.1   Development of the research focus

My previous work on sound-to-picture centred around creating a descriptive system that would enable the analysis of film sound tracks in terms of the *modality* system (Constantinou, 2002; cf. Kress and van Leeuwen, 1996; cf. Van Leeuwen, 1999). The aim was to contribute to the existing categories of visual semiotics and sound semiotics in order to construct an analytic framework with which to comprehend the ways in which reality was constituted audio-visually. This was achieved, however, in a relatively context-independent manner, with only ad hoc discussion of the situations in which sound-to-picture made meaning, which was intended to

account for the invoked *coding orientation*. I concluded that work with the statement that, clearly, context (and genre) were factors in need of elaboration, for without insight into the social situation that sound and image are integrated to operate within we cannot make confident claims about how a particular audio-visual configuration 'constructed reality'.

The present work aims to engage in depth with the notion of context in order to get at what is peculiar to non-fiction sound-to-picture. I do not claim to 'right the wrongs' of my previous work, but rather to extend the research into sound-to-picture in a more general way – beyond the idea of reality construction and modality, into an area in which the integration of sound and image can be described in terms of how it achieves the rhetorical goals of non-fiction and fiction filmmaking. In the present work, I want to take a step back, so to speak, to begin with traditional and technical resources of sound-to-picture and to allow a theory of audio-visual semiosis to emerge from analyses of the ways in which those resources are deployed in audio-visual film texts. While I do not claim to be the first to attempt a 'generalised' social semiotic approach to film (cf. Thibault, 2000; Iedema, 2001), the primary focus here is on explicating comprehensively sound's role in audio-visual integration, and as such it constitutes a deviation from existing approaches. The first brush strokes of a semiotic theory should be broad enough to allow scope for further development. The following work represents this conviction.

1.2 Research questions

At this point, we can specify two central research questions that guide the inquiry of this research. They are:

- What functions does sound-to-picture perform in audio-visual semiosis?
- What semiotic roles does sound-to-picture fulfill in non-fiction film?

The second of these questions will entail, as we will see in the thesis, an engagement with non-fiction as 'context of situation' and 'context of culture', in order to identify the emergent constraints on sound-to-picture. In addition to these

questions, the thesis also aims to understand the following: How can technical and traditional resources of sound-to-picture be deployed in social semiotic analysis?

1.3   Methodology

In this section, I discuss the underlying assumptions of the descriptive and analytic approach adopted in the thesis. I also explain precisely why a social semiotic approach is considered a powerful method for engaging with sound-to-picture in non-fiction film.   Finally, I discuss the mode of audio-visual transcription employed.

In this thesis I adopt a social semiotic approach to sound-to-picture. In its focus on non-linguistic modes of meaning making, the present research is closely aligned with the project of multimodal discourse analysis (hereafter MDA). As I discuss in more detail in Chapter 3, social semiotics (e.g. Hodge and Kress, 1988; Thibault, 1991; Lemke, 1995; Kress and van Leeuwen 1996) has, over the past couple of decades, attempted to engage theoretically with non-linguistic semiosis, often with success. I will not discuss these attempts any further here, except to point out that they have laid the foundations for systematic research into multimodality. As I argue in more depth in Chapter 2, the work of several scholars (Van Leeuwen, 1991, 1999; Iedema, 2001; Thibault, 2000; Baldry and Thibault, 2005) has demonstrated how sound may be incorporated into existing social-semiotic frameworks. For this reason, in addition to other reasons that I attend to in Chapters 2 and 3, social semiotics – and its aims at 'making strange' the familiar (*ostranenie*, for the Russian Formalist school) – informs the analytic and theoretical engine of this research.

In general terms, the method and analytic procedure that is adopted to investigate non-fiction film sound-to-picture is based on the notion of film as audio-visual *text*, and on the practices of *production* involved in the creation of such texts. The approach incorporates practical, technical and historical perspectives on sound-to-picture into a social semiotic model for analysing and understanding how sound is configured in audio-visual texts. This not only enables a richer understanding and

appreciation of audio-visual semiosis than would be possible by attending, by purely 'formal' semiotic means, to audio-visual texts. That is to say, the technical and traditional resources outlined in Chapters 4 and 5 do not function merely as 'context' or 'thick description' for an otherwise formal analysis. The incorporation of a practice-based perspective into a social semiotics of sound-to-picture also provides the *basis* on which analysis, and interpretation, may proceed. By attending primarily to the deployment of technical resources in audio-visual semiosis, the resources become semiotised as descriptive tools. In addition, and related to this point, the use of technical terminology performs the dual function of introducing analysts and non-professional listener-viewers to the practices that are central to the production of audio-visual texts, and enabling sound practitioners to gain 'access', by way of familiar terminology and concepts, to otherwise unfamiliar analytic territory. In short, it goes some way toward bridging the gap between practice and theory, and thus between practitioners and academics.

I do not aim to provide a descriptive account of the conventional uses of sound in non-fiction filmmaking; in other words, a taxonomy of sound-to-picture. I believe such an enterprise would be of value, but my aim here is different. That aim is to address the differences in the *functioning* of deployments of sound in fiction and non-fiction filmmaking; that is, to aim to understand the ways that the contexts of fiction and non-fiction *constrain* what meanings can be made with sound-to-picture. I start from the assumption that there are no 'formal techniques' proper to non-fiction that are not practiced in fiction filmmaking, including those of sound-to-picture. On this view, typologising the practices of sound-to-picture in non-fiction would yield a descriptive account, but could not illuminate what functions of sound were proper to non-fiction. That is not to say that there is no descriptive aspect to this research. Indeed, a prerequisite of analysing how the different contexts of fiction and non-fiction constrain the semiotic potential of sound-to-picture is the description of how sound is configured as a component in the audio-visual textuality of film. To that end, two analytic dimensions are specified on which to first analyse then compare the functional differences in deployments of sound-to-picture across the two contexts. These were themes that emerged from of

preliminary analysis of a wider data set. Those dimensions are the construction of *agency* and *subject position.*

While the textual strategies of constructing agency and subject position are specifiable dimensions of filmmaking practice that effectively narrow the focus of the research to a manageable task, they are also much more than that. They are ways of understanding the semiotic potential of sound-to-picture, for those textual strategies are at the core of media constructions of power and solidarity. With agency the media point the finger. They specify not only the causes of an event, but also the effects: with whom do we side? At issue in the construction of subject-position is the ways in which we as listener-viewers are enabled access to, and precluded from 'getting close' to, certain represented participants, actions and so on.

1.3.1 Data

The audio-visual texts selected reflect the analytic focus: they are representative of fiction and non-fiction texts. The texts presented in the analyses are four of a corpus of 30 film texts that were analysed (not to the same extent as those presented here) in a preliminary fashion. From this initial engagement with the corpus, two themes emerged: the audio-visual construction of *agency*; and that of *subject position.* The four texts selected for further analysis – two fiction and two non-fiction – were chosen because they exhibited complex constructions of subject position (*Stalker*; BBC News 24: 'America Under Siege') and interesting constructions of agency ('America Under Siege'; 'Thank You For Smoking';*Commandante*). The non-fiction texts comprise: *Commandante* (Oliver Stone, 2003); and a TV news text originating from the September 11th 2001 BBC News 24 broadcast of 'America Under Siege', a critical reflection on the attacks on the World Trade Centre and the Pentagon. The fiction texts are *Thank You For Smoking* (Jason Reitman, 2006); and *Stalker* (Tarkovsky, 1979). Additionally, further analyses were carried out in the later stages of research, and these were to serve as a point of comparison with the focal data (*The Corporation*, Mark Achbar and Jennifer Abbott, 2003; *9/11*, Jules and Gedeon Naudet, 2001).

9

1.3.2 Conditions of analysis

I engaged with the film texts in several acoustic environments, through professional 'flat-response' monitors (Alesis Monitor 1) as well as through professional headphones (AKG K271). The frequency response of the monitors enabled the detection of very low frequencies and very high frequencies, though in no cases were the sound tracks of a quality that could not be identified on domestic hi-fi equipment. Listening through headphones enabled the detection of very subtle stereophonic information, in addition to some rather more obvious stereo effects, such as 'panning'.

1.3.3 Transcription

The transcription follows a very simple procedure. In contrast to the kind of transcription developed by Thibault (2000) and Baldry and Thibault (2005), the method of transcription adopted here is based on *verbal* description of the sound and image tracks. The use of a notation system of the kind advocated by those authors mentioned above (see Chapter 2 for further discussion) is not necessary in the present thesis for two reasons. Firstly, the transcription procedure involves specifying the technical resources identifiable in the data. These are already 'shorthand' terms: 'panning', 'reverb', and so on. Secondly, the use of verbal description of the sound track enables *accessibility* – a desired outcome of the thesis. There is simply no need for an elaborate system of notational symbols for performing the kind of analysis that I do in this thesis. Symbolic notation introduces a further level of abstraction from the phenomenon itself. I discuss this in more detail in Chapter 2, in relation to Baldry and Thibault's (2005) work. Below, I briefly describe how the transcription is performed.

The data file is loaded into *VirtualDub*, a freeware PC application that enables sophisticated audio-visual editing. *VirtualDub* affords a precision with which the analyst is able to engage with the data frame by frame. The video clips are played, and the salient points of synchronicity between sound and image (POSs) are noted.

Once the duration of the text has been established, image frames are 'captured'. There is no system for deciding beforehand the length of the intervals between the captured frames. This is instead determined according to the audio-visual 'density' of each text. Once the interval duration is decided, however, an equal length between the frames is strictly maintained.

The frames are arranged vertically on the page (landscape orientation). To the left of the images is the 'time' column, specifying the duration of the text in units of one second. To the right of the images is the 'sound track' column, in which I specify the technical resources deployed and the POSs between sound and image tracks. The transcripts are to be read from top to bottom, and synchronicity is to be construed 'horizontally'.

1.3.4   Challenges to audio-visual analysis

Despite the above rationale (and that argued in Chapter 3) for adopting a social semiotic perspective, there remain a number of challenges for the researcher to overcome in pursuit of the analytic 'stability' necessary to developing a theory of audio-visual semiosis. Questions at the fore include: what should constitute the basic units of multimodal analysis?; and, what role are descriptions of perceptual properties of semiotic phenomena to play in the construction of a multimodal analytic framework? Since such concerns are at the core of contemporary multimodal theory, it follows that in theorising the role of sound in the audio-visual representation of factual film the researcher is faced with the same issues.

The role of perceptual description

This is particularly the case with regards to the question of perceptual description, since a fundamental concern in many discussions on the semiotics of sound is precisely that of sound's temporality (van Leeuwen, 1999; Thibault, 2000). How, more precisely, are we to 'capture' and typologise sound when it is a widely accepted empirical fact that sound is fleeting? We are unable, unlike with images, to 'take a long, hard look' at sound, because we experience it as ephemeral and

11

dynamic. There are two ways of responding to this concern. The first is to argue that the semiotician is no more required to explain how typologising a continuously varying mode is possible, than a linguist would be if confronted with a similar question regarding how she is able to typologise the sound stream of speech into the discrete units known as phonemes. Such practice is widely uncriticised – to the point that such typologies appear as 'natural' and inherent – yet it remains a cornerstone of modern linguistics.

But my response to concerns over the relationship between perception and semiosis is not only to dismiss them by likening apparent 'blind spots' in the practices of linguistic and non-linguistic semiotics. It is also – and this is the second response – to point out that, as in linguistic practice, typologising a phenomenon that displays continuous variation (see Lemke, 1999) is a necessary abstraction that serves as a heuristic principle: a basis on which analysis and interpretation *through the medium of language* can proceed. As above, this means that the typologies proposed in an auditory semiotics, as in linguistics, do not derive from distinctions that are objectively 'there' in the phenomenon of sound, or indeed in our experience of it, but instead must be understood merely as 'footholds' necessary to semiotic analysis and interpretation. In short, typologies are 'ways of seeing' particular phenomena. As such, a semiotic typology of sound would not be *bound* to all of the laws of human perception – such as sound's continuous, topological nature – but would be free to connect with those perceptual observations that are useful to comprehending how sound can mean in a given context.

As Murray Schafer (1977) and Theo van Leeuwen (1999) have argued, the meanings that can be made with sound, while socially, culturally and historically dependent, often have close ties with our perceptual experience. For example, van Leeuwen (1999: 12-34) extends Schafer's (1977) ideas and proposes a semiotic framework in which sound is configured in terms of auditory perspective: on this view, a 'soundscape' can be considered separable into the three layers of *figure*, *ground* and *field*. Such a typology is convincing exactly because it appeals to our own (not only aural) perceptual experience of space and sound; but it is not *bound* to the laws of aural perception since it imposes strictly bounded categories on our continuous, topological experience of sound. In our experience of the world, of

course, there are no such sudden jumps in our perception of sound and space, and thus no such objective distinctions between what we experience as *ground* and what as *figure*. Rather, in using the primarily typological meaning system of language in theorising topological experience, we necessarily carve what is essentially a continuum into discrete units. Ultimately, for the reason that a typology of a given social use of sound is essentially a *translation* of the sonic mode into the linguistic, the linguistic differentiation of reality into discrete units is an enabling condition of a semiotics of sound.

As I have been arguing so far, a semiotic study of sound's role in audio-visual factual film need only make connections with certain aspects of human aural perception. It is simply unnecessary to take an 'all or nothing' position with regard to incorporating aural perception into social semiotics for the reason that a social semiotic theory of any cultural phenomenon is not a theory of perception *per se*. Rather, it is a theory intended to illuminate how a particular culture understands and makes meaning with a specific resource – a theory of semiotic perception, so to speak, which of course would be selective in the connections it makes with human aural perception.

Units of analysis

The second question concerns how we should define the basic unit of analysis in a semiotic theory of sound. As Nicholas Cook (1998: 43) argues, this should not be answered in a definite fashion, since to define a singular analytic unit is to presuppose the overall functioning of a multimodal text – a situation best avoided at the early stages of investigation into a relatively new area of semiotic study. As Cook (1998: 143) points out in his provocative exploration of multimedia, *Analysing Musical Multimedia*, 'the identification of the functional components [...] becomes not so much a prerequisite for analysis, but rather an analytical outcome'. However, there is a need to assemble an analytic toolkit, no matter that these tools are 'merely' heuristic. Accordingly, analysis can only proceed on the basis of well-defined heuristic categories through which to 'reconstruct' the data under discussion, helping to illuminate certain desired aspects of that data.

There is a tendency in film- and media studies to unquestionably allow the

13

visual track to determine the analytic units: for instance, the use of the visual 'shot' as a governing entity for all other semiotic resources studied (cf. Thibault, 2000; Iedema, 2001). Similarly, when utilising modern multimedia technology to facilitate semiotic analysis, the tendency might be to segment the text into 'frames', perhaps the result of a desire to be as micro-analytically rigorous as possible given the availability of technologies which afford such precision. But in multimodal texts, which are not necessarily visually oriented, such visually-biased practice might hinder insight into the composition of the text under consideration.

Cook's (1998) favoured type of unit is the *instance of multimedia* (hereafter IMM). However, this is a rather generally specified unit: its temporal boundaries are unclear and it does not specify the precise moments at which sound and image appear to meet – the audio-visual 'hits', as Cook terms them. For this reason, and due to the dynamic nature of audiovisual media, the units investigated in this thesis are time-based: the texts are segmented into recognisable *points of synchronicity* (POS; Chion, 1994). The criteria for what determines a POS are not based on an 'objective' measurement, but are based on an *intuitive response* to the data as an experienced practitioner and listener-viewer. This is not to be understood as a privilege of sound professionals. Salience is cultural, and recognition of the most salient moments of sound-image alignment is constrained in similar ways for both professionals and non-professionals alike. As Baldry and Thibault (2005: 183) argue, 'it is not necessary to adopt totally externally or objective measurement criteria. Instead, criteria based on perceptual salience are preferred'. As they go on, multimodal transcription, and thus analysis,

> is not concerned with etic criteria of an objective physicalist nature as obtained by some kind of mechanical measuring apparatus. Rather, *perceptually salient units* must be discovered and determined, as Pike (1967: 37) expressed it, during the preparation of the transcription. The analyst of multimodal texts is thus interested in how perceptually salient features in such events contribute to the meaning-making process of that event. In this *emic* point of view, the analyst is concerned with the identification of units that are perceptually and semiotically salient for the members of the culture in question. This is a consequence of the fact that multimodal transcription is *meaning-based*. Given that meaning is always relative to an observer or participant – an agent – it follows, of course, that the meaning-making patterns in the text can be construed in different ways

by different *participant-observers*.

Therefore, rather than analysing the sound track for its objective 'peaks' and 'troughs', and mapping where these correspond with salient visual events in the image track, identifying POSs not only *can* be an interpretative practice, but actually *ought* to be.

1.3.5 Analysis, subjectivity and praxis

The approach adopted here is an entirely qualitative one. The analysis of the interaction of sound and image in audio-visual texts involves a very detailed analytic engagement with the data in order to attend to the multiple dimensions of how both sound and image – as material phenomena and semiotic modes – may integrate to make meaning. A quantitative method may be useful in attending to the acoustic characteristics of sound, such as how the frequencies of the identified sound events are distributed and configured in a sound track. It may also be crucial in measuring such frequencies precisely, by for instance engaging in spectrum analysis. But given the focus here on salient perceived effects of the sound track, and the salient differentiation of sound characteristics in a mix of sounds, the precision of such measurement would be unnecessary.

The type of audio description that is deployed in this thesis is concerned to establish categories that deal with how a sound is used in an audio-visual context. Perceptions of such 'use' are constrained (not in a negative sense of the word) by the practice-oriented descriptive framework (see Chapter 5) that precedes the analysis. Armed with technical resources of sound-to-picture, such as *equalisation*, the analyst approaches the data with the differentiation among frequency bands (here, categorised simply as *low*, *middle* and *high*) serving as a 'guide' that to a large degree *structures* their perception of the sound track. As the framework is developed in future research, the need for more precise measurements will be more pronounced. At this stage, however, it is sufficient that audio description is informed by less 'delicate' measurements. Another reason for allowing the

'subjective' interpretation of analysts and listener-viewers guide the description is that such delicate differences in acoustic character may not be perceivable by the average listener-viewer. In deciding not to deploy precision equipment in measuring frequential characteristics of sound, I aim to both sharpen the listener-viewer's perceptual facility and to attend to an assumed, already extant facility by constructing quite 'basic' analytic and descriptive tools.

There are, of course, areas in which precision equipment may be of no use to analysis. For example, differentiating the sound track into distinct perspectival planes can only be accomplished subjectively. I do understand this as problematic to a degree: after all, how generalisable can such a procedure be if we cannot be sure that others will interpret a sound track in the same, or at least a similar, way? However, such a differentiation reflects my own engagement as a critical listener-viewer with a given audio-visual text. The aim is not to be in a position to claim that such a sound track *is* structured in a particular way *objectively*. Rather, it is to construct a framework that may be deployed by others 'as is' or modified to a small or large degree, that will enable them to engage confidently with the integration of sound and image. It is, in essence, an approach to the semiosis of sound-to-picture that enables further meanings to be made, either in the form of audio-visual texts (as might be one 'response' from a practice-oriented reader) or in the form of semiotic interpretation and discussion.

Where this might raise the objection that this essentially subjective basis is a failing of social semiotic methods for analysing non-linguistic phenomena, I consider this both a trait of all qualitative analysis and a *benefit* of a social semiotic approach. As Iedema (2001: 186) agrees,

> Social semiotics promotes detailed analysis, but its starting and end points are about situated praxis [...] It acknowledges that the analyst's own reading position is likely to guide her or his interpretations, but it sees that as a strength rather than a failing. Analysis is a sociopolitical relevance, not some theoretical abstraction.

This view is shared by Thibault (1991: 4), for whom social semiotics is to be considered 'an intervention in the theory and practice of semiotics'. As he puts it,

> such an intervention starts from the praxis-oriented view that our practice as analysts and theorists of the social meaning-making practices and their textual products in our own and other social semiotic systems is itself a set of social meaning making practices just like those we study and analyse. We are not 'above' or 'external to' the meanings and social practices that constitute the 'object' of our theory making. [...] Our theories and analytical practices are a part of the contextual relations and dynamics of the social meaning making practices we analyse. To theorise about or to analyse these means to interact with them.

Indeed, my own status as a sound practitioner must be understood not only as 'informing' my approach in a variety of ways, across all dimensions, but also as *constitutive* of the ways in which I analyse audio-visual texts and theorise audio-visual semiosis. For instance, the analytic procedure developed and deployed in Chapters 6 and 7 is undoubtedly influenced by the ease with which I am able to differentiate between sound and image as *media*. I present such differentiation as fairly unproblematic, and demonstrate the basic methods for 'prising apart' sound and image meanings. That is not to argue that such a procedure is simple. In fact, it can be rather complicated and, at times, laborious. Rather it is to argue that my own practical background in sound-to-picture entails that I distinguish the sound and image tracks of an audio-visual text primarily on the basis that, in technical practice, sound and image *are* easily differentiated.

Additionally, the descriptive framework and analytic approach I adopt has relevance to my identity as a practitioner, to the extent that the framework is itself a 'textual product' of meaning making, and represents how *I* approach sound-to-picture in practice. The intuitive readings of audio-visual texts I proffer in the analytic chapters, too, represent an 'interface' between my practical orientation to sound-to-picture and my analytic engagement as a listener-viewer. That is not to claim that I am always critically engaging with audio-visual texts. If that were the case, I would not be able to maintain a sensitivity to the emergent semiotic totality of a film or television programme. I would not feel moved, persuaded, perplexed, anxious or intrigued by the integration of sound and image if I were always ready

17

with a repertoire of analytic categories at hand, 'typing' my experience of a film into discrete units. It is rather to claim that a practical perspective is always guiding my interpretations and engagements with film and television programmes, and that my critical approach in this thesis, as well as in my everyday existence, is necessarily inextricable from this orientation.

1.4   Structure of the thesis

In this final section, I present an outline of the remainder of the thesis, with reference to the main functions that each chapter fulfills.

In Chapter 2, I present a review of the relevant literature. Given the focus on the *practice* of sound-to-picture, in addition to audio-visual *text*, this review entails a discussion on two, quite distinct literatures. The first part discusses the relevant works of the, broadly defined, film studies literature. This is to present the concepts that will be pertinent to the analyses: *point of audition, point of synchronicity* and *synchresis*. Importantly, the discussion introduces the notion of auditory perspective in relation to point of audition. The section ends with a brief discussion on the value of trade literature to the research focus. The second part focuses on the relevant works within the field of MDA and social semiotics that aim to engage analytically with sound-to-picture and audiovisual semiosis. The aim of this section is twofold: to situate the present thesis within the field; and to provide a critical discussion of previous social semiotic research on sound and image.

Chapter 3 deals with the theoretical foundations and provisions of social semiotics and MDA, that enable the description and theorisation of sound-to-picture and audio-visual semiosis. After a discussion on social semiotics, I turn to explore the concepts of mode and medium, ultimately defining how each will be deployed in the thesis. I then move onto a discussion of the relata of audio-visual semiosis: that is, presentational, orientational and organisational meaning. Within this discussion, I describe how sound may produce such meanings. Once explained, I turn to defining the key relational mechanisms by which sound and image integration may be described and analysed. Finally, I present the concepts of 'context of situation'

18

and 'context of culture', concept I invoke in order to explain the emergent constraints on audio-visual semiosis in non-fiction film.

Chapter 4 is concerned to present those concepts pertaining to the traditions of sound-to-picture. These are formulated as 'resources', and described in their applicability to a semiotics of non-fiction sound-to-picture. After demonstrating how we may conceive of traditions and conventions as resources for making meaning, I move onto the first core section: an explication of 'sync-sound'. Here, the semiotic potential of synchronised sound and image is explained. I then discuss the central features of early sound-to-picture: narrative versus topographic sound; the perceptual fidelity model (thus contextualising POA sound). Finally, I discuss two 'Voices of resistance': the practice of asynchronous sound advocated by Eisenstein, and 'direct sound'. All the above issues are related explicitly to non-fiction film sound-to-picture.

In Chapter 5, I present the technical and technological resources of sound-to-picture that are central to the analyses. This is an 'extended glossary' of techniques and technologies of sound-to-picture, a chapter that may be returned to with ease during the analytic chapters. I categorise the resources under the 'four dimensions' of the sound track: Temporal, Spatial, Frequency Range, and Dynamic Range. The resources are: Editing, Panning, Phase-inversion, Room tone, Reverb, Equalisation and Loudness level.

In Chapter 6, I turn to examine the fiction-film sound track in terms of how agency and subject position are constructed by means of sound-to-picture. I deploy the resources of Chapters 4 and 5 as part of a descriptive framework, and the conjunctive relations presented in Chapter 3 as a means to explain the ways in which sound and image are configured to accomplish: (1) the construction of an audio-visual metaphor, and the specification of a human agent; (2) the construction of a disorienting POA. Additionally, the chapter demonstrates the proposed analytic procedure for doing sound-to-picture analysis. I conclude that certain ambiguities of perspective, and thus subject position, are 'resolved' by the narrative fiction film

context of situation.

Chapter 7 is concerned with sound-to-picture in the non-fiction film, in particular how agency and subject position are constructed in a TV news text and a documentary film. Again, the technical resources and conjunctive relations are deployed in analysis, this time in terms of (1) how a human agent is specified as a cause of human suffering; (2) how identification with the sufferers is realised through POA; (3) how a subject's 'interior state' is constructed; and (4) how 'subjectivity' and 'identification' is constrained by non-fiction as context of situation. I conclude with the statement that in order to properly understand audio-visual semiosis – and thus sound-to-picture – in non-fiction film, we must attempt to construct a comprehensive model of the manifest contextual constraints on meaning making.

Chapter 8 takes up the task proposed in the previous chapter, and attempts to build a model of non-fiction film as context of situation in order to delineate the functioning of sound-to-picture in that context. The definitive features of the (prototypical) non-fiction film are presented, and then formulated as 'situation' – that is, they are categorised as 'field', 'tenor' and 'mode'. Once accomplished, I deploy the model in comparative analyses of fiction and non-fiction constructions of subjectivity, attempting to illuminate how configurations at a metafunctional level co-pattern with those at a 'situation' level. Finally, I determine that in order to understand just how non-fiction as 'situation' is maintained as a predictable and regular *type*, an understanding of the context of culture is necessary. To that end, I briefly describe how such maintenance is achieved through institutional constraints and trade principles.

In Chapter 9, I conclude the thesis by way of a discussion on the limitations of the analytic procedure and methodology. I also point to avenues of further research that have arisen in the course of the project. Finally, I evaluate the contribution that the thesis makes to the MDA field.

Chapter 2: Review of the literature

2.0 Introduction

Having established the aims and objectives of this thesis in Chapter 1, in the present chapter I investigate the relevant literature concerning sound-to-picture. The literature can be divided into two 'camps': first, those writings from the film-studies and trade literature, concerned with a broad array of subjects. These are, for example, the practical, institutional and technological dimensions of sound-to-picture, almost entirely concerned with fiction filmmaking. The second comprises the research conducted within the fields of social semiotics and multimodal discourse analysis (hereafter, MDA), which takes as its main focus of study how a given semiotic 'mode' (see Chapter 3) is deployed to make meaning in a particular context. This second 'camp' includes many attempts at incorporating an understanding of sound into its already established analytic and theoretical frameworks, but only few *committed* attempts at explicating the sound track's role in a multimodal context.

Following a brief introduction to the topic of sound's status in research on multimodal semiosis, I present a discussion on the film-studies literature, broadly defined, that concerns sound directly. This section will be brief since its central topics are discussed in some detail in Chapter 4 in the context of an exposition of relevant sound-to-picture traditions. In the final section, I provide an account of the most pertinent research from the social semiotic and MDA fields, delineating the territory that the present thesis aims to develop.

The literature on non-fiction film (including only a small handful of works on sound) will be discussed in Chapters 7 and 8.

2.1 The status of sound

The very fact of this thesis' specific focus on how sound is used to function in audio-visual media suggests that it has not been considered – in the MDA research literature – to the same degree as the visual (e.g. Kress and Van Leeuwen, 1996, 2001), or indeed other modes, such as gesture (e.g. Martinec, 1998).

Presently, in the area of the semiotics of sound design and music in multimodal contexts we have only Van Leeuwen (1999), Iedema (2001), Cook (1998), Baldry and Thibault (2005) and my own work on sound in TV and film (Constantinou 2002). While the under-representation of certain other potential modes of semiosis might be expected, the dearth of research into *sound* in multimodal contexts is particularly striking for two reasons. Firstly, there are no existing semiotic 'grammars' that can be easily tailored to fit sound; sound is a complex and problematic phenomenon that cannot be understood simply by analogy to, say, linguistic or so-called 'visual grammars' (cf. Van Leeuwen, 1999; Kress and van Leeuwen, 1996; O'Toole, 1994). Since this is the case, it is imperative to develop a medium-specific way of thinking about sound.

Secondly, sound is central in our contemporary media environment – in film, TV, websites and in 'real' spaces such as museums and domestic settings – and has been throughout the last century. My call for sound to be explored in a critically engaged and systematic way should not indicate that I believe its study more important than other potential semiotic modes – smell, for instance – but rather that I believe the sonic to be as central to today's mass media as the visual.

Having argued that point, I should state that there *is* a veritable well of research and writing on sound in film and animated cartoon in the 'film studies' literature (though very little of it deals with non-fiction, as I discuss in this chapter). As Sergi argues (2004: 58)

> The sheer amount of writings on and around film sound that has appeared over the Internet and in 'technical' magazines is quite overwhelming. From countless accounts of aesthetic and technological contribution of individual filmmakers to a staggering drive for historical accounts of film sound in all its facets, this enthusiastic, at times crusade-like attitude of all those involved has developed a remarkable resource for scholars of sound and can now count on a steady readership.

However, as I argue below, many of the concepts and observations presented in that body of work are derived impressionistically. Moreover, they are extremely wide-ranging, and the engagement with any one orientation to sound (semiotic, for example) is sporadic. Therefore, it is not so much a general lack of material on sound that is of concern here. Rather, it is that there are too few examples of

research into sound that aim to catalogue systematically and critically how meaning is made in the context of the image. The task for a researcher oriented in social semiotics and multimodal discourse analysis is to incorporate the existing ideas into a more systematic study of sound in an audio-visual context.

The final point I want to make here is that sound's take up in the academic literature – whether in the fields of film-, media- or communications studies – has its corollary in the status of sound in the film and television industries. In both contexts, there is no prominent discourse on the role of sound-to-picture, both in terms of its semiotic functioning and its aesthetic contribution to filmmaking practice. It is also true that even though there is a huge amount of labour involved in production and post-production sound-to-picture, the *economic* status of sound in relation to other filmic processes reflects how it is often regarded in the context of the total film effect (Sergi, 2004). As mentioned in the introduction, sound-related tasks are considered by the Directors' Guild of America to be 'technical' only, and consequently no sound-related occupation is allowed to appear in a film's head credits, alongside the more 'creative' roles of film editing and production design. Correspondingly, sound is considered to perform mainly 'background' duties; it is, in a Hallidayan (1982; Iedema, 2003) sense, *automatised* i.e. backgrounded in order to augment the foregrounding of other modes (e.g. visual or spoken interaction; cinematography).

2.2 Sound in the film-studies literature: relevant concepts

This section will deal with those writings in the film-studies literature that address sound-to-picture as practice and aesthetic effect. For my immediate purposes, this will include cultural-historical research; treatise contained in trade journals; and the (often psychoanalytically oriented) studies on sound-to-picture concerned to illuminate its latent potential as a semiotic mode (though such social semiotic terminology is not deployed).

Sergi (2004: 56) identifies two 'busy periods' in the study of film-sound: the first concerns the so-called 'coming of sound' to film in the 1920s and 1930s; the second

23

concentrates on the 'Dolby period' – what Charles Schreger (1978) terms 'the second coming of sound'. This second period is characterised by an increased creativity in Hollywood film sound practice, comcomitant with the collapse of the studio system and the new-found independence of filmmakers and editors from the Classical narrative-film paradigm. Accordingly, the most interesting and useful accounts of film sound are concerned with either early film sound (spanning the 'silent' film era to the 'transition period'; approx. 1907-1938) and the post-Classical output of Hollywood film and European cinema (from the 1970s onward for Hollywood; European cinema was generally less dependent on the Classical paradigm far earlier).

The most prominent scholars of film sound are Michel Chion, Rick Altman, and, more recently, James Lastra. Here, I want to make some general points about their research and its most salient concerns. The writings that are most pertinent to the present thesis are those concerning the early period (Altman, 1992; Lastra, 2000) and the post-Classical period (Chion, 1994).

2.2.1 Chion

Michel Chion's (1994) book, *Audio-Vision: Sound on Screen* has become something of a classic in recent years. Chion himself has become, in Sergi's (2004: 67) words, 'one of the most outspoken theoreticians on sound'. Chion's take on film-sound is a highly original one, eschewing the cultural-historical register adopted in much writing on film sound (e.g. Altman, 1992; Lastra, 2000). His broad focus is the *phenomenon* of sound-to-picture, most importantly the perceptible integration of sound and image by means of audio-visual synchronisation. As he states in the outset of *Audio-Vision*, the main objective of his work 'is to demonstrate the reality of audiovisual combination – that one perception influences the other and transforms it.' (1994: xxvi) Two important concepts are introduced and developed in Chion's work, that are to enable understanding of, and analysis of, the integration of sound and image in film: *synchresis* and the *point of synchronicity*.

For Chion, synchronisation of sound and image is instrumental in the listener-

24

viewer's acceptance of a film's overall 'reality' effect – he terms this the *audiovisual contract*, the willingness to accept the truth of whatever sound has been joined with whatever image (by 'truth' I mean that what we hear is a consequence of what we see). Along with this comes the perceptual effect of synchronisation: *synchresis* (Chion's amalgam of *synthesis* and *synchronisation*). For Chion, synchresis 'is an important factor in film in how it manages to glue together entirely unlikely sounds with unlikely images' (1994: 54). As he expands, *synchresis*

> makes dubbing, postsynchronisation, and sound-effects mixing possible, and enables such a wide array of choice in these processes. For a single body and a single face on the screen, thanks to synchresis, there are dozens of allowable voices.
> (1994: 63)

The same holds true for all other types of sync-sounds, including footsteps (almost always re-recorded) and other sound effects. This potential for verisimilitudinous illusion is the basis for much manipulation of film and television sound today.

The other contribution Chion makes that is relevant to this thesis is in his concept of *point of synchronicity* (POS). POS refers to the salient moments at which sound and image 'meet in synchrony'. The examples of POSs that Chion provides are as follows:

- As an unexpected double break in the audiovisual flow, a synchronous cut in both sound and image track. This is characteristic of external logic, frequent in *Alien* for example.
- As a form of punctutation at the end of a sequence whose tracks seemed separate until they end up together (synch point of convergence).
- Purely by its physical character: for example, when the synch point falls on a closeup that creates an effect of visual fortissimo, or when the sound itself is louder than the rest of the soundtrack.
- But also by its affective or semantic character: a word of dialogue that conveys a strong meaning and is spoken in a certain way can be the locus of an important point of synchronisation with the image.

(Chion, 1994: 59)

The issue arises, at this point, of the subjectivity involved in identifying POSs. For

Chion, the salience of POSs 'generally obeys the laws of Gestalt psychology' (1994: 58). As I deploy it in this thesis, a POS may be clearer to some listener-viewers than others. As a professional sound practitioner, I have developed a variety of quite intuitive procedures for identifying POSs. This may seem problematic (cf. Baldry and Thibault, 2005: 183; see Chapter 3 for discussion). However, the analyses of Chapters 6 and 7 go some way towards demonstrating what may constitute a POS in audio-visual analysis. These chapters reflect the thesis' aim to contribute to the development of an audio-visual 'literacy', and so to train listener-viewers to identify POSs themselves. Often, POSs can be identified by firstly noticing clear, punctual visual events. Given the average listener-viewer's facility with identifying the events of visual data, the non-professional may find this a reliable way of mapping out the sync points of an audio-visual text. In those cases where the POSs under discussion do not entail overt *punctuality* – for example, a visible gunshot or punch – I hope to demonstrate the range of more subtle types of POSs that may be analysed: for example, moments when subtle changes in an otherwise steady 'stream' of sound co-occur with obvious or subtle visual edits.

The main criticism of Chion's work, which I share to an extent, has been its lack of methodological and analytic rigour. Although he does propose some quite extensive suggestions for analysing sound-to-picture, many of Chion's semiotic readings of audio-visual texts are made entirely impressionistically. Chion is certainly poetic in his insights and hypotheses, but they remain only that: hypothetical propositions in need of theoretical flesh and empirical bones. Yet his descriptions and theories of the interaction of sound and image resonate profoundly with *practitioners*, such as the renowned sound designer and film editor, Walter Murch, who contributed the foreword to the English translation of *Audiovision*. The support of practicing sound editors lends Chion an invaluable level of credibility within that community and indeed within the, broadly defined, 'academic community'. *Audio-Vision's* presence is almost guaranteed at the top of reading lists for the numerous 'Sound Design' courses emerging in the UK and US. There is, of course, disagreement within the academic community over how one should take Chion (that is, how his ideas may be deployed in other's research). The central reason for his success is primarily one

of accessibility; an accessibility certainly not hindered by Chion's independence from existing specialist terminologies and theories. So whereas academics consider Chion's work too independent from existing theoretical and methodological frameworks and procedures, it is this very independence which renders Chion's ideas accessible.

My own opinion as regards Chion's value in socially oriented research is that his ideas resonate so deeply with practitioners (myself included) that they deserve to be considered, deployed and tested in analyses. Moreover, a critical audio-visual analysis will benefit from 'mapping out' the points of contact between Chion's concepts and those of social semiotic research. This is one way in which Chion's work contributes to this thesis, and one way in which I hope to continue his general project of increasing audio-visual literacy.

2.2.2 Altman

Rick Altman is perhaps the most prolific researcher on film sound theory and history. His contribution, as it is relevant to this thesis, concerns Altman's historical scholarship and a particular concept that he introduces called *point of audition* (POA) sound. In Chapter 4, Altman's historical observations are, with James Lastra, at the basis of the discussion on traditional resources. I will not deal with those here. His notion of POA sound, however, along with Chion's criticisms of it, will be outlined below.

POA

As Altman formulates it, POA sound represents the aural perspective of a specific character (or set of characters). It is, in short, the aural equivalent of the point of view (POV) shot. Both POV and POA are concerned with the representation of a character's subjectivity. Altman gives the example of the Hollywood sound-film *Only Angels Have Wings* (Howard Hawks, 1939), in which the actors, Cary Grant and Jean Arthur, watch – and listen to – another character's plane flying off into the distance. The sound of the plane decreases in loudness and an accompanying increase in reverberation signifies – on the perceptual fidelity model – the

characters' auditory perspective. While scale matching insists on the strict matching of image and sound perspective, POA sound constructs a perspective from *within the diegetic frame* of the film image (1992: 60). In Altman's words, POA sound

> relates us to the narrative not as external auditors, identified with the camera and its position [...], nor as participant in the dialogue [...], but as *internal* auditor. (1992: 60; my emphasis)

The implications of POA sound on the 'subject positioning' of the audience member are profound. The listener-viewer is positioned as one who is 'asked not to hear, but to identify with someone' who will hear for her. (Altman, 1992: 60) As will be demonstrated in later chapters, this technique of constructing 'identification' with a character within a diegesis, whether fictional or non-fictional, has a major role to play in contemporary sound-to-picture practice. For the reason that the construction of auditory perspective is less noticeable to the average listener than that of visual perspective, POA sound – like the point of synchronicity – is a powerful method of *naturalising* identification (e.g. sympathy) with certain characters and actions to the exclusion of others.

But, as Williams (1980: 58) argues, each sound recording has an in-built point of audition. In his words,

> My contention is that in sound recording [...] the apparatus performs a significant perceptual work for us, isolating, intensifying, analysing sonic [...] material. It gives an implied physical perspective on [...] sound source, though not the full, material context of everyday [...] hearing, but the signs of such a physical situation. We do not hear, we are heard. More than that: we accept the machine as organism, and its attitudes as our own.

Whether we are to identify with a particular character by means of POA sound, or whether we are positioned as voyeur, as in the modern, synthesised Hollywood film sound track, perspective is encoded into sound recording so that the listener-viewer is always 'positioned' at some particular point in relation to the represented sounds. For this reason, POA may be used to refer to not a 'preferred' point of audition, as in

that of a character in the diegesis, but rather a necessary, *spatial* one that may have nothing to do with a represented character. Chion's (1994: 89) formulation of the distinction is concise and useful here:

> Let us first note that the critics have come up with the concept of a point of audition based on the model of point of view. Here begins the problem, since cinematic point of view can refer to two different things, not always related:
>
> 1. The place from which I the spectator see; from what spatial location the scene is presented – from above, from below, from the ceiling. From inside the refrigerator. This is the strictly *spatial* designation of the term.
>
> 2. Which character in the story is (apparently) seeing what I see. This is the *subjective* designation.

As might be obvious, the spatial POV shot – given that it is the default position – is far more commonly 'deployed' than the subjective POV shot. Which is to say, a shot must always signify 'spatial location'. The spatial POV shot 'rests on the possibility of inferring fairly precisely the position of an "eye" based on the image's composition and perspective' (1994: 90). Yet there are stark differences between a *spatial* POV shot and a *spatial* POA sound. The difficulties with the *spatial* POA definition stem from differences in sound and image representation and perception. POV is an indices of camera position, made explicit by the visual framing of a film image, and is therefore a *definite* point in space with which the viewer may identify and, from which, relate to the visual action. POA sound, on the other hand, offers no such unambiguous point. Chion – to quote him at length – offers the explanation that

> the specific nature of aural perception prevents us, in most cases, from inferring a point of audition based on one or more sounds. This because of the omnidirectional nature of sound (which, unlike light, travels in many different directions) and also of listening (which picks up sounds in the round), as well as of phenomena involving sound reflection.
> Consider a violinist playing in the centre of a large round room, her audience grouped in various places against the wall. Most of the listeners, even those standing at diametrically opposite points of the room, will hear roughly the same sound, with slight differences in reverberation. These

differences, related to the acoustics of the space, are not sufficient to locate specific points of audition. Every *view* of the violinist, on the other hand, can immediately situate the point from which she is being looked at. (1994: 90-1)

In its spatial sense, then, we cannot refer to a *point* of audition, but rather 'a place of audition, or even a zone of audition' (1994: 91). A 'special case' of *subjective* POA sound which Chion identifies concerns sounds that may – in perceptual reality – only be experienced at close range. Perceiving 'indices of proximity' in sound tracks do not require a corresponding image track for ther listener-viewer to be able to identify a clear subjective *point* of audition. But, by and large, subjective POA sound, through which may identify with a visualised character in the diegesis, is not determined by acoustic characteristics – reverberation level, or even placement within the stereo field – but instead by the image composition: 'for it is the image that always creates the [subjective] point of audition, which in this case is worthy of the term *point*' (ibid.)

For the above reasons, auditory perspective cannot in itself be marshalled into making the kinds of meanings that, as many have argued (Kress and van Leeuwen, 1996; O'Toole, 1994; Thibault, 2000), representation of visual perspective may afford. Given the very vague sense of location and distance that auditory perspective can provide, the question arises: why deploy it as an analytic concept in this thesis? There is one main reason: the analytic focus vis-à-vis auditory perspective is squarely on how it is configured to interact semiotically with visual perspective. While precise points of location and distance cannot be constructed or construed with sound, there are sometimes very obvious differences between sound and image perspectival information in an audio-visual combination. It is these which constitute the focus of those parts of this thesis concerned with perspective.

In this thesis, I deploy POA in both its *spatial* and *subjective* definitions. Chion's distinction is even more significant in relation to non-fiction film, in which POA sound will very rarely 'attach' to a visually represented social actor. In this thesis, my use of the term POA is concerned to explain how the subject-positioning of the listener-viewer is constructed by means of auditory perspective. POA, then, is a

30

realisation of auditory perspective.

Auditory perspective as it is formulated in this thesis, concerns the construction of, as I have argued, a point (or 'place') of audition. This has its foundation in Gibson's (1986) theory of visual perception (deployed by Baldry and Thibault [2005: 192] in their framework for analysing visual perspective in images, but not extended there to sound). Here, I will quote Baldry and Thibault's (ibid.) gloss on Gibson's theory of the 'optic array':

> the ambient optic array is a structured arrangment of stimulus information that specifies the environment of the observers who inhabit it (Gibson, 1986 [1979]: 63). To be *ambient* at a point of observation that could be occupied by an observer means to surround that point. A point of observation is a point in the ecological space-time of the observer from which its environment can be observed. When a position is occupied by an observer [...] the *ambient array* is modified in such a way that it provides information about the observer as wellas information about the environment that surrounds the observer [...]

Clearly, this is foundational to the present discussion, as POA sound – whether considered as specific as a 'point' or as general as a 'zone' – is deployed in this thesis as an analytic tool to enable discussion on how the listener-viewer is made to *relate* spatially to sounds in a delimited auditory array (2005: 210). Auditory perspective is the concept that encompasses the configuration of listener-viewer and the sounds events of a film sound track.

2.2.3 Trade literature

Finally, there is a mine of valuable historical information contained within the pages of the film industry journals of cinema's early period, ranging from silent-film sound practices to later sound-film techniques. As Lastra (2000: 157) reminds us, the duties of sound technicians of the sound-film 'conversion' era were split between producing film sound tracks and writing treatises on how sound should be deployed in relation to the film image. As Altman suggests (1992: 49), it is proper to define these technicians as also *theoreticians*, given their insitigation of and contribution to the formative debates of the period.

31

The trade press from the early 'silent' period and from the 'transition' period was filled with articles on all aspects of sound-to-picture, written from representatives of both sound and image camps. This offers not only a rare insight into the process by which standards of realism and other representational principles are established and changed, but also into the construction of human standards of perception. As Wurtzler argues, the well-documented debate over 'scale matching' between sound and image (see Chapter 4 for discussion) offers 'a privileged historical instance in which can be read the process of establishing realistic representational conventions.' And as Altman (1992: 55) understands the same instance, this privilege allows us insight into a moment when human perception itself was in the process of a major shift: '[w]e often give lip service to the notion that cinema teaches us to see and to hear, that the media determine our very notion of reality [yet] we are rarely privileged to isolate the moment when and whereby our perception changes.'

The trade literature provides much support for the analyses of the present thesis, but it should be made clear that, where deployed, the accounts of early practitioners are sourced from secondary literature – primarily Altman (1992, 2005) and Lastra (1992, 2000). This is not only to state that I have not personally trawled the archives for such accounts, but it is also to remind that the accounts I do deploy have appeared to me as relevant on the basis of *prior* contextualisation by Altman and Lastra. Additionally, their arguments have informed to a great extent the approach adopted here to explicating and formulating the resources of tradition. For that reason, Chapter 4 owes much to these authors' rigorous and creative research.

2.3   Sound in the MDA literature

This section deals with the quite limited number of research attempts into sound in the MDA field. A scan through some of the last decade's work reveals only a few committed attempts at theorising a semiotics of sound and sound-to-picture. Often, in multimodality circles, lip service is given to sound, and the need to incorporate it more fully into existing theoretical frameworks, yet it has become something of an 'elephant in the room': difficult to reason with and impossible to ignore.

Consequently, one will most often find sound given a mention in the introductions of academic papers – as one mode among many that MDA can encompass methodologically.

That so few have attempted to analyse sound in social semiotic terms is understandable, of course: sound *is* 'difficult'. Its most prominent terminologies have primarily been associated with the quantitative technicalities of acoustics, the physical orientation of psychoacoustics, or the specialist language of musicology, a field in which the aspects of sound-as-medium that are not easily notated are ignored (cf. Middleton, 1990: 104-5). Or else such terminologies are not systematic, remaining impressionistic without recourse to an explicit theory or method.

Below, I discuss those attempts that aim to theorise sound as a semiotic mode, capable of being integrated with other modes.

2.3.1 Van Leeuwen

Of those who have made attempts to theorise sound in social semiotic terms, the most famous has been Theo van Leeuwen's (1999) *Speech, Music, Sound*. In this work, Van Leeuwen attempts to make several connections between theories of musicology, treatise on film sound, the 'soundscape' studies of Murray Schafer (1977), and social semiotics. Its relevance to the present thesis is primarily in its theory of auditory perspective, which it synthesises from the theories of practitioners such as Walter Murch, concerning the differentiation of a sound track into figure, ground and field, and the sociologist, Edward Hall (1964), concerning social distance between people. This is an extension of Van Leeuwen's work, in collaboration with Gunther Kress, on visual perspective in static images (Kress and Van Leeuwen, 1996).

Perspective

Auditory perspective, as deployed in this thesis, mainly concerns the *distance* of sounds from the point of audition (see section 2.2.2 above) represented in a sound recording. As Van Leeuwen (1999: 14) argues, 'there is no equivalent to the 'frontal'

or 'side on' angle in sound. Sound is a wrap-around medium.' Accordingly, perspective in terms of recorded sound is limited to the *lateral* dimension (sounds arranged in a stereophonic array) and that of *depth* (sounds arranged in terms of figure, ground and field). 'Surround sound' and techniques of 3D simulation and reproduction (such as binaural recording) are not considered here, as they do not constitute the most prevalent reproduction environments of film.

Below, I briefly gloss Van Leeuwen's attempt at theorising auditory perspective. Firstly, to quote Murch (Paine,1985: 357) on conceiving the perspectival structuration of the sound track prior to recording the sounds that will constitute each layer:

> The [...] thing is to think of the sound in layers, break it down in your mind into different planes. The character lives near the freeway, so you've got this generalised swash of traffic sound, but then occasionally a plane flies over: these are long, atmospheric sounds. On top of these you then start to list the more specific elements: the door closes, the gunshots, the bats that live in the attic − who knows? Isolated moments. Once you've done that, once you can separate out the backgrounds from the foregrounds, and the foregrounds from the mid-grounds, then you go out and record [...] Since each of the layers is separate, you can still control them, and you can emphasise certain elements, and de-emphasise others the way an orchestrator might emphasise the strings versus the trombones, or the tympani versus the woodwinds.

Van Leeuwen proposes that the foreground, mid-ground and background be thought of in terms of the terms *figure, ground* and *field.* What will be *figure* is to be construed as the 'focus of interest'; *ground* is the 'setting or context'; and field denotes the sounds that constitute the 'soundscape' of the listener, though, as Baldry and Thibault (2005: 212) put it, 'the listener is not expected to orient to them or to take up an explicit evaluative stance on them.'

These perspectival layers, as Van Leeuwen (1999: 23) argues, are *realised* by 'the relative loudness of the simultaneous sounds'. However, this explanation of the realisation of perspective simplifies how the layers may be differentiated in sound-to-picture practice. While it is true that loudness levels are key to such differentiation, there are a number of other resources of sound-to-picture that are

34

deployed to perspectival ends that need to be explicated. I have described these to a degree in previous work (Constantinou, 2002), where I explain that manipulation of audio frequency by means of equalisation is often used to distinguish relations between sounds events in a sound track. For example, by diminishing the high frequencies of a sound, its 'directionality' (high frequencies have a 'narrower' range than mid or low frequencies, the latter of which are entirely omnidirectional; see Moylan, 1992: 26) is curbed while its perceived loudness will be as only slightly lower. The effect would be something like hearing loud voices or music through a separating wall. I discuss equalisation in more detail in Chapter 5.

Furthermore, in many sound tracks, differentiation by means of perceived loudness – or indeed other technical means – is not so easily achieved. In the sound tracks of broadcast news 'location reports', sounds often compete in the middle band of the frequency range. In such cases, we may differentiate among layers only in terms of what is considered as significant information. In addition to this contextual consideration, sounds that are louder are not always considered as figure, even in such cases that they are also represented as close by means of an increase in high frequencies. As Moylan explains, the human ear becomes desensitised to sound with high pressure levels that are sustained for a long duration. The perceived loudness of such sounds diminishes gradually, allowing other, more punctual sounds to be heard. Additionally, listening fatigue may also affect the 'ear's ability to detect new sounds at frequencies within the frequency band [...] where the high sound pressure level was formerly present'.

Such considerations must be considered in adopting Van Leeuwen's theory of auditory perspective, as it is explicitly concerned with the *perceived* loudness of sounds in a sound track. The contingencies of sound playback in cinema theatres – particularly considerations loudness – will affect perceived differentiation between sounds, to the extent that that any semiotic theory of auditory perspective should always consider such variables explicitly.

Conjunctive relations in audio-visual texts

Van Leeuwen's contributions to semiotic concerns of sound and image have been most fully explored in an early work (1991) that engages with audio-visual conjunction. In 'Conjunctive structure in documentary film and television', Van Leeuwen is concerned to theorise the conjunctive relations of contiguous images in an edited sequence, in addition to the 'vertical' intersemiotic relations of voice-over commentary and images in documentary film. Below, I outline the article's most relevant contributions.

Van Leeuwen applies the logico-semantic relations of systemic functional theory (Halliday, 1985: 306-7) to image-image and image-sound (visual-verbal) relations. As he argues, relations of semantic expansion (elaboration, extension and enhancement) are applicable to audio-visual texts. He avoids the 'danger of projecting linguistic structures into the domain of the visual' (1991: 77) by employing a functional-semantic metalanguage to analyse both images and verbal narration. As he puts it,

> It is unreasonable, for instance, to speak of 'visual clauses' or 'verbal shots'. It *is* reasonable to expect that each semiotic, in its own way, can realise the kinds of logical relations which the culture generally, or which some socially defined domain within the culture, allows to be made actions, propositions and so on. (ibid.; my emphasis)

Van Leeuwen's focus on the verbal aspect of film sound tracks, however, renders this article only of limited relevance to the present research. His main finding is that verbal-visual conjunction in documentary audio-visual texts is mainly *elaborative* – it deals with explaining the image's meanings. But he also notes that elaboration functions in both directions – from image to sound, and sound to image. The dynamic interaction between sound and image tracks make this possible – as Van Leeuwen explains: 'Sometimes the text makes a point which is then elaborated by the visual track, sometimes the image makes a point which is then elaborated verbally.' (1991: 101) In the present thesis, I also deploy the logico-semantic relations of expansion to audio-visual texts, but whereas Van Leeuwen allows for these to be realised retroactively, the focus here is squarely on the simultaneously

aligned sound and image tracks. That is, the relations as they are applied in this thesis are deployed to account for 'vertical' audio-visual semiosis, without consideration of how sound meanings may expand the meanings of an image presented earlier.

Its limited relevance to the research aside, Van Leeuwen's article is a milestone in the theorisation of audio-visual semiosis. It certainly is the first attempt of its kind, introducing the notion that logico-semantic relations could not only be applied to moving images and image editing, but also *intersemiotically* – between the sound and image tracks. To a large extent, the present research rests on this foundation.

2.3.2 Iedema

Though not widely published in the field of MDA, Rick Iedema has made a small yet generally useful contribution to discussion on sound in audio-visual texts (Iedema, 2001). Most pertinently, this work – a book chapter entitled 'Analysing Film and Television: a Social Semiotic account of *Hospital: An Unhealthy Business*' – engages with the use of images, sounds and verbal narration in non-fiction film. While his comments on sound remain at a general level, his is an early attempt at theorising how a film sound track may be conceptualised metafunctionally. The relations that enjoin sound meanings to image meanings are also generalised, however, so the direct relevance of his work is limited here. I will briefly gloss the most pertinent aspects of 'Analysing Film and Television'.

Although I discuss metafunctional meaning, and how sound may fulfill it, in more detail in the Chapter 3, it is worth rehearsing Iedema's comments here, in order to gain a sense of how the present thesis fits in with his work. Firstly, he mediates on how music and sound track may fulfill presentational meaning (in his terms, *representation*):

> Is it ethereal church-like music suggesting spirituality and 'higher values' [...], or is it electronic and avant-garde creating associations with technology and science? Is the soundtrack about office noises, such as the shuffling of paper or the creaking of chairs, or is it about ward noises, the squeaking wheels of hospital beds, footsteps echoing through corridors, or the tinkle of

surgical tools and the bleeping of heart monitors? And is the verbal or speech track about patients, their suffering and their well-being, or is it about money, budgets and schemes aimed at balancing the books?

In this way, Iedema describes how the sound track is concerned to represent 'aboutness'. As for orientational meanings ('inter-personalising'), he – following Van Leeuwen (1999: 28) asks:

> What degree of social distance is construed between sound and the listener-viewer? Are we more likely to be impressed by the hustle and bustle of a busy hospital ward, by the swooshing pulsings of a patient's blood rate and by the technical bleeps of a heart monitor, than by the more domesticated noises coming out of the meeting room – the rustle of papers, the squeaking of a felt pen on a whiteboard, the measured speech, the creaking of office chairs, or the rattle of typing on a keyboard? The busy noises are more likely to inter-personally appeal to us than the controlled sounds coming from the meeting room.

Iedema does not make the same mistake that Van Leeuwen does regarding which sounds are likely to be salient in a sound track. That is, Van Leeuwen depends on an 'ideal' perception with which we can predict that other listener-viewers will differentiate a sound track on the same or similar lines. Iedema, instead, suggests that a perceived *excitement* will determine which sounds we attend to (or which 'appeal'), thus introducing a social dimension that circumvents the issue of auditory perception.

As for organisational meaning, Iedema does not offer much more than a brief note about how editorial rhythm 'interweaves speech, sound, movement, image editing and macro textual structuring' (2001: 192). Here, Iedema misses a chance to discuss 'continuity sound', a concept he introduces earlier in the chapter. This is a minor fault with what is otherwise a provocative and informed attempt at theorising the sound track in terms of metafunctional meaning. (It is actually only one of a few attempts at doing so. I discuss the others below.)

While Iedema's elaborations of the sound track's metafunctional possibilities (of course, in specific relation to one non-fiction film text) are useful – especially for

those not 'sonically-oriented' – he only makes a general statement concerning how the metafunctional configurations of a sound track may co-pattern with that of a synchronised image track. As he puts it:

> The point of asking all these metafunctional questions is that their answers will give us a sense of what kinds of patterns prevail. More importantly, they may give clues about how the various patterns enrich each other. For example, we know that the documentary is about doctors and hospital administrators. We also know that the orientation of the documentary is to favour the clinicians, with their longer exposures, their more dynamic appearances, and so on. In consonance with this, the textual organisation [...] is such that each of the administrative-financial sequences (whose orientation is formal and rationalising) is both preceded and followed by a highly dramatic and inter-personalising sequence about one of the patients. *The point is that these organisational, orientational and representational patterns and choices enhance and reinforce each other*. (2001: 192-3; my emphasis)

This is an insightful statement in that it does suggest an 'easy way into' questions of social semiotics and audio-visual textual analysis. Yet it inevitably suffers from a generality that could be avoided. Prior to this, Iedema presents a very detailed analytic framework concerned with understanding the hierarchy of filmic units: frames, shots, scenes, sequences, generic stages, and 'work as a whole'. Yet he does not deploy this hierarchy in the analysis, causing one to wonder why he did not instead attempt to account for the audio-visual relations that might help to determine how 'redundancy' is achieved. (Iedema does later acknowledge that redundancy is an implicit principle in his analysis.)

It may be unfair to move on without mentioning that Iedema's chapter is part of an 'introductory' text book (Van Leeuwen and Jewitt, 2001), and for that reason the chapter's generality is a necessity. My criticisms here are motivated not by a sense of frustration at 'what might have been', but a sense of opportunity. In other words, what Iedema leaves unattended in his chapter, I intend to fulfill in the present work. The later focus on understanding the ('logical') relations between sound and image in fiction and non-fiction film texts is a direct response to Iedema's chapter, which I first encountered at the outset of my research. In short, Iedema's generality inspired me to delve deeper into the connective functioning of audio-visual texts.

2.3.3 Baldry and Thibault

Finally, I introduce the work of Paul Thibault and Anthony Baldry (Thibault, 2000; Baldry and Thibault, 2005), whose research into 'multimodal transcription' extends way beyond the scope that the phrase suggests. In fact, its value for the present thesis does not concern transcription at all, but rather the framework that it establishes for analysing the relations of sound and image in film texts.

Thibault's (2000) chapter, 'Multimodal textual transcription of a television advertisement: theory and practice', was one of the very first attempts at theorising the integration of sound and image in a systematic way; i.e. one that attends to sound and image as semiotic systems that we combine to make meaning. Thibault's attention to the sound track, in particular how it integrates with the image, is unprecedented. Van Leeuwen's (1991, 1999) work touched on some of the ways in which sound and image can be meaningfully combined, but Thibault's focus on the 'multiplicative' functioning of combining semiotic modes, (whereby '[t]he meaning of the text is not the result of merely adding the meanings of one resource – the soundtrack, say – to those of another, such as the visual image [but] is the result of the ways in which the combination of the two – sound and image, for instance – produces a new patterned relation which cannot be reduced to the sum of the two, seen separately') speaks to the urgent need for a general model of sound-image interaction in the multimedia and hypermedia age. In Lemke's (1998a) words, 'this combinatorial semiotic principle provides not just a theoretical framework, but an analytical engine for investigating multimedia semiotics.'

Baldry and Thibault (2005) enlarges the scope of Thibault (2000) beyond TV advertisements. It is concerned with a wider range of multimodal text types including print media, web pages and film texts. Much of the analytic tools of Thibault (2000) concerning the sound track are included and developed in Baldry and Thibault (2005). For that reason, the following discussion concerns – and the following crticisms are directed at – the latter text.

As I discuss in detail in Chapter 3, Baldry and Thibault present an insightful approach to differentiating the metafunctional work of a sound track. It involves, primarily, understanding that human perceptual facility is at the heart of the presentational ('experiential') potential of the sound track, in that an acoustic 'stream' of information must be at first differentiated into sound events or objects. Their explanation of how sound may fulfill orientational and organisational meanings is dealt with further in Chapter 3, so I will not rehearse it in this section. Additionally, the authors attend to the logico-semantic categories of expansion: elaboration, extension and enhancement. They suggest how these may be used in multimodal analysis by means of demonstrating how they can be applied to image editing, but unfortunately not to the sound track.

At this point, I will move on to my presenting criticisms of Baldry and Thibault's approach. I then identify how the present research aims to deploy and develop their concepts and theories concerning TV and film sound tracks. Where it might be expected that Baldry and Thibault make an invaluable contribution is in the difficult and complex task of transcribing film and television sound tracks. Indeed, this is certainly the case in their take up of Gregory's (2002) notion of the metafunctionally determined *phase* as the basic analytic unit. However, the utilisation of notational symbols for describing sound entails that they introduce yet another level of abstraction from the sound track, and another symbolic code with which the reader of such transcripts must then acquaint herself. It should be clear that devising a consistent terminology and conceptual system for understanding the semiotics of sound-to-picture is already an abstraction from sound-as-mode and sound-as-medium. To then suggest a further abstraction into the visual symbolic mode seems unneccesary at best, and at worst a major hindrance to analysis. After all, we should attempt to stay as close as possible to the *linguistic* ideal: that is, of a mode as its own metalanguage. It is perhaps more useful that we stop at one level of abstraction. In short, their transcription system is too technical to be generally useful. In terms of this particular thesis, what using their system has revealed is that a simple, verbal description of the sound track that uses a technical vocabulary is the most 'transparent' representation of an audio-visual text and how it unfolds in

time.

However, my main criticism of their work is that Baldry and Thibault downplay the technical, practical dimensions of sound-track construction. Given the dearth of research on the functioning of the sound track in the field of multimodal discourse analysis, their approach may be defensible: after all, they have at least systematically constructed a way of attending to the sound track, and understanding how it functions as part of the larger multimodal whole. But it remains a serious problem that the authors approach the sound track from the sole perspective of a 'reader', even a critical one. Without attending to aspects of the *production* of film and television sound-tracks, a rich resource for understanding the functional role of the sound track is disregarded. There are strong constraints on the choices that sound practitioners must make, and these derive, in large part, from the technical resources that have evolved in the context of sound-to-picture. Of course, the array of available technologies is itself interrelated with the relevant contexts and genres that those technologies have evolved to produce and distribute. But, at any rate, technical and technological constraints may be identified, and these may help not only to inform and support *readings* of multimodal texts, but also they will provide extant terminologies that, if deployed in social semiotic theory, will play a huge role in bridging the gap between multimodal practice and multimodal theory.

I believe their decision was made on the basis that adopting production practices as the context for social semiotic analyses of particular media has been a recurrent criticism of the work of Kress and Van Leeuwen (1996, 2000) and Van Leeuwen (1999), who attempt in those works to engage with image and sound, respectively. Putting aside the issues of whether criticisms of 'production determinism' are valid in those cases, it should be made clear that the systemic-functional theoretical framework that provides the foundation for Kress and Van Leeuwen's work, and, by extension, that of Baldry and Thibault (2005), is itself based on the notion of 'insider knowledge', and on experience of how humans *use* language in practice. It is a theory constructed by humans who use language, and it deploys language as its own metalanguage – that is, it uses the very practice it explains to describe itself.

42

Put another way, the systemic-functional basis of contemporary multimodal discourse analysis was itself constructed from a 'production' perspective, or at least with a great deal of insight into production practices. The present work differs on just that respect: it aims to introduce technical resources of sound-to-picture into multimodal discourse analysis.

2.4   Conclusion

The present research aims to develop the work reviewed above in the following ways. Firstly, it incorporates practical, technical and historical perspectives on sound-to-picture into a social semiotic model for analysing and understanding how sound is configured in audio-visual texts. As argued above, this dimension is certainly missing from most accounts of sound in the MDA field. The technical resources of sound-to-picture (see Chapter 5) are useful heuristic 'tools' for audio-visual analysis. Yet these categories also have the added value of being a significant part of the trade and film-studies discourse on sound-to-picture. Integrating those resource categories   into a social semiotic framework, therefore, opens up a dialogue between a practice-orientation to sound-to-picture and a critical and analytic one.

Secondly, the present thesis aims at a level of analytic specificity that is mostly absent from the MDA literature on sound and image integration (though it is present in analyses of other intersemiotic relationships, e.g. image and text [cf. Martinec and Salway, 2005]). While Van Leeuwen (1991) does not directly deal with the kind of audio-visual interaction that constitutes the focus of the present work, in that his aim is to construct a characteristically detailed and rigorous framework for analysing *verbal-visual* conjunctive relations in audio-visual texts, he does lay the foundation for sound-to-picture analysis as I perform it here. Additionally, the generality of Iedema's (2001) analysis of audio-visual relations – though necessary to its 'handbook' context – leaves unanswered questions concerning how audio-visual redundancy is 'textured'. Yet, like Van Leeuwen's (1991) work, it also provides a useful platform on which to build a more elaborate account of the co-patterning of sound-meanings and image-meanings. Baldry and

Thibault's (2005) attempt describing the intersemiotic functioning of sound and image perhaps provides the most specific account of how sound and image may be integrated semiotically, yet its neglect of *practices* of sound-to-picture means that it cannot account for the ways (other than loudness level) in which sounds may be made more or less salient. The present thesis, in its incorporation of technical and traditional resources, is able to provide a perspective on audio-visual semiosis that can enable practice-oriented descriptions to be co-deployed and integrated with social semiotic descriptions of intersemiotic functioning.

Finally, much of the work on sound-to-picture has been predominantly concerned with the *fiction* film, and the practices of fiction *filmmaking*. As I discuss in Chapters 4, 7 and 8, non-fiction film sound practices are in urgent need of social semiotic analysis. Indeed, a commonsense understanding of the differences between non-fiction and fiction might rest on a distinction of practice: that is, the techniques deployed are considered specific to each context. As I show, this notion is something of a fiction itself, since the tools for sound-to-picture practice are clearly transgeneric. Of the few works that have dealt with documentary (Ruoff, 1992; Nichols, 1991) and TV news (Batcho, 2005) sound, not one of them attempts to understand how audio-visual meaning making may be 'explained' semiotically. The present thesis is an attempted rectification of this situation.

Chapter 3: Theoretical foundations

3.0 Introduction

I stated in the previous chapter that there is a great potential and opportunity for social semiotic theorisation of sound-to-picture. As I argued, this is not merely due to a gap in the literature but, rather, to two more significant factors: firstly, the 'provisions' of existing research on the semiotics of sound, image and action (Van Leeuwen, 1999; Iedema, 2001; Kress and Van Leeuwen, 1996; Martinec, 1998; Martinec and Salway, 2005), and, secondly, the connections that can be usefully made between that research and the diverse literature on film and film sound (e.g. Eisenstein, 1943; Weis and Belton, 1985; Altman, 1992; Chion, 1994; Lastra, 2000). In this chapter, I will focus on the extant provisions of social semiotics and multimodal discourse analysis and discuss the concepts and theories that have a direct bearing on the subsequent argument and analyses of the thesis. The aim in the later analytic chapters will be to utilise the traditional and technical resource categories of sound-to-picture, as presented in the next two chapters, with the social semiotic concepts presented in the present chapter. It is argued that such a synthesis of bodies of knowledge should enable the construction of a richer, more practice-oriented theory of audio-visual semiosis in non-fiction film.

3.1 Social Semiotics: Representation as work

While traditional semiotics (Saussure, 1974[1916]; Peirce, 1966; Eco, 1976) has concerned itself with formal description of the sign, either as the dyad 'signifier' and 'signified' (in Saussure's theory) or as the triadic composition of 'representamen', 'interpretant' and 'object' (in Peirce's), social semiotics instead foregrounds *semiosis* – the act of making meaning. Saussure himself stressed that his proposed 'semiology' was a science of the 'life of signs in society', yet traditional semiotics has tended to neglect the social dimension to Saussure's teachings. Instead, it has tended to concentrate on theorising the complex meaning relations between signs, thus abstracting signs and sign systems from their actual use in context. As Lemke (1995: 102) argues,

Social semiotics offers the view that socially meaningful *doings* constitute cultures (social semiotic systems), that cultures are systems of interdependent, socially meaningful practices by which we make sense to and of others, not merely in explicit communication, but through all forms of socially meaningful action (speaking, drawing, dressing, cooking, building, fighting etc.). Sign systems are abstractions from such practices (e.g. linguistic signs from speech), and they change as social practices change.

Social semiotics, along with critical linguistics (e.g. Kress and Hodge, 1979; Fowler, 1991) and critical discourse analysis (e.g. Fairclough, 1989; Van Dijk, 1993), emerged as a response to the formalist semiotic approach to meaning. The view of representation adopted in this thesis derives from social semiotic theory (Halliday, 1978; Lemke, 1984, 1995; Hodge and Kress, 1988; Thibault, 1991): that it is most profitably considered as both process and product.

As my use of the word 'role' in the formulation of this thesis' title implies, I concur with proponents of social semiotics that representation is an activity: meanings are *made* by people, they do not simply exist in a culture. The formal systematisation of sign systems, as Lemke (1995:157) has argued, can help to achieve only a partial understanding of how we mean, since '[a]ll formal systems, all meaning relations, are immanent in and enacted by our actions, by what we do in using them'. In following this central tenet of social semiotic theory, I explicitly presuppose that all media and semiotic modes are resources that we, as sign-producers, engage with to perform some kind of *work*, some specific semiosic labour, that contributes to the construction of meaning, and to the production of texts.

Moreover, such semiosic work is always done and such texts are always made – whether explicitly or implicitly – in the service of particular social, political and cultural interests. As Lemke (1995: 9) argues:

[...] all meanings are made within communities and [...] the analysis of meaning should not be separated from the historical, cultural and political dimensions of these communities.[Social semiotics] is useful for studying meaning in a way that then enables us to see how the meanings we make function to sustain or challenge the relationships of power in our communities.

46

In Thibault's (1991: 6) words, social semiotics 'seeks to show how [...] meaning making practices function in ways that enact, maintain, reproduce, and change the social semiotic system or some part of it.' The interests mentioned above, for Thibault, are an inherent part of the 'social semiotic system', as are all things social. They will be many, and differently motivated, and although we may identify *dominant* interests, contesting interests will co-exist in a social semiotic community, served by co-existing and differing resources (and configurations of resources), modes (and combinations of modes), and media for making meaning.

Semiosis, motivation and history

Such semiosic labour plays a large role in the 'shaping' of semiotic resources over time, so that what we *do* with a resource impacts on its evolution. There is a transformational, historical dimension to the work of representation: such work acts as a constraining and determining factor on the development and evolution of a mode or medium, and thus the kinds of meanings that can be made in a social semiotic community. Furthermore, social, cultural and political interests may change over time, and the media and modes which serve them will, necessarily, co-evolve in relation to such changes (Lemke, 1995: 102). If those interests do not change, then their corresponding semiotic resources might instead develop to better serve them.

To demonstrate briefly how a medium can be mobilised to fulfill certain social, cultural and political requirements, and therefore to show how the evolution of a medium and/or mode is not at all arbitrary but *motivated*, I will use the example of how the sound-film enabled propagandists in Nazi Germany to make more effective kinds of meaning than those that could be made previously. In his essay, 'Film as a Weapon', Nazi propagandist Fritz Hippler (1931) meditates on the power and utility of *images with sound* as a major resource in the propagandists' arsenal:

> If one compares the directness and intensity of the effect that the various means of propaganda have on the great masses, film is without question the most powerful. The written and spoken word depend entirely on the content or on the emotional appeal of the speaker, but film uses pictures, pictures that for almost a decade

47

have been accompanied by sound. We know that the impact of a
message is greater if it is less abstract, more visual. That makes it
clear why film, with its series of continually moving images, must
have particular persuasive force.

Here, Hippler reduces the "persuasive force" of film to the visual's capacity for
concreteness and specificity, and to film's capacity for presenting images
"continually". The sound track being mentioned only in an oblique and offhand
manner gives the impression that Hippler considers its value to inhere in its basic
capacity to enable *speech* to be reproduced. Film is powerful and effective for
propaganda, then, for Hippler, because it adds a visual dimension to the already
successful, emotive mode of speech. More precisely, it contributes a specificity that
complements the abstract (albeit, persuasive) nature of spoken or written language.
In short, Hippler believes that because of this coupling of speech and images, the
sound-film would help propagandists to persuade more effectively than if they were
to use speech alone. And not only those who were involved directly with the
governments of their respective countries, such as that of Nazi Germany, were led
to conceive of film in such an overtly and politically influential way. Upon the
advent of sound in motion pictures, Harry Warner, the Hollywood film mogul, also
believed that the sound-film could serve the purposes of propaganda (in the less
negative sense of the term):

We think of the film as the greatest of all the media for propaganda [...]
(with sound, it) may even serve to eliminate war among the nations.
(Sperling and Millner, 1994: 121; original parentheses)

Much has been written concerning the modern Hollywood film as propaganda
(Monbiot, 2002; Constantine, 2003), and indeed such a form evolved directly from
the culture and industry of the sound-film's early years (Carter, 2004), including
necessarily the German and Soviet cinema's overtly propagandist orientation to the
film as a semiotic and cultural form. Since then, the means by which films are
produced – i.e. the knowledges, techniques and technologies crucial to their
production – have not drastically altered at all, but have been developed and honed
in the service of a semiotic form which is essentially identical to its forebears. For
example, most contemporary cinema (fictional and non-fictional) promotes spoken

dialogue as the central tenet of its philosophy of sound, as the-early sound-film did with the first 'talkies' (e.g. *The Jazz Singer*, 1927). In addition, while the *stereophonic* sound-film became a reality in the 1950s (Belton, 1985), the positioning of speech within the stereo field has in large part remained identical to the first monophonic sound films; that is, dialogue is almost always positioned in the *centre* of the stereo field.

As the above example demonstrates, the co-evolution of sound and image as audio-visual representation is *motivated*; it is no mere 'accident' of history. It is an exclusive relationship that has been developed and maintained in order to meet certain requirements of the past century. The needs of the Nazi propagandists that the evolution of audio-visual semiotic resources fulfilled were to establish a way of communicating

(1) that was less overt and direct than written or spoken language, but still efficient and effective;

(2) that subtly generated second-order, connotative meanings, though appearing to operate only according to a system of first-order signification;

(3) that drew on a variety of communicative modes, engendering a situation in which the meaning of any one instance of audio-visual representation cannot be located in or reduced to any *one* of the modes or media that texture a communicative event.

In order to introduce the relationship between semiosis and social interests and motivations, I have formulated it above in a somewhat simplistic and misleading way. As suggested in the quotation from Hippler, where he appears impressed by the sound-film as a potential propagandist medium, the relationship between semiosis and the interests it serves cannot be posited as a simple linear, one-way, deterministic one. Rather, as Thibault suggests (1991: 6), they are a part of one overarching "social semiotic system" or "culture" (Lemke, 1995: 102) and as such are involved in a complex and reciprocal relationship in which the resources of a

medium or mode are as much a determinant of the interests which they appear to "serve" as vice versa. In the case of the Nazi propagandists, their decision to use the sound-film as a medium for propaganda was in response to their perception of its success as an effective medium for persuasion. In turn, their deployment of and commitment to the sound-film helped to establish its place as a *mass* medium - both for overtly propagandist purposes, and for cinema, more generally.

3.2 Multimodal semiosis: interaction of modes and media

3.2.1 Mode and medium

Perspectives on communication have, until recently, tended to focus solely on one semiotic mode: language. In the humanities and social sciences, this monomodal view of human interaction has been the prevalent one, and even when discussion of the visual mode has become more common, it has often been studied in isolation from other modes (e.g. Berger, 1972). However, proponents of a multimodal view of communication (e.g. Kress and Van Leeuwen, 1996, 2001; Thibault 2000; Iedema, 2001, 2003; O'Halloran, 2004, 2005; Lemke, 1995, 1998, 2002) take the position that meaning making never occurs with the use of only one semiotic mode. A multimodal view aims to establish that meaning is made through a plethora of semiotic resources available in an instance of communication: visual, verbal, graphological, gestural, auditory, and so on. This is reasoned on the basis that we

> cannot make meaning that is construable through only one analytically distinguishable semiotic resource system. Even if for many purposes we analytically distinguish the linguistic semiotic resource system from that of depiction or visual-graphic presentations, and both from others such as the music-sound system or the behavioural-action system, the fact that all signifiers are *material* phenomena means that their signifying potential cannot be exhausted by any one system of contrasting features for making and analyzing meaning.
>
> (Lemke, 2002: 302; my emphasis)

On this view, not only are meanings always necessarily made multimodally but the separation of modes into distinguishable, bounded units is made entirely on heuristic grounds (Norris, 2004; Constantinou, 2005). This implies that any 'integration' between semiotic resources that occurs in an interaction cannot be

founded on an inherent, pre-discursive (Thibault, 1991) division between modes. As proponents of socially-oriented semiotics, such as critical discourse analysis, multimodal discourse analysis and sociolinguistics, argue, languages and other semiotic modes are analytical abstractions from material processes and phenomena (see Lemke, 1998, 2002). Hence, we can view the system of written language as an abstract typological ordering of the continuous, topological sound-stream of speech (Lemke, 1999) and as consequently not accounting for much of the materiality and physicality involved in the act of speaking: voice quality, intonation, visual-gestural information, and so on. Accordingly, sound-as-mode, as deployed in this thesis, refers to the general system of semiosis that has co-evolved with sound-to-picture practice, and draws upon the resources of sound-as-medium (cf. Stillar, 2005). As Stillar (2005) defines it, modes are the recognisable ways in which the resources of a medium have been deployed over time. As I have discussed elsewhere (Constantinou, 2005: 609), modes derive from the evolution of media according to the social-cultural interest in using a particular medium to perform, fulfill or enable a particular social function (Kress and Van Leeuwen, 2001).

The above definition of mode as an abstract resource for semiosis, not a physical or material one, is posited in contrast to that of 'media', which is defined by Kress and van Leeuwen (2001: 21) as '[the] material resources used in the production of meaningful products or communicative events. "Media" can be both materials and tools (e.g. airwaves and radio)'. Norris (2004) explicitly calls for our conception of mode to remain open, yet the same needs to be argued for the concept of 'media'. 'Media' certainly can be defined as the material means with which meanings are made; but another dimension of understanding is made possible by opening the concept up to other interpretations. As Jonathan Sterne (2003) argues in his book *The Audible Past*, most studies that engage in the process of mediation and its products are let down by a rather murky and tangled conception of *medium*, which for him is

> a recurring set of contingent social relations and social practices, and contingency is the key here. As the larger fields of economic and cultural relations around a technology or technique extend, repeat and mutate, they become recognisable to users as a medium. A medium is therefore the social basis that allows a set of technologies to stand out as a unified thing with

clearly defined functions. (Sterne 2003: 182)

For Sterne, then, a medium is, in Lukàcs' (1971) terms, a relationship between people that has acquired a 'phantom objectivity', a certain 'thingness' (Sterne 1999: 504). It is not a physical or material basis with which meaning is made and modes evolve – such as a technology or set of technologies. Rather, a given medium – television, in Sterne's (1999) research – *stands for* the plethora of relations between people, and the 'phantom objectivity' that those relations attain imbues that relationship with 'an autonomy that seems so strictly rational and all-embracing as to conceal every trace of its fundamental nature: the relation between people' (Lukács 1971: 83, quoted in Sterne 1999: 504).

While in social semiotics a distinction between medium and mode is often maintained on a physical or material basis (Kress and Van Leeuwen, 2001; Jewitt, 2004), Sterne's definition highlights the social maintenance of the boundaries around different media, thereby negating any difference between how we choose to differentiate media and how we differentiate modes. As I have argued, the semiotic modes that are deployed 'within' those media also take on a phantom objectivity, and their boundaries become reified such that modes are often seen as identical to the media that 'house' them. Essentially, the differentiation of one mode from another, and thus the reification of that distinction, is achieved as a determining and determined part of the processes for the reification of media that Sterne describes.

However, even if we accept Sterne's formulation of medium, it does not follow that we must dismiss the idea that media are also material. Nicholas Cook (1998) meditates on the question: What is a medium? He finds value in a definition offered by art-critic, Jerrold Levinson, who denies that the term is 'equivalent to *material or physical dimensions*' (1984: 7; original emphasis). 'Rather', he continues, 'by a medium I mean a developed way of using given materials or dimensions, with certain entrenched properties, practices, and possibilities.' (Levinson 1984: 7, quoted in Cook 1998: 262)

The definition of 'medium' I deploy in this thesis incorporates both perspectives, but is *primarily* concerned with the material dimensions of media. While Levinson's definition of medium is attractive, if deployed here there would be redundancy between the 'entrenched properties, practices, and possibilities' of a

medium, and the definition of mode offered above, that it is a way of using a particular medium to fulfill certain social functions in particular social contexts. Additionally, focussing on the material aspects of a medium encourages one to recognise the ways in which deployments of sound-as-medium may *overflow* the functions that sound-as-mode fulfills. In other words, where the use of a particular technical resource – e.g. reverb – has evolved mainly to represent spatial characteristics, as in its deployment as a signifier of cavernous space when the synchronised image depicts such a space, the characteristics of the technology may appear to exceed such a basic narrative function. As I show in Chapter 6, if older reverb technologies – such as spring or plate reverbs – are experienced nowadays, their 'simulations' of real acoustic space may seem unconvincing. In films in which they are used as signifiers of acoustic space, the very specific character of the technology *overflows* the narrative functions they are deployed to fulfill. This overflow may be semiotised as a signifier of a particular point in history (cf. *provenance* – Kress and van Leeuwen, 2001: 23); for example, the use of spring reverbs in early TV sci-fi programmes. Or it may simply remain at the level of affect, until it is shaped and mobilised into performing some semiotic work either in practice or analysis. As I discuss in Chapter 8, there is always a certain excess specific to sound-as-medium and image-as-medium, which may be understood in Barthes' (1977: 52-53) terms as the 'third' or 'obtuse' meaning: signification that may not destroy narrative, but can often subvert it.

Given the above difficulties in conceiving of semiosis as multimodal – that to make such a statement one *apparently* treats the differentiation between modes as inherent in those modes – how then can qualitative inquiry benefit from a multimodal perspective on semiosis? In particular, if we are to consider multimodality as merely a heuristic, where is the 'stability' on which we are to base the kinds of convincing, empirical analyses and interpretations that are intended to illuminate the meaning of an instance of semiosis by way of modal differentiation and, hence, configuration? In short, there can be no stability, at least not in a positivistic sense. All we have to support our typological differentiation of modes is our knowledge of their history, and that of their associated media: that is, the often

complex trajectory through which a mode has been deployed as a separate entity to another mode. One such process that Halliday (1975, 1992) identified in relation to what can be called 'modedness' is what he termed *semogenesis*. As Lemke (1998a) expounds:

> By a general process, common in development and evolution, and sometimes called *semogenesis* (Halliday 1975, 1992), when an original unity becomes semiotically construable as being composed of separable units, those units become free to combine in new ways, making new meanings possible, and the culture is then free to evolve new restrictions on these combinations, endowing them with yet another dimension of meaning (as "indices" of some situation-type or subcommunity; Lemke 1993b). When writing is distinguished from drawing, it becomes possible to compose multimedia texts, incorporating both semiotic modalities, and allowing us to *multiply* the meanings made with one by those made with the other, provided our community has established conventions for how to read such multimedia genres.

On Halliday's view, a semiotic 'unity' is something that is culturally and historically contingent. It follows that construing an apparent unity as divisible into 'separable units' is also contingent along the same lines. The present thesis rests on this very premise, and indeed the analytic framework that I introduce over the following chapters relies on the heurism involved in the notion of modedness as license to perform the kinds of fine-grained semiotic 'incisions' that make a thesis such as this possible. After all, identifying the relationships between modes is as much a matter of maintaining previously erected boundaries around modes as it is a matter of challenging them through the creation of new distinctions. The social semiotic analytic process, then, is primarily an interpretive one (Iedema, 2001: 200). It cannot hope to *establish* the meaning or functioning of a textual instance, but can only offer interpretations on the basis of a descriptive account of how a semiotic resource has been deployed across contexts.

Such an impressionistic approach in the context of the 'social sciences' may appear to be a hindrance. But weighed against the consequences of *not* taking a multimodal perspective, the interpretive dimension of multimodal semiotic analysis appears as far less obstructive. One consequence, for example, is that if we do not look at interaction through a multimodal lens we preclude the identification of

certain semiotic processes and manifestations in a given text. Moreover, taking a multimodal view does *not* preclude the recognition of the kinds of processes that would be identified should one choose a monomodal view; for example, we are still able to identify linguistic processes at play in a text if we look at that text multimodally. In fact, multimodal analysis might help us to understand, through a mode's integration with other modes, how a singular mode functions in a text, for the reason that much of the meaning made through, say, language is accomplished in collaboration with other, non-verbal modes, such as gesture. To reiterate, the reduction of communication to a single mode (which still constructs heuristic boundaries as in multimodal analysis) certainly *does* preclude particular semiotic processes from being identified.

3.2.2 Semiotic Functions

The material basis of all meaning making, and all semiotic systems, entails that combinations of different modes and media involve the bringing together of material phenomena that always maintain their individual 'identities' (cf. Cook , 1998) at the same time that they are integrated to some social semiotic end. As Lemke argues (2002: 303), different semiotic modes are incommensurable, a fact that may be demonstrated by considering translation from one mode to another:

> Even though a culture may create conventions about how, say, a painting is to be described in words, or commented on in scholarly fashion, or how a mathematical equation is to be graphically represented, text, image, and other semiotic forms are sui generis. No text *is* an image. No text has the exact same set of meaning-affordances as any image. No image or visual representation means in all and only the same ways that some text can mean. *It is this essential incommensurability that enables genuine new meanings to be made from the combinations of modalities.* (my emphasis)

Given this incommensurability, in order to claim that sound and image *combine* or *integrate* to make meaning we must specify some common semiotic dimensions between the two modes. As Lemke (2002: 303-4) asks, '[a]t what level of abstraction can we say that images [...] and other semiotic productions make meaning in the same way?' Following Halliday (e.g. 1978, 1985), Lemke (1995:

41), and Kress and Van Leeuwen (1996), we may posit that we use language and other semiotic modes to make meaning in three, simultaneous ways. As Lemke (1998a) explains: '[w]hen we make meaning we always simultaneously construct a "presentation" of some state-of-affairs, orient to this presentation and orient it to others, and in doing so create an organized structure of related elements". These three dimensions of semiosis – presentation, orientation and organisation – are central to the analytic framework proposed in the present thesis.

The principles are generalised versions of Halliday's metafunctions (ideational, interpersonal, textual), rendered less specific to language and more amenable to application to other semiotic modes (Lemke, 1995: 41). To name a few of the most famous applications beyond the linguistic: Kress and van Leeuwen (1996) have produced an extensive theory of how images may be understood as fulfilling these three functions. Martinec (e.g. 1998) has explored how human gesture and action can also be theorised along metafunctional lines. O'Toole (1994; 2004) has extended the theory to 'displayed art' and architecture. Only recently has sound been theorised according to presentational, orientational and organisational meaning (Van Leeuwen, 1999; Thibault, 2000; Baldry and Thibault, 2005). I will briefly discuss these in more detail and explicate how sound may fulfill such functions.

Presentational

Every act that we construe as meaningful constructs 'semiosic reality' by means of presenting some state of affairs; processes; circumstances; and by classifying participants and relations between them. In language, what is termed propositional content would be categorised within this category: i.e. the natural and social world. In images, actions and relations between participants may be realised by vectors (Kress and van Leeuwen, 1996: 42).

Because sound is perceived more or less as a 'stream' – as 'acoustic flux' (Baldry and Thibault, 2005: 210) – with only certain sound events being salient and easily distinguishable, it is not surprising that few have formulated how it may make presentational meanings. The 'average' listener-viewer has more facility with delineating and identifying visual objects within an image than with specifying the

events of an auditory array. As sound designer Tomlinson Holman (1997: 41-42) argues:

> [A]nyone can tell the difference [on screen] between a table and a chair. People do not have the same facility with the variety of sound objects present in films without training, and this is a core idea. It explains why sound is so valuable to film makers.

Stereophonic production/reproduction has made this a more manageable task, as sounds may be positioned in discrete places across the lateral dimension of the stereo field. Yet, by and large, sounds' transient physical nature, in contrast to the 'static' image, precludes everyone but sound professionals and audio afficianados from being able to differentiate among the edited and mixed sounds of a film sound track or music production. Such differentiation is crucial to construing the presentational meanings made in a given sound track. Thibault (2000) and Baldry and Thibault (2005: 210) argue that

> [e]ach source and the component of the acoustic flux that corresponds to it is a specific event in the overall array. Further, the assigning of different parts of the acoustic wave to different informational sources entails that listeners construe meaningful relationships among different sources [...]'. The meaning that we construe in these informational sources and the relations among them are not reducible to the kinds of phenomena studied by acoustic physics. Aside from subjective aspects of perception, social practices and cultural values also play their part in shaping how we perceive acoustic phenomena.

They go on to note that the 'construal of acoustic information as different classes of objects suggests an experiential [presentational] dimension' (ibid.). In film and TV texts, the synchronised image contextualises the acoustic stream of the sound track and helps the listener-viewer to construe the sonic components that redound with the image on some specifiable dimension (e.g. we may differentiate one 'set' of footsteps from a mix of many sets because only one person's steps are represented in the image track).

Orientational

Each semiotic act constructs a two-fold orientation: to its presentational content and

to the addressee(s). This is where evaluations are made towards presentational meanings and attitudinal stances are taken towards potential addressees (Lemke, 1998a). Visual representations make orientational meanings by deploying the resources of perspective, point of view, and techniques that manipulate the 'reproduction' aspects of image production – such as 'grain', colour, and in moving images, frame-rate and slow-motion.

The function of language and other semiotic modes to orient – to 'enact our relations' (Martin and Rose, 2003: 6) – is central to deployments of sound-to-picture as a resource for 'positioning' the listener-viewer to the represented participants and actions in the image and sound track. Such deployments, as I demonstrate in the analyses, include methods for augmenting or maintaining auditory perspective (via the notion of point of audition; see Chapter 2). According to writers on film sound such as Williams (1980) and Altman (1992), it is an inherent function of sound reproduction to represent not only the *event* (the presentational function of sound), but also a social relation to that event. Questions include, for example, how is *distance* established between listener-viewer and audio-visually represented participants and actions? And in stereo representation, where are participants and actions positioned *laterally* in relation to the centre of the stereo field (the 'reference point' for the listener-viewer – cf. Tagg, 1991: 60).

Additionally, technical resources such as reverb, which can create very precise kinds of acoustic spatial effects, have developed to present an attitudinal standpoint towards the presentational meanings of the sound track (and, of course, towards those of the synchronised image track). The application of reverb to so-called 'dream sequences' in fiction film is a key example. Though I do not deploy it in this thesis, the theory of audio-visual modality (Constantinou, 2002) is one useful way of attending to the orientational functioning of the film sound track.

Organisational

Organisational meaning concerns how a given 'stretch' of text is constructed meaningfully. It concerns organisation in terms of part/whole relations, cohesion, foregrounding and backgrounding, and so on. In language, organisation is about connecting the semantics of one clause to another, in order to create meaningful

wholes. In this capacity, organisation performs an enabling role (Halliday, 1978: 141; Lemke, 2002: 304): it constructs relations between the presentational and orientational meanings. In images, we make organisational meanings through visual composition: we organise the visual text 'into elements and regions' and we may 'link disjoint regions by such features as colour and texture' (Lemke, 1998: 94).

Sound, as Baldry and Thibault (2005: 211) argue, may 'form parts of larger wholes on the basis of relations of foregrounding, backgrounding, spatial location, distance from the listener, relations of dependency with other events, and so on [...]. [Sound] therefore [has] properties of textuality as part of a larger Gestalt to which they belong.' Organisational relations between foreground and background in a sound track functions to differentiate between those 'planes', but it also may serve to relate sound events that occur in each.

The most fundamental organisational function of sound-to-picture, however, is the temporal relations between sounds and images. This function is realised technically by *synchronisation*: what was once the mechanical temporal alignment of sound and image media, but which is nowadays integrated into a single digital code. Sync-sound (see Chapter 4) is the *sine qua non* of sound-to-picture. In audio-visual media, sound and image dynamically interact (and are configured to interact) to make meaning. Moreover, the prevalent use of sound and image in audio-visual media is one that provokes the perception, and interpretation, of both media as *unified*; that is, as a composite text, irreducible to its component modes (cf. Thibault, 2000; Constantinou, 2002; Lemke, 1999). As Lemke (1999: 173) remarks:

> [T]here is clearly something about the dynamical coordination in time (on the millisecond scale) of acoustic and visual input which inclines us to integrate them functionally; that is, we make unified meanings from them, treating them as in effect unitary rather than distinct phenomena.

In this thesis, I attend to presentational and orientational meanings primarily. This is for no other reason than conceptual scope. Organisational meanings of sound-to-picture will be investigated more fully in future research.

Central to this thesis, and one of its main contributions, is the notion that the resources that may fulfill presentational, orientational and organisational functions may be specified technically and traditionally. That is, I begin with the conventions of sound-to-picture that have evolved since cinema's early period, and couple them

59

with technical (and technological) 'realisations', such as *reverb* (to represent space, distance etc.) and *equalisation* (e.g. to represent social relations in terms of distance and 'accessibility'). Incorporating technical terminology into a semiotic framework helps to close the gap between practice and theory, empowering both analysts in their command of the practices behind the construction of texts, and practitioners, making comprehension of how certain of their practices can be 'explained' theoretically far less problematic than if 'removed' academic terminology is constructed and deployed.

3.2.3 Contextualising relations

Multimodal semiosis is fundamentally about how the semiotic resource systems of language, images, sounds, gestures, and so on, are configured and combined in terms of presentational, orientational and organisational meaning. Social semiotics understands these three categories as the 'common denominators' among different semiotic modes (Lemke, 2002: 303). Moreover, it insists that in any given text, identifying the presentational, orientational and organisational meaning of only one semiotic mode (e.g. the visual) is insufficient to making claims concerning that text's meaning making operations. Social semiotics sees meaning not as inherent in wordings or 'soundings', but rather as meaning *potential*. Meaning is made only in some specified context. As Lemke (1995: 166) explains,

> We say that when an act occurs it occurs in some *context*, and that 'its' meaning depends in part on what that context is. Better to say that we make the act meaningful *by* construing it in relation to some other acts, events, things (which we then call its contexts). The relations we construct to some (and not other possible) contexts select and emphasise some of the possible meanings of the act.

In Cook's analysis of a television advertisement (1998: 83), he argues that if one were to experience the sound track in separation from the image, the 'message' made via the audio-visual coupling is unlikely to be construed in the same manner. Precisely, there are an 'indefinite number of dimensions of variance' embodied by the sound track, and the salience of only a few of those aspects is generated through an 'intersection' with the 'more overt' and narrower meanings of the images and

60

verbal narration (Cook, 1998: 83). Music, and other media, Cook proposes, have a potential for signification 'much broader' than the meanings that can be actualised in any specific context (1998: 83). From this perspective, Cook's concept of signifying 'potential' is identical to the social semiotic notion of *meaning potential* (Halliday, 1978; Lemke, 1995). What matters, for Cook,

> is not so much the signification that is 'in' the sound, but rather the potential for signification that it may support by virtue of specific intersections with other media. Seen thus, signification becomes a function of context; it is, in a word, performative. (1998: 83)

The 'specific intersections' that serve to make defined and delimited meanings from the semiotic potential of a medium (here, music) are, in social semiotic terms, the relations between a semiotic act and the context(s) that it invokes and that determine what function that act is to fulfill in a semiosic instance. In Lemke's (1995: 167) words:

> Meaning consists in relations and systems of relations *of* relations. These relations are basically *contextualising* relations; they tell us what the contexts are in relation to which an act or event has its meanings in our community. They specify what the *combinations* are that an event of a given type can belong to, *and* what the kinds of events are, the sets of alternative events or acts of the 'same' kinds, that can make up the various types of combinations.

Furthermore, as mentioned in section 3.1 above, the contextualising relations, the 'specific intersections', made with other contexts (most pertinently, other *media*), are precisely that: *constructed* by an agent, not only formally extant. While we can define a system of meaning relations as a formal system (and a 'fiction', in Halliday's [1978: 109] words), the actualisation of the meaning potential of that system – the links to contexts that function to *specify* what we mean in a given semiosic act – is *performed* by a sign-maker working within particular constraints.

Multimodal semiosis, then, rests on the notion that meaning is made by means of co-contextualisation (e.g. Thibault, 2000). This indicates that multimodal analysis must proceed by specifying the contextualising relations that obtain between two or

more modes. It involves examining the redundancies among the presentational, orientational and organisational meanings of several modes that can be seen as integrated in a given textual instance. In this thesis, I examine how the semiotic potential of sound is actualised through a dual contextualisation: first, how sound and image co-contextualise one another; second, how the functional configurations of sound and image redound with fiction and non-fiction film as *contexts of situation*. A crucial further question concerns how those contexts are regulated and maintained *as types*; this requires an investigation into the practices that constitute fiction and non-fiction as contexts of *culture* (see Chapter 8). As Baldry and Thibault (2005: 2) put it, semiosis is 'contextualisation all the way up, and all the way down.'

I will address the first type of contextualisation here; that between sound and image meanings. The analysis of audio-visual semiosis requires the specification of *contextualising relations*. In this thesis, following Van Leeuwen (1991, 2005) and Martinec and Salway (2005), I adapt the logico-semantic categories of Hallidayan systemic-functional grammar (cf. Halliday, 1985:225-241). Specifically, I am interested in understanding how the presentational and orientational meanings of sound and image are made to combine meaningfully. Orientational meanings are interdependent with presentational and organisational ones, so specifying the presentational, *semantic* relations between sound and image is a necessary 'way in' to understanding the basis on which audio-visual orientational configurations make meaning. The categories I will deploy are relations of conjunction: they are the means by which the meanings of sound and image are 'expanded' (cf. Baldry and Thibault, 2005: 235). The relations are: *elaboration, extension* and *enhancement*.

Elaboration, for Van Leeuwen (2005: 222), involves one unit of information repeating or restating the information of another, 'for purposes of clarification'. This concept is relevant here as means of illuminating how one mode *specifies* the meaning of another in a multimodal complex. This suggests that it serves a similar function to Barthes' (1977) *anchorage* – the specification of meaning in one mode by another mode; in Barthes' terms, anchorage functions to 'remote control' the reader to particular, preferred meanings of a polysemous text (1977: 44). Barthes'

focus is the specification of meaning in images by text, but the concept – like elaboration – is here adapted to synchronised sound and image relations. *Extension* is deployed here as a means to describe the *addition* of meaning from one mode to another. In the sense it is deployed here, extension functions *logically*: that is, information in 'the second of two items of information gives a reason for, a condition of, or a comparison with the information in the first item' (Van Leeuwen, 2005: 223) Finally, *enhancement* relations between sound and image serve to qualify the presentational meanings of either mode in terms of circumstance: that is, place, time and reason/purpose (Martinec and Salway, 2005: 349). Enhancement may also specify *causal* relations between aligned modes.

Van Leeuwen (1991, 2005) is concerned with the conjunctive relations between syntagmatic units (images), and as such the notion of status – i.e. which unit is primary, which dependent – is resolved fairly simply by appealing to the temporal 'unfolding' of audio-visual film. We watch a film from beginning to end, and succumb to its temporal logic. In deploying elaboration, extension and enhancement in analysis of synchronised sound and image, I face similar problems to those who engage analytically with image-text relations (cf. Barthes, 1977; Martinec and Salway, 2005). This is the problem of the *relative status* of modes in a semiotic ensemble. In terms of images, there are no temporal, linear relations to determine 'dependency' (Baldry and Thibault, 2005: 235); we can determine this only by investigating the 'clauses' of both image and text, and then reasoning their relations of complementarity or dependency. With sound-to-picture, however, the temporal dimension, coupled with strongly conventionalised modes of combining sound and image, renders relative status far more circumscribed. To expand, at a given point of synchronicity (see Chapter 4) between sound and image we must be able to identify relative status in order to interpret emergent audio-visual meaning. Questions arise such as: do sound meanings *extend* image meanings, or vice versa? How, specifically, is the relative status of sound and image in a synchronised audio-visual configuration realised in a given audio-visual text? My response to the problem of status is this: the convention of synchronisation between sound and image has evolved to suggest a basic causal relationship between them, specifically positing image as the cause and sound as the effect (indeed, that might be the origin

of the term 'sound effect'). There are instances when this convention is challenged –
such as the visual appearance of bullet holes on a wall, synced with the sound of
gunshots – but as long as a represented visual action is understood as a
sound emitter, a closely synchronised sound is likely to be understood as an effect
(see Chapter 4).

3.2.4   Context: Situation and Culture

As regards the second type of contextualisation, there is a need to firstly understand
how the choices made at the level of tri-functional meaning are related to contexts
of situation. The 'situation', for Halliday, is where text 'comes to life' (1978: 109). It
cannot, however, be described in terms of a concrete environment, but rather as the
general *type* of situation that functions as a determinant of audio-visual meaning
making. As discussed in Chapter 8, contexts of situation are semiotic constructs,
and can be understood in terms of three social dimensions; in Halliday's (1978,
1985) and Gregory's (1967) terms, field (the social action), tenor (the role structure)
and mode (the symbolic organisation). Furthermore, there is a synomorphic
connection between situation and meaning making, to the extent that dimensions of
situation may be understood as *enacted by* and *constraining* the configurations of
the semiosic functions (respectively, presentation, orientation and organisation) and
thus the meanings, that can be made in a given social situation. While Halliday
(1978: 110) states that field, tenor and mode are 'general properties of the situation
[that] collectively function as the determinants of text, in that they specify the
semantic configurations that the speaker will typically fashion in contexts of the
given type', it must be made clear that any distinction of text from context is
analytical and formal, and should not be reified. Reification of a text-context
distinction encourages the notion of a one-way determinism, most commonly from
the 'level above' to the 'level below': that is, from social situation to text. The use of
the Hjelmslevian term 'realisation' to explain the synomorphy between text and
context is often misconstrued in such a deterministic manner – i.e. language as the
realisation of situation. Yet realisation is a two-way process, a 'dialectic in which
texts enact, create, and produce their context of situation as much as they are
determined by them' (Thibault, 1991: 123). Accordingly, *constraint*, as the term is

deployed in this thesis, is not to be understood as a one-way restriction on semiosis. Instead it characterises a dual aspect of the *realisation* relationship between the text we construe as focal, and the text we construe as *context*. Constraint, then, both restricts action, and thus semiosis, and it *compels* us towards certain choices. As Thibault (ibid.) puts it,

> the contextual variables of field, tenor and mode are often seen as determinants of the linguistic features selected in texts at the level 'below'. The top-down emphasis on these variables as determinants of the semantic selections in texts tends to reify the social context of situation in such a way that it is made out to reproduce itself independent of specific textual productions. A too normative, top-down model of the text-context relation cannot adequately conceptualise the text as the site of a plurality of overlapping and contradictory social discourses and semantic registers [...] Texts cannot be reduced to the necessary properties of an a priori social situation. This would imply that texts are no more than the mere appearances of an underlying (social) reality that they 'represent' or 'refer to'.

Such a 'top down' view cannot account for changes in the semiosic system – for example, the emergence of new registers in the context of non-fiction film. A description of non-fiction film as context of situation, therefore, cannot provide an explanatory model of *why* certain semiosic choices are made with sound-to-picture. However, if we look to the level of context of culture, we can begin to see how situations are regulated *as types*. Contexts of culture provide certain semiosic norms, and norm violations, that are characteristic of situation types. As I argue in Chapter 8, an understanding of contexts of situation enables the *description of* the manifest constraints on semiosis; attending to contexts of culture, on the other hand, enables us to *explain* those constraints, to the extent that we can describe how they are instituted. We must look to 'culture' in order to determine how 'situation' is itself constrained (perpetuated, genericised etc), and thus how the force of such constraint cascades 'down' to semiosic choice in the production and reception of individual texts. Both 'levels' of context need to be attended to in order to comprehend the social semiotic environment in which non-fiction film operates, and which governs the expectations that constrain the meanings that may be made with non-fiction sound-to-picture.

In this thesis, I do not engage explicitly with the notions of genre and intertextuality. For Thibault (1991) and Lemke (1995, 1999) these are fundamental to understanding how situation relates to culture, and thus for understanding how situations are recognisable as types. The main reason for this omission is one of scope: given the tasks of this thesis, introducing genre and intertextuality as analytic mechanisms would entail extending the research concerns beyond the acceptable limits for a doctoral thesis. However, the notion of genre, at least, is present implicitly in the concept of institution (see Chapter 8). Institutions are the sites of genre production, and may be understood as at one level of abstraction from genres (cf. Lemke, 1995; Stillar, 2005; Graham, 2006: 180). Constraints and regulations on semiosis that can be identified as institutional can of course be understood in terms of genres. In Chapter 8, I deploy the notion of *action genre* (Lemke, 1995: 31-2) as an intermediary concept between situation and culture.

3.3 Conclusion

In this chapter, I have outlined the aspects of social semiotic theory that are crucial to the analytic framework of this thesis. In the next two chapters, I turn to examine the types of resources that will be implemented in the analyses of audio-visual texts of Chapters 6 and 7.

Chapter 4: Resources of sound-to-picture 1 traditions and histories

4.0 Introduction

This chapter will demonstrate that the meanings that are made with sound-to-picture are dependent on a history of, often competing, traditions. From the standpoint of a contemporary observer of film style, it is perhaps easy to think of film sound tracks as homogeneous in function, style and approach. By extension, it is also easy to think of the most common approach – that is, as an adjunct to the image that aims to serve narrative intelligibility – as somehow 'natural'. Without attending to the rich history of film sound, and the practices that have evolved from those pivotal events, we continue to *naturalise* the ways in which sound is deployed in contemporary fiction and non-fiction filmmaking. The aim of this chapter is to 'make strange' (*ostranenie* in the terminology of the Russian Formalists) the familiar conception of film sound, in order to contextualise the analyses of the functioning of non-fiction and fiction film sound tracks that occurs in the analytic chapters. In short, the aim is to *de-naturalise* the current prevalent mode of sound-to-picture, so that we can more fully comprehend its functioning.

The following discussion focuses largely on early sound-to-picture practice. This might seem a strange choice given the focus of the analytic chapters on modern and contemporary film. The logic in this choice is that the conventions of filmmaking relevant here – that is, sync-sound, narrative integration, point of audition, and so on – were established during the early period of film, some of them even during the silent-film era. It makes sense then to attend to the early practices of sound-to-picture in order to situate those practices, and the resources that I claim they make available, in the moments of their establishment. As mentioned above, such a focus helps us to understand the explicit function that those resources were intended to fulfill.

Furthermore, this chapter's objective is to describe the practices, principles and traditions of sound-to-picture in *emic* terms. As discussed in Chapter 3, this thesis' emphasis on sound-to-picture as a *practice* is designed to serve the social

67

semiotic project of understanding meaning making as an activity, not as a pre-structured 'code' (cf. Halliday, 1978; Hodge and Kress, 1988; Kress and Van Leeuwen, 2001). To that end, I provide a discussion on the technical, technological and semiotic resources, and the meaning-making potential, of sound-to-picture, both in the *in vivo* discourse of sound practitioners – the 'view from the trenches', as Sergi puts it (2004: 73) – and in the intellectual and academic discourse surrounding their practice (see Chapter 2 for an overview of this reflexively oriented literature on film sound as practice and text). Not only will this approach provide the means for a far richer understanding of the social semiotic practice of sound-to-picture than can be provided through text analysis alone (cf. Cottle, 2003: 5), but it will also prepare the reader for the analyses of Chapters 6 and 7 by explicating concepts and terminology in use there.

In what follows, I discuss the traditions and conventions that, it can be argued, have a bearing on contemporary sound-to-picture practice – in particular those practices apparent in the analyses of later chapters. To the extent that they can be separated analytically from traditions and conventions, I present, in the next chapter, the technical and technological resources that have become established in that practice. The technical resources form the conceptual basis for the subsequent analyses.

4.1   Precepts for sound-to-picture: traditions, conventions and influences

The scope of the present focus on resources for sound-to-picture is not only inclusive of those related to the technical practices of sound editors and designers. It also includes the *conventional* resources that derive from the evolution of the sound track as a site of semiosis, since the controversial 'coming of sound' in the early 20th Century to the fully integrated digital workflow of sound and image in the 21st. The resources I present in this chapter, both the apparently concrete (technologies) and abstract (techniques), are precepts that 'guide' sound professionals in the construction of the film sound track. To offer an example, one such precept is the conceptualisation of sound-track structure as *dialogue, music* and *effects* (hereafter *DME*).

Conventions and traditions as resources: *Dialogue, Music and Effects*

The *DME* structure originally referred to the three-channel master mix of Academy Mono (or Academy Curve) films, a Hollywood standard since 1938 (Sergi, 2004: 14). Since then, because of the national specificity of the 'talkie', in comparison to the international appeal of the 'silent' film, the music and effects (*M & E*) mix was introduced as a partial remedy to the threats of economic decline that the talkie brought with it (Sergi, 2004: 14). The *M & E* mix, in contrast to *DME*, includes everything *but* the dialogue track so that dialogue can be dubbed in different languages without having to re-create the music and effects tracks in the process.

Apart from fulfilling an obvious economic role, the continued tripartite structuration of the sound track, as dialogue, music and effects, enacts a certain audio-visual aesthetic and tradition, and reflects a specific way of understanding the semiotic role of the sound track in film. In Hollywood cinema, since the beginnings of the sound film to the present day, the voices of protagonists are made distinct from other perceptible sounds in the sound track, such as effects and music, in order to serve an aesthetic in which the intelligibility of dialogue, and thus narrative, is paramount (Doane, 1985: 58). The *DME* model serves this dominant mode of filmmaking via its function as a practical precept for maintaining the professionally ascribed functional boundaries around dialogue, music and effects.

Although *DME* is no longer bound to the 1938 standard of a three-channel mix, the differentiation of the sound track as dialogue, music and effects – what sound designer Larry Blake calls 'the three basic food groups of film soundtracks' (Blake, 1999) – continues to be the dominant conceptualisation of sound-track structure for sound practitioners. Its presence can be identified throughout all the stages of production and post-production sound: location recording, the selection and editing of sound effects, dialogue editing, and final mixing. In particular, the process of mixing (also known as 'dubbing' or 're-recording'), involves attending to the aural relationships – of loudness, timbre, spatial positioning – between all three components: dialogue, music and effects. For the reason that the final mixing process involves organising the relations between all three elements, it is the practice in which the functional *discreteness* of the roles of dialogue, music and effects is most easily seen. 'The mixing process,' as contemporary Hollywood sound

designer Gary Rydstrom explains (Sergi, 2004: 173),

> is about [...] focusing the audience attention on what's important at
> any given time. [...] It's a dance really between the different elements
> that are available on a sound track and how you play them [...] How
> you play music and dialogue and sound effects moment to moment
> really affects the effectiveness of a scene, the drama of a scene, so I
> think it's probably misunderstood by a lot of people as being a fairly
> technical exercise of just getting everything at the right level and
> you're done *but it comes down to a constant choice being made of
> what elements to hear, how much of them to hear.* (my emphasis)

However, the elements in a sound mix, in Hollywood filmmaking at least, are not of
equal status, and the choice of which are to be heard in a mix at any given moment
is constrained by the parameters of dialogue-oriented narrative filmmaking. In
another Hollywood sound designer's words,

> Everything [all sound effects and music] is balanced against the
> dialogue. The dialogue is the key because that's where the information
> is – that's the story, so people have to hear that. (LoBrutto, 1994: 179)

As Lastra describes it (2000: 92-122), narrative filmmaking brought with it a
particular and rigid hierarchical structure for the representation of elements in the
image composition, and, consequently, that of elements in the sound track. Given
the centrality of narrative in contemporary filmmaking, dialogue and other voices –
in the *DME* model – are invested with an entirely different functional status to
other identifiable elements of a film sound-track.

The *DME* sound-track structure is commonly thought of as a fundamental trade
principle, and as an established filmmaking standard. Like other standards of
filmmaking practice, the *DME* sound-track structure 'pre-exists' film production
and post-production, and its potential status as one option among many is rarely
questioned or considered . The most obvious reason for this is that the dialogue-led
narrative film has dominated filmmaking practice since the late 1920s, leading it to
become the 'natural' position on filmmaking from which other approaches are
thought to 'deviate'. Accordingly, the main function of the *DME* standard is to

70

ensure that the exposition of narrative, which in the modern sound film is in large part performed by dialogue, is clear at all times (Doane, 1985; Comolli, 1980) .

But the success of the *DME* paradigm, evidenced by its prevalence throughout the history of American and European cinema, is not only attributable to the dominance of narrative filmmaking. It is arguably also the result of human perceptual facility. As Chion (1999: 5) argues, 'in every audio mix, the presence of the human voice instantly sets up a hierarchy of perception.' This perceptual dominance of the voice in the listening-viewing experience (*vococentrism* in Chion's terms), as O'Brien (2005: 117) speculates, may be unavoidable '[g]iven the physiology of human audition'. This perceptual notion, coupled with the dominance of dialogue-led narrative film, explains why it is particularly unusual for contemporary filmmakers to integrate voices, sounds of the diegetic world and music in such as way as to challenge the established functional hierarchy of the dialogue-centred sound track, and especially to blend them into an undifferentiated whole[2].

However, although this structuration of the sound track is an established standard that still dominates contemporary narrative filmmaking, we *can* identify narrative-oriented films and filmmaking practices that deviate from that tradition. Such identification is crucial because it helps to illuminate the presence of *choice* that exists in a given filmmaking culture, at a given point in time. The *DME* tradition can be characterised as a *resource* for the reason that certain examples emerge from the cultural context of contemporary filmmaking that are exceptions to the dominant tradition, and that have, so to speak, *used* that tradition by subverting it, or some aspect of it. American filmmakers David Lynch and Robert Altman are two germane examples to invoke as regards the existence of choice in the established tradition of *DME*. In films such as *Lost Highway* (1997), Lynch composes a relationship between sound and image that is arguably more predicated on an emotional audio-visual congruence than designed for efficient narrative exposition, using 'sound effects the way most filmmakers use music, to help the audience understand the emotional state of character' (Dancyger, 1996 :175). Given his particular audio-visual aesthetic, it is not unusual that Lynch chooses to challenge the *DME* paradigm by foregrounding music and sound effects over

dialogue, and it is not coincidental that the traditional idea of narrative intelligibility is decisively spurned by Lynch in that film, the plot of which 'loops back on itself in an unending nightmare of psychosis' (Brophy, 2004: 156). Dialogue in *Lost Highway* is infrequent, monosyllabic, and, crucially, often submerged under the oppressive drones and tones of the onscreen spaces. In the protagonist's apartment, for instance, the sound of that space – what sound practitioners term 'room tone' – is represented as 'thick, fibrous, lichenal' (2004: 156), and builds into

> a dense bed of rumbling tone which lugubriously bleeds into the air-conditioned atmosphere of the cinema itself. [...] *Lost Highway*'s sound design *privileges* the very substance which the sound track [conventionally] figures as noise. (2004: 156; my emphasis)

In other words, Lynch's privileging of the elements of *DME* structure that are conventionally considered ancillary to the dialogue track, and the resultant 'positioning' of dialogue as a sound effect, amounts to nothing less than an inversion of the usual functional proclivities of the *DME* model.

Robert Altman, and his right-hand sound man, Richard Portman, for their part in challenging the *DME* paradigm, created many sound tracks that subverted the established analytic differentiation between dialogue and effects. This practice is often considered as co-evolving with certain technological and technical developments in production and post-production film sound, behind which Altman was a key figure. During the early 1970s, Altman helped to introduce multi-track recording techniques into film sound practice. In order to exploit the key affordances of multi-track recording technology that had already been established in popular music recording – i.e. greater control over the many parameters of recorded sounds such as relative loudness, timbre, dynamics and spatial characteristics – Altman developed a multi-channel radio-microphone system for individually recording the voices of actors. In addition to greatly changing aspects of film production by freeing the actors from the 'tyranny of the microphone' boom system, enabling a certain degree of improvisation[3] (Altman, 1985: 49), Altman's system also impacted hugely on post-production, by liberating the sound-mixing personnel from the usual constraints of production sound – such as the recordist's choice of microphone placement which determines to a large extent the relative

loudness of speaking characters – affording the luxury of sonic manipulation of *individual* voices. With the use of this system, as Rick Altman explains (1985: 50),

> Each track can then be dealt with [in post-production] separately in any of the ways in which sound signals have traditionally been handled (filtered, reverb added, amplified, etc.), so that the final mix can do anything from reproducing the exact sound actually heard from a specific point to constructing a highly contradictory set of signs which utterly splits the hearing subject. By manipulating his sound, [Altman] is in a position to manipulate his auditor independently from his spectator.

But Altman's film-sound innovations, as regards his challenging of *DME*, are too easily characterised as co-evolving with his technological and technical contributions. Such an explanation can be misleading, implying that both the filmic idea and technical 'solution' were exactly contemporaneous. In fact, as Schreger argues (1985: 349), 'the idea preceded the technology' by a significant number of years. Even in certain of Altman's pre-1970s films, such as *Countdown* (1967), Schreger identifies the genesis of Altman's contribution to film sound, arguing that, in those films, we can observe the director in the process of developing 'the practice of overlapping dialogue – or as he calls it, "live sound effects"' (Schreger, 1985: 349). As Schreger continues, Altman was primarily interested in 'dialogue simultaneously spoken by two or more characters: a wall of sound, a Tower of Babel, a film of bit players trying to get their two cents' worth in.' (1985: 349) By deploying dialogue as both carrier of story information and sound effect, Altman had begun to challenge the functional boundaries of the *DME* paradigm[4].

Altman first deployed his multi-channel recording system in *California Split* (1974), in which, as Brophy writes, the voices of the two lead actors 'are more oral performance than scripted dialogue' (2004: 51). Most importantly, as Brophy (2004: 51) goes on to elaborate,

> the soundtrack does not privilege Charlie and William [the two protagonists] above others. The sound of a babbling bartender will be just as loud as Charlie when he is delivering an important piece of dialogue; injection by an off-screen extra will dogfight with William's banter.

In adopting such an approach, Altman can be viewed as attempting to subvert the differentiation that the *DME* model proposes – especially that between dialogue and 'ancillary' effects. In later films, such as *The Company* (2003), a film which takes the practices of New York's Joffrey Ballet as its subject, Altman deploys dialogue as an 'atmosphere track': voices constantly overlap and, as in *California Split*, off-screen voices are presented at ostensibly the same level of loudness and clarity. In terms of the *DME* model, voices in *The Company* are therefore accorded the status of sound effects, albeit quintessentially human ones. At the same time, the sounds of the dancers' footfalls are foregrounded in the sound mix, rendering them perceptibly distinct from other sound events, so that what is usually considered, at best, an ancillary sound effect is here promoted to a central role. In this sense, Altman proposes a configuration of dialogue, music and sound effects that – because it challenges the semiotic status of one above the others – diminishes the expected centrality of the central characters, which in a general sense reflects his 'ensemble' approach to filmmaking. In a culture that has evolved around the notion of narrative intelligibility, as realised in large part through dialogue, it is perhaps surprising that Altman's films have found such a wide audience, since listener-viewers are required to accept that the sound of the human voice (including that of the main characters) can legitimately escape its conventional expository function. In addition, if we accept the notion of a vococentric bias in human perception, audiences are challenged to interpret a film such as *The Company* counter-intuitively: the 'hierarchy of perception' that most sound tracks enact and perpetuate is avoided by Altman, leaving listener-viewers to realign their expectations, and allowing the very status of central characters to be challenged, if not entirely subverted, by the particular mix of loudness levels in the sound track.

According to Philip Brophy (2004: 1), the notion of 'the sound track' is a 'chimera of cinema': 'It is sound *and* noise; noise *and* music; music *and* speech; speech *and* sound. At no point can it be distilled into a form which allows us to safely state its essential quality'. The limited number of established ways in which sound is actually deployed in modern filmmaking is the result of the institutionalisation of precepts for sound-track construction, achieved, for example, through the standardisation of technology and technique as a remedy for the

economic concerns of early Hollywood cinema (Sergi, 2004: 13). The prevalence of certain modes of audio-visual integration – and ways of conceptualising the sound-track's role – in filmmaking helps to naturalise those modes and perpetuate their status as 'givens'. As the remainder of this chapter will demonstrate, the widespread use of these modes, if not arbitrary, represents only a few of the many ways of differentiating and organising the perceptible features of the sound track.

4.2   Sync-sound: synchronisation as tradition

Although primarily understood as a concern of the history of technology, the synchronisation of sound and image is also a semiotic concern. In other words, not only can synchronisation be understood in terms of the audio-visual relationships it *makes possible*, but, crucially, it can be understood as 'any fixed or purposeful relationship between sound and image' (Lastra, 2000: 94). As it is deployed in this thesis, the concept of synchronisation is relevant mainly in terms of the *perceptible convergences* of the sound and image tracks. The phrase 'sync-sound' denotes the practice of deploying sound in a 'punctual' manner, such that sounds are perceptible as being synchronised with events represented visually. Following Chion (1994: 58), the phrase 'point of synchronicity' (hereafter POS) will be used to refer to those instances of convergence. The analyses of the subsequent chapters will rely heavily on this concept. In addition, 'sync-sound' stands for the many possible audio-visual relationships that the evolution of sound-image synchronisation has engendered; '[f]rom this perspective', as Lastra argues, '"lip-synch" is only one, rather banal, possibility.' (2000: 94)

As I demonstrate in this section, the uses to which sync-sound is put in contemporary filmmaking have their foundations in cinema's early period – ironically, the so-called 'silent era' of film. The resources of sync-sound have been additionally developed during the early years of the sound-film (late 1920s to the mid-1930s), when the dialogue-led narrative film became the dominant paradigm. The modern sound-film – as indebted to the classical sound-film as it appears to be – has further developed an audio-visual 'vocabulary', especially of sync-sound, that is, on the one hand, a refinement of classical film's resources, and on the other, an attempt to liberate the many sedimented ways of thinking about sound-to-picture

practiced in classical filmmaking, so that wholly new audio-visual resources are able to evolve (cf. Lemke, 2002: 302). In particular, filmmakers such as Andrei Tarkovsky, Jacques Tati and David Lynch have explored ways of doing sound-to-picture that call into question the conventionally assigned functions of the classical film sound-track, often by means of synchronising sound and image in a manner that 'eschews traditional narrative sound mixing in favour of the bizarre, the numinous, and the sacred' (Altman, 1992: 176). Many filmmakers in the contemporary period have followed these filmmakers' leads, continuing to use sync-sound in ways that challenge both the preconceptions and creativity of the sound practitioners involved, and that of the audiences, who also are 'trained' to conceive of film sound in particular ways. What is clear is that to more fully comprehend the complex functions of modern and contemporary uses of sync-sound, in both fiction and non-fiction filmmaking, its early documented deployments must be investigated.

The sound of 'silents': sync-sound in early cinema

The phrase 'silent film'[5], as historians of film sound have pointed out, is a glaring misnomer, given the fact that sound has had its place in film exhibition since the early 20th Century. As Lastra identifies (2000: 98), it has become 'one of the great revisionist clichés of film studies' to point out that the silent film was never silent, yet it bears repeating here. As the cliché tells it, prior to the standardisation of synchronised sound in the late 1920s, live-performance music and sound effects were a commonplace in silent film exhibition (Altman, 2005). However, that account presents an over-simplified inversion of the lay notion of the silent film (Altman, 1996: 656-658). In point of fact, as Lastra argues, sound prior to 1910 was often 'an integral part of the nickelodeon program *except* at the time during which the film was projected' (2000: 98; original emphasis). As he continues, sound's most important function in early cinema was as a means of 'attracting attention to the nickelodeon itself' and whether 'in the form of phonograph, barker, drummer, or piano, [it] primarily served the function of hailing potential patrons.' (ibid.) However, it is crucial to point out that in certain theatres music did indeed accompany only the film, remaining silent at all other times (Lastra, 2000: 99) – although, as Altman argues (1996: 677), before 1910, audiences were not led to

expect that films would be accompanied continuously by music. What arises from this account is that, even while the cinematic *image* had a uniform character (Altman, 1992: 45), no *single* approach to sound and musical accompaniment was dominant nor standardised in early cinema exhibition (Lastra, 2000: 99). At that point in cinema's history, without an established standard for how to deploy sound in silent film exhibition (not *production*, it must be remembered), a number of very different approaches to sound accompaniment were able to co-exist. It would take until the late 1920s before the filmic uses to which synchronisation was put, and the functions it was to fulfill in modern and contemporary American cinema, were ostensibly established. Below, I discuss the functions, effects and practices of sync-sound, and the resultant relationships of sound and image, that were initiated in the early period of cinema.

4.2.1 Functions, effects and practices of early sync-sound

Prior to the institution in the late 1920s of several systems for synchronising sound and image, the possible modes of audio-visual integration were many and varied. From approximately 1907 to 1926, several accepted notions of 'matching' between sound and image – what it meant for sound and image to 'go together' (Lastra, 2000: 94) – were evidently in use. Also during this period, several recognisably distinct types of audio-visual text can be identified – the most salient being the 'illustrated song' film, the sound-accompanied slide lecture, and the multi-shot narrative film – that brought with them distinct modes of sound-image interaction. With the tendency towards narrative filmmaking after 1910, a general notion of audio-visual 'matching' began to evolve that aimed to fulfill the overarching narrative functions of the emerging dominant film form. This is what Lastra refers to as 'narrative integration': the functioning of both sound and image to fulfill narrative ends. This section describes the practices of early sync-sound that can be considered resources for contemporary film sound, in both non-fiction and fiction filmmaking. These practices are organised under the following functions of early sync-sound:

i   added value;

ii   sound as performance, image as text;

iii   practices of 'realism'.

i   Sync-sound as 'added value'

As with the musical accompaniment of film, sound practitioners (such as drummers, pianists and prop men) would perform to a pre-recorded image assembly, and, depending on the year or programme of a film's exhibition, would not aim to discourage the audience from giving their performance a certain level of attention.

> Like the barker, but in concert with the screen, the singer, pianist, and drummer directly acknowledged their audiences, and even elicited response through cleverness, ingenuity, and above all spot-on performances stressing synchronism. (Lastra, 2000: 104)

Perhaps inevitably, the live performance of sound-to-picture led to a situation in which the skill and virtuosity of a performance attained a value additional to that of the film it accompanied. In the slide lecture, for example, approaches to sound accompaniment would emphasise the precision and timing of sound performances, placing 'an unusual emphasis on performance, discursivity, and "liveness"' (Lastra, 2000: 103). During this period, as Lastra continues, 'both lecturers and "prop men" understood "synchronisation" as a category of *evaluation*' (2000: 103; my emphasis). The live accompaniment of later, narrative-based films continued with an emphasis on the accuracy of a sync-sound performance. In Lastra's words,

> One measure of the sound practitioner's success seemed to be exactness of his or her timing during performance. These and other sound practices stressed performance to such an extent that audience attention often seemed split between the world on the screen and the performers in the theatre. This *dual focus* fostered the tendency to evaluate all sound, but particularly 'synchronisation,' in terms of accurate performance.(2000: 104; my emphasis)

Exploiting the situation, many companies produced films that provided the opportunity for sound effects, the advertising, aimed at exhibitors, making this affordance explicit (2000: 102). Exhibitors recognised in the practice of live-sound

78

performance an opportunity to attract more patrons to a picture show, and 'promoting the sound component of the motion picture show seems to have been good business for exhibitors, even if manufacturers had already begun the drive toward narrative integration and image dominance.' (2000: 101). Yet although it is clear that exhibitors and performers did aim to foreground the added value of sound – its 'production value' – the virtuosity of live performers attained a certain value that cannot be *reduced* to the commercial concerns of the film industry of the time. It should not be forgotten that professional integrity and craftsmanship were also key factors in the development of a style of sound accompaniment that valued diversity and precise performances (2000: 107). Practitioners such as the Lyman Howe troupe were often lauded for extremely sophisticated performances in which the *practice* of doing live sync-sound constituted a significant part of the critical focus. Howe, in particular, went to great lengths to ensure a performance that would impress an audience, but insisted on foregrounding the performance and the performers (pianist, drummer and vocalist), attempting to establish their practice as a legitimate 'art', and not as mere 'hidden' accompaniment (Lastra, 2000: 103).

The performers' timing aside, exhibitors, critics and audiences found value in the diversity and skill of a performance, which was essentially based on the choice of sound effects in use. Certainly, the perceived *innovation* of the performer would constitute much of the value of live film-sound, and the novelty and range of the chosen sound effects would be at the foundation of critics' praise. As Lastra states (2000: 104), contemporary critics 'often praised drummers for the sophistication of their "props" and "traps" ['contraptions'] and commended the tendency to look for every opportunity to supply sound effects.' But as the narrative film took precedence, the performer's aural interpretation of the film image came to be judged on a criterion of *appropriateness*. Such appropriateness was based on criteria relevant to narrative continuity: that sound effects, and music, should 'fit' with the psychological character of a scene, not detract from or counter the image, or disrupt the linearity and continuity of the narrative. The specific functions of narrative necessitated a hierarchisation of elements in the image, and therefore acted as a constraint on the kinds of sounds that were to be selected or foregrounded in

synchronisation with the image. Specifically, a character- and plot-oriented film would require that these elements are treated consistently; that is, that a central character's presence in a shot or scene is always foregrounded or backgrounded appropriately in terms of the place of the shot or scene in the film's plot.

However, prior to and during the transition to the narrative film, the 'appropriateness' of sound effects and music was defined by far broader, non-narrative criteria. Perhaps inadvertently, the values attached to sound in that context were concerned as much with the *repleteness* of the sound performance – that all visual features were attributed sounds – as they were concerned with the burgeoning dominance of narrative integration. During the silent, live performance era, the appropriateness of this primarily *topographical* interaction with image (Lastra, 2000: 96) was predicated on a general movement towards creating an aural 'world' that matched the image in terms of complexity and depth (Lastra, 2000: 98). This approach to according all perceptible visual action and objects with sound and musical effects would later come to be concerned with a – rather complex – notion of 'realism'.

After the transition was more or less complete – establishing narrative films that comprised pre-recorded and synchronised images *and* sound as the norm – film sound continued to emphasise its presence as a significant and legitimate 'partner' to the filmic image, thus performing a demonstrative function in communicating the sound track's production value. Examples of sound effects 'for sync's sake' abound in the early period of synchronisation (1926 onwards), and trade journals of the time are replete with examples of films 'whose chief purpose seems to be providing the opportunity for sound "effects," which served not in support of the narrative but as attractions themselves.' (Lastra, 2000: 103) Indeed, at times it seems that although the narrative film form was dominant, the affordances of synchronised sound, and the will of the producers, studios and exhibitors, led to a filmic realisation that chose to forego the narrative 'mandate' for sound's deployment, instead offering a veritable showcase for the technology, technique and virtuosity of sync-sound. One such example is a sequence in the Vitaphone film *Old San Francisco* (Crosland, 1927), which is, this sequence aside, a 'standard "silent"

melodrama' with conventional narrative uses of sound based on character and ethnic difference. The film, to quote Lastra at length,

> ends with a spectacular earthquake lasting several (ca. eight) minutes chock-full of synchronised sound effects. [...] [T]he earthquake's spectacle is unwarranted by the film's narrative economy. Earlier parts of the film are sporadically accompanied by diegetic sound effects often tied to dramatic narrative points. [...] The earthquake sequence, however, contains the film's overwhelming majority of sound effects, and does so in an excessive manner.
>
> The non-stop symphony of crumbling buildings, running victims, roaring fires, and Anna May Wong's climactic scream offer a density of effects unheard in silent film, yet not entirely without precedent. Like some predecessors, this spectacularisation of sound was not primarily narrative in motivation, nor essentially "realist". Indeed, it stalls the plot, and its length is narratively unjustified. *It does, however, foreground the technology of sound recording and sound reproduction, offering its perfect performance for appreciation and admiration. The sound does not support the narrative here, but the narrative supports the gratuitous display of sound.*' (2000: 119-120; my emphasis)

The resemblance this example bears to later Hollywood film sound is striking, particularly the sound tracks of those films made between the 1970s 'Dolby era' and the present day (Sergi, 2004). Emphasising the technologies of film sound is clearly one objective of post-classical Hollywood sound track composition, although perhaps one would expect that, in contrast to the films of the late 1920s, there is nowadays far more subtlety to how this objective is realised. Certainly, sound-track construction has become highly sophisticated, in terms of the technical execution of sound editing and mixing, the dynamic range and frequency response of theatre playback systems, and the institutionalisation of technological standards that contributes to the 'transparency' of workflow – three aspects that have benefited from the transition since the 1990s to digital systems for film sound. However, the concept of 'production value' is very much still a prominent and conspicuous feature of Hollywood filmmaking, particularly in its 'blockbuster' output. Even if *Old San Francisco's* 'gratuitous display of sound' can be considered, with hindsight, a rather obvious attempt at attracting attention to the technical possibilities of sound-to-picture, the difference between early and later synchronised filmmaking style is

primarily due to the technical execution of its sound design, not the extra-narrative 'spectacularisation of sound'. In fact, I would argue that the central reason we understand lengthy action sequences and gratuitous displays of special effects – including sound effects – as being more narratively integrated in contemporary filmmaking than in early cinema is that narrative filmmaking has evolved over a century to fully encompass and integrate *opportunities* for such spectacles, leading us to understand those spectacles as somehow inherent to certain Hollywood film genres (I am thinking of action-oriented films and, since the Dolby era, comic book adaptations).

Outside of the present-day Hollywood system, in European cinema for instance, we can generalise less about the represented added value of sound. Moreover, its presence in the sound tracks of non-fiction films is even harder to identify. Yet, it should be accepted that this practice of foregrounding the 'added value' of sound has had an impact on a wide range of contemporary filmmaking practices, including documentary film, television news and 'Reality TV'. Significantly, as I argue in this chapter, the notion of sync-sound as added value is difficult to dismiss as irrelevant to the aims of, for example, contemporary non-fiction filmmaking for the precise reason that such filmmaking practice is not wholly separate from the institution of fiction-film production. In fact, in the recent uprise in high-profile documentary films – in particular, the Hollywood-backed *Faranheit 9/11* (Moore, 2004), *Supersize Me* (Spurlock, 2004), and *An Inconvenient Truth* (2006); and the BBC's *Planet Earth* (2006) and *Walking with Dinosaurs* (1999) – the distinction between fiction and non-fiction filmmaking, in terms of their 'production ecology' (Cottle, 2004) are increasingly hard to maintain. *Faranheit 9/11*, in particular, deployed a post-production crew that are extremely active in Hollywood fiction-film production. For example, its supervising sound editor and re-recording mixing engineer, Garry Rizzo, has also worked on *The Incredibles* (2004), *Hellboy* (2004), *Black Hawk Down* (2001). (Following this trend, *An Inconvenient Truth*'s post-production sound crew was headed by Skip Lievsay, whose other credits include *A Scanner Darkly* [2006] and *Men in Black II* [2002]). This is not to suggest that Rizzo would adopt the same sound-production rationale for documentary film sound as he would for that of a feature film. Having

said that, the boundaries around fiction and non-fiction are also difficult to maintain in terms of formal and aesthetic features; as Noel Carroll argues,

> [t]he distinction between nonfiction film and fiction film cannot be grounded in differences of formal technique, because, when it comes to technique, fiction and non-fiction can and do imitate each other. (Carroll, 1996: 286)

It is to be expected, therefore, that there are many production and post-production practices common to both fiction and non-fiction films, and that sound track construction is likely to be one of those practices, though the commonalities might manifest themselves in ways that are hard to compare without serious investigation.

ii   Sound as performance, image as text

The split between the audience's attention – or its 'dual focus', as Lastra puts it (2000: 104) – towards the images on the screen and the performers in the theatre is crucial to understanding the variety of types of sound-image interaction that were practiced prior to the 'narrative integration' of the pre-sync era, and certainly prior to the industry-wide adoption of synchronisation as a technological standard at the end of the 1920s. As mentioned above, narrative filmmaking established ways of deploying sound that were far narrower in scope than in cinema's early period in that they were overwhelmingly concerned with the basic needs of storytelling: character, action, plot and continuity. Before the narrative film began to dominate filmmaking and exhibition practices, the simultaneous presentation of visual text and sound performance in the theatre, and the kind of engagement that this required from audiences, helped to generate a range of ways for sound to engage with the image that were mostly disregarded with the fully-fledged narrative film. However, during the transition to narrative film, ca.1910, exhibitors and performers *resisted* the apparent mandates of the narrative text and continued to practice a mode of sound accompaniment that (1) addressed audiences directly, acting as a mediator between the film and the cultural specificities of the audience; (2) oriented itself towards the text in ways that were autonomous – at least thematically, if not temporally – of it and its producers, even going so far as to disrupt the narrative

continuities that were present in the film text (Lastra, 2000: 105). These two main facets of sound-to-picture performance should be seen as fundamental to the presence of non-narrative modes of sound-to-picture in the early period, particularly because they continued to be practiced during the period in which the filmmaking industry aimed to forcefully establish (mainly via the trade press) explicit standards of sound performance that were to be coherent with the narrative aims of the film text (Altman, 2005; Lastra, 2000). The later technological solution to the purposes of narrative integration – synchronisation – was the culmination of this period, and its ultimate arrival and subsequent precedence over filmmaking practice should indicate just how forceful the drive towards narrative integration actually was.

A notable example of the 'direct address' of early sound accompaniment, and its relative freedom from the (often) narrative logic of the film image, is the practice of 'funning' or 'kidding' a film. Being one of the 'last arenas of exhibitor control', as Lastra writes, sound

> allowed various theatres not only to appeal to the religious, economic, and ethnic particularities of their audiences but offered the possibility of ridiculing the film, if it pleased the locals. (2000: 106)

The practice of 'funning' (Lastra, 2000: 106; Altman, 2005: 237), as a specific part of the more general practice of audience 'direct address', is perhaps the most remarkable of all live-accompaniment of silent narrative film, since it illuminates just how independent a sound and music performance could be from the film image. Unlike other live sound practices which merely diverted the audience's attention away from the film and its narrative, funning went to great lengths to connect with the image-narrative, in a satirical manner, by providing (mostly amusing) *commentary* on the action depicted on-screen:

> [Funning] satirised scenes or entire films through musical puns, commenting on the picture through the title, lyrics, or melody of the accompanying music. [...] Such techniques catered to particular audiences and their prejudices and/or drew attention to the musician's cleverness or stupidity, but did so to the detriment of the film, whose uniformly coherent address was hopelessly fractured. (2000: 112)

The authorities and trade press who were at this time attempting to standardise musical accompaniment, thus 'robbing it of any nonsanctioned voice' (2000: 112), nevertheless considered catering to 'the ethnic, class, and religious particularities of the audience' an essential function of a distinguished sound performance. Particularly in the smaller theatres and in 'ethnic enclaves [...] these techniques continued to provide a non-narrative or meta-narrative pleasure while implicitly cementing a bond between performer and audience. To the consternation of manufacturers, this community often came together at the expense of the film.' (2000: 113) In other words, because 'direct address' practices of sound remained so important to exhibitors, producers, and the rest of the emerging film industry, they left film exhibition practice open to the kinds of performance practice that would threaten the coherence of the very product they hoped sound accompaniment could serve.

Not only did the direct address of sound threaten to undermine the narrative continuity of a film, it also ran counter to certain other aims of producers and exhibitors. In terms of musical accompaniment, pianists' 'careful punning' of a film would elicit an appreciative response from its audience. Such an outcome was of value for performers (see *i* above), as funning drew attention to their craft, and also communicated their ability to define the functional scope of film-sound performance, and the power of sound and music – the media they expertly harnessed – to drastically alter (by *recontextualising*) the film's meaning as constructed by aspects of the image.

As one might expect, the value of funning for performers, and the audible excitement it generated from the audience, was not shared by film producers and exhibitors. For them, funning the film by musical and sonic means undermined 'filmic intentions', and with the rising interest in narrative films from both producers and audiences, exhibitors and producers wished to focus audience attention on the screen, as a space in which meaning was made and the narrative was conveyed, and divert it from the very visible, decidedly extra-narrative, performers (Altman, 2005: 226). However, as will be explained below, the general desire was to conceive of narrative films *with* sound accompaniment, resulting in a strategy to establish certain standards of sound and musical performance that would bolster, not counter,

the film image and its narrative aims, and to make its presence in the theatre far less visible.

Turning now to the discussion of documentary filmmaking, it can be argued that the above characterisation of sound's direct address can be understood in terms of Bill Nichols' (1991) concept of the 'reflexive mode' of documentary representation. As Nichols argues, this mode 'gives emphasis to the encounter between filmmaker and viewer rather than filmmaker and subject' (1991: 60). Specifically, in foregrounding this particular interaction, reflexive filmmaking aims to expose the usually effaced conventions of filmic representation – particularly those of the other modes of documentary representation that Nichols typologises: expository, observational and interactive. In doing so, it explicitly problematises the 'reality' that other kinds of documentary representation claim to reveal. Reflexive documentary, as Nichols explains (1991: 33),

> arose from a desire to make the conventions of representation themselves more apparent and to challenge the impression of reality which [the expository, observational and interactive documentary] modes normally conveyed unproblematically. [...] [I]t uses many of the same devices as other documentaries but sets them on edge so that the viewer's attention is drawn to the device as well as the effect.

Drawing attention to the construction of a mode of representation clearly aims to *speak directly* to audiences about the way in which they attend to both non-fiction and fiction filmmaking. The deployment of sound to perform this function, either alone or with the image, is one way in which the – marginal and performative – direct address of early cinema has continued in spite of the widespread institution of the classical narrative-film paradigm. However, since documentary films that adopt the reflexive representational mode are comparatively uncommon – the observational and interactive styles being the most prominent in contemporary documentary filmmaking – it is not surprising that sound tracks that serve to reflexively critique, satirise or parody documentary conventions (whether conventional relationships between sound and image or the conventions of either the sound or image track separately) are rare also. The rather specialist case of the

animated documentary film (see Ward, 2005: 82-99 for discussion) offers one example of a documentary form that does communicate direct address – but this occurs primarily through the (obviously constructed) image, not through the sound track. In animated documentaries, such as Tim Webb's *A is for Autism* (1992), 'the very constructedness of the animation forces the 'reflection' on form and meaning that is central to a 'reflexive' mode, something that sometimes gets lost in the seductively mimetic world of live action'. (Ward, 2005: 89) *A is for Autism*, for example, deploys a sound track that includes the verbal accounts of several sufferers from autism, while the image track reproduces and animates the drawings and paintings of these people. So while the sound track has an indexicality which is unproblematic (in the interactive tradition of filmmaking [Nichols, 1991]), the shifting, 'squiggly' forms of the image track appear to function as illustrations of the sound track's verbal accounts (2005: 94). This inversion of the usual functions of the documentary-film image track (which, in terms of the sound track being clearly primary in the sound-image relationship – in the sense that the image track depends on the verbal testimony as an already recorded 'event' – has a close affinity with the sound-dominant 'illustrated song' and slide-lecture of early cinema exhibition) problematises the nature of the relationship between sound and image in non-fiction filmmaking in a way that is obviously reflexive. While it should be maintained that the image track is the primary site of direct address in animated documentary film, it is also fair to argue that the sound track – in staying its ground as a site of testimony in Nichols' sense of an 'interactive mode of representation', and not functioning as a representational site of subjectivity, as the image track certainly does – plays a significant role in what is a sophisticated audio-visual critique of non-fiction film conventions.

Aside from the animated documentary, in which reflexivity is in operation whether an explicit intention of the filmmaker or not (as Ward argues, 'you cannot have an animated film which is anything less than completely "created"' [2005: 85]), and in which the sound track, even when maintaining its indexical status, contributes significantly to an animated film's reflexive, direct address, other examples of documentary sound-tracks that practice direct address can be found in parodies of television programme genres. Falling under a category of 'satirical

documentary', television programmes such as *Brasseye* (Morris, 1997) and *The Day Today* (Ianucci, 1994), which are both parodies of non-fiction film types – current affairs and broadcast news respectively – deploy sound in precisely this 'reflexive' manner, and thus in a similar way to early, more musically oriented, funning practices. In several interviews with politicians and celebrity figures, *Brasseye*'s makers radically reconstruct the subject's speech via digital editing techniques, 'putting words in their mouths' and making fun of the conventions of the medium of television, the techniques of its producers, and the genres of news and current affairs. It might seem fair to argue, since both sound and image are edited simultaneously in these programmes, resulting in an obviously fractured image continuity, that this does not constitute funning in the traditional sense because both sound and image tracks bear the scars of the severe editing, whereas in early funning practices the image is left intact. However, because it is the editing of *speech* here which is primary in the 'funning', there is obviously a very close resemblance to early modes of kidding the image.

iii   Practices of realism

Establishing the notion of realism in sound-to-picture was a critical development in the live sound and musical accompaniment of early cinema. What came to be understood during this period as the 'realistic' use of sound effects, it can be argued, both emerged *from* the direct-address tradition of sound performances, and became oriented *against* that tradition. In the first sense, the accurate and skillful synchronisation of sound effects (which I will refer to as 'spotting') to all perceptible visual features in order to create an impressive and complex sound track – an essential characteristic of early audio-visual realism – developed from a context in which sound's performative dimension was foregrounded. What later came to be perceived as a realistic relationship of sound and image was first complicated by the 'dual focus' (on sound-performance and image-text) that direct address functioned to encourage. In other words, while many agreed that comprehensive spotting of sound effects resulted in a more successful approximation of perceptual reality than previous deployments of live sound, the foregrounding of the performative dimension of sound accompaniment meant that

this dual function of sound challenged the spectator's experiential, perceptual immersion in a given film that was considered necessary to a fully successful 'reality effect'. And so, in the second sense of realistic-sound's orientation *against* a performance's direct address, the performative aspect of sound accompaniment – the main hindrance to the establishment of realism as the general aim of early sound-to-picture – had to be backgrounded, and hidden from the audience's purview. The initiation of 'behind-the-sheet' techniques during the late aughts (ca. 1908) attests to the perceived incompatibility of direct address and verisimilitudinous realism (Altman, 2005: 236), and emerged in response to a desire to exploit what producers, filmmakers and sound practitioners felt to be a latent, inherent 'property' of sync-sound: the realism 'that synchronised sound effects are so good at producing' (2005: 236).

What this rather simplistic, linear account of the emergence of realism as a precept for sound-to-picture does not make clear is the fact that 'realism' as a concept held different meanings at different points in the evolution of the early sound-film (Altman, 2005: 245). So while it may seem appropriate to talk about realism as a stable notion for sound practitioners and critics of the early cinema period – for instance, in terms of fidelity to perceptual reality, as with the *phonographic imperative* model (discussed below) – there were, in fact, several distinct notions of realism that were operational during the establishment of narrative filmmaking as the dominant form, and during the first few years of sound-film practice proper. During this time, practitioners and critics worked together in their attempt to define notions of realistic ways to 'match' sound and image. Particularly in terms of sound practitioners' engagement with the image – that is, how they chose to interpret and 'score' the image for sound accompaniment – distinct notions and practices of audio-visual realism were in use, as several scholars of early film sound have identified and explored (e.g. Altman, 2005; Lastra, 2000). The purpose of this section is to expound these practices, and to establish them as resources for sound-to-picture beyond their historical context. Following a brief discussion on sync-sound and realism, *topographical, narrative* and *phonographic* models of audio-visual realism are investigated in terms of how they impacted on

the early practices of sound practitioners, and, briefly, their influence over later practices of sound in fiction and non-fiction filmmaking is discussed.

'Tin-sheet thunder and coconut shell hooves': realism and synchronisation
In more ways than one, human perception became the model for many of the approaches to film sound that existed in cinema's early period. In particular, early notions of realism, and the strategies employed to achieve them practically, were built on knowledge of how humans aurally experience the world. As I discuss later in this section, this perceptual model formed the basis for the beginnings of the sound-film proper, in the sense that the model of sound recording that governed early understandings of how sound and image should go together was founded on the established practices of phonography, and its criteria for what counted as a faithful representation of perceptual reality. Here, though, I discuss the basic potential for realism that synchronisation held in terms of a very basic perceptual model: the expectation of hearing a sound when seeing an image of an event that we *know* causes a perceptible sound.

For film sound theorist Michel Chion, synchronisation of sound and image is instrumental in the audience's acceptance of a film's overall 'reality' effect – he terms this the *audiovisual contract*, the willingness to accept the 'truth' of whichever sound has been joined with whatever image (by 'truth' I mean the notion that what we hear is a natural consequence of what we see). Along with this comes the perceptual effect of synchronisation: *synchresis* (Chion's amalgam of *synthesis* and *synchronisation*). Synchronisation, for Chion, 'is an important factor in film in how it manages to glue together entirely unlikely sounds with unlikely images' (1994: 54). As he elaborates (1994: 63), synchresis

> makes dubbing, postsynchronisation, and sound-effects mixing possible, and enables such a wide array of choice in these processes. For a single body and a single face on the screen, thanks to synchresis, there are dozens of allowable voices.

The implication in this definition is that synchronisation *produces* audio-visual realism. It does not perform at the service of a realism that exists simply in utilising the 'correct' sound for an onscreen event (Lastra, 2000: 147); close synchronisation

90

of sound effect and visual event is crucial to the production of realism. Trade journals of the early period attest to this very claim, as Altman has argued, quoting an article from 1908:

> Many exhibitors are prone to overlook the importance of music as an essential to the success of the show. Even a piano properly handled, as regards the various situations in the story being portrayed on the canvas, will be found to be an almost complete equipment. When people see an object fall, a natural tendency makes them listen for some sound as it strikes, and a thump on the proper key of a piano will give all the realism necessary.('Exhibitions with Sense', *Views and Films Index*, 4th Jan, 1908; quoted in Altman, 1996: 698)

As Altman goes on to comment (1996: 698-699), modern readers would find this passage perplexing since music and sound effects are conflated into one filmic phenomenon. For Altman, the author's point is precisely to functionally define 'music and the human voice as sound effects'. This explains how the author is able to make the, perhaps equally surprising, claim that 'realism' can be generated by the synchronisation of a piano note with an event onscreen. 'In 1908,' he surmises, 'film music clearly remains under the spell of the magic of synchronisation' (1996: 699). In the historical context of the late 'aughts', the treatment of sound effects and music as if they were *functionally equivalent* (Altman, 1992: 113-125; Bordwell et al, 1985: 117-127) is a consequence of the effect of synchronisation's 'magic': it makes little difference exactly *what* the source of a sound effect is, as long as it conveys a sonic character appropriate to the visual event it accompanies and, importantly, is 'tightly' synchronised with the image. This same phenomenon is explained by Lastra (2000) in the context of a discussion on 'fidelity' of a sound to a visual source. As he concludes (2000: 147)

> Decades of tin-sheet thunder and coconut shell hooves prove [...] that fidelity to source is not a property of film sound, but *an effect of synchronization*. A gun firing on the screen accompanied by any brief, sudden, explosive sound *produces* the effect of source; it doesn't require it as a pre-condition.

The basic potential of sync-sound to 'glue together entirely unlikely sounds with unlikely images', in Chion's (1994: 54) words, is crucial to understanding the

91

potential of sync-sound to *construct* reality, rather than 'merely' reproduce it. In the crudest sense, a 'gap' that exists between sounds and images not of the same pro-filmic source is 'closed' by the effect of synchronisation. In making a paradigmatic substitution of an expected sound for one that is entirely *un*expected, the audio-visual closure that a point of synchronicity produces does not quite efface the work that might be done in modifying the effect of realism – as in Chion's definition of sync-sound's potential. In fictional genres, such as comedy, much of the humour of hearing, say, a 'honk' synchronised with a person slipping and hitting the floor derives from our expectation of an (appropriate) accompanying sound being met, for a brief moment, and then just as swiftly realising that our expectation has been mocked. However, apart from these *obviously* comedic deployments of sound, the audio-visual 'gap' *can* be closed to such a degree that the effect produced through close synchronisation is not so obviously constructed. This subtlety of sync-sound can occur even in comedic forms: Chion uses the example of Jacques Tati's *Mon Oncle* (1957) to describe how a

> little boy's footsteps on the cement in the yard make a pleasant and concrete rustling, while those of his father, a large, uptight, and unhappy man, only produce a thin, unrealistic 'ding'. (1994: 115)

In this example, even with as pedestrian a film-sound effect as human footsteps, the difference between the effect of synchronisation of the boy's footsteps and that of his father's contributes to the construction of certain aspects of their respective characters. And yet it does so in a manner that we could argue is subtle and 'naturalised'. Were it not Tati under discussion – a filmmaker notorious for using sound in ingenious yet obvious ways – and were it not Chion analysing his work – a theorist whose expertise is precisely in the area of film-sound – we might even expect that, to the average listener-viewer, the 'closure' is complete and the work of synchronisation completely effaced. Put another way, sync-sound and its ability to produce realism, can and does go largely unnoticed. In the semiotician Roland Barthes' (1977) discussion of 'trick effects' in the realm of photography, he argues that the power of such effects as bringing together an American Senator and a Communist leader in a way that they appear in conversation, as he identifies in a

92

photograph in the American press in 1951, derives from the fact that those effects 'intervene without warning in the plane of denotation; they utilise the special credibility of the photograph – this [...] being simply its exceptional power of denotation – in order to pass off as merely denoted a message which in reality is heavily connoted; in no other treatment does connotation assume so completely the 'objective' mask of denotation' (1977: 21). What Barthes identifies here concerning a 'naturalisation' on the denotative plane of 'trick effects' can be extended to the present discussion of sync-sound, for alterations and manipulations in the sound track of an audio-visual point of synchronicity also 'utilise the special credibility' of the medium of audio-visual film in order to naturalise the combining of sounds and images through synchronisation. (The potential of sync-sound to naturalise sound-image combinations is explored further in the two analytic chapters.)

As this suggests, effacement of the work of sync-sound is an extremely valuable resource for filmmakers and sound practitioners. Indeed, those genres of film that do not generally aim their deployments of sync-sound towards comedic effect, such as the documentary film, *do* attempt to exploit the realism effect of sync-sound and engage in manipulations of synchronised sound in ways that are not immediately noticeable. In non-fiction filmmaking, the use of sync-sound to alter the pro-filmic matching of sound and image in insipid ways is, in the main, considered as 'tampering' with the very reality that non-fiction film claims it is privileged to engage with. It should be obvious why this might be the case. However, 'tampering' can occur in ways that filmmakers do not consider as such, for instance the commonplace, apparently only technical, adjustments in timbre and dynamics that are achieved with frequency-equalisation and audio-compression tools. In such a case, the actual source of the sound need not be fully *substituted* in order to create effects that, sometimes, radically alter how an audience perceives and interprets a moment of sound-image synchronisation. As I will go on to discuss below, such concerns have been at the core of disagreements and misgivings over synchronisation – the 'coming of sound', as it is commonly referred to – since its very beginnings in the late 1920s, even in the arena of fiction filmmaking, let alone non-fiction film and the expectation of a 'close' engagement with pro-filmic reality that it generally invites (Nichols, 1991: 29; Plantinga, 2005). Particularly in France,

the early 1930s engendered practices that aimed to counter the 'manipulation' of reality that sync-sound apparently encouraged, chief among them the practice of 'direct sound' (*le son direct*). This practice will be explored in detail below, but here let it simply be suggested that synchronisation opened up the possibility of 'choice' in terms of which sound to combine with which image. It is this presence of choice that is both at the heart of those misgivings towards sync-sound, and at the core of the semiotic possibilities that film sound practice has gone on to nurture and refine as crucial to its definition as a *creative* activity, not merely a technical one. Proponents of direct sound clearly saw the choice implicit in synchronisation practices, and interpreted that very choice as a problem, not a virtue. For Chion (1994: 105), direct sound was 'not only the sole morally acceptable solution in filmmaking but also the one that simplifies everything, since it eliminates the problem of having to make choices.' What direct-sound practitioners treat as a problem is fundamental to the practices of other film sound practitioners, and this fact should go some way towards characterising these very different, almost polarised, orientations to film sound.

The ability to produce a sense of realism remained at the core of further attempts to develop mechanical systems for sound-image synchronisation. Key to their success was their ability to match the precise timing of a live sound performance, since it was this feature of synchronisation that early critics and audiences identified as crucial to the effect of realism. In other words, although a criterion of accurate synchronisation is certainly not sufficient in present-day judgments of audio-visual filmic realism, *early* realism was apparently determined on the basis of such a feature, and it is this knowledge which enables us to understand differences between earlier and later notions of realism. Once sync-sound became valued in live performances as key in the production of realism, the practice developed to achieve two things: (1) the improvement of the accuracy and precision of synchronised performance; and (2) to evolve beyond the basic criterion of accuracy to become concerned with constructing a complex 'sound world' that would simultaneously impress audiences and further develop filmic realism by 'producing the effect that sounds emanated from sources on the screen' (Lastra, 2000: 106).

While one can argue that the practice of sync-sound itself was sufficient in convincing audiences that sounds emanated from objects onscreen, and thus of the audio-visual 'reality' of a film, it is clear that a much closer approximation of perceptual reality would entail that sound effects are given to *all* visual objects that would produce sound in human experience of the 'real' (pro-filmic) world.

Topographic and Narrative modes of engagement

As suggested above, '[r]ealism connected the realm of sound performance *with the demands of narrative integration* by producing the effect that sounds emanated from sources on the screen' (Lastra, 2000: 106; my emphasis). However, prior to the establishment of an institutionalised mode of sound accompaniment that revered the narrative structure of the filmic image, the effect that onscreen objects and actions produced sound in the diegetic world of the film image was used to *realistic* but non-narrative ends. Following Lastra (2000) and Altman (2005) this thesis distinguishes between *topographical* and *narrative* modes of engagement, as concepts that deal with sound performers' engagement with the filmic image at different points in early-film history. As Altman (2005: 245) argues, the strategies for realism in the 1900s – the nickelodeon period – and those of the 1910s can be distinguished in the following terms:

> During the aughts, realism was sought in what we might term a positive or additive manner; that is, exhibitors went out of their way to add anything to the presentation that might increase realism, including multiple sound effects, voices behind the screen, or synchronised phonograph recordings. During the 1910s, the process was reversed as exhibitors were instead instructed to avoid anything that might undermine realism. From this standpoint, many earlier practices were found lacking. Even though each earlier synchronised sound event added realism in a punctual [and topographic] manner, the overall effect was now perceived to threaten [narrative] realism. The campaign to standardise sound thus concentrated instead on a series of 'dont's,' each one designed to avoid jeopardising diegetic coherence.

As I have indicated in this passage, Altman asserts that sound-image 'realism' in the 1910s was concerned with narrative, whereas the the earlier, nickelodeon period's 'punctual' approach to sync-sound is more concerned with a realism that is

95

topographically produced. A topographic engagement of sound with image can be defined as an approach that is concerned with treating each shot as an event in itself, rather than as one component in a larger narrative unity. In the early period, such treatment meant that the phenomenal specificity of each shot was attended to and augmented with sound, and the image 'scanned' for objects and actions that would produce sound (Lastra, 2000: 96). Although often the performativity and direct address of a performance was a major motivating factor for performers wishing to engage in a topographic manner with the image, a commitment to the uniqueness of each shot – especially in terms of its represented phenomenality and space – was an overarching mandate for those performers (though it was not an explicit, trade principle, as the narrative 'mandate' certainly was[6]).

In its comprehensiveness, the topographical mode of engagement, and the version of realism it produced, clearly drew considerable attention to the *performance* of sound. In the narrative-oriented 1910s, any aspect of a sound performance that drew attention to itself *as* as performance, in Altman's view, was doomed to be downplayed. As with other new technologies, Altman explains (2005: 245), early cinema 'sought to capitalise on its novelty by accentuating the apparatus'. Later, while 'striving to establish a new classicism, cinema would increasingly seek to conceal its devices'. (Altman, 2005: 245) So not only were performers to be hidden 'behind the sheet' in efforts to maintain audience attention on the screen and those elements central to narrative (characters, actions and continuity), but also the repleteness of the performer's engagement with the image, no matter how diegetically faithful, was to be reined in so that 'realism' was not perceived as *constructed*.

Whereas a topographical approach is defined as one which prioritises the uniqueness and phenomenality of a shot, scanning the *entire* image, and 'spotting' for sound effects accordingly, the narrative film demanded a particularly different way of attending to the screen, entailing a 'radical redefinition of the cinematic image' for both spectators, performers and producers alike (Lastra, 2000: 127). Specifically, this redefinition required that the onscreen actions and events (that

were to be accorded sound effects or musical accompaniment) be spatially hierarchised in order to fulfill the 'pluripunctual'[7] functions of narrative continuity. Accordingly, the aim of sound accompaniment, and the measure of success of a sound performance, gradually shifted from a *topographic* relationship to the image, in which the *entire* image was scanned for events and appropriate sounds performed in synchrony (Lastra, 2000: 96), to bolstering a hierarchical structure that prioritised those elements of an image composition central to narrative exposition – character, action, and spatial continuity between the contiguous shots that constitute a scene (Iedema, 2001; Nichols, 1991). This tradition emphasised, what Lastra calls, 'a rigid hierarchy in providing sound effects, separating the image into zones of importance and of unimportance.' (2000: 110). As he puts it, 'a gun wielded by the villain was an appropriate cue for sound effects, but a tiny background canary was not.'

A narrative mode of sound-to-picture, therefore, can be defined as an approach to sound-effect performance that aimed to augment the already encoded *hierarchies* of narrative-film image composition. This approach was oriented decisively against topographic sound-image engagement, with the trade press strongly urging performers to follow a strict code of performance (Altman, 2005: 245) that would not threaten the intended narrative realism of a film by disturbing its encoded 'spatial hierarchies'. Ultimately, with the help of a mode of sound-image engagement that would bolster, not disrupt, that hierarchisation, the narrative film shifted the attention of both performer and spectator away from the phenomenal specificity of each image, and towards groupings of shots (scenes) that made meaning in their combination and texture. In doing so, it also created a notion of realism that was essentially determined by the abstract thematic concerns of plot, rather than the concrete specificity of each shot; or, as Lastra succinctly puts it, 'the relevant terms of comparison are uniqueness versus recognisability or event versus structure[8]' (2000: 139)

As Roland Barthes (1977) argues in the context of a discussion on the phenomenal, material specificity of the film image, attending to the 'grain' of each filmic moment 'structures the film differently' (Barthes, 1977: 64). As mentioned briefly above, for

97

producers of the late 'aughts' the main problem of a topographical engagement with the image was that such practices could 'destroy the intended spatial hierarchies of the image.' (Lastra, 2000: 105) In the trade press of the time, rigorously investigated by Lastra, accounts abound of

> overactive drummers and effects men who, in an attempt to 'enhance' a picture, go to absurd lengths to punctuate every image with a sound. One writer describes lecturing with a film only to be interrupted by the repeated sound of a bird whistle. Perplexed as to its supposed source within the image, he asks the effects man and is pointed to 'a diminutive canary in a tiny wooden cage on a top shelf at a far corner of the room,' whereupon he 'wallops' the whistler and proceeds. (2000: 105)

The explicit concerns of producers – as documented in the contemporary trade press – over sound's ability, in the topographical mode, to 're-structure' the image by, for instance, foregrounding elements that were not necessary to narrative exposition and backgrounding those that were *not*, clearly attest to the perceived power and signifying potential of sound when synchronised with the filmic image. A key consequence (perhaps an *aim*) of the rise of narrative filmmaking in the early 20th Century was the diminished status of the phenomenal – or sensory – effects of filmic representational events, of both image *and* sound. For example, with the narrative film's dominance, cinematography became concerned with representing each filmic moment as a 'piece in a puzzle', a unit that made sense only in terms of its place in an assembly of other units. Lighting a main character in order to foreground her is one identifiable key method of maintaining the salience of that character across any number of shots and scenes, thus fulfilling a requirement of narrative continuity and character-centered filmmaking. Prior to the institutionalisation of the narrative film, however, cinematography, like photography, was concerned with aspects of the image that were deemed irrelevant to the objectives of story-telling, but still very much germane to what was understood as 'cinema'. These aspects diminished in importance in early cinema practice after the narrative film took hold.

As with the typographical particularities of printed text which are of little importance when compared with a text's semantic signification, the film image

became a relatively transparent signifier of a verbalizable event, rather than an 'event' in and of itself. Whatever it thereby lost in complexity and 'grain', it compensated with a gain in legibility and signifying stability.(Lastra, 2000: 96)

Exploring the notion of the phenomenal specificity of the filmic image as 'grain', in his essay *The Third Meaning*, (1977) Barthes discusses what he calls the 'obtuse' meaning in film. The obtuse meaning, for Barthes, is that which overflows the 'obvious' (*obvius*) meanings of the 'informational' and 'symbolic' levels (1977: 52-53), and therefore cannot be verbalised or otherwise represented (re-presented) through language: the third meaning is 'a signifier without a signified' (1977: 61). In filmic terms, the 'third meaning' is the individuality and phenomenal specificity of the shot, or the 'grain' of the filmic moment (cf. 'the grain of the voice', in Barthes, 1977) and it is in this sense that Barthes notion of the obtuse is germane to the present discussion on topographical and narrative modes of sound-image engagement (and to the perceptual fidelity model of realistic representation discussed in the next section). In particular, Barthes' (1977: 60) argument that if we remove the third or obtuse meaning, 'communication and signification [of narrative and symbolism] still remain, still circulate, still come through' resonates profoundly with the notion that the narrative film, rigidly conceived as plot and character exposition, had little or no use for the phenomenal, material specificity of the individual shot, and that sound was thus enlisted to augment the communication of narrative, allowing its non-narrative functions to become redundant. This necessarily entails that phenomenal specificity is not crucial to the functioning of the narrative film, for 'without [the obtuse meaning] I can still state and read' a film (1977: 60).

The third meaning is key to another, related notion of Barthes': that of 'the filmic' (1977: 64-65):

> The filmic is that in the film which cannot be described, the representation that cannot be represented. The filmic begins only where language and meta-language end. Everything that can be *said* about *Ivan [the Terrible]* or *[Battleship] Potemkin* can be said of a written text (entitled *Ivan the Terrible* or *Battleship Potemkin*) except this, *the obtuse meaning* [...]

99

As he elaborates, the filmic is extremely rare in contemporary film. Its rarity can be explained in terms of the prevalence of narrative filmmaking: '[f]orced to develop in a *civilisation of the signified* [the narrative realm], it is not surprising that (despite the incalculable number of films in the world) the filmic should still be rare' (1977: 65; my emphasis). Furthermore, and most pertinent to this thesis, the filmic can most powerfully be brought out with the deployment of sound[9]; specifically, with the possibilities of *a*synchronous audio-visual montage which Eisenstein championed. In Eisenstein's own words,

> [T]he basic centre of gravity [...] is transferred to *inside* the fragment, into the elements included in the image itself. And the centre of gravity is no longer the element "between shots" - the shock – but the element "inside the shot" - the accentuation within the fragment [...] (quoted in Barthes, 1977: 67)

In stark contrast to Eisenstein's notion of montage, the rise of the narrative film meant that visual filmic devices 'such as dissolves, fades, and wipes' lost their phenomenal or sensory appeal – their obtuse meaning was rendered redundant – because they began to be determined by narrative functions, and were instead mobilised in terms of 'their ability to signify certain kinds of spatial and temporal articulations in a consistent manner.'(Lastra, 2000: 96) Similarly, sync-sound very materiality was at risk of being diminished with the institution of narrative-oriented sound practices, because narrative subservience overshadowed any non-narrative effects that the sound track might have generated. Implied here is the notion that sound – and any other representational mode that involves a sensory dimension – necessarily overflows the demands and constraints of narrative, and therefore cannot be reduced to its functions. Indeed, even when seemingly fulfilling a narrative function as mundane as maintaining intelligible dialogue, 'sound may simultaneously be performing other, nonnarrative or even nonrepresentational ones'. (Lastra, 2000: 97) In terms of dialogue recording, functions that are unnecessary to the aims of narrative include providing 'access' to the phenomenality of an image or series of shots by means of representing its 'spatial signature' (Altman, 1992: 61). Beyond non-narrative functions, dialogue recording can also bring with it non-representational effects, such as the sensory qualities of the human

100

voice. Although it can be argued that these qualities can be easily enlisted to serve narrative and representational functions, this does not preclude them from producing effects that are *beyond* representation. Indeed, in stark contrast to other filmic aspects, sound practices survived the transition to narrative integration more intact than other elements precisely because, as Lastra puts it, 'its functions and effects were only occasionally narrative' (2000: 97).

The above claim that sound's 'functions and effects were only occasionally narrative' (2000: 97) is not only to argue that there are always an untold number of uncodified functions of sound at play in narrative film, ones that could not possibly be identified predictably or reined in and placed under filmmakers' control. It is also to assert that non-narrative functions of sound – as far as the standardisation of reception environments will allow – have been, and are in the process of being, codified across the multiple domains of contemporary audio-visual practice, such as music videos, computer animations, video games, and what we might recognise as the 'traditional' audio-visual film, both fiction and non-fiction. As I demonstrate in the later, analytic chapters, these non-narrative and sometimes non-representational functions of sound that were practiced in topographic engagements with the image, and have co-evolved (though 'in the margins', as Lastra reminds us) with narrative modes of sound accompaniment, perform functions and produce effects that are central to contemporary filmmaking, and contribute greatly to the audio-visual semiosis of particularly non-fiction film genres.

The 'phonographic imperative' model

In contrast to the narrative and topographical approaches to (and definitions of) audio-visual realism, the *phonographic imperative* model is a concern of sound recording and reproduction, and therefore emerged later in cinema's evolution. Like the practice of phonography, its primary concerns are with perceptual fidelity – a fact that was difficult to assimilate into a production model for the early sound-film. The emergent conflict between the Hollywood mode of narrative realism and the perceptual-fidelity mode rigidly adhered to by sound practitioners (who arrived to filmmaking from the industries of radio, phonography and telephony) can be understood as one founded on a difference in aesthetic and representational

principles. Questions of perceptual fidelity were largely irrelevant to the sound tracks of the narrative fiction film, as was the case throughout film's history prior to the coming of sound. As O'Brien argues,

> a film's impression of story-world wholeness did not require the reproduction of an extra-filmic reality. In fact, more often than not, a film's diegetic world amounted to a construction whose impression of coherence rested on technical ruses of one sort or another, from the virtual geography of the eyeline match to showstopping *trucages* ('tricks') and special effects, as well as tried and true costcutting methods such as trompe l'oeil scenery and day-for-night shooting. (2005: 109-110)

In essence, the notion of realism and the representational principles that were shared by the majority of sound engineers in the late 1920s and early 1930s were founded on an 'inappropriately literal notion of realistic duplication', making it extremely difficult for those principles to be incorporated into the dialogue-led narrative objectives that the film industry were intent on perpetuating. (Lastra, 2000: 170) In basic terms, the phonographic imperative is to *reproduce* the spatio-temporal specifics of a musical performance 'as if heard from the best seat in the house' (2000: 139). Lastra defines the difference between the phonographic concern with fidelity, and the 'telephonic' concern with dialogue (narrative) intelligibility, in the following way:

> The 'fidelity' approach assumes that all aspects of the sound event are inherently significant, including long or short reverberation times, ratios of direct to reflected sound, or even certain particularities of performance or space. The 'telephonic' approach, not literally limited to telephones and voices, assumes that sound possesses an intrinsic hierarchy that renders some aspects essential and others not. (2000: 140)

The phonographic model of reproducing the experience of occupying 'the best seat in the house' suggests a singular *perspective*: one seat, and one pair of ears. In fact, whether it is the best or worst seat represented, the phonographic objective was to maintain fidelity to that one perspective. The fact that film-sound recording emerged out of the music industry[10] and therefore, at least initially, adhered to the phonographic imperative, meant that fidelity to the presumed perception of an auditor present at an original performance – a 'witness' to the event – entailed a

strict commitment to reproducing that auditor's aural perspective. In terms of perception, recreating perspective involves attending to aspects of sound production such as frequency 'curve' and the ratio of direct to reflected sound that an auditor would experience if present at a musical or pro-filmic event.

Once the perceptual fidelity model was introduced into the audio-visual arena, concerns over the representation of aural perspective were understandably thrown into relief, since the multi-shot narrative film posed the challenge of matching the sound track to a variety of different visual perspectives on the same pro-filmic event. Specifically, 'the single most important question occupying Hollywood technicians during the late twenties and early thirties was this: what relationship should obtain between image scale and sound scale?' (Altman, 1992: 46) The answer given by proponents of the phonographic imperative, and one they thought obvious, is that those scales should be 'matched' to respect 'the bodily integrity of the supposed "observer".' (Lastra, 2000: 160) On this logic, scale matching thus entails listening to and viewing a pro-filmic performance from a singular position (2000: 160) In practical, but crude, terms, 'scale matching' entails ensuring that long shots are systematically matched with distant-miking techniques, and close-up shots with appropriately 'intimate' microphone placement. As Joseph P. Maxfield, a prominent sound technician of the early sound-film, puts it, the aim of the perceptual fidelity approach, and the aim of scale matching, is to obtain 'a sound record which correlates with the picture in such a manner that a member of the audience is given the illusion of being an actual spectator in the scene'. (1930: 85; quoted in Altman, 1992: 49)

However, scale matching, as with the phonographic imperative in general, was the victim of a conflict of interests between, on the one hand, an industry whose representational norm was daily life as aurally perceived by a single auditor, and, on the other, an industry with clear narrative exposition as its measure of representational success. There were, for Wurtzler (1992: 97), 'those who advocated a strict matching of sound perspective with image scale, and, on the other hand, those who advocated a sound track of relatively uniform sound level that privileged

103

the intelligibility of the voice over the matching of sound and image perspective.' Effectively, this apparent conflict between the phonographical model and the established demands on sound of narrative filmmaking practice can be explained concisely as a practical incompatibility of perceptual fidelity and narrative (particularly *dialogue*) intelligibility.

From the particular standpoint of advocates of the narrative intelligibility model, the problems with the fidelity model's insistence on scale matching can be reduced to two related concerns: the impact of perceptually faithful sound-image relations on texture or continuity, and the challenge of representing dialogue intelligibly. As those advocating scale matching argued, the 'illusion' of cinema depended on the perceptually faithful reproduction of a filmed event and space. On this view, a commitment to representing the phenomenal specificity of the individual shot, and therefore an insistence on matching sound perspective to each shot, is a primary concern. As Altman (1992) and Lastra (1992, 2000) argue, the 'invisible observer' model's insistence on a committed sound-image perspective matching seriously threatened the model of realism that narrative filmmaking had begun to establish more than fifteen years prior to the sound-film. More precisely, the texture and continuity of multi-shot edited scenes were threatened by jarring ruptures between the component shots of a scene caused by the sudden changes in sound perspective that occur when the listener-viewer experiences *aurally* the jumps in perspective that were (and are) commonplace in film image-editing. Such sudden breaks in sound texture that result from trying to marry the strict matching of sound and image scales with a narrative-continuity approach to film editing were 'clearly inappropriate to the more homogeneous character of classical representation and narration that assumed *compromises* between literal fidelity and narrative clarity such as instituting narratively determined spatial hierarchies between the sharply focused foreground and the comparitively indistinct background.' (Lastra, 2000: 161; my emphasis)

With the severe breaks in continuity that scale matching caused came the related concern of how sudden changes in perspective impaired the much-vaunted

intelligibility of dialogue. The system for sound recording that respected scale-matching that was initiated (by Maxfield) was to record with a single microphone maintaining a constant relationship, not to the actors, as in the intelligibility approach, but to the camera's line of sight. Thus, when an actor moved towards or away from the camera, the correct loudness level and ratio of direct to reflected sound would be automatically inscribed to the sound recording (Altman, 1992: 50). However, as with the discontinuities that manifested from the distinct sound-image relationships *between* shots, the recording of dialogue was also subject to these shifting perspectives. Whereas the single-mic technique in a close-up shot would result in clear dialogue (as long as the actors did not speak away from the camera-mic setup), a shift to long-shot perspective would introduce a perceptible amount of reverberation and thus threaten intelligibility (Lastra, 2000: 161). In addition, in terms of dialogue recording, the very shift of perspective *itself* would introduce an aural discontinuity that would disrupt an audience's immersion in the dialogue-oriented narrative experience.

Although, as Maxfield insisted, intelligibility could be maintained by matching sound-image scales with the single-microphone technique (1992: 58), it is clear that the narrative film's model of intelligibility differed radically from that of the fidelity model. For narrative filmmaking, a model of intelligibility that derived from the theatre was evidently the standard. As one exhibitor head puts it, the greatest advantage of the sound-film was the

> ability to present every word so clearly and distinctly that no one need strain to hear what is being said, at least when recording and reproducing is properly conducted. A whisper is clearly audible from the front row in the orchestra to the last row in the balcony' (Franklin, 1930; quoted in Altman 1992: 58)

This model of theatre is one explanation for the sound cinema's insistent position on intelligibility when confronted with the sound community's incessant calls for fidelity; for Altman, it is 'because cinema continued to find in the theatre a long-consecrated code of reality applicable to audiovisual narratives.' (1992: 59)

Whereas it is clear that for filmmakers – as in the theatre – dialogue was not 'just

105

another sound', it is just as clear that, from the standpoint of early sound recording's phonographic orientation, the spoken voices of actors were sounds that must be treated on a par with other sounds in a recording. That is not to suggest that proponents of the fidelity model considered dialogue as equally important to other sounds within a narratively determined hierarchy. Indeed, as mentioned above, Maxfield and others claim intelligibility as one aspect of fidelity to a pro-filmic event. Rather, what was resisted was the 'artificial' foregrounding (by means of an adjustment to sound recording technique) of certain, narrative sounds at the expense of others. In other words, the sound technician saw it as the job of others to set up the pro-filmic scene so that the sound technician could record that event faithfully. (It should also be noted here that staging a pro-filmic scene for perceptually faithful recording was not at all problematic for sound engineers.)

The dominance of the objective of intelligible dialogue over considerations of reproducing aural perspective from a definite point in space, prior to Maxfield's single-mic solution, led to the implementation of a multiple-mic set up. During that time, the multiple-microphone technique became established, briefly, as the best method of ensuring an intelligible recording of dialogue. Whether performed live during a take, or later in post-production, a sound technician operating a mixing desk would choose the microphone source that provided, at any point during a take, the most intelligible sound. For the community of practicing sound technicians, the central problem with the narrative mandate for dialogue intelligibility was that the aural perspective resulting from the telephonic model's multiple-microphone approach is *indefinite*. As John L. Cass, a technician of the period, and staunch proponent of the fidelity approach, describes the effect:

> The resultant blend of sound [...] may not be said to represent any given point of audition, but is the sound which would be heard by a man with five or six very long ears, said ears extending in various directions' (Cass, 1930: 325; quoted in Altman,1992: 49)

For Cass, the problem of the multiple-microphone, narrative approach to a sound recording of intelligible dialogue was that contemporary, narrative image-editing practices 'forced' the spectator to 'jump from a distant position to an intermediate

106

position, and from there to close-up positions on important business,' while the sound track, mixed from differently placed microphones, would reproduce (since that was the aim of the fidelity approach) a variety of 'indefinite' perspectives that neither matched image scale nor maintained an otherwise fixed perspective (1992: 49).

It is clear that for Cass, and indeed the model of perceptual fidelity in general, the heart of the tension can be defined as that between *reproduction* and *representation*. For the sound technicians who had made their way to the film industry via phonography and radio, the narrative film's objective of *constructing* aural 'reality' could not be reconciled with their understanding of sound recording as a *reproductive* practice. However, as above, it was 'reality construction' *by means of sound recording technique* that sound technicians countered, not staging the event appropriately to narrative aims. Indeed, sound practitioners were certainly not against 'manufacturing' the pro-filmic scene by encasing whirring cameras and other noise-producing equipment in soundproofed boxes in order to get a 'good' recording.

With the single-mic technique established, an early practice of the sound-film – the multi-camera, live take – engendered the practice of deploying a continuous sound take in synchrony with a sequence of different shots, thus sacrificing scale matching by producing a combination of a single, definite aural perspective and multiple visual perspectives. Clearly a conceptual forerunner of the continuity function of the classical and contemporary sound track, this technique involved 'a single microphone system synchronised to two or three cameras', enabling complex image-editing to be anchored in a continuous sound track (Altman, 1992: 51). In later classical filmmaking, and continuing into the present day, this aim at sound continuity (as a solution to the lack of a 'grammar' of shifting aural perspectives, as was established in image editing: the 180 degree rule, for example) remains a standard method for constructing an audio-visual sense of temporal linearity, though since live takes of entire scenes are not the norm – except in live television broadcasts – aural continuity is entirely synthesised either from recordings made

separately from those of the pro-filmic event, or edited from within spoken lines of dialogue. The aim of the dialogue editor, for example, as well as to piece together intelligible, naturalistic-sounding takes, is precisely to 'situate' those often severely edited fragments of dialogue within a 'real' space. To that end, what sound practitioners call 'room tone' is deployed to produce a sense of continuity in the dialogue and therefore to efface the work of editing and the construction of the sound track (Doane, 1985).

Although strict scale matching was ultimately subjugated to the overwhelming concerns of narrative filmmaking and techniques for constructing foreground/background hierarchies in the sound track, it did partially survive in the development of one particular technique for audio-visual perspective that will be central to later discussion on non-fiction film sound practices: that is, *point-of-audition* (POA) sound (Altman, 1992: 60). As Lastra's puts it, the acoustic specificity of a particular pro-filmic space that was central to the standard of perceptual fidelity 'was quickly pressed into a carefully defined, and carefully delimited, role. It was soon harnessed almost exclusively to a particular character within a fictional world'. (2000: 142) POA sound is discussed in more detail in Chapter 2.

In the conflict – both representational and professional – that emerged between dialogue intelligibility and perceptual fidelity can be recognised a close similarity to the tension that existed in earlier cinema practice between topographical and narrative approaches to matching sound and image. Where topographic modes of engagement with sound involved a commitment to treating the shot as an autonomous semiotic unit, without regard for the discontinuities it caused between shots, so did those acting under the banner of 'perceptual fidelity' in later film practice attempt to treat each shot as an individual event without care for the narrative hierarchies encoded in the image. And where the narrative mode of sound-image engagement necessitated servitude to those image hierarchies, so did the aim of 'intelligibility' disregard the phenomenal specificity of each shot and the

phonographic imperative's requirement that the 'scales' of synchronous image and sound be matched on a criterion of fidelity to the audience's perception of image perspective.

As with the continuance of a topographic mode of engagement in contemporary film-sound practice, as I have argued above, the perceptual fidelity model too has continued into the present-day filmmaking context, despite being subjugated, within a decade, to the 1930s overwhelmingly powerful drive towards narrative filmmaking and the insistence on a model of 'sound space' prestructured by concerns over the intelligibility of dialogue. Most obviously, the treatment of dialogue as 'just another sound effect' by American filmmakers such as Robert Altman (discussed earlier in this chapter) easily suggests a resurgence of, at least some aspects of, the phonographic imperative as it was operationalised in the world of film sound. But less obviously, fidelity to the pro-filmic event has maintained an influential position as a viable filmmaking method in the tradition of *direct sound*. In the final section of this chapter, *direct sound* is one of two practices I consider as running counter to the conventions that became established in the early sound film. Both the direct-sound tradition and the Soviet filmmakers' call for asynchronous sound and image are discussed below, and characterised as 'resistant' practices; that is, practices that evolved as a reaction to the establishment of certain conventions, such as the *constructed* space of the narrative film, and the capacity of sync-sound for creating a convincing illusion of reality. These two traditions have remained marginal, yet their presence as resources in contemporary fiction and non-fiction filmmaking cannot be denied.

4.2 Voices of Resistance: Direct Sound and Asynchronism

Although direct sound and the asynchronism advocated by early Soviet filmmakers do not share a historical commonality either in terms of their specific periods of development or their central tenets of practice, they *do* share a similar reactionary impulse: they are both responses to the mainstream institution of synchronised sound and, as Altman has put it, the realism 'that synchronised sound effects are so good at producing' (2005: 236; my emphasis). While they share this commonality

of purpose, they respond to the perceived negative potential of sync-sound in polarised ways: that is, whereas direct sound reacts by insisting on a commitment to the *original* sync-sound of location sound recordings, the Soviet school argues against that very practice, favouring entirely *a*synchronous sound and image. In other words, where one saw that 'reality' is most faithfully reproduced through a strict adherence to an original temporal co-ordination of sound and image recording, the other viewed sync-sound's 'reality' as a representational *effect* best avoided in film practice.

Direct Sound

In basic, technical terms, direct sound (*son direct*) involves recording sound and image simultaneously. In more complex, political terms, direct sound involves nothing less than the repudiation of the artifice of American and European mainstream filmmaking practice. It is, then, both an aesthetic and technical approach to film sound, and, by this point in the chapter, a familiar *ideological* standpoint. For embedded in the practice and tradition of direct sound is the very tension between *fidelity* and *intelligibility* that is evident in the differences between the narrative and topographic modes of engagement of early 'silent' cinema, and in the conflict that later arose between the phonographic imperative and dialogue intelligibility in narrative-filmmaking. As in those contexts, at issue is the *reproduction* of a unified subject position, as in the perceptual fidelity model, versus the *construction* of an indefinite position, as in the intelligibility model.

The practice of recording sound and image co-spatially and contemporaneously was initiated in the very early period of the sound-film. In Hollywood, it constituted general practice from the late 1920s until it was abandoned in 1932. By this time, in contrast to its early phase as 'recorded theatre', American filmmaking had become a veritable process of *assembly* – the piecing together of sound and image 'fragments' in order to construct a legible, coherent text. In France, on the other hand, these multi-track techniques and re-recording practices were not practiced until almost a decade after they were succeeded in Hollywood. (O'Brien, 2005: 2) Even then, as O'Brien argues, direct sound 'remained unusually common in French filmmaking.'

110

(2005: 2) In the 1950s and 1960s, direct sound experienced a resurgence in French filmmaking practice, in part a result of the development and accessibility of portable sound recording equipment. But whereas in the 1930s, direct sound appeared, to film modernists at least, as a mainstream rejection of the aesthetic, 'art house' principles of late silent film (O'Brien, 2005: 8), later direct sound 'became essential to the styles of the most ambitious alternative filmmakers of the time' (2005: 158). As O'Brien continues,

> the young auteurs of the postwar decades [...] can be said to have adopted the earlier techniques in light of a self-conscious reflection on direct sound's artistic possibilities, which in key respects were alien to pre-war culture. (2005: 158-9)

Whereas the first incarnation of direct sound became associated with 'aesthetic backwardness' because of its servitude to the simulation of 'theatrical or radio-broadcast liveness', its later incarnation – which concerns us here – made explicit its 'stylistic implications' and thus emphasised its artistic potential. (O'Brien, 2005: 162) Furthermore, postwar direct sound was recontextualised as a moral choice, not 'merely' an aesthetic one; that is, a reflexive exploration of alternatives to the mainstream sound-cinema initiated in the late 1930s. As O'Brien argues concerning direct sound's new status, 'a manifesto-like statement can be found in Straub and Huilet's claim[12] that "direct sound is not merely a technical decision but a moral and ideological one"' (O'Brien, 2005: 163; Straub and Huilet, 1985: 152).

As mentioned above, Chion (1994: 104) has argued that direct sound is widely considered 'not only the sole morally acceptable solution in filmmaking but also the one that simplifies everything, since it eliminates the problem of having to make choices.' Here, Chion's explanation of the simplification offered by direct sound suggests a mainly *practical* simplification. What Chion does not explicitly recognise (or chooses to avoid) is that the 'problem' of the presence of choice in sound practice – indeed in any representational practice – is a moral one as well as a practical one. Put another way, direct sound is the 'sole morally acceptable solution' exactly *because* it eliminates choice as a factor of filmmaking practice. Choice, then, for proponents of direct sound, introduces the potential for

111

'corruption' into filmmaking practice, whereas *son direct* serves to maintain a commitment and 'respect for the shot' (as Alan Williams has put it [O'Brien, 2005: 98 fn.]).

Yet, despite the militant strictness apparent in discussion of direct sound, the situation in practice is quite different. As Chion argues (1994: 119),

> [i]n France, where just about everyone toots his or her horn for direct sound, most filmmakers still go ahead and partially or wholly redo sound after shooting anyway; ordinarily, though, this fact is hidden like a shameful thing.

While Chion's generalisation might be unfair, his observation does raise an interesting, though unintended, point: even when direct sound is performed in a 'pure' manner – that is, without any post-synchronised treatment – the *pro-filmic* event can be prepared so that a particular kind of audio record is achieved. Such issues of 'staging' have been raised in the earlier discussion on the phonographic imperative in early sound-film practice, in which the notion of 'fidelity' to the pro-filmic event enabled a certain latitude in practice, whereby a variety of technical 'solutions' circumvented the problem of simultaneously remaining faithful to the event *and* achieving dialogue intelligibility. In other words, while advocates of 'strict' direct sound reject the manipulation of sound in post-production, they *do* exercise control through production considerations, thus shifting manipulation over aspects of the sound track from post-production to production. As Altman (1992: 173) observes,

> '[d]irect' sound [...] has been treated by people who should know better as an unmediated form of film sound recording, as if a preference for sync recording of location sound could possibly circumvent the mediation implicit in the choice of a particular microphone, location, or volume level.

In the specific practice of documentary filmmaking also, directors who deploy a direct-sound methodology are able to ensure that dialogue and narrative intelligibility will be maintained by making appropriate technological choices. Although Ruoff, in an essay on documentary sound, makes the general claim that location sound recording in observational documentaries 'does not clearly

differentiate foreground and background spaces; rather, all sounds compete together in the middleground', he goes on to describe precisely how this is circumvented, in a case study of the documentary series *An American Family* (PBS, 1973):

> Shotgun microphones are frequently used in documentary productions precisely because they allow for a choice of narrative information and raise the ratio of direct to reflected sound [eliminating reverb], thus isolating sounds in the environment. Directional microphones enable recordists to place certain sounds in the foreground while relegating other sounds to the background. (1992: 225)

In short, and in Chion's words, '[t]he notion of direct sound involves no less of a reconstruction (even if by simple subtraction) than the notion of postsynchronised sound. (Chion, 1994: 106) There are those, however, who have attempted 'pure' direct sound, eschewing the possibilities for staging the pro-filmic event and for insuring a narrative-oriented spatial hierarchy (proponents of the *Dogme 95* manifesto, for example). In such cases, as Chion argues,

> [t]he result is totally strange. Is this because the spectator isn't accustomed to it? Surely. But also because reality is one thing, and its transposition into audiovisual two-dimensionality (a flat image and usually a monaural soundtrack), which involves radical sensory reduction, is another. What's amazing is that it works at all in this form. Indeed, we tend to forget that the audiovisual tableau of reality the cinema furnishes us, however refined it may seem, remains strictly (on the level of reproduction) that with which a sketched reproduction of a human, with a circle for a head and sticks for the arms and legs, is to an anatomical drawing by Albrecht Dürer. There really is no reason for audiovisual relationships thus transposed to appear the same to us as they are in reality, and especially for the original sound to ring true. (Chion, 1994: 96)

Implicit in these quotations is the degree to which direct sound attempts to serve the functions of both a 'fidelity' mode of sound-image engagement and a narrative 'intelligibility' one. The use of directional microphones, coupled with the practice of staging the pro-filmic event to exhibit the desired sonic characteristics, allows the sound recordist and filmmaker to rest safe in the knowledge that continuity between shots, though not ensured to the same degree as in the Hollywood model, will not be entirely jeopardised by adopting a direct-sound approach. As long as the

recordist maintains the desired foreground/background hierarchy by positioning directional microphones appropriately, and as long as the continuity of an image sequence is not disrupted by 'anecdotal' sounds that are discontinuous, the continuity so central to the narrative sound-film since its inception can be produced.

Perhaps most importantly, 'authenticity' – the *raison d'être* of direct sound – is maintained by striking a delicate balance between foregrounding dialogue in a recording and maintaining fidelity to the specific acoustic space and its ambient sound characteristics. As Marie argues: '[d]irect is really a manifestation of a new modality of voice recording in film' (quoted in Ruoff, 1992: 223). As this suggests, direct sound has evolved to become a way of doing dialogue-oriented film-sound that can still lay some sort of claim to perceptual fidelity. Authenticity, then, becomes an signified *effect*. In non-fiction filmmaking, in particular, the authenticity of direct sound – what Nichols (1991: 184) discusses as a certain aspect of 'documentary realism' – is a product of its evidentiary function: as with location *image* recording, direct sound entails an indexicality that 'testifies to presence'. Location recording in non-fiction filmmaking reassures the listener-viewer that 'the filmmaker was there, the evidence proves it' (1991: 184). However, this should not suggest that a technique such as direct sound can be deployed as the sole marker of a film's representational status as fiction or non-fiction. As Nichols argues, just as the indexical function of the film image does not guarantee its historical authenticity – only 'the bond between image and what was present before the camera' – direct sound, as an example of realist style, is also more a guarantee of a 'historically real recording of a situation or event, whatever its status' than one of an event's status as historically real (1991: 184-5). As Nichols (1991: 185) elaborates:

> [s]igns of presence – of recognisable people, places, and things, of familiar
> sounds and images; signs of incomplete control over what occurs of how is
> unfolds – imperfect framing, missing elements of action, loud background
> noises – such signs may be less evidence of the historical world than of the
> real recording of a world whose status as representation remains open to
> question and debate. These signs testify to presence, but not necessarily to
> the presence of historical reality. They more properly testify to the presence
> of the recording apparatus and the reality of the recording process which we,
> often on faith, assume to have occurred in the face of pell-mell contingency.

Understandably a mainstay of the reflexive mode of documentary film (see section ii above), the signification of the artifice of filmmaking is understandably less desirable in fiction film. However, direct sound is extremely popular in modern and contemporary French filmmaking. Given the risk of revealing the 'illusion' of film that is entailed by using direct sound, why is direct sound deployed by fiction filmmakers? What is at the heart of their apparent recklessness? Nichols argues that fictional realism has been most vaunted for its ability to efface itself, to enable 'unimpeded access to the world of representation' (1991: 180). In other words, deploying direct sound (as part of a more general 'realist' approach) in fiction filmmaking does not entail testifying 'to the presence of the recording apparatus and the reality of the recording process' because, as Nichols puts it:

> [T]he centripetal force of the narrative draws such signs of authenticity into the woof and warp of the story; the location becomes one more signifying element, more or less well motivated in relation to the plot. (1991: 181)

From this perspective, the multi-track methods of Hollywood sound practice and the *construction* of reality they are brought in to serve are redundant, since the very context of the fiction film ensures that the narrative governs the audience's interpretation of a particular sound-image relationship, whether it be composed of direct sound, with its mostly indeterminate sound 'effects', or the constructed sound track, with its non-dialogue sounds being mostly post-synchronised and fully determined. I will return to this issue in Chapter 8, where it is argued that the distinct materialities of sound and image media, and their culturally ascribed indexicality, ensure that there is always a certain amount of functional 'overflow' at stake in their use.

The contextualisation of the signification of 'direct sound' as a *feature* of a film, suggests that discussions concerning how to distinguish non-fiction and fiction film (as form, not practice) should not centre on techniques employed and their historical contexts of deployment. We cannot posit direct sound as a marker of functional differentiation between non-fiction and fiction since it is deployed extensively in both forms (especially in France in which, the practice has 'endured [...] to the point of defining the national film style' [O'Brien, 2005: 11]). As Noel

Carroll argues with respect to formal technique as a distinguishing feature,

> [t]he distinction between nonfiction film and fiction film cannot be grounded in differences of formal technique, because, when it comes to technique, fiction and non-fiction can and do imitate each other [...] The distinction between nonfiction and fiction, therefore, does not collapse with the recognition of stylistic correlations, since the distinction never rested upon such formal or technical differentiae in the first place. (Carroll, 1996: 286-287)

The relationship between filmic techniques and the fiction/non-fiction distinction will be discussed in the later analytic chapters.

Asynchronous sound

As I have discussed above, the audiovisually represented world can be 'de-realised' in the narrative mode of film sound, by mismatching sound and image scale. Though this is not the explicit aim of the dialogue-oriented narrative mode, it has been that of certain early filmmakers. Orson Welles, for instance, created films in which sound and image are closely synchronised, but spatially mismatched. This aural suggestion of simultaneous connection and disconnection to the represented world of the image 'negates reality' (Mintz, 1985: 289) and 'makes [the listener-viewer] vaguely uncomfortable, slightly dislocated, usually without our knowing why' (1985: 290).

As Mintz goes on to suggest (1985: 290), Welles' technique of sound-image de-realisation can be understood in terms of audio-visual *counterpoint* – albeit a 'harmonised' one – and therefore resonates with the theories of audio-visual montage that Eisenstein and certain other Soviet Filmmakers championed since the very beginning of the synchronised sound-film. The crucial difference between Welles' technique and a Eisensteinian counterpoint – a difference that Mintz's qualifier ('harmonised') is supposed to indicate – is that where Welles' spatially 'disrupted' sound-image relationship is closely synchronised, Eisenstein called for entirely asynchronous sound.

To clarify, asynchronous sound does not refer to sound that is *out of sync* with the

images – as would be the case if a scene was shot with direct sound, and then the sound and image tracks were misaligned in post-production. Rather, the phrase asynchronous sound (*non-sync*, for filmmaker Alberto Cavalcanti [1985: 110]) denotes sound that was *not* synchronised with the image during production. The asynchronous sound track, therefore, is an entirely constructed one, yet not constructed towards the 'slavish imitation of naturalism' (Pudovkin, 1985: 86), as is the case in the majority of sound films until the present-day.

That the call of certain Soviet filmmakers was political and reactionary is evident in their joint publication of the *Statement on Sound* in 1928. This veritable manifesto, authored by Eisenstein, Pudovkin and Alexandrov, proposes that the technical advance of the sound-film, as the *Statement* reads, 'is being employed in an incorrect direction' (1985: 83), citing the 'talkie', in particular, as a prime example, and disparaging it as merely filmed theatre, therefore lacking artistic merit in any regard. Specifically, although synchronised sound does offer 'potentialities' in terms of advancing the film as an art form, its employment in the cinema primarily as a means for reproducing speech led Eisenstein and his cohorts to argue that the place of sound in the cinema will most probably 'proceed along the line of least resistance, i.e. along the line of *satisfying simple curiosity.*' (1985: 83; original emphasis) This satisfaction, according to the *Statement*, is to be found in the sound-film's basic capacity for 'providing a certain illusion of talking people, of audible objects, etc.' (1985: 84)

The most profitable way in which to explore the artistic possibilities of the sound-film – to avoid creating audio-visual 'illusions' – was to deploy an asynchronous sound-track. As the *Statement* demands (1985: 84):

> THE FIRST EXPERIMENTAL WORK WITH SOUND MUST BE DIRECTED ALONG THE LINE OF ITS DISTINCT NONSYNCHRONISATION WITH THE VISUAL IMAGES. [...] [O]nly such an attack will give the necessary palpability which will later lead to the creation of an ORCHESTAL COUNTERPOINT of visual and aural images.

Counterpoint, for the Soviets, is the key element of *montage*, the basic principle of

which, as Eisenstein expounds (1943: 14), is that 'two film pieces, of any kind, placed together, inevitably combine into a new concept, a new quality, arising out of that juxtaposition.' In response to the coming of sound, Pudovkin – co-signer of the *Statement* – came to extend the notion of montage to be inclusive of, what he termed, 'vertical' montage: 'the montage of sound and picture' (Cook, 1998: 85). For Eisenstein, a non-synchronous, contrapuntal approach to film sound held the greatest potential for developing film as an art form: 'art begins when the creaking of a boot on the sound-track occurs against a different visual shot and thus gives rise to corresponding associations' (quoted in Barthes, 1977:61).

Sound was thus incorporated, at least theoretically, into the practice of montage. As Eisenstein argued, it was not that the introduction of film sound itself must be resisted. Rather, it was that the practice of montage, so central to film and the Soviet identity, should not be compromised by sound. Indeed, sound was considered an easily integrated component of montage:

> There is no fundamental difference in the approach to be made to the problems of purely visual montage and to a montage that links different spheres of feeling – particularly the visual image with the sound image[13].

The damage that synchronous sound would do to montage was argued on the basis of a central principle of montage: *neutralisation*. For Eisenstein, montage was able to function successfully because the photographic images it deployed – the montage 'pieces' – were divorced from 'all surrounding reality, transforming 'nature' into a block of material (a complex sign, if you will), which the editor can use in assembling a work.' (Weis and Belton, 1985: 76) Sound, however, seriously threatened that process 'by restoring power and autonomy to the photographed object, increasing 'the independence of its meaning' and thus 'its inertia as a montage piece'. (1985: 76) In other words, in the primarily visual approach of montage (sound is the *addition* here), the introduction of synchronous sound 'restored' *context* to otherwise *decontextualised* ('neutralised') images, that could, and did, serve the purposes of montage well. Indeed, a contrapuntal use of sound

> will prevent the sounds and images from further realising the world from
> which they derive and enable them to function as neutral fragments of
> material whose potential is realised in a montage sequence. (Weis and
> Belton, 1985: 86)

Not only did the *Statement* reveal that the Soviets believed sound to offer entirely
new ways of making meaning with the montage form, but also it was considered a
solution to certain problems that were a consequence of the increasing complexities
of theme and story. Such solutions could not be found in the resource of the image
alone. In sound, he identifies the potential to overcome two particular 'impasses'
that presented themselves during attempts to *visually* solve the ever-growing
narrative complexities: the first impasse concerned the subtitle and the difficulties
experienced in attempting to incorporate it into montage composition (here he cites
'breaking it up into phrases and even words, increasing and decreasing the size of
type used, employing camera movement, animation and so on' [1985: 84]); the
second concerned montage components ('pieces') that were intended to perform an
explanatory function, such as close-ups that ultimately 'burden the montage and
retard the tempo' (1985: 84). In the words of the *Statement*:

> Sound, treated as a new montage element (as a factor divorced from the
> visual image), will inevitably introduce new means of enormous power to
> the expression and solution of the most complicated tasks that now oppress
> us with the impossibility of overcoming them by means of an imperfect film
> method, working only with visual images. (1985: 84-5)

As suggested above, the asynchronous techniques of the Soviet filmmakers have
had a profound impact on film style up until the present-day, for example in the
'expository' documentary in which, as Nichols notes, 'non-synchronous sound
prevails' and the image track serves 'as illustration or counterpoint' (1991: 34). The
dominance of the narrative film, however, has meant that many of the filmic
instances in which the principles of audio-visual montage manifest themselves do
so in the context of an audio-visual style that depends to a large degree on
deploying sync-sound – particularly speech – in order to make its narrative
meanings. In other words, as with the continued presence of a topographical mode
of sound-image engagement, and likewise with that of the perceptual fidelity
model, the massive influence of narrative filmmaking has resulted in these practices

continuing only 'in the margins'. But not only the margins of the film industry – in which, incidentally, narrative still appears to rule the roost. Such practices also continue on the periphery of the narrative film style. Like the notion of documentary realism in fiction filmmaking, the apparently *contra-narrative* practice of asynchronous sound become subject to, what Nichols calls narrative's 'centripetal force'. In short, asynchronous filmmaking techniques become 'one more signifying element, more or less well motivated in relation to the plot.' (1991: 181)

But as with the treatise and theoretical writings that proliferated among early sound practitioners in Hollywood concerning the 'proper' manner in which to combine image and sound perspectives, Eisenstein's notion received mainly theoretical treatment. Indeed, not even Eisenstein himself ever fully practiced asynchronous sound. However, that it can be recognised to a degree both in the Soviets' sound films and in others' is sufficient reason to include it here as an influential *practical* tradition, not merely a theoretical one. Indeed, its presence can be identified in present-day filmmaking: significantly, asynchronous deployments of sound are today deployed in genres which aim towards facticity, such as broadcast news and documentary film (as I will demonstrate in later chapters).

4.3   Conclusion

This chapter has presented an account of the traditional resources of sound-to-picture. These will be deployed in the analysis. In the next chapter, I present those resources of technique and technology, which are fundamental to the descriptive and analytic framework that I develop in this thesis.

Chapter 5: Resources of sound-to-picture 2
Techniques and Technologies

5.0   Introduction

In this chapter I present an overview of the technical resources of sound-to-picture. It is artificial, of course, to separate techniques and technologies from their historical context. However, it is necessary to create such a division in this thesis, for the reason that much of the analytic terminology is grounded in technical practices. The following discussion should be considered as an extended technical glossary that will enable a better understanding of the subsequent analytic chapters. In what follows, I attend to the resources of sound that are deployed in the analyses, and categorise them under, what Hollywood sound designer Tomlinson Holman (1997) terms, the 'four dimensions' of the sound track: Temporal, Spatial, Frequency Range, and Dynamic Range.

Below, the immediately relevant technical resources are outlined within the context of the dimensions which they manipulate. The organisation of the resources within such categories is, of course, highly artificial. For instance, reverb is as much a matter of the temporal as the spatial dimension, in that the prolongation of sound that reverb effects involves imperceptible, multiple repetitions (cf. Lacasse, 2000). Panning a sound, also, involves time as much as it involves space. For this reason, such categorisation is performed only to enable a certain presentational clarity in the organisation of the information.

5.1   Temporal

5.1.1   Editing

'Editing' denotes the practices of sound-to-picture that involve the isolation of and rearrangement of sounds in time. The practitioner, as Moylan (2000: 312) argues, 'can physically hold time in their hands and move it around. [...] The sound can then be changed and reordered by physically altering the storage medium itself (as in cutting analog tape), or by altering the way the storage medium reproduces the sound (i.e., replaying a portion of a digital recording/sound file).'

All sounds in fiction and non-fiction films are edited in some manner and to

some degree. In fiction film of the Hollywood paradigm, often an entire sound track is created from discrete units of sound of short duration. Dialogue editing, for instance, involves the intricate assembly of a variety of performances in order to construct a coherent and effective performance that is germane to the narrative. It usually entails the critical selection and rearrangement of phrases, words, phonemes and even breaths, and concerns issues of prosody, rhythm, and so on. Additionally, each phrase or sentence created by dialogue editing must 'work' with the choices made in the picture edit (whether we are able to see an actor's lip movements, for instance) which do not conventionally incorporate concerns of sound into its practices. Sound effect editing proceeds in a very similar manner, in that the editing operations involved are designed to efface the work performed. As Doane (1985 54) puts it, 'in an industry whose major standard, in terms of production value, might be summarised as "the less perceivable a technique, the more successful it is", the invisibility of the work on sound is a measure of the strength of the sound track.' A contemporary sound track may comprise an untold number of edits, and individual units of sound, yet, to the average listener-viewer, their perception of a sound track is as of a sonic unity.

In an entirely digital environment, precision of editing takes on new meaning. The ability to 'magnify' the temporal unfolding of a sound event increases resolution enormously. The execution of timing, and the intricacy of composition, is of a different order to editing in an analogue environment, in which the physicality of the medium of magnetic tape, and the technology that enables an exacting manipulation of recorded sound, is a strong 'canvas constraint' (cf. Bateman et al, 2000) on the kind of editing that may be accomplished. But it is no doubt that this is also tied to economic constraints, as cutting and splicing tape is a destructive, costly, and time-consuming practice.

In the case of digital editing, we are able to perform very precise edits (including of course alignments of sound and image) not only because of an affordance of the resolution of computing technology, but also because of its very non-linearity and non-destructiveness. Given this technological affordance, digital audio editing allows the practitioner to attempt the execution of complex edits without fear of wasting valuable time by having to make further analogue transfers,

or by attempting a creative configuration of edited sound and image that might ultimately be considered unsuccessful. It is very common in contemporary sound-to-picture practice to provide several sound designs for filmmakers to choose from. While this is technically possible in an analogue environment, the general economic constraints of filmmaking, and the budgets, are the same, precluding the possibility for the kind of experimentation with sound that is possible with digital audio.

The basic resources of digital editing that enable the practitioner to engage and manipulate sound are essentially identical to those that exist with analogue editing. Fades (increases and decreases in sonic amplitude), crossfades (simultaneously decreasing amplitude of one sound while proportionally increasing that of another), cuts/splices (the basic operation of isolating audio of a certain duration, in order to move it in time or to repeat it elsewhere), and even replaying a sound backwards or altering its speed – all are available resources in both the analogue and digital domains. That this is the case is telling: arguably, sound-to-picture practice has remained ostensibly unchanged since the beginning of multitrack analogue sound recording (see Chapter 4) because digital media have been adopted primarily for the efficiency they afford the film and television industries. In which case, it is not surprising that digital audio has not so much revolutionised the *semiotic* practice of sound-to-picture, as it has enabled existing practices to be performed with far more efficiency. This is not only to argue that the turnaround time for a Hollywood film is much faster with the introduction of digitally-oriented practices; but also that its sound tracks are more *dense* because of the affordances of the technology and the media of digital audio and digitised image (from analogue film) or digital image (entirely recorded with digital cameras).

Of course, the above observations are of the functioning of sound in the audio-visual film industries. There, we might expect that – given the prevalent view of the role of sound-to-picture, and thus the status of its practice and practitioners (see Chapter 1) – the evolution of audio editing would follow a predictable course, constrained by the interests of dialogue-oriented narrative filmmaking. However, even within this mainstream context, we may identify instances of sound-to-picture innovation. In David Fincher's recent films – *Fight Club* (1999) and *Zodiac* (2006), for example – he and sound designer Ren Klyce go to great lengths to incorporate

experimentation with sound in a contemporary narrative film context. In *Zodiac*, the voice of the mysterious 'character' of the Zodiac – which is never fully visible – comprises the voice recordings of several actors, spliced together in an alternating fashion to form an apparently unified, though disorienting, performance. As Klyce recalls,

> [Dialogue editor] Gwen Whittle and I thought it was a pretty bad idea and that it would never work [...] Splicing from one actor to another? Forget it! We said, 'Okay, let's just do it and show David that it can't work.' So Gwen started to do this — and it worked! It was *so* weird.

(Jackson, 2007)

Yet, clearly, in *Zodiac*, a broad narrative function is fulfilled by innovative editing: that is, the apparent experimentation is in the service of the general aims of the dominant film industry. This kind of experimentation with editing is usually the province of (broadly defined) electronic music, such as electro-acoustic music, and (for want of a better word) 'experimental', often non-narrative films. Indeed, as Ren Klyce has pointed out (personal communication), the limits of his own approach to film sound – while bounded by 'narrative integration' – are expanded by his previous work outside of the film industry, such as electro-acoustic music production, TV commercials, music videos. These are areas which are arguably less restricted in terms of the possible sound and image alignments, and the edited configurations of sounds. The *sine qua non* of musique concrète, for example, is the decontextualisation and recontextualisation of so-called 'found' sounds, by means of analogue – and thus linear – sound editing. Such a practice is symptomatic of the post-war drive towards creating new possibilities for artistic expression. The manipulations in pitch, duration and amplitude of sound that defined musique concrète are fundamental to the practices of modern and contemporary sound-to-picture; as Walter Murch (Chion, 1994: xvi) reflects:

> Recording magically lifted the shadow away from the object and stood it on its own, giving it a miraculous and sometimes frightening substantiality. [...]

124

The tape recorder extended this magic by an order of magnitude, and made it supremely democratic in the bargain, such that a ten-year-old boy like myself could think of it as a wonderful toy. Furthermore, it was now not only possible but easy to change the original sequence of the recorded sounds, speed them up, slow them down, play them backward. Once the shadow of sound had learned to dance, we found ourselves able to not only listen to the sounds themselves, liberated from their original causal connection, and to layer them in new, formerly impossible recombinations [...] but also – in cinema – to reassociate those sounds with images of objects or situations that were different, sometimes astonishingly different, than the objects or situations that gave birth to those sounds in the first place.

Yet, as Murch goes on to identify, while the incorporation of the techniques of musique concrète into the context of narrative fiction filmmaking engenders the expansion of semiotic resources, it also inevitably delimits the semiotic potential of those resources, in that each choice made must fulfill and exhibit narrative motivation. Editing for sound-to-picture, therefore, cannot call attention to itself; the impact on the conventional practices of film sound editors is massive, as one of the first 'lessons in sound' a novice learns is how to edit 'cleanly', by rapidly fading up on the attack transient of each sound event, so that the 'square-curve jumps' (Kittler, 1999: 118) that occur when splicing into a sound are avoided, and thus the medium (digital or analogue) is not foregrounded (or 'deautomatised'; Thibault, 1991: 162).

But such resources did not evolve merely to fulfill an effacing function. As Kittler (1999: 118) argues (following Bischoff), many of the techniques, particularly fading sounds 'in and out', developed in the solely sonic context of the radio play, were modelled on dramatic techniques that derived from the visual practices of silent film. In his words:

[W]hen pioneers of the radio play such as Breslau's Walter Bischoff were looking for genuinely "radio-specific" (*funkisch*) means of expression, they studied the parallel medium of silent films and considered only the fade out, not the cut, as a possible model: " The man working the amplifier," as Bischoff argued in *Dramaturgy of the Radio Play*, "is in charge of a function similar to that of a camera man. He fades in and out, as we say in the absence of a radio-specific terminology. By slowly turning down the condenser at the amplifier, he lets the scene, the finished sequence of events, fade into the background, just as he can, by gradually turning the condenser

up, give increasing form and shape to the next acoustic sequence."

5.2 Spatial

Stereo production technology offers the sound designer 'space' as a practical and semiotic resource (cf. Pasquariello, 1997: 99-100). This has implications for naturalistic representations, as it contributes to the verisimilitude of a simulated third dimension. It also provides (particularly with 'surround sound' systems) for realistic off-screen sound effects, signalling a world outside of the physical limitations of the screen. It is no accident of history that stereo sound reproduction was introduced at the same time as the first widescreen formats in the 1950s (cf. Belton, 1992). One of these formats, *Cinerama*, brought with it the first 'magnetic' stereo sound. The most important benefit of this was an increased sense of realism; as the blurb read:

> [Y]ou see things the way you do in real life – not only in front of you as in conventional motion pictures, but also out of the corners of your eyes [...] you hear with the same startling realism.
> (circa 1952, quoted in Belton, 1992: 159)

5.2.1 Panning

In the first days of stereo in the cinema, all (diegetic) sounds followed spatially the action on the screen; sound effects *and* dialogue were often panned according to the actor's or sound-emitter's position in the frame. As one film reviewer (circa 1953) enthused:

> [S]ound seem[ed] to come from the exact point of origin – made it appear as if the words spoken by each actress came from her lips, giving the whole scene a life-like quality.
> (1992: 160)

This practice, however, was phased out since – rather than adding a naturalistic sense of realism – it drew attention to the technology itself. (Whether this was a factor of it being a *new* technology is still a source of debate among film-sound critics).

The production of Disney's *Fantasia* in 1939 was the site of origin of the 'pan pot': a device that encouraged ease of control over the positioning and

movement of sounds in the stereo field. The name – and function – of the pan pot, derive clearly from effects possible in the image. Motion of sound events within a stereo field, called 'panning', is the functional equivalent of lateral movement across the cinema screen. It was Disney's chief engineer, William Garity, who was to fulfill the challenge of simulating this movement. As Kay et al recount,

> [Garity] determined that fading between two speakers might be able to create this illusion. A special 3-circuit differential junction network was created to accomplish the task. Thus Disney could use it to mix down to a three track master. They called this device "The Panpot". Later a two channel Panpot was created to vary the ratio between close and reverberant sound while maintaining constant level. Mixdown required six mixers to control the various Panpot in real time. The mix was performed much like a modern film scoring session with [composer] Stokowski conducting pans and level changes. Stokowski had marked all of these level and pan changes measure by measure in his musical score.

The dynamic, kinetic function that panning serves is central to contemporary filmmaking. Yet it also serves a static function in the composition of sound events within the lateral space of the stereo image. In many cases, particularly in fiction filmmaking, certain sounds are positioned at particular points in the stereo field. Dialogue that serves an narrative expository function, for instance, is invariably positioned in the centre. This is also the case in popular music production, in which stereo centre has evolved to act as the listener's 'reference point' (Tagg, 1991: 60). In film-sound practice. This convention is termed 'monocentric panning' by the musicologist Philip Tagg (1991: 60). Lacasse (2000: 140) compares this convention with a visual concept: Marcelin Pleynet's 'monocular perspective' in painting (deriving from the notion of Renaissance perspective). In making this connection, Lacasse links the monocentrism of popular music production with the 'circumscription of the subject' of Pleynet's 'monocular perspective'. Paraphrasing Bertrand Augst (1976: 98), Lacasse argues that the stereo compositions of most sound texts 'are based on the principle of a fixed point in relation to which the objects heard are organised' (2000: 140).

5.2.2 Phase-inversion

> The ability for humans to localise sounds is one of the survival mechanisms
> we have retained from our ancient past. [...] Humans use differences in the
> same sound wave appearing at the two ears for the accurate *localisation of
> direction*. Interaural time differences [...] are the result of the sound arriving
> at each ear at a different time. A sound that is not precisely in front or in
> back of the listener will arrive at the ear closest to the source before it
> reaches the furthest ear.
> (Moylan, 1992: 24-5)

Localisation of sounds within the stereo field is a key concern of sound-to-picture.
As discussed above, sounds can be positioned in very precise points in the stereo
field, enabling spatial relationships to be established between sound events, and
between the stereo image and the listener-viewer. Moylan terms these *point
sources*: sound sources that occupy a 'specific point in space; narrow in width, and
precisely located in the sound stage' (1992: 51) However, sound events may be
rendered as indeterminately positioned: or as *spread images*, in Moylan's (1992: 50)
terms. *Spread images* can vary greatly in apparent width – from slightly wider than
a *point source* to occupying the entire stereo field. In both cases, however, the
perceived source is unclear, preventing a listener-viewer from localising a sound
with any degree of certainty. In sound-to-picture, the facilitation of such uncertainty
is achieved through the manipulation of the phase relations of a stereo sound.

Both *point sources* and *spread images* are types of stereo imaging that
Moylan refers to as *phantom imaging*:

> *Phantom images* are sound sources that are perceived to be sounding at
> locations where a physical sound source does not exist.[...] through the use
> of phantom images, sound sources may be perceived at any physical
> location within the stereo loudspeaker array, and up to 15° beyond the
> loudspeaker array. (ibid.)

As Lacasse (2000: 187) argues, this 'particular configuration is mostly used for its
euphonic quality, but can also give rise to contrasting effects'. Indeed, *spread
images* can be further effected by modifications in stereo *phase* relations, creating a
generally disorienting listening experience. The concept of phase is concerned with
the reproduction of a stereo signal, and the timing relations of the two sound signals

that constitute it (Yewdall, 2003: 501). Slightly modifying the start time of a signal by a matter of milliseconds creates a dramatic change in how those signals interact, rendering perception of localisation highly ambiguous. As I argue in Chapter 8, inverting the phase relationship of an existing stereo source – or creating a *spread image* by duplicating a mono source, panning each 'version' left and right in the stereo field, and off-setting the start point of one version by a fraction of a second – can be deployed to 'internalise' reproduced sound, by precluding the listener-viewer from localising it as a sounding 'object'. Accordingly, such effects are common in representation of hyperreal experiences or dream sequences in fiction film.

5.2.3 Spatial signature: Atmosphere and 'room tone'

Stereo also provides the foundation for creating realistic simulations of room space. Sonic atmosphere is often recorded on location, but is usually augmented, if not entirely replaced, with library sound effects at the post-production stage. These contain stereo recordings of different rooms' atmospheres (room tone), and can be edited into a sound track easily. The effect is a realistic sense of the visual space on-screen. Room tone can also be generated by a room simulator, a digital audio processor based on reverb technology (see below), and acoustic spaces can be created from scratch, so to speak.

5.2.4 Spatial signature: Reverb

Reverb denotes the practice of augmenting recorded sound with simulated spatial characteristics. In the digital domain, reverb is essentially a manipulation of time, in a very similar manner to *echo* or *delay* technologies. Often supposed to be synonyms, reverb and echo are distinct effects. Yet it is true that they have much in common, at least at the level of auditory phenomena if not always at that of sound processing. As auditory phenomena, reverb and echo are both the result of multiple acoustic *reflections* of sound. Any difference between reverb and echo is emergent, not inherent, since reverberation is the effect of a multitude of *imperceptibly* rapid acoustic reflections, while echo is that of a series of reflections that *are* perceptible.

In analogue reverb technologies, such as spring or plate reverbs, the effect is

a result of *vibration*. In spring reverb units, a sound is sounded via a coiled spring, the ensuing vibrations of which are amplified. In contrast to digital reverb, analogue reverb does not represent the process of reverberation naturalistically. Rather, spring reverbs have a distinct sonic signature which is, nowadays, construed as a specific to a particular historical period. Their use in contemporary films and other media, therefore, is largely to index or connote that period, and the values that most immediately attach to it. As I argue in Chapter 8, the specific character of spring reverb, if heard nowadays, is thrown into relief, exceeding the narrative functions is may have been deployed to fulfill in the first instance. Plate reverbs are more sophisticated (and a great deal more expensive) than spring reverbs, yet they too cannot reproduce a naturalistic sense of space if compared to digital reverb (and thus the current standard of naturalistic representation of space).

For as long as filmmakers have been concerned with constructing a verisimilitudinous diegetic reality, they have been concerned with the issue of how to represent the space of the diegetic 'world' with sound. In the physical world, every space is an acoustic space: there are no acoustically 'dead' rooms (apart from the anechoic chamber). Consequently, in the production of a film sound track – where the location atmosphere has not been recorded or will not be used – producing a sense of space is key in convincing the listener-viewer of the verisimilitude of a scene, and reverb is often added to this end. It is significant that this effect is often *synthetically* (re)created after filming has completed, at the post-production stage. Production sets will often incorporate 'sound blankets' to absorb natural reverberation, so that simulated space can be accomodated later. Hence the most effective signifiers of spatial versimilitude (stereo and reverb) are, conventionally in Hollywood filmmaking at least, *synthetic* ones. A sense of real acoustic space is *constructed* layer by layer and not captured by a microphone (cf. Williams, 1980, for discussion).

With the use of reverb, sound practitioners can construct realistic acoustic environments for artificial or synthetic sounds to inhabit. Thus 'unreal' or 'untrue' (to the physical world) sounds can be rendered and contextualised as of the physical world (e.g. digital sounds in an otherwise analogue environment). The table below (from a reverb unit instruction manual) illustrates that the technical parameters of

reverb can be – and are encouraged to be – utilised to *convince* the listener of a sense of acoustic space.

Large	*Halls and other large spaces*
Big Hall	A large hall reverb similar in size to Carnegie Hall.
Church	A large, highly reverberant space similar to a medium sized cathedral.
Long Silk	A long unnaturally silky decaying [re]verb. Also a bit spacey.
Medium	*Rooms and Chambers*
Empty Room	The room you just emptied before you paint the walls. Bright and pretty reverberant.
Garage	An empty two-car garage, bright, reverberant and a bit "ringy".
Stairwell	A dark, concrete stairwell.
Small	*Room Reverbs*
Living Room	The room you got for when you entertain your in-laws. Slightly large, medium dark from all the furniture, carpets and drapes.
Men's room	Similar to *Bathroom* but a bit bigger and a bit more reverberant.
Real Room	The name says it all. Sounds like a real, medium small room. Good on guitars and drums among other sounds.
Roomish	Sounds something like a smallish room.
Store Room	A medium, dark room, filled with lots of cardboard.

Figure 1: Selected parameters from the *Soundscape®* reverb chart. Chapter 15: 9-10)

The fact that, with reverb, acoustic space can be synthetically created suggests that it is used to mimic or to amplify physical reality, to represent it as *more real, less real* or *unreal* (cf. 'Long Silk' parameter in Figure 1; see Constantinou, 2002). Thus, reverb or room simulation can de-realise the visual representation of the world on-screen by augmenting it with an apparently incongruous sonic 'room atmosphere', as in the films of Orson Welles in which 'sound is temporally synchronised with its source, but [...] mismatched in terms of space' (Mintz, 1985:

131

289).

More pertinent to discussion of non-fiction sound-to-picture is the use of reverb to construct a *spatial signature*: sound's 'auditory fingerprint' that is contingent on the physical characteristics of a space (Altman, 1992; Truppin, 1992: 241). It is, in short, the *sound* of a space. Signifiers of spatial signature are, for Truppin, reverb level, loudness ('volume'), frequency and timbre, and these 'allow auditors to interpret the sound's identity in terms of distance or the type of space in which it had been produced and/or is being heard.' (1992: 241)

Finally, there is the issue of representing distance with reverb. As Rick Altman argues (1992: 61), and Chion (1994), sounds represented with low reverberation levels are perceived as close to the listener. Furthermore, as Altman goes on to argue,

> it is also sound spoken *toward* me rather than away from me. Sound with low reverb is sound that I am meant to hear, sound that is pronounced *for me*. (ibid.)

As I demonstrate in Chapter 6, reducing reverberation (by dubbing voice in post-production, for example) is key to producing an effect of directionality – or 'for-me-ness' (Altman, 1992: 61) – and presence. Low-reverb sounds appear as directed *toward* the listener-viewer, while also generating a sense of intimacy with speaking actors. Such an effect is achieved often in concert with the deployment of *equalisation*: that is, a degree of control over the timbre of recorded sound.

5.3   Frequency range

The human ear can detect sonic frequencies from as low as 20hz (20 cycles per second) to as high as 20khz (20, 000 cycles per second). Any sonic representation of reality which aimed at being *naturalistic* would therefore benefit from reproducing this frequency range as closely as possible. Cinema audio technology has developed to the point where frequency range is not limited. Most mainstream feature films nowadays utilise as much of this range as possible, paying particular attention to emphasising the lower range (bass) and the higher range (treble).One of

sound-to-picture's most utilised resources, the frequency range of audio can be intricately manipulated so as to construct or augment the representation of 'soundscapes' in film and TV sound tracks. It can be manipulated in practice by the use of *frequency equalisers, filters* and audio processing tools such as *exciters*. This technology often allows control over the entire range of reproduceable frequencies.

5.3.1 Equalisation

Equalisation is a key resource in sound-to-picture, serving a primarily *sonic shaping* role. Filtering serves a very similar role, but with greater potential for isolating aspects of the sound spectrum and thus modifying a sound in terms of its salient frequency and resonance characteristics. Somewhere in between equalisation and filtering is parametric equalisation: again, a way of controlling the frequency character of a sound, but this time also in terms of the *breadth* of a frequency band. This enables the practitioner to focus on a very specific part of the harmonic structure, so that she can make salient or diminish very precise aspects of the original sound.

Crucially, the manipulation of audio frequencies is conventionally deployed to fulfil one function in particular: augmenting and diminishing the 'presence' of sound. This is accomplished by means of manipulating the perceived timbre of sound, such that certain effects are achieved: sounds can be rendered as close or distant from a listener; and therefore as more or less foregrounded. As Moylan explains (1992: 28),

> Two impressions lead to the perception of the distance of a sound source from a listener: (1) the ratio of direct sound to reverberant sound; and (2) the primary determinant, the loss of low amplitude (*usually high frequency*) partials from the sound's spectrum with increasing distance [...] (my emphasis)

And, of course, this works the other way around: high frequencies *increase* the closer a listener is to a sound emitter, and the fewer physical objects there are to obstruct the sonic path. Though I do not set out to discuss them in any detail in this thesis, psychoacoustic principles are *mediating* principles between the physical character of sound production and our interpretation of its semiotic potential. But

133

while psychoacoustics is not discussed here explicitly (aside from reference to human perception of sound) the roles such principles conventionally fulfill in sound-to-picture is central to our understanding of how a sound track functions in an audio-visual context.

For example, in consonance with the psychoacoustic principle that we hear higher frequencies the closer an emitter is to us, the manipulation of high frequencies has the potential to render a sonic representation more intimate and realistic. As Chion (1994: 98) points out:

> It is particularly through gains in high frequencies that sound has progressed in definition; high frequencies reveal a new multitude of details and information, contributing to an effect of greater presence and realism.

This is the case in television commercials, as well as in films. The voice-over of a luxury ice-cream commercial is often intended to appear close and sensual, so as to connote seduction, the closeness of breath which can only be represented by high frequencies (cf. Van Leeuwen, 1999:133).

In other cases, high frequencies are *reduced* in the television sound mix, and a greater distance from the sound-emitter is signified; this renders the affected parts of the sound track muffled, like hearing voices through a separating wall. The reduction of high frequencies, then, especially on the human voice, erects a kind of wall between represented speaker and listener-viewer. Whether it is a consequence of poor recording conditions or not, in news broadcasts it is significant that the recorded voices of war correspondents are often muffled and more distant (perhaps, socially and geographically). Disregarding discussion of technological constraint at this point, it suffices to state that the reduction of high frequencies can be deployed to signify social separation and distance.

5.4   Dynamic range

5.4.1   Loudness level

Perhaps the most basic of the resources presented here, loudness refers to the amplitude level of a sound signal. It is important here because, as with equalisation,

the process of increasing and decreasing loudness can be used to represent a sound's motion towards and away from the listener-viewer (or rather the POA coded in the relationship of sound and image scale). I have already attended to the use of audio fades in the discussion on *editing* above. There it was suggested that increases and decreases in loudness evolved in sound-specific media (i.e. the radio play) to function in narrative terms that had previously developed in the context of the silent film.

Loudness is also a crucial notion in sound mixing. The relative amplitude levels of sounds in a multi-channel mix comprise often very abstract relations between sound events – particularly in electro-acoustic music production. In sound-to-picture, however, the relations are most often motivated by the logic of the narrative. As Gary Rydstrom has described it (quoted in Chapter 4), the process of mixing sound for film (Sergi, 2004: 173),

> is about [...] focusing the audience attention on what's important at any given time. [...] It's a dance really between the different elements that are available on a sound track and how you play them [...] How you play music and dialogue and sound effects moment to moment really affects the effectiveness of a scene, the drama of a scene, so I think it's probably misunderstood by a lot of people as being a fairly technical exercise of just getting everything at the right level and you're done but it comes down to a constant choice being made of what elements to hear, how much of them to hear.

In mixing popular music, the relations between sounds, while largely considered non-representational, are also highly conventionalised. Lead vocals, for instance, usually hold a privileged place in the mix, and are often far louder than other sounds that emit similar frequency characteristics. This is challenged by many recording artists yet it is a model that is exhibited in the most commercially successful pop songs (cf. Middleton, 1990). Other regularities of configurations of loudness levels include: the subjugation of rhythm guitars to lead guitars (in guitar-oriented rock and pop); the relative low level of 'backing vocals' to lead vocals; the foregrounding of drums/percussion and bass instruments in dance-oriented productions. Being deployed with regularity reveal much about the culture (of

135

production and reception), and such standardisation is often reasoned with respect to psychoacoustics, and even oral narrative.

6.0   Conclusion

In this chapter, I have outlined the technical and technological resources available to the sound practitioner in performing sound-to-picture. The preceding can be considered itself a resource to be returned to at any point during the subsequent analytic chapters. As we will see in Chapters 6 and 7, the technical resources of sound-to-picture are central to audio-visual semiosis, particularly to fulfilling the narrative and expository aims of fiction and non-fiction film.

Chapter 6: Analysis 1 – The sound track in fiction film

6.0   Introduction

In this chapter, I engage in detailed analyses of two fiction film texts to illuminate how sound and image are aligned and integrated in technical and semiotic terms, and to identify the specific contribution that sound makes to audio-visual meaning making. The examples provided here are chosen on the basis that they illuminate the contextualising relations (Lemke, 1995; Baldry and Thibault, 2005) that make audio-visual meaning-making possible. Most importantly, they illustrate the semiotic functions that are peculiar to the sound track by demonstrating just how it functions *in the context of* the audio-visual totality. This is a crucial step in explicating the meaning-making resources of sound-to-picture, and deploying the resources of tradition and technique (presented in Chapters 4 and 5) in the analyses is a central move in this endeavour. Moreover, the examples are not only chosen to highlight what semiotic functions are peculiar to the sound track; they are also intended to foreground what is germane to *fiction film* sound-to-picture. This will prove crucial in the next chapter, where *non-fiction* film sound-tracks are investigated for their functioning.

Before moving on to the analyses, a brief note is in order on the sequencing of the extracts in this chapter. Although the first example analysed in this chapter is an instance of audio-visual metaphor, and thus gives rise to fairly explicable figurative meanings, this example is not representative of the other texts deployed. In fact, the co-dependency of sound and image involved in audio-visual metaphor is not representative even of mainstream narrative filmmaking practice. As Lastra identifies (2000: 121), the *co-articulation* (in our terms *co-contextualisation*) of sound and image, by means of the resource of sync-sound, is most evident outside the classical narrative film form, in the audio-visual isomorphisms of 'mickey moused' cartoons, the decidedly anti-Hollywood (but not anti-narrative) works of Vertov and Eisenstein, and the post-classical output of Larry Gottheim and Peter Kulbelka. In such films, he argues, a 'marginal but powerful tradition of synchronisation' is evident. Such 'truly' metaphorical instances are analysed here not only because they are important and interesting instances of sound-to-picture in

themselves, but also because they effectively demonstrate the *co-contextualising* functioning of synchronised sounds and images (Thibault, 2000: 362). In other words, metaphor requires that its constituent components (*A & B*) equally contribute to the production of an entirely separate meaning (*C*); thus, in an audio-visual realisation of a metaphor, sound and image are implicated in an interdependent relationship (cf. O'Halloran, 1999), putting rest to the assumption that film sound *necessarily* functions to buttress meanings already 'encoded' in the image. Indeed, an implication of identifying such interdependency is that, as Cook argues (1998: 115), it then makes little sense to presume any one mode in a multimodal ensemble as primary in making meaning – a tendency that has haunted many theoretical perspectives on film semiosis, and equally the domain of music criticism (see Cook, 1998 for discussion).

Finally, it should be made clear at the outset that the first example in this chapter is accorded a greater depth of discussion than the others because it not only aims to present the analysis of the extract but also seeks to demonstrate the analytic procedure – thus establishing a template for the remaining analyses of this chapter and the next. Because of its dual function, the example below describes clearly and reflexively the *stages* of the analytic procedure. Since there is no need to repeat this procedure for the other analyses, they are presented more concisely.

6.1   Constructing an agent with sound-to-picture

Extract 1: *Thank You For Smoking (Dir: Jason Reitman, 2006)*

The central objective of the analyses in this thesis is to explicate the semiotic resources of the sound track, not to present a theory of audio-visual metaphor. However, in the case of *Thank You For Smoking,* in which a metaphor is constructed equally from sound and image tracks (as in linguistic metaphor, one is *tenor*, one *vehicle*), there is a need to firstly discuss this metaphorical construction and the context it both produces and enacts, before going on to describe how the sound track has been configured to extend, enact, embody or even distort this metaphor.

TYFS is a light-hearted, contemporary Hollywood film that utilises the very real corruption evident in the tobacco industry as the source of a fictional diegesis. The endeavours of the central character – the chief lobbyist for 'Big Tobacco', Nick Naylor – are played out in a plot that places moral obstacles in his path, which he repeatedly and competently surmounts through his facility with rhetoric, and his earnest belief in the virtue of argument. The excerpt analysed in this section appears within the first minutes of the film, in a sequence in which the character and his occupation (as 'the Sultan of Spin', as his pre-teen son reverently puts it) are explained, serving to contextualise the subsequent plot and other diegetic dimensions.

6.1.1 Audio-visual description

The image track depicts the spin artist – our protagonist, Nick Naylor – stood behind a podium, giving a speech at a press conference (see Figure 1). The first 3 seconds of the clip is presented in extreme slow-motion, as Nick's head turns from the extreme left of the visual frame to the centre, at which point his gaze is toward the camera, addressing the viewer directly. The sound track, during this time, presents only Nick's narration ('I get paid to talk') and the music that is threaded throughout the larger sequence from which the present clip is extracted. Once Nick's head turn arrives at the frame's centre point, the audio-visual metaphor under discussion here is initiated. In the image track, the flash of a camera bulb initiates the radically sped-up repetition of Nick's head turn from extreme left to centre, at which point the image is slowed down to the initial playback speed, just in time for Nick's slow-motion wink towards the camera. His head then continues to turn past this centre point. In the sound track, the sound of a shotgun being cocked is synchronised with the beginning of the sped-up visual motion, followed by machine-gun fire that pans the stereo field apparently corresponding to the turning of Nick's head in the image, but exceeding that motion by continuing past the centre point of the stereo field. Nick's slow-motion wink to the camera is synchronised with the sound of a grenade explosion, the first of a sequence of three such explosions (though the last two are not synchronised with a clear gesture in the image track). These explosions rhythmically initiate the end-phase of the clip.

139

6.1.2 The metaphor

The metaphor that emerges from the integration of sound and image can be described as 'SPIN IS A DEADLY WEAPON', or 'SPIN ARTISTS ARE DEADLY'. The metaphor comprises what is, I will argue, clearly a visual tenor and an auditory vehicle: the 'practice of spin' is depicted in the image, and the concept of 'deadly weapons' or 'lethal assault' is produced by the sound track.

6.1.3 Deployment of Resources

The analysis begins with an interpretation – based on the total audio-visual composition – of the functions that certain technical resources are deployed to accomplish in the text, and then moves on to determine the roles each component plays in that semiotic totality. This enables the identification of the functional limitations of the sound track in the extract.

There are three main technical resources of sound-to-picture deployed in the construction of this text: *sync-sound*, *panning* and *equalisation/reverb* (these two are considered together since they perform the same function of augmenting audio presence). These resources are used as a means to co-contextualise the sound and image tracks, and play a key role in specifying the metaphor 'SPIN IS A DEADLY WEAPON', serving to disambiguate it from an array of other possible readings. This specification is achieved by means of three specific *points of synchronicity* (POSs); *panning* of the sound that constitutes the second POS; and increased presence of the mid-range frequencies and a high ratio of direct to reflected sound (a high level of sound 'absorption' in Van Leeuwen's [1999] terms) achieved through *equalisation* and low *reverb* levels. Editing of the sound track is discussed, but this is not a separate resource as such. Rather, in the sense it is used here, editing is a practice of sync-sound.

Sync-sound

The POSs that occur in the sequence serve to establish Nick as the *agent* of the weapon fire. These are:

- POS 1: the salient and precise synchronisation of the sound of the shotgun

being cocked with the camera flash and the sudden change from slow to
sped-up motion in the image;

- POS 2: the ensuing machine-gun fire synchronised with the sped-up head
  turn, lasting the time it takes for the head-turn to reach the centre point of
  the image;

- POS 3: the synchronisation of the first explosion sound with the slow-
  motion wink in the image track.

In order to demonstrate how these POSs are crucial to the construction of the audio-
visual metaphor, it is necessary to consider a realisation of the metaphor in which
no perceptually salient POSs are evident. If one imagines an *asynchronous*
relationship of sound and image (see Chapter 4) then at least three meanings other
than the specified metaphor appear possible. The first plausible meaning is the
metaphor 'the press conference is a war zone'. Without synchronising the sounds to
specific visual actions and a specific source, there is no single agent specified. In
other words, if they were not linked by some representational device, such as a
POS, to the actions of a speaker then the sounds of gunfire and explosions would
not directly signify *assault*.

Time	Visual track	Visual Description	Sound Track	Narration
0.0				
1.0		Head movement, left to right, in very slow-motion	Music throughout	VO: I get paid to talk
2.0				
3.0				I don't have an MD or law degree.
4.0		Bulb flash. Head begins the sped-up movement from extreme left.	[POS1] Cocking of shotgun, followed by [POS2] machinegun fire PANNING from left to right, and continues until	
5.0		then slowed down again to a slower than normal speed to catch actor's wink, facing the camera, head on. Continues to talk and move head past the centre point in very slow motion.	a sequence of 3 explosions occur, the first [POS3] closely SYNCHRONISED with the actor's wink,	I have a Bachelor's in kicking ass and taking names.
6.0			the last two with no clear synchronised gesture in the image.	
7.0				

Figure 1 Extract from *Thank You For Smoking* (Jason Reitman, 2005)

The second meaning that might be generated without sync-sound is that Nick is defending himself against an army of reporters. This meaning arises because, once the POSs are disregarded, the weapon sounds become either *off-screen* or *non-diegetic* sounds; without a specified visual agent, the sounds appear as from an unseen or unknown source. As Chion discusses (1994: 72-5) the listener-viewer must therefore attribute those sounds, and the actions they represent, to some other, implied, agent: in this case, either an off-screen battalion of reporters, or a spatio-temporally distinct, non-diegetic source. (In both the off-screen and non-diegetic readings of the sound track, the 'press conference as war zone' and 'defending against an army of reporters' meanings are produced.)

The final possible meaning resembles a *simile* more than a metaphor: 'SPIN *IS LIKE* A DEADLY WEAPON'. Without the specific POSs to establish Nick as agent of the assault, there is a larger 'interpretive space' (Cheong, 2004: 164) to accept the correspondence of sound and image as engendering the 'IS LIKE' of a simile, rather than the 'IS' of metaphor proper. The specific temporal alignment of lip movements and gestures with the punctual precision of the weapon sounds specifies that it is the *practice of spin* (public speaking) which is directly associated with the potentially *deadly weapons* of the sound track. This direct association, it can be argued, is grounded in the very recognisable convention of 'lip-sync': the temporal matching of lip movements and vocal articulation. 'Lip-sync' – as a signifier of conventionally precise synchronisation – is also crucial in disambiguating the first of the alternative meanings described above. By using the oldest, most banal (Lastra, 2000: 94) device of synchronisation, and thus the most recognisable to the average viewer, the weapon sounds and the spin-speak are *directly associated*. In doing this, the meaning of the sounds of gunfire, and that of the movements and gestures in the image, are clearly specified: the sounds are not those of the *battlefield* of the press conference, as such, but rather the sounds of a directed *assault*; and the gestures in the image are not those of a character effortlessly and justifiably defending himself in battle, but are the *sources* of violence.

In the above three possible meanings generated, the imagined lack of specific,

salient POSs results in a greater potential for other metaphorical meanings to be generated. Once the actual clip is recalled, with its precise and salient POSs, it can be seen that synchronisation can function to *directly associate* the thematic meanings of the sound and image tracks, by positing a visual source as the *agent* of the sync-sounds. In other words, the POSs of the clip function as the 'is' of literal metaphor. Moreover, the roles that both sound and image play in the metaphor are determined by the convention of lip-sync. It is precisely because of the expectations that this convention sets up – that the moving lips *produce* the voice – that directs the reading of the protagonist as agent of the weapon fire, and not vice versa. Another convention of sync-sound also serves as a point of reference here: the tradition of 'mickey-mousing'. The phrase refers to the close synchronisation between visual movement in the film's diegesis and (usually) *punctual* sounds in the sound track that developed in early animated cartoons; for instance, when Mickey takes a few steps, the accompanying music punctuates each of them (Cook, 1998: 179). Both lip-sync and 'mickey-mousing' specify a synchronised visual source as the agent of a sound, and not vice-versa. It is this apparent 'inherent property' of sync sound (Altman, 2005: 236) – though here it is argued that it is at least partially *conventionally* derived – which enables the functional designation of the image track as tenor and the sound track as vehicle.

Returning to consideration of the 'original' version of this extract – that with precise POSs between sound and image – another related function of sync-sound that arises is the editing of the sound effects. In the edited presentation of the cocking of the shotgun, the machine-gun fire, and the explosions, there are absolutely no overlaps. For the metaphor to work, the listener-viewer is required to form an association between 'speaking' and 'firing'. Through an absence of overlapping, the discrete editing of the sounds enacts *monophony*. Representing that voice as *monophonic* – that is, as human – contributes towards the greater goal of understanding Nick as the sole source – the sole 'voice' – of the weapon fire.

The POSs also establish a number of *associated* meanings in the audio-visual realisation of the metaphor. These can be considered as elaborations – in the form

of *specification* – of the metaphor (Halliday, 1985; Van Leeuwen, 2005: 222). Sync-sound simultaneously enables the production of the metaphor proper, and that of further specified meanings. For example, the sound of an explosion synced with the visual action of a wink extends the overarching audio-visual metaphor to make the meaning that spin comprises not only speech or oratory skills, but also *non-verbal* devices. POSs like this do not function as standalone metaphors – they do not make meaning that is not construed as part of the overarching meaning of metaphor of the extract. For example, 'a wink is deadly' may be metaphorical, but makes little sense on its own, and far more sense as an *elaboration* of the metaphor 'SPIN IS A DEADLY WEAPON'.

Finally, the salient POSs not only function to specify the metaphor with extreme precision, but they also serve a broader narrative role. Given that the sequence under analysis is part of a larger sequence that functions as an exposition of the central character and his profession, it is reasonable to argue that the POSs, in situating Nick as the agent of weapon fire, perform the function of *embodying* the emergent meanings of the metaphor in the central character. In this way, the character is able to be developed and the ensuing plot contextualised. This embodiment sets up what can be described as a *metonymic* relation between sound and image – this will be developed in the second part of this chapter.

Panning

Although the sync-sound is sufficient in disambiguating or specifying the metaphor, there is another technical resource deployed here that performs further specification. The technique of *panning* a sound across the stereo field (see Chapter 5) is deployed here to correspond with the turning of Nick's head. Panning redounds functionally with the deployment of sync-sound, and is not in fact *necessary* in the construction of the metaphor. Furthermore, matching the panning of a sound with the movement of a visual object is dependent on synchronisation. For these reasons, analysis of the deployment of panning in this extract is a valuable avenue to pursue. That is to say, while the *kinetic function* of panning clearly contributes to the designation of Nick as agent, in that the movement of the machine-gun fire across the stereo field is synchronised with the movement of Nick's head in the image, it

145

also evidently performs other functions that overflow those necessary to the audio-visual metaphor (as explained below).

As in the deployment of sync-sound above, the panning of the machine-gun fire to match Nick's head movement contributes to the embodiment of the metaphor in Nick. The panning achieves this by indicating that the deadly assault is emanating from him, a particular character in a specific story that is about to be told. The use of panning in sound-to-picture is often to match an on-screen motion to a sound in the the sound track. By moving a sound across the stereo field, in synchrony with a moving object in the image, a spatio-temporal association is established between the two. In this example, however, the panning is temporally associated with a turning head, but is spatially associated with a far wider distance.

Why the incongruity? It can be argued that there is not an attempted approximation of human aural-visual perception here – as in other instances of panning to match objects on-screen in sound-to-picture (e.g. the close matching of a person's footsteps as they traverse laterally the on-screen space) – but a formal parallelism that indicates an *abstract* relationship between the movements of sound and image. In other words, there is an audio-visual relation of *chiming*: the movement of the turning head on-screen aligned with the panning motion of the machine-gun fire in the sound track is related formally as a parallelism in order to exploit 'the opportunity for a juxtaposition of two ideas' (Whittock, 1990). Through the mechanism of chiming, the panning in the sound track and the head turn on-screen represent the process of 'firing in an arc' – a recognisable deployment of machine-gun fire, and one that fulfills the task of systematically killing every living being in its range. In this way, the panning *elaborates* the metaphor 'SPIN IS A DEADLY WEAPON' by recontextualising the speaker's head turn (his 'aim' of the mouth/weapon) *as* a strafing machine-gun. We are asked to interpret the firing-spin metaphor *in terms of* the easily evoked practice of firing in an arc. It does not achieve this by producing a perceptually faithful spatial effect, but by presenting the correspondence of physical motion of both mouth and gun-barrel as an abstract connection. This linking of movement in the abstract both bolsters the metaphor by maintaining a strict separation of physical realities and it *enacts* the metaphorical functioning.

146

Equalisation and Reverb

Also further specifying the metaphor, equalisation is deployed here to increase the mid-range frequencies of the weapon sounds, so that they are rendered more 'present' to the listener. In addition, *reverb* is pertinent here but only in its absence: a low reverb level, as with increased mid-range, increases the sense of acoustic presence. It is argued here that both of these technical resources contribute towards positing Nick as the agent of the weapon fire, because, in deriving a functional meaning from a perceptual truism, humans experience sounds with increased mid-range and low reverberation as being *directed* towards them (Altman, 1992: 61). When interpreted in the context of the image – that is, Nick's facing the viewer, and even directly addressing her at one point – there is a clear redundancy between the functions of sound and image that amplifies the meaning of being directly addressed.

6.1.4   Isolated functioning of sound

In the above section, the functioning of sound-to-picture was discussed in terms of the audio-visual production of a metaphor, and therefore was reasoned in consideration of the sound and image tracks in synchrony. That is, the first step of the analysis was to describe the functioning of the intersections of the sound and image tracks in the context of an explicit, technical understanding of how each track is configured to formally align with the other. The next step is to understand that interaction in terms of *co-contextualisation*: how sound and image contextualise one another to produce a composite meaning (Thibault, 2000: 236; Lim, 2004: 178). In order to do this, it is necessary to understand the above analysis in terms of the meaning-making functions: *presentational*, *orientational* and *organisational* (Lemke, 1995; see Chapter 3 for discussion) because the thesis adopts the heuristic that it is at the level of *functional meaning* that interaction may occur between the sound and image tracks to make *audio-visual* meaning. Categorising the emergent analytic themes under the functional headings *presentational*, *orientational* and *organisational* involves distinguishing what the sound and image tracks are doing *in isolation* from each other – that is, they must first be treated as autonomous semiotic entities in order to understand their

147

functional integration (what Chion [1994: 187-8] terms 'the masking method'; Cook [1998: 134-5] advocates the same approach). In practical terms, the analyst is required to 'take a step back' from her understanding of the semiotic *totality* of the audio-visual text, and to attempt to understand the limits and extents of what each mode can accomplish functionally. In short, this analytic procedure makes it possible to identify the individual contributions that the resources of the sound track identified above make to the audio-visual text.

Table 1 (below) presents the the noted events of the sound and image tracks in terms of the semiotic functions they perform.

Functions	Sound	Image
Presentational	Process: A sequence of Shotgun cocked; Machine gun fire; 3 Explosions. Technical resource: Basic **editing** Process: Firing from left to right. Technical resource: **Panning**	Process: Turning head from left to right; Speaking; Winking. Camera flash.
Orientational	Perspective: [Figure] Sounds are represented as *close* Technical resource: **Equalisation**; low **reverb** level Perspective: Machine gun fire passes by the listener-viewer's 'point of reference' (i.e. centre point of the stereo field). Technical resource: **Panning**	Gaze: Direct address: wink and 'demand' to camera. Motion: Dynamic relation between viewer and protagonist, as he turns his head towards and past the viewer Distance: Medium close shot
Organisational	Composition: [Sequencing] Sequencing of sounds, with no overlaps. (A monophonic 'sentence'.) Technical resource: Basic **editing**	Composition: Slow- and fast-motion. (The move from slow to fast motion is co-deployed with the camera flash to initiate the metaphor's realisation.) Camera flash at metaphor's initiation.

Table 1: Functional meanings in the sound and image tracks

Sync sound

The function of sync-sound is not recognised until its contribution is realised at the audio-visual level. In other words, sync-sound is a latent feature of the sound track,

and only becomes manifested as a resource when considered in terms of its contribution to the audio-visual whole. As argued in Chapter 4, 'sync-sound' is not synonymous with 'synchronisation' because it is being characterised in this thesis as a *practice*: a technical and traditional resource of sound-to-picture which sound practitioners *use* in the construction of a film sound-track (which of course they do not do without reference to the film image). Practitioners do not use the resource of *synchronisation*, as that term refers to an enabling condition of sound-to-picture; it is therefore not *directly* available as a resource in making audio-visual meanings. The term 'sync-sound' denotes, then, the practice of *exploiting* the enabling condition of sound-to-picture, in order to accord some aspects of the image with sound and others not. It is problematic, therefore, to discuss sync-sound in isolation from the synchronised image track, which is why the resource does not appear in Table 1.

Panning

In contrast to sync-sound, as can be seen in Table 1, panning clearly functions semiotically when considered in isolation from its synchronised visual correspondent. Panning functions as a resource for making *presentational* meanings in that it enacts the process of firing from left-to-right. However, panning is not *sufficient* to the presentational meaning of left-to-right machine-gun fire because it is only a way of *kineticising* sound in the stereo field, and has no sonic content in itself. The only presentational function we can attribute to the panning in the extract – of course, considered separately from the audio-visual semiotic totality of the text – is that it represents the motion of left-to-right.

Panning also performs an *orientational* function in the *TYFS* sound track: in its movement between the extremes of the stereo field, the sound being panned is situated dynamically in relation to the listener. In other words, the spatial relationship (in terms of lateral distance) between sound and listener is not a stable one, but rather unfolds over the duration of the machine-gun fire. The represented action that panning signifies, when considered in isolation from the audio-visual context, is therefore *strafing*. Moreover, the particular movement involves passing the centre-point of the stereo field – what has been described as the listener's

'reference point' in the context of popular music production (Tagg, 1991: 60). Though it is not music but film sound-track production under discussion here, the notion of the centre as a 'reference point' is evident in film sound practice. In terms of the *DME* model discussed in Chapter 4, the centre of the stereo field is the position at which dialogue is most conventionally placed. As discussed in Chapter 5, this convention is termed 'monocentric panning' by the musicologist Philip Tagg (1991: 60). Following Bertrand Augst (1976: 98), Lacasse argues that the stereo compositions of music production and film sound tracks 'are based on the principle of a fixed point in relation to which the objects heard are organised' (2000: 140). It is precisely this reference point which the panning of the machine-gun fire traverses, situating the listener as one of its many targets or victims.

Equalisation and Reverb

In this extract, the 'for-me-ness' (Altman, 1992: 61) of low reverb sound is augmented by rendering the sound with presence in the mid-range frequencies. In addition, the sounds are *recorded* using close miking techniques. Both techniques serve to establish that the sound of the machine-gun fire is represented as *directed* towards the listener. This is obviously crucial in consideration of the functions of panning because, since 'panning' denotes only motion and direction, panning must be considered in conjunction with the presentational content (weapon fire) and the acoustic character (directional perspective) of the sound.

Editing

This category is concerned with the manner in which sounds have been sequenced. It is not concerned with the *order* of sounds; that is, their sequentiality. Rather the term 'editing' is used here to discuss the character of the linear transitions between sounds that are presented in sequence. As argued above, the discrete editing of the weapon sounds in the *TYFS* extract represents the monophonic character of the human voice, but considered in terms of the isolated sound track it can only be argued that there is a basic *organisational* function being fulfilled here by the discrete editing, and that is to link the three weapon sounds into a non-verbal 'sentence'.

6.1.5 Sound-to-picture: Contextualising relations of sound and image

Once an understanding of how a technical or conventional resource functions as an autonomous semiotic entity is reached, it is then possible to discuss how it functions in the context of the image. That is, it becomes possible to talk about that resource as a *semiotic* resource – a means for making audio-visual meanings with sound-to-picture. As Cook argues (1998: 135), the analysis of individual media is no more that a starting point in analysis: a 'precondition of multimedia analysis proper'. Analysing sound and image in isolation from each other, and without consideration of prior knowledge of how they function in combination, is a reliable means for illuminating the limits of what each achieves functionally. The next step of considering those now isolated components in combination, is a useful method of comprehending the divergence 'between the [text] as a whole [...] and the sum of the effects of its constituent media' (1998: 135). In this stage of the analysis, the discrete functioning of the resources of the sound track that have been 'isolated' above are considered in terms of functional redundancies (Lemke, 1995: 167; Iedema, 2001) between the presentational and orientational meanings of the synchronised sound and image tracks.

In this analytic stage, only one resource of sound-to-picture constitutes the focus. This is because an exhaustive reading of how the functions of all the resources interact and cross-multiply is potentially massive, and would exceed the scope of this chapter. In this section, therefore, the presentational and orientational functions of *panning* are considered in terms of how the sound and image tracks co-contextualise each other.

Panning in TYFS: Presentational meaning

As mentioned above, without attending to the functional redundancies between panning and synchronised actions in the image, it can only be asserted that the panning in this extract presentationally 'means' 'firing' and 'movement'. In truth, panning cannot make the presentational meaning 'firing' since it is a technical resource for moving and positioning sounds in the stereo field, and has no *sonic* content of its own. For that reason, the term as deployed here is *inclusive* of the sound that it functions to mobilise.

151

Sound: Panning	Image	Audio-visual Functions
Presentational	*Presentational*	*Presentational*
Firing motion from left to right	Head turning left to right	The process of firing in an arc.
Orientational	*Orientational*	*Orientational*
Passing by listener's reference point	Passing by viewer's reference point (determined by the 'demand' of the gaze, and the wink)	Situates Nick as the agent of the firing; Situates the listener-viewer as a target, by passing by the centre of the stereo field (the audience's reference point).

Table 2: Functions of *panning* in an audio-visual context

When comprehended in the context of the image's functions, it can be argued that the functions of panning and those of the visual actions are redundant with each other, 'producing' an overall audio-visual meaning: 'firing in an arc'. Specifically, the direct redundancy between the presentational functions of panning and those of Nick's head turn in the image track indicates that a *co-contextualising* relationship obtains at a presentational level. This is reasoned on the basis that neither sound nor image track can make the audio-visual meaning 'firing in an arc' on its own. This audio-visual meaning is a result of the *multiplying effect* (Lemke, 1998a; Baldry and Thibault, 2005: 18) of the combination of the sound and image resources for presentational meaning.

This may be understood in terms of relations of conjunction, by which meanings are 'expanded' (Halliday, 1985; Martinec and Salway, 2005): *elaboration, extension* and *enhancement* (see Chapter 3). The presentational meaning 'firing in an arc' is constructed by means of audio-visual *extension* relations (cf. Baldry and Thibault, 2005: 235; Van Leeuwen, 2005: 230). Precisely, these are relations of *addition*. In Van Leeuwen's words, 'the second of two items of information gives a reason for, a condition of, or a comparison with the information in the first item' (2005: 223) Here, sound and image are in a complementary, paratactic relationship; that is, one of *relay* (Barthes, 1977: 41). In Barthes' words, relay can be defined as follows:

text [...] and image stand in a complementary relationship; the words, in the same way as the images, are fragments of a more general syntagm and the unity of the message is realized at a higher level' (ibid.)

Given the status of the text as audio-visual metaphor, the metaphor itself is the 'higher-level' message, and sound and image provide 'different, but semantically related information' to realise that message (Van Leeuwen, 2005: 229) . As I have argued, the metaphor is made possible by a convention of sync-sound: that is, 'lip-sync', which specifies Nick's physical act of speaking as *tenor*, and the gun-fire of the sound track as dependent on that act – as *vehicle*. While extension denotes the addition of meaning from a 'secondary' item of information to a 'primary' one, extension (and relay) also entails that both primary and secondary meaning are 'expanded' semantically.

The *extension* relations in the *TYFS* extract can be described as follows: the panning of the gunfire *extends* the meanings made in Nick's head turn, *adding* to its presentational meaning of 'turning and speaking simultaneously' (in order to address an audience comprehensively) to make the meaning 'firing about (rotating on) an axis'. In its turn, the pivoting, talking head *extends* the presentational meaning of the panned gunfire, *adding* meaning by transforming the gunfire-in-motion into gunfire emanating from an agent. Given the metaphorical basis of the text, when both sound and image tracks are interpreted as a semiotic unity, they redound completely with (co-contextualise) each other, with no 'overflow' of presentational meaning in either.

There is, put simply, a transformation of the primary meaning – here, the image of Nick's head turn and speech – by the secondary meaning – the panning gun-fire. This is sufficient to fulfill the audio-visual metaphor 'SPIN IS A DEADLY WEAPON'. Yet, a further function is performed by the fact that a concrete, visualised character (Nick) is depicted as the *agent* of spin, and therefore the agent of the gun fire. This is an *enhancement* relation operating inversely: a specification of cause and effect from image to sound that specifies the gun fire as the *effect* of Nick's speech, thereby positing Nick as the *cause* of it.

Panning in TYFS: Orientational meaning

But what of the redundancies between the *orientational* dimensions of the panning and those of the dynamic depiction of Nick's addressing of his audience? At first glance, these too appear to be in a 'congruent', 'attitudinally reinforcing' relationship (cf. Royce, 1998): both the panning of the machine-gun fire and the turning of Nick's head can be described as passing by the listener-viewer's 'reference point' of the centre of the stereo field and image, respectively. Moreover, as demonstrated above, we know that together the panning and the head-turn function to situate Nick as the agent of the firing. But, as mentioned in that discussion, the specific motion of the panning is spatially incongruent with that of the head-turn: the panning *exceeds* the lateral scope of the visual gesture. This issue was resolved by arguing that there is no need for a direct and exact spatial correspondence between the respective presentational functions of sound and image since an abstract relationship is sufficient (even *preferred* in the construction of an audio-visual metaphor) in conflating the two movement-meanings into a unified presentational meaning. However, that is not the end of the matter: in consideration of *orientational* meaning, it can be seen that the action of passing by the centre point in both sound and image is dealt with in starkly different ways by each mode, giving rise to a drastic functional differential between the two. Precisely, the panning of the gun fire traverses the *entire* stereo field by the time Nick's pivoting head reaches the centre point. Furthermore, at this point in the image track, the playback speed returns to the original slow-motion, in time for Nick's direct gaze and wink to the camera. What arises here is that while the panning machine-gun fire indicates that the listener-viewer herself has already been subjected to Nick's weapon assault, in the image track the co-ordination of the second slow-motion phase with Nick's direct gaze and wink at the camera suggests that the she is 'spared'. This is a huge divergence in meaning: the orientational meanings of the sound track are not entirely redundant with those of the image track. There is, in Lemke's (1998) terms, a *cross-modulation* of orientational (and presentational) meanings across the sound and image tracks.

What, then, are the semiotic implications of such a divergence between the orientational meanings of the sound and image tracks? It appears that while the

efficacy of the audio-visual metaphor depends on the listener-viewer construing Nick's head turn and the panning of the gun fire as a semiotic unity, the requirements of the diegesis are that the listener-viewer is able to orient to Nick simultaneously in terms of distaste and sympathy. In other words, for the narrative to work – for its readers to get drawn into the diegesis and *care* about the central character – the listener-viewer must understand Nick as dangerous, but not so dangerous that we would not want to 'get to know him' better. The functional differences between the sound and image – specifically the panning's orientational meaning of including the listener among its victims, and the image's orientational meaning of sparing the viewer, and indeed directly addressing her with a 'demand' (Kress and Van Leeuwen, 1996: 120-122) and a wink – might therefore be considered as producing a 'mixed message'. However, it is more likely that there is an overarching narrative function being fulfilled here, one that requires both that (1) the meaning of the metaphor is unambiguous (a metaphor cannot function ambiguously) in order to establish the deadly power of spin ('deadly' in terms of Big Tobacco) and thus orient the listener-viewer to the lethal potential of the particular character of Nick; and that (2) the listener-viewer is drawn into wanting to find out more about his life as a spin artist through being oriented to him as a likeable and jocular person. Not a mixed message, then, but instead a *dual orientational function* is being performed here through the integration of the functional meanings of panning and those of the image resources.

In other words, what might have been firstly understood as attitudinal dissonance (Martinec and Salway, 2005: 340; Royce, 1998), now, in the context of the contemporary Hollywood narrative film, can be understood as 'resolved' in the fulfillment of an overarching narrative function – one that is perhaps crucial to a text-type that aims to represent its central character as, in equal measures, someone to be avoided and someone to be attracted to. Indeed, other aspects of Nick Naylor's character exemplified throughout the film – his earnest conviction in the virtue of his profession, his role as unorthodox yet attendant father – lend credence to this interpretation of how the orientational functions of panning in the sound track, and direct address and slow motion in the image track, recontextualise one another to

155

create a disjunctive meaning that is resolved at a higher level by the demands of the Hollywood narrative.

Panning, then, can be deployed to simultaneously fulfill the distinct yet allied roles of the modern film narrative. The semiotic potential of panning is that – in combination with a similarly-oriented visual motion – it can be used to make overt presentational meanings (through a basic thematic redundancy) that, in their prominence, overshadow the semiotic divergence emergent at the orientational level. It is at the orientational level that the relations between listener-viewer and text producer, and the producer and its thematic content, are constructed and construed (Lemke, 1995, 1998).

6.2   Constructing a subject: Spatial configurations of sound and image

In the following analysis, configurations of represented space and perspective are investigated in order to understand the semiotic functioning of the technical resources that accomplish them. In addition, the aim is also to gain an understanding of how complex configurations of POA in the sound track and POV in the image serve a narrative role in the context of fiction filmmaking.

Extract 2: *Stalker (Andrei Tarkovsky, 1979)*

In *Stalker*, director Andrei Tarkovsky creates an extremely unsettling and disorienting diegetic atmosphere. Though an example of the science-fiction genre, *Stalker* does not present the listener-viewer with the kind of extravagant audio-visual design that has become associated with that genre since the late 1970s. Rather, the source of the phenomenal tension central to its status as sci-fi can be located in the disjunction between the space of the image and that of the sound track. As Mintz (1985: 289) has argued concerning the films of Orson Welles:

> The sound is temporally synchronised with its [apparent] source, but at the same time mismatched in terms of space (that is, apparent distance and surroundings). As a result there is a tension created between the space and the sound; between our aural and visual perceptions. *If this tension remains*

156

unresolved, a partial fragmenting of our senses takes place. Sound becomes disembodied and takes on a force and presence of its own. (my emphases)

As she continues,

[T]he mismatch makes us vaguely uncomfortable, slightly dislocated, usually without our knowing why. A sort of floating tension is created that can be used, by the filmmaker, in directing audience response. (1985: 290)

As I demonstrate below, the *disorienting* effect of Stalker is determined, in part, by constructing a listening-viewing subject position that is highly ambiguous (cf. Truppin, 1992). As in Mintz' argument concerning spatial mismatch in the films of Orson Welles, the central argument in the present analysis is that the disorientation and ambiguity emergent in the *Stalker* extract cannot be located in either sound or image, but is rather the result of the intersection of divergent meanings that are constructed in each.

This semiotic interaction is characterised here in terms of an interaction of presentational and orientational meanings across the sound and image tracks. Analysing the complex configuration of these meanings – here, discussed in terms of represented *space* – is one useful way of understanding how the effect of disorientation is constructed. Additionally, as in the previous analysis, a core objective here is to explicate how a particular technical resource is deployed in the construction of audio-visual meaning. This is an invaluable task because it can help to sensitise readers to the semiosis of sound-to-picture, both perceptually (so that they can identify the different resources deployed) and semiotically (so that they are able to make interpretations of the role that a particular resource, or configuration of resources, plays in the audio-visual semiosis of film). Below, the sound and image tracks of the extract are categorised in terms of the semiotic functions that two particular resources of sound – *reverb* and *equalisation* – are mobilised to perform. From this, we are able to see clearly the role that sound plays in the extract, and the workings of audio-visual co-contextualisation.

157

6.2.1 Audiovisual description

As the camera pans across the debris-ridden shallow waters approaching the mouth of the cavern, the sound of highly reverberant lapping water is heard, indices of the cavern's acoustic space. A change in image perspective then occurs, the camera pulling back out of the cavern's mouth, while the Stalker and Writer slowly exit, walking towards the camera. However, auditory perspective does *not* change along with the camera, and the sounds of heavily reverberant drips of water and footsteps continue into the represented exterior. However, as the Writer and Stalker speak, their voices are rendered non-reverberant and at a very close auditory perspective. Furthermore, as the Stalker sits down on a rock, the fabric of his clothes rustles and this, like the voices, are extremely close-sounding and non-reverberant. As the Writer walks further towards the camera, his footfalls on water are as reverberant as they were in the cavern. The audio-visual effect of the complex configuration of visual and auditory perspectives is that the listener-viewer is constantly unsure of how they are being positioned as subject.

6.2.2 Deployment of resources

Crucial to the conventional project of situating the image in an apparently concrete space is the use of the sound track as a site of spatial signification. As Arnheim (1933) argues, 'sound arouses an illusion of actual space, while a picture has practically no depth' (quoted in Fischer, 1985: 232).On Arnheim's view, it is not surprising then that the potential of recorded sound to simulate the acoustic characteristics of a space has been traditionally marshalled into situating the diegetic world of the image in the context of an acoustic environment. As Mintz argues,

> We are not limited, aurally, as we are visually, by the flat screen. If the sound track of a movie accurately conforms to the behaviour of natural sound in space, we receive aural cues which determine distance, direction, and, to a certain extent, the surroundings of the source. (Mintz, 1985: 290)

Whether achieved in surround sound, stereo, or monaural sound tracks, the practice of deploying the sound track to achieve spatial contextualisation – an aural 'world' –

may be accomplished through the suggestion of 'off screen' activity (not explored in this thesis) and the use of reverb and equalisation.

Sync-sound

Though much of the sync-sound in this extract occurs off-screen, in the footfalls of the Writer for instance, the POSs can be identified through perceiving his overall gestures. That is not to say that we can identify close synchronicity, as that would require seeing the contact between foot and ground. What is important here is not precision of synchronisation, but the character of the sound synchronised with the footfalls. Though the disjunction of sound and image space is recognised as continuous throughout the extract, the other POSs important here are those audio-visual moments that *reveal* that disjunction through their perceptual salience.

The three POSs relevant to the analyses are:

- POS 1: The Stalker and Writer exit the cavern, and the interior cavern sounds continue;

- POS 2: The Stalker sits down and his clothes rustle;

- POS 3: The Writer begins to walk and his footfalls are heavily reverberant.

Reverberation

As discussed in Chapter 5, manipulating reverberation is a technique so central to sound-to-picture practice, and so prevalent to cinematic experience, that its deployment is usually not obvious unless attending critically to a sound track. As it is deployed in this extract, reverb accords a *spatial signature* to the footfalls and sounds of water that function in terms of what is usually called 'ambient sound' (here, 'spatial sound', following Truppin [1992]). As that term suggests, such sounds are intended to serve as background, most commonly to a foreground that serves an overt narrative function – such as dialogue. At the first POS, the sound of water dripping that has served a spatial signification role *inside* the cavern, continues *outside* the cavern. Whereas inside the cavern the reverberation level of the sounds seem entirely congruent with the image space – its presentational

Time	Visual Track	Visual description	Sound track	Dialogue
0		Camera pans over debris under lapping, shallow water	The sound waterfall recedes, and reverberating footfalls in shallow water are mixed to the fore.	
8				
16		The Writer and Stalker exit cavern.	[POS 1] Drips of water are prominent and reverberating as in a cave. Footsteps reverberate also. Voices are extremely close, and non-reverberant.	Writer: There he is!
24		They meet the Professor.		
32		The Stalker sits down.	[POS 2] The Stalker's clothes' rustle as he sits. The sounds are rendered 'dry' and close through EQ, and treated with no reverb.	Professor: I'm certainly grateful to you that you, but…
40		The Writer walks down a level.	[POS 3] The Writer's footfalls are heavily reverberant, indicative of the cavern from which they emerge.	Stalker: How did you get here? Professor: Mostly I had to crawl up here on my fours.

Figure 2. Extract from *Stalker* (Andrei Tarkovsky, 1979)

meanings (as indices of a particular space) redounding with the presentational meanings of the image – *outside* the cavern the reverberation jars noticeably with the represented exterior space. Moreover, the absence of reverberation on the overdubbed voices and clothes' rustles (POS 2) forms a definite contrast to the continuing reverberant water sounds and footfalls (POS 3).

Equalisation

In *Stalker*, the acoustic character of voices, footsteps and clothes' rustles – signifiers of human activity – is pronounced in the mid- to upper-mid range. This contributes a great deal of presence, and serves to foreground those sounds. Yet, sounds that are conventionally backgrounded, such as water drips and footsteps, are also emphasised in the same way. Putting aside, for a moment, the *reverberation* of the footsteps and drips, and the differentiation that reverb establishes between reverberant and non-reverberant sounds, the fact that human voices and movements sound with very similar frequency characteristics to sounds of water and footsteps (which are also signifiers of the ground's surface) means that the conventional foreground/background division between dialogue, sound effects and spatial sound is abandoned. Sounds that are conventionally rendered as 'support' for the central roles of dialogue (at least in the modern and contemporary narrative film) are rendered through equalisation as foreground sounds – and thus the attention that is conventionally given over to what occupies the foreground (whether in visual or auditory semiosis) is necessarily shared by those usually discreet sounds. Once we consider, again, the reverberation that differentiates the footfalls and water drips from other sounds in the *Stalker* extract, it can be noticed that the shared increase in frequential 'presence' of the otherwise differentiated sounds renders their spatial differences extremely prominent. In other words, the differences in spatial signature between the voices and footsteps are thrown into relief.

6.2.3   Isolated functioning of sound

In this section, the sound and image tracks are analysed in terms of presentational and orientational functions, and the technical resources that are deployed in

161

achieving them. In order to do this, it is necessary to 'prise apart' the semiotic functioning of the sound and image tracks, so that the meanings made in each mode can be explicated, and thus the configuration of sound and image to accomplish the highly ambiguous 'audiovisuality' of the extract can be understood more comprehensively.

Reverberation and Equalisation

As is clear from Table 3, reverb is used to contextualise the dripping water and footsteps in terms of a specific physical space. But it is *not* deployed to situate the voices and other 'Foley' sound effects (such as clothes' rustles) in the same acoustic space. This negation creates a sense of 'for-me-ness', which Altman argues results from representing sounds with little to no reverb (1992: 61). The rendering of the voices as very close and present through equalisation (as Figure) places the listener in close proximity to the speaking characters. As argued in the *TYFS* analysis, this works in conjunction with low reverb level to suggest sound that is directed toward the listener (Altman, 1992: 61).

Functions	Sound	Image
Presentational	Process: Footsteps through shallow water, located inside a cavernous space. Process: Dripping water, cavernous spatial signature. Process: Voices/Dialogue, situated in a non-reverberant space. Technical resource: **Reverb; Equalisation.**	Process: Transition from interior (cavern) to exterior. Circumstance: Interior cavern then exterior forest.
Orientational	Perspective: [Figure *and* Ground simultaneously] Reverberating footsteps and dripping water heard as *close* but as in an interior, cavernous space. Technical resource: **Reverb; Equalisation** Perspective: [Figure] Voices and clothes' rustles are represented as 'dry' and very close, without reverberation. Technical resource: **Equalisation; Reverb**	Motion: Camera trucks backward out of the cavern, at the characters' pace. As it moves outside, the Professor is revealed at the left of the frame. Distance: Medium to Medium Long Shot.

Table 3: Presentational and Orientational functions in the sound and image tracks

As well as performing an orientational function, the sounds of space and distance signified through reverb and equalisation also perform a clear presentational function. Spatial signature is deployed as an indices of a particular diegetic space. That is, it constructs a presentation of 'states of affairs' (Lemke, 1995: 41) that serve an important narrative role.

Within the sound track, therefore, it might appear that there is a very pronounced sense of acoustic ambiguity: the listener-viewer is not placed in a definite position in relation to the characters and environment; they sound as if in two distinct spaces (Chion, 1994: 189). However, the fact that the sound track can

be construed as positioning the listener in close proximity to the speakers, so that nearby reverberating sounds could still be heard from the cavern, offers some resolution to this tension. It is only when considered in synchrony with the image track that a drastic tension might be thought to arise, given the fact that the the image does *not* position the viewer in a close relationship with the characters. However, as will be argued below, the close sounds of the voices, and the images they are synchronised with, serve as a reference point – a *conventionally* determined position of 'normality' from which this particular use of spatial sound deviates.

6.2.4 Sound-to-picture: Contextualising relations of sound and image

The disorienting effect of *Stalker* can be understood in semiotic terms of the *co-contextualisation* of image-meanings and sound-meanings. In *Stalker*, the characteristics of the sound track do not align with those of the image in an expected manner. Below, I describe the audio-visual functioning of spatial sounds and voices (in terms of reverb and equalisation); that is, how the semiotic functions of sound and image are integrated in the audio-visual text.

Spatial sounds: Presentational and Orientational functions

As they are presented in Table 3, the presentational functions of the spatial sounds of reverberating drips and footsteps and those of the represented image space are in a non-redundant relationship. As a matter of convention, sounds such as these have evolved to fulfill an ancillary role in narrative: they are deployed to denote basic information about the spatial setting of a narrative (Altman, 1992; Van Leeuwen, 1999). For Chion (1994: 75), spatial sound (in his terminology 'territory sound') is

> sound which envelopes a scene and inhabits its space, *without raising the question of the identification or visual embodiment to its source*: birds singing, churchbells ringing. We might also call them territory sounds, because they serve to identify a particular locale through their pervasiveness and continuous presence. (my emphasis)

164

Furthermore, their presentational function is usually conventionally redundant with that of the other ancillary sounds, such as clothes movements; that is, they are indices of the *same* diegetic space. In this extract, though, that is not the case, and the emergent contrast between the spatial signatures of the various sounds foregrounds the fact that these are not to be construed as occurring in what the listener-viewer has come to consider the same diegetic space. But as much as the spatial sounds are made prominent, it is also claimed here that the manipulations of spatial signification, that are fundamental to the unsettling atmosphere, are very subtle and function 'covertly' as background. In part, this is undoubtedly the result of a lack of perceptual facility with sound among lay listener-viewers (cf. Holman (1997: 41-42). But the subtlety is also due to the low-level of attention conventionally given over to spatial sounds in the construal of a film narrative. In Iedema's (2003) and Halliday's (1982) terms, their presence and significance is 'automatised', because they have evolved to perform ancillary semiotic duties. As Iedema argues (2003: 40), the foregrounding of one mode (in narrative film, dialogue) is often accompanied by or perhaps even *achieved* through the backgrounding of another (here, spatial sound). Consequently, spatial sound's conventionally assigned role as ancillary support (at least in narrative filmmaking practice) is such that they are construed by the listener-viewer as bearing little significance to the unfolding narrative. As Mintz (1985: 292) argues:

> Most of the time the kinds of sounds that define space are sound effects and background noise. Straight dialogue usually drawn and holds our attention away from the spatial dimension: we are more concerned with what is being said than the relationship between the source and space. Dialogue has the effect of taking us out of space and placing us in the realm of ideas.

For listener-viewers, then, there is a clear tension between what they have *learned* to 'do' with the presentational meanings that spatial sounds make, according to the conventions of the live-action narrative genre, and how spatial sound is *actually* treated in the *Stalker* extract (and the film in general). This tension remains unresolved, perhaps, because the presentational meanings of dialogue divert our attention from those of the indexical spatial sounds (Mintz, 1985: 292) and

therefore distract the listener-viewer from emergent visual-aural inconsistencies at a presentational level. In sum, the apparently disjunctive relationship between the presentational meaning of the spatial sounds and those of the image space is one cause of the audio-visual disorientation in the *Stalker* extract.

While attention is certainly drawn to the apparently incongruent spatial sounds, as to the presentational meanings generated through dialogue, it is also diverted by the conventional *status* of spatial sounds as ancillary sound effects; as sound that does not 'raise the question' of its source (Chion, 1994: 94). The listener-viewer is perpetually undecided whether to categorise spatial sounds as having narrative significance (foreground) or not (background); or as serving a *primarily* presentational or orientational function. The contextual determination of the relative status of presentational and orientational meaning in semiosis is at the heart of the process of distinguishing what is to count as narrative foreground, and what as background. In Lemke's words (2002: 304-5):

> I believe that it is customary in our culture to pay conscious attention primarily to presentational meanings, to orientational ones only in special circumstances, and to organisational ones only if you are a professional user of the medium. We rely on familiarity with genre conventions to automate our use of organisational and orientational cues and to allow us to proceed directly to presentational information.

The narrative fiction film context constrains the listener-viewer's construal of the hierarchical structuration of the sound track. The *orientational* meanings produced through the complex configuration of aural and visual perspectives are backgrounded in our interpretation of the film, subordinated to the presentational meanings of dialogue and action in both sound and image tracks. In the *Stalker* extract, the reverberant sounds of water and footsteps do not redound orientationally with the perspective constructed in the image, nor with the varying distances between image frame and characters. The consequence of this is that the listener-viewer's subject position in relation to the characters and environment is split. In other words, while there is a clear and singular *visual* perspective (POV) at

166

all times, the sound track makes it difficult to identify a single, precise *point of audition* (POA) constructed in the audio-visual design. For instance, even though the reverberating water drops and footsteps were heard previously in synchrony with the image from within the cavern, it cannot be claimed that the listener-viewer 'remains' in the cavern as the characters exit it. But that is exactly the point. This indefinite position of the imagined 'listener', at once both inside and outside the cavern, is key to the construction of a diegetic world in which beliefs are continually tested. In Truppin's words (1992: 235), 'the use of ambiguous sound plunges the audience into a never fully resolved struggle to believe in the diegesis, much as the characters struggle with their own ability to have faith'.

The fact that there is no visual source for the drips renders those sounds as *off screen*, and so diegetic (Chion, 1994: 73). However, because their acoustic character is not of the environment imagined to constitute the off-screen space, the sounds is rendered *non-diegetic*: in other words, the sound of dripping water occupies the same relationship to the diegesis as voice-over and music. Understood in this way, the reverberant drips occupy an indeterminate position as simultaneously part of the narrative world represented in the image, and spatio-temporally detached from that very world. The visualised source of the similarly reverberant footsteps, however, partially resolves the ambiguity, in the sense that the water drips can be construed as in the same diegetic space as the footsteps. Also contributing to disambiguating the diegetic status of the spatial sound is the fact that the same reverberant acoustic character is experienced just prior to the exit from the cavern.

In terms of the spatial sounds' effect on the visual diegesis, we are led to understand that if the sounds *are* to be considered off-screen and diegetic, then the represented image space is itself indeterminate: we cannot be sure that what we are seeing (a naturalistic visual setting) is as it appears. The reverberant footsteps, occurring approximately 40 seconds after the exit from the cavern, re-establish this point, recontextualising the image in terms of a cavernous space. The water drips, and the reverberation as an 'independent' acoustic character, function as metonyms

167

for the image's cavern, and the Zone's many other catacombs and tunnels.

In terms of the contextualising relations specified in Chapter 3, we may understand the interaction of presentational meanings of the sound and image track in terms of elaboration and extension. While inside the cavern, the relation between the spatial sounds and the diegetic visual space is one of elaboration. At no point is the camera positioned inside the cavern; the listener-viewer actually only sees the characters enter and emerge from a dark, shadowed space. The presence of the Stalker and the Writer inside it is merely suggested by the image editing, but it is further *specified* by the spatial sounds. The sound elaborates on the meaning of the image 'by a more detailed description of it' (Martinec and Salway, 2005: 342). We may also understand this relation is terms of metonymy. There is a part-for-whole relation constructed in the visual editing: a shot of the characters entering the cavern, and a shot of them exiting it. But the sound track also functions metonymically in relation to the image, by providing dripping water (part) as a key spatial signifier of cavernous space (whole).

Once the characters exit the cavern, however, the relations between sound and image are disrupted. The impression of an incongruous sound-image relationship can be understood in terms of *extension*: once outside the cavern, the presentational meanings of the spatial sounds add further *adversative* meaning to those of the exterior image. Linguistically, this may be expressed as 'The image is of the exterior, *but* the sound is of the interior'.

The Zone, then, through its sonic signature, is all-pervasive. When the two characters emerge from the cavern to meet with their companion, there is a residue of reverberant, *interior* sound that 'bleeds' into the *exterior* environment. The convention of *continuity sound* is crucial here, especially because it is deployed to unconventional ends. Deployed conventionally, continuity sound establishes a consistent auditory space that extend across visual edits, in order to situate the discontinuities of space and time engendered by film editing (and sound editing if it follows the visual cutting loyally, as in the *direct sound* tradition) within a comparatively continuous sonic context that serves to efface the diegetic

discontinuities. As Altman (1992: 62) describes it:

> The image displaces us incessantly, offering us diverse angles on objects located at radically different distances. Our voyeurism consists precisely in this mobility. Yet we flit about at our own peril, constantly risking dizziness. Just as we are about to lose our balance, however, the sound track holds out its hand, offering continuity of scale as an effective stabiliser.

However, its deployment in the *Stalker* extract is such that the spatial sounds of the cavern, which are diegetic sounds when inside it, very subtly continue upon exit, rendering those same sounds as straddling the distinction between non-diegetic and off-screen sounds.

Voices: Presentational and Orientational functions

In Table 3, we can identify an apparent non-redundancy of orientational meaning between the close perspective of voices and the varying perspective of the image. That is, the entirely non-reverberant voices are not considered redundant with the space of the forest represented visually, and importantly with the varying distance between the image frame and speakers. What this reveals, however, is not that voices and visual space in the *Stalker* extract do not congrue at all on an orientational level. Rather, it reveals that there is something amiss in how the relationship between sound and image is being 'reconstructed' in analysis. This can be compared to the issues that arise in reconstructing the modality parameters of sound and images once comprehended in isolation from one another. As with audio-visual modality (Constantinou, 2002), at the heart of this issue is the functioning of a higher-order context, governing how apparent spatial 'mismatches' are to be interpreted (resolved by implementing Bernstein's notion of 'coding orientation' in audio-visual modality analysis).

On a criterion of scale-matching – that is, matching the auditory and visual perspectives in terms of the image frame – the protagonists' voices and the spatial

169

sounds appear, independently, incongruous with the image space. Throughout *Stalker*, the sounds of voices are rendered close and are distinct from the space of the image – their spatial signature is noticeably incongruent with the image space and represented distance of the speakers. But perceptual incongruities abound in film sound, and have done since the beginnings of the sound-film. Indeed, the conflict that arose between sound technicians coming from the phonographic and telephonic industries and film producers and cinematographers precisely turned on a disagreement over whether film sound should aim towards *perceptual fidelity* or a *narrative intelligibility* (see Chapter 4 for discussion). What then does it mean to talk about perceptual incongruities between sound and image in film when narrative intelligibility *necessitated* perceptual oddities in the representation of an aural-visual world? (The reader might recall here early film sound-engineer, John Cass's description of the experience of hearing as if one had 'five or six very long ears, said ears extending in various directions.' [Altman, 1992: 49]) As argued in Chapter 4, the practice of matching sound- and image-scale is unconventional (in early, modern and contemporary filmmaking), so cannot serve as a useful measure for comprehending sound-image relations.

In narrative filmmaking, the long-established *DME* structuration (see Chapter 4) of the sound track has 'trained' audiences to understand that voices heard as in close proximity (and centred in the stereo field) can exist independently of the represented distance (and lateral position) of speakers represented visually. As Lastra (1992: 76) argues,

> Schematically, dialogue recording tends almost uniformly, from the early thirties on, to minimise the amount of reverberation, background noise and speech idiosyncrasies, as it simultaneously maximises the 'directness' or 'frontality' of recording, and the intelligibility of the dialogue. Even when a speaker appears to turn away, a high level of direct sound often implies that he or she is still speaking 'to' the auditor, because speech is understood not simply as an abstract sound, but as a sound with a specific function – a narrative function.

At the same time, audiences have learned to interpret (and thus necessarily *perceive*) all diegetic sounds *other* than dialogue as bearing some spatial relation with their visualised sources. It is this convention which determines that we should treat the apparent non-redundancy between the spatial signature of voices and image space with extreme analytic caution.

So while a perceptual tension can be identified between the spatial signatures of voices and *spatial sounds* no *semiotic* tension emerges between them because, conventionally, the aim towards intelligibility in narrative filmmaking established that very perceptual tension as a valid trade-off in the endeavour to create dialogue-oriented narrative films. To be clear, what I am arguing here is that there *is*, in fact, a redundant relationship between the orientational functions of voices and represented visual space. This is because the representation of voices as proximous to the listener-viewer has co-evolved with the varying perspective of visual representation in narrative filmmaking. In other words, the relationship of close voices with a range of represented perspectives in the image is a relationship that is itself in a redundant (meta-redundant) relationship with the context of narrative filmmaking. In short, the combination 'close voices and varying visual perspective' *produces* and *enacts* the context of 'narrative film'. This is a co-contextualising relationship of the sound and image functions.

In sum, the tension between sound and image arises between the spatial sounds and the image space. This incongruity is to some extent 'disguised' by the *conventionally* congruent (functionally redundant) relationship of a fixed aural position and a varying image perspective, which serves to orient the listener-viewer to the film as a member of the narrative film genre. The overarching narrative film context cannot fully resolve the spatial incongruity, but it does provide a frame with which to understand the status of the reverberant drips as in an indeterminate position as simultaneously a part of the narrative world represented in the image, and spatio-temporally detached from that very world.

But the above analysis suggests a sense of semiotic 'closure' to the audio-visual

experience of *Stalker* that is far from hitting the mark. Rather, it can be argued that while there are co-contextualising relations between voices and image space that function to resolve some of the audio-visual ambiguity of the combination of spatial sounds and image space (they at least limit the number of diegetic functions that the spatial sounds occupy; i.e. either off-screen or non-diegetic), they are made to uncomfortably co-exist (in audio-visual synchrony) with *disjunctive* relations between the spatial sounds an image space. To repeat Truppin's words, this unresolved ambiguity is perfectly appropriate to *Stalker*'s diegesis, because it 'plunges the audience into a never fully resolved struggle to believe in the diegesis, much as the characters struggle with their own ability to have faith'. Furthermore, it is precisely because *Stalker*'s audio-visual relations do not fully meta-redound with a recognisable context that its diegesis and audio-visual representation is construed as disorienting. In Lemke's (1995: 128) terms, the 'meaning non-relations' between the combination 'spatial sounds/image space' and a possible, definite context that might *resolve* the audio-visual disjunction indicate meanings that are yet to be made; in his words, they are 'gaps that are not even seen as gaps'. Or, in Bernstein's (1982) terms, they are meanings that are 'yet-to-be-voiced'.

6.3   Russian Version 5.1: the recontextualisation of Stalker

Conceived as an opportunity to rectify certain perceived 'errors' with Stalker's original monaural sound track, what is called 'Russian Version 5.1' on the 2000 Tartan DVD release is nothing less than a total recontextualisation of *Stalker*'s audio-visual design. Whether heard in 5.1 surround sound or two-channel stereo, the film is transformed by the deployment of an arsenal of contemporary (Hollywood-derived) conventions for constructing film sound tracks. Technically speaking, the 'rectification' involves treating the original monaural sound track as the central element (for that is where the dialogue is located) while adding entirely new sounds, utilising the stereo field, and mixing the sound track in such a way that many of the original track's sounds are apparently reconfigured in terms of the role they play in synchrony with the image. Semiotically speaking, the tension that arose in the original version between the (presentational and orientational) meanings of the spatial sounds and represented image space is, in the new version, partially

resolved.

In this section I want to focus on the explicit ways in which the 'Russian Version' (which sounds suspiciously American) recontextualises the film's diegesis by reconfiguring how spatial sounds and image space are integrated. The transformations that occur are discussed in terms of the configuration of semiotic functions, as in the previous examples, but also in terms of two types of transformations that routinely occur in, what Van Leeuwen and Wodak (1999), and Van Leeuwen (2005), discuss as 'the recontextualisation of social practices': that is, addition and deletion. Crucially, these terms are congruent with those associated with metaphorical relations in cinema, such as metonomy and synecdoche (in which a 'deletion' is always necessary) (Whittock, 1990).

As I demonstrate below, the transformations of addition and deletion, and so on, are invaluable concepts with which to analyse audio-visual recontextualisation. This, as will be shown, is mainly due to the fact that, given the opportunity to compare two radically different sound tracks for the same film, the semiotic processes that have occurred in what is obviously a staged process (by some 30 years) are immediately apparent. Moreover, their effect on the audio-visual meaning making of the film is drastic, and so usefully demonstrates the significance of the choices made in the original sound track, and the degree to which the sound track plays a role in constructing perspective.

Presentational and Orientational functions

The procedure followed in the above analyses is not adopted here in full, for the reason that it is only necessary to describe and analyse the differences between the original and new versions of the sound tracks, and to consider the new sound track's functioning in the context of the image. Here, I describe the major differences between the versions.

Functions	Sound	Sound (Russian)	Image
Presentational	Process: Footsteps through shallow water, located inside a cavernous space. Process: Dripping water, cavernous spatial signature. Process: Voices/Dialogue, situated in a non-reverberant space. Technical resource: **Reverb; Equalisation.**	Process: Footsteps through shallow water, located inside a cavernous space. **Diminished in presence.** Process: Dripping water, cavernous spatial signature, **but diminished in presence.** **Process: Lapping water from brook.** Process: Voices/Dialogue, situated in a non-reverberant space. Technical resource: **Reverb; Equalisation.**	Process: Transition from interior (cavern) to exterior. Circumstance: Interior cavern then exterior forest.
Orientational	Perspective: Reverberating footsteps and dripping water heard as *close* but as in an interior, cavernous space. Technical resource: **Reverb; Equalisation** Perspective: Voices and clothes' rustles are represented as 'dry' and very close, without reverberation. Technical resource: **Equalisation; Reverb**	Perspective: Reverberating footsteps and dripping water **heard as *close* but lower in the mix. *Panned* left of the stereo field.** **Lapping brook presented as *close*. Foregrounded in loudness, and emphasised presence through equalisation. Also a stereo recording that occupies the breadth of the stereo field.** Perspective: Voices and clothes' rustles are represented as 'dry' and very close, without reverberation. Technical resource: **Reverb; Equalisation**	Motion: Camera trucks backward out of the cavern, at the characters' pace. As it moves outside, the Professor is revealed at the left of the frame. Distance: Medium to Medium Long Shot.

Table 4: Comparison of original and Russian Version (modifications in bold type)

As the characters exit the cavern, the spatial signature of the reverberant dripping water of the original sound track is diminished by the *addition* of much less-reverberant lapping water. Given the affordance of stereo space, this sound is of a stereo recording that is presented as spanning the full breadth of the stereo field. In synchrony with the image, the perceived distinction between interior and exterior space is increased by this addition of sounds and resources of sound-to-picture, and they function to offset the original spatial divergences of sound and image. Specifically, the non-reverberant, close-sounding lapping water constructs a POA that is unambiguously *outside* the cavern; because this is so definite, the listener-viewer is 'with' the Professor far earlier than in the original. The lapping water, having no apparent reverb added, and being composed of higher frequencies (Chion, 1994: 98), are rendered closer than the reverberating drips to the listener-viewer. As Chion (1994: 92) explains, high frequencies are central in denoting the close proximity of a sound emitter because they 'travel in a more directional manner than the low [frequencies]'. In Van Leeuwen's terms (1999: 22-3), the directness of the sounds mean that they assume the foreground, further entailing that the reverberating water is itself *backgrounded*.

Furthermore, the stereo potential of the new version brings with it the affordance of lateral space. The monaural signal of the original version is itself recontextualised in the stereo space of the 'Russian version'. In particular, the mono signal is able to be placed at any position in the stereo field, and to *move* within that field. The reverberating water, is not only backgrounded by the addition of the close-perspective lapping water, but also by its position at the left-hand side of the stereo field. Whereas in the original the incongruence of spatial sounds and image space was emphasised by the (necessary) occupation of the centre of the stereo field – the same location as the dialogue – in the new version, the status of the reverberating drips as 'territory sounds' (Chion, 1994) is far more definite. Footsteps, too, are transformed by their new-found stereophonic mobility, and are often matched (as they are at approximately one minute after this extract ends)with the lateral position of the on-screen actors. They follow the actors' movements, serving to 'bind' the sound to the diegesis. In the original, sounds of footsteps are bound to the diegesis

only by synchronisation, and the reverberation accorded to them there that is so presentationally and orientationally incongruent emphasises the basic disjunction: both of the space, and not.

The mono soundtrack that is now only one element of the new sound track can then be kineticised and spatialised in new ways. Footsteps and voices, for instance, follow the actor's lateral movements. However, in the main, the dialogue remains centred as it necessarily was in the original sound track. Against the background of a choice of positions (nonetheless constrained by the dialogue being bound to the other sounds in the original) for the dialogue in the stereo field, the decision to centralise it coheres with the original decision to render the voices as in conventional narrative film. This is because monocentric panning (Tagg, 1991) functions as a reference point for the spatialisation of the other sounds in a sound mix. As such, this too is a convention of narrative filmmaking, and one that in the new Stalker sound track serves to anchor the listener-viewer in the recognisable and comfortable context of the contemporary narrative film.

Through the deployment of equalisation, much of the presence of the sound track – in particular the spatial sounds – is diminished. Whereas in the original they were pronounced in the mid- to upper-mid range (the 'vocal register'), in the new version they are rendered as spatial sounds conventionally are – as part of the territory (Chion, 1994).

Through the addition of certain acoustic features, and thus the meanings that they were deployed to make in the original version, the new version attempts to resolve the original audio-visual ambiguities. Given the more unified audio-visual space of the new version, and the more stable auditory perspective represented, it can be argued that the constructed listener-viewer is positioned in a more definite way. In other words, there is less of a difficultly in believing in the diegesis (cf. Truppin, 1992), because much of the fragmentation of perspective that arises through presentational and orientational disjunctions across sound and image in the original version is resolved to a unified POA: an imaginary listener-viewer within the

diegesis of the forest.

The redundancies between the functional meanings of the sound and image tracks enact a context with which to interpret those meanings that are *not* redundant. The varying image perspective that co-contextualises with the close and fixed auditory perspective of the voices (as *Figure*) enacts and produces a narrative film context that forms a primary frame through which to interpret the emergent semiotic disjunctions of the audio-visual spatial configurations.

But such a reading is perhaps only possible upon technical semiotic analysis of how sound and image are configured in *Stalker*. For the 'lay' viewer, the enactment of the narrative film context renders the identification of disjunctive audio-visual space unlikely. It is this subtlety of auditory perspective, particularly in the context of narrative filmmaking, that defines its power as a resource of sound-to-picture.

The difficulties that arise in engaging with the use of sound-to-picture in *Stalker* should go some way towards indicating just how unconventional a sound track it has, and thus illuminates the conventional deployment of audio-visual relations in terms of spatial sounds and represented image space. That is, through engaging with *Stalker*, it can be seen that constructing an ambiguous audio-visual space does not serve the aims of conventional narrative filmmaking – i.e., to ensure that the listener-viewer can *believe* in the diegesis. A fragmented listening-viewing position is by no means unusual in fiction filmmaking, as has been argued above. What is unusual is that the listener-viewer is suspended between two 'territories' (Chion, 1994). Such an indeterminate position raises issues that most narrative filmmaking has traditionally tried to efface (see Chapter 4 for discussion). For this reason, it could be argued that Stalker deploys the representational conventions of narrative filmmaking as a resource for disorienting the listener-viewer by presenting a narrative-filmic world to her that is modified in a subtle yet critical way.

6.4   Conclusion

The examples in this chapter are unconventional in their audio-visual design. For this reason, they are not exactly representative of fiction film sound. However, they

are chosen precisely for their unconventionality because the audio-visual strategies they enact serve to illuminate what *is* conventional about film sound; that is, they emphasise the conventions that usually go unnoticed in non-professional engagement with audio-visual film exactly by subverting them. As in the example of dialogue, music and effects (DME) discussed in Chapter 4, it is when conventions are deployed to ends that ultimately throw those conventional practices into relief that they can be characterised as *resources* for making audio-visual meanings.

From a practice-oriented perspective, sound-to-picture, in fiction filmmaking, is concerned with elaborating the synchronised image; with *specifying* it through according sync-sound to certain visual events that are deemed crucial in fulfilling the overarching aims of a particular genre practice. In the case of narrative fiction filmmaking, as discussed in Chapter 4, the image is itself the product of the semiotic processes of continuity scripting and cinematography, and thus is already encoded to fulfill narrative functions (such as continuity between contiguous shots). But, given the 'added value' of sound that has continued to influence deployments of sound-to-picture since the early 'cinema of attractions' (Lastra, 2000; Gunning, 1986), it is not surprising that sound, if it is to perform any function at all, should be marshalled into bolstering the narrative practices of the image – particularly lighting, foreground/background composition, continuity, and so on.

In collaboration with the image, the sound track is deployed to perform, what can be termed, 'specification operations' that are invaluable in filmic story-telling. The consistent foregrounding of the voices of central characters is but one of these operations. Others, discussed above, include constructing stable *points of audition* from which to identify with certain characters or around which to voyeuristically hover ('anchored' by the sound and comfortable in the notion that the reverbless dialogue is intended for you, the listener-viewer [Altman, 1992: 60]).

In *Stalker*, the Russian Version of the sound track performs such operations, and provides a recognisably fiction-filmic audio-visual perspective from which to become involved with the diegesis – as an experience that affords the continuity and transparency of story-telling. In the original version, however, the sounds that are made salient through the practices of sound-to-picture disorient the listener-

viewer through offering a drastically fragmented POA, that obstructs her from fully involving herself in the diegesis; this is accomplished by *not quite* fulfilling the expectations that the genre of narrative filmmaking 'solicits and elicits' (Graham, 2001, 2006). Such an example of narrative sound-to-picture illuminates the extent to which the listener-viewer is positioned by means of sound-to-picture practices, such as reverb, equalisation and mixing, and by the resultant semiotic integration of represented sound and image spaces and perspectives. Indeed, the comparison of the original and new version of the *Stalker* sound track foregrounds just how effective sound is in orienting the listener-viewer to the image, and the diegesis. Crucially it also emphasises just how the boundaries around the context of fiction are maintained through the patterned deployments of audio-visual resources. When the 'critically deaf' (Brophy, 2000) listener-viewer is confronted with an audio-visual experience that, on the surface, produces and enacts a recognisable context, such as fiction film, but that remains somewhat disorienting or confusing, the genre boundaries are in the process of being expanded to include new possibilities of mode practices.

Chapter 7: Analysis 2 – The sound track in non-fiction film

7.0   Introduction

In this chapter, I examine the non-fiction film sound-track in terms of how the resources of sound-to-picture are deployed and to what ends. Specifically, as explored in Chapter 6, the use of certain practices to construct *subject position* and *agency* in non-fiction film are explored in order to comprehend what practices are germane to non-fiction sound-to-picture. The extracts deployed here are sourced from contemporary non-fiction films, which, I argue, deploy the technical and conventional resources of sound-to-picture that fiction filmmaking utilises, but to different semiosic ends. In consonance with Carroll (1996: 286), the transgenericity of the resources of sound-to-picture across fiction and non-fiction contexts is accepted as a heuristic for analysis. To repeat his words,

> [t]he distinction between nonfiction film and fiction film cannot be grounded in differences of formal technique, because, when it comes to technique, fiction and non-fiction can and do imitate each other.

On this basis, then, the aim is to understand not the differences and commonalities of the 'formal techniques' deployed in non-fiction and fiction (although that may be interesting in itself). It is instead to determine how very similar practices are deployed to make very different meanings across both contexts. This, of course, necessarily involves attempting to understand the relationship that obtains between semiotic practices and contextualisation – to which I turn in the next chapter.

In what follows, I explore the ways that sound is used in TV news and documentary film to achieve two representational strategies: to ascribe *agency*, and to construct a *subject position*. This entails an understanding of the textual aims of these two types of non-fiction film, in order to comprehend how these aims are realised in the practice of sound-to-picture. That will be the task of the next chapter, in which I turn to discuss the themes and connections that arise in the analyses.

7.1 'America Under Siege': the BBC's reflection on the events of 9/11

Extract 3: *BBC News 24: September 11th, 2001*

At the outset of the 9/11 coverage, the TV news channel *BBC News 24* broadcast images recorded by 'helicam'. At times, this footage was noticeably *silent*, perhaps a decision to divorce the images of the collision from the specific position signified by the sound of a helicopter engine. Seven or so hours after the events, the channel broadcast a sixty-minute 'special' hosted by David Dimbleby, with an intro sequence featuring footage of the second collision. This time, however, the footage had been radically altered in two ways. Firstly, new sounds were added to the previously silent film: a background sonic atmosphere of 'emergency signifiers'; and, secondly, certain visual effects were added.

In the addition of sounds and image effects to the footage of the second plane hitting one of the World Trade Centre towers, this brief 8-second audio-visual introduction to the BBC's roundtable discussion on the day's events in an obvious audio-visual *recontextualisation* of the events. Not only is this true by virtue of the fact of the immense amount of semiotic labour given to the image editing, but it is also true of the wholesale *addition* of sounds where none existed previously. In this sense, it resembles the sonic recontextualisations of the image in archive footage that occur regularly in non-fiction film. One cliché relevant here is the addition of camera sounds – whirrs, flash bursts, clicks – to footage of press photographers at work. Another is the addition of a projector sound to silent film footage, which positions the listener-viewer as spectator at an early picture show. This knowledge of the addition of sounds in the *Siege* extract offers us a rare opportunity to examine how TV news practice deploys sound to semiotic – particularly *orientational* – ends.

7.1.2 Audiovisual Description

In the image track, the aircraft careers towards the North Tower, briefly disappearing behind the South Tower. In the sound track can be heard the constant 'whirr' and 'whine' of a helicopter, the beginning of which is in synchrony with the

start of the visual stream of images. At the point at which we assume the collision has occurred, there is a very brief deepening of the image contrast and brightness and a subsequent 'flash' on the left side of the screen (the effect of which is produced by the brief contrast/brightness intensification). The flash is made all the more salient by the brief deepening of contrast just before it. In the sound track, the same helicopter sound as previously heard continues. Next, in the image, the plane emerges from the other side of the tower. This is visually represented by a sharp return to the 'normal' contrast/brightness of the sky, and a 'fireball' beginning to form at the point of exit. Simultaneously, the text 'America Under Siege' begins scrolling across the screen. In the sound track, a salient, brief noise is heard in synchrony with the sudden return of contrast and the start of the scrolling text. The collision becomes the point at which the flash subsides, the text begins to scroll, the fireball begins to grow, and a sound with a sharp attack transient occurs. Furthermore, following this moment, the 'collision sound' decays, and a number of other sounds begin to emerge: various emergency signifiers such as 'walkie-talkie' voices, and sirens at a street-level aural perspective. The rise in pitch of the siren on the sound track occurs in apparent synchrony with the growing of the fireball on the visual track.

7.1.3   An intuitive reading

The extravagance of the audio-visual design of this extract indicates immediately that it is not a news *report*. Other indications of its status as an instance of the TV news genre, however, are present: the clock at the bottom left of the screen, the channel 'ident' at the top left. It is impossible to determine whether someone seeing this clip for the first time would be confused by its content, but it is assumed here that if a given listener-viewer was already aware of the events of 9/11 then they would have seen this same footage many times during the day, and would be primed to read the *Siege* extract as itself offering an 'interpretation' of the events. The content of the sound track, as well, is typical of the associated footage of the events that had also been repeated incessantly throughout the day on every 24-hour news channel. But their presence in synchrony with the image of the collision is

striking. That is not to say that the combination seems illogical, but rather it is to argue that the configuration of images depicting the exact moment of the second attack with sound of its aftermath is noticeably aiming to communicate something to the listener-viewer that is not just a matter of *presenting* the event. Instead, the on-the-street sound track seems to orient the listener-viewer to the human-social consequences of the attack. The scrolling text 'America Under Siege' contributes much to this interpretation and the listener-viewer is oriented toward the events as a media event that they *must already know* has occurred. Specifically, an intuitive reading reveals that the synchronisation of the attack and its aftermath – that is, action and reaction – is geared to provide some commentary on the events; perhaps that what we see in the image is the cause of what we hear in the sound track from the sync-sound on. Indeed, commentary is precisely the function of the ensuing programme: a critical reflection on the events, from the vantage point of a number of hours after the attacks, that aims to discuss the reactions that are possible. Simultaneously, the clip does semiotic work that is surplus to the assumed presentational function of the TV news report: it *positions* the listener-viewer in relation to the events depicted in a way that connects overtly with the action-movie (or 'disaster movie') genre. As a number of authors have argued, since the 1980s, the tendency of newsmaking towards entertainment has involved a number of apparent linkages with the genre of fiction filmmaking (Postman, 1985; Tumber, 1999). The repeated statements by reporters that the events were 'like a movie' are clearly represented in the configuration of sound and image in this clip, which appears to perform a 'distancing' role from the events themselves, so that the media can take time to analyse the events and their implications, and even to speculate on the identity of the agent(s).
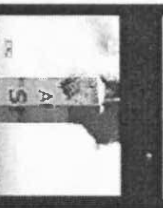
Time	Visual track	Visual Description	Sound Track
0.0		The plane flying towards the tower from the right.	Helicopter whirr and rotating blades.
1.5		The plane disappears behind, and a brief flash is seen.	
3.0		The flash subsides and an explosion is seen on the left side as the plane exits. Text begins to emerge from the tower.	[POS 1] The 'crackle' of a walkie-talkie, with a sharp attack transient.
4.5		The text 'America Under Siege' emerges.	[POS 2] Voices as if through walkie-talkie. Voices at street level.
6.0		The ball of fire slowly grows and rises. Image contrast deepens on the growing ball of fire.	[POS 3] Rising siren, continues until end.
7.5			

Figure 1 Extract from *America Under Siege*, BBC News 24, 11/09/2001

7.1.4 Deployment of resources

Sync-sound

Aside from the fact that the image and helicopter sound begin at the same time, there are three salient POSs functioning here:

> POS 1: The explosion in the image and the sound track's walkie-talkie crackle.
>
> POS 2: The initiation of the scrolling text and the sounds of voices through a walkie-talkie and street-level voices.
>
> POS 3: The growing ball of fire in the image and the rising siren in the sound track.

These serve to establish the audio-visual 'phrasing' of the sequence (Chion, 1994), and function as linking mechanisms between the sound and image tracks. It may not be the case, however, that POS 2 and POS 3 would be considered salient were POS 1 not present. In other words, it is entirely possible that 2 and 3 depend on 1 for their salience, given that POS 1 is a punctual and definite POS. That is, its precise temporal alignment with the emergence of fire upon the plane's exit, coupled with its brevity as a sound effect, encourages the listener-viewer to understand it as sync-sound – sound that is purposely combined with a represented visual event. In any case, there are three visual events that are associated by means of sync-sound to three, apparently discrete, sounds events.

POS 1 is most striking because it offers a sync-sound effect at precisely the point we would expect to hear one if, indeed, this were a movie. Not only that, but it provides a 'pivot' from the helicopter sound to the 'on the ground' sounds of human response. The fact that the sound of the POS 'belongs to' the second set of sounds suggests that synchronising it with the plane collision was perhaps opportunistic. At any rate, it provides the means for an efficient configuration of audio-visual

meanings – that is, a representation of the basic fact of the attack, evidenced in the image (notwithstanding the image modifications), coupled with a representation of human response and reaction, evidenced in the sound track.

Editing

The sound track displays only one edit that we can be certain of: the *transition* from helicopter to street sounds. This composition is telling, as the edit – the shift to an 'on the ground' array of sounds – is synchronised with the depiction of the collision. In terms of the intuitive reading offered above, this point of transition signifies the moment as that when human response is instigated.

7.1.5 Sound-to-picture: Contextualising relations of sound and image

Isolating the functions of sound from its audio-visual context is not possible in this analysis because the only resource under discussion here is sync-sound. As was argued in the previous chapter, this is a category only realised in an audio-visual context. For that reason, here I move directly on to discussing the deployment of sync-sound in terms of how its presentational, orientational and organisational functions are configured in a redundancy relationship with those of the image.

Functions	Sound	Image
Presentational	Process: Helicopter engine and rotating blades. Crackle of walkie-talkie. Rising emergency siren. Panicked voices Technical resource: **Sync-sound**	Process: Plane flying into the tower, erupting in flames. Image darkens then a brief white flash, left of frame, at moment of impact. Scrolling text: America Under Siege Ball of fire upon the plane's exit.

Functions	Sound	Image
Orientational	Perspective: [Figure then Ground] Helicopter as if very close, with no reverberation. Voices and walkie-talkie as very close, no reverberation. Street-level. [Ground] Siren as distant but loud. Reverberant. Street-level. Technical resource: **Sync sound**	Motion: A static position. Distance: A far distance, but with no obstructions. Modifications: [Brightness] deepened contrast; bright flash; [Colour] deeper red on ball of fire; [text] scrolling, coloured text.
Organisational	Composition: [Point of transition] Walkie-talkie crackle acts as a point of transition from the helicopter perspective of the moment prior to the collision to the street perspective representation of voices and sirens. Technical resource: **Sync-sound**	Composition: [Point of transition] Deepened contrast and bright flash. Initiation of scrolling text.

Table 5: Functional meaning in the sound and image tracks.

Sync-sound: configurations of presentational and orientational meanings

At issue in consideration of sync-sound here is the redundancy between the walkie-talkie sound and the collision it is synchronised with. In terms of presentational meanings, the sound and image are redundant only on a criterion of time: that is, they are represented simultaneously. While this may seem a superficial redundancy, it is in fact crucial to the meaning making of the clip in that the sync-sound of the walkie-talkie initiates a stream of audio containing other sounds of human response: sirens, voices and so on. Therefore, sync-sound should be discussed here in terms of both the 'collision' sound effect, and those sounds that unfold *after* POS 1 (the collision point).

I argue here that sync-sound is performing two related functions. Firstly, it is deployed to suggest that the 'emergency' is the result (effect) of an *intentional* act (cause). This is accomplished through temporal co-ordination of sync-sound and the

text and visual events in the image. In concert with the text 'America Under Siege', sync-sound here plays a role in establishing that the ensuing sounds of emergency are the result of the actions of a specific *agent* ('terrorist')'. Without the scrolling text, the sync-sound can only make the meaning that the collision itself, not a human agent, is the cause of the emergency in the sound track. And while – without the sync-sound – the collision and text *can* specify a human agent and a victim ('America'), the sync-sound presents concrete human response and, crucially, a *human* victim, not the abstraction of a country name, nor an economic symbol. The configuration is achieved through expansion relations of *elaboration, extension* and *enhancement* (Halliday, 1985: 225-241; Martinec and Salway, 2005; Van Leeuwen, 1991; 2005: 230) between sound and image; image and text; and text and sound.

This kind of 'participation analysis' (Martin and Rose, 2003) is crucial to understanding how certain represented participants come to be construed as agentive in multimodal texts. Analysing how agency is constructed is a key step in understanding the semiotic means with which the media – in both fiction and non-fiction – assign blame to individuals, groups and happenings.

Secondly, sync-sound positions the listener-viewer to the event in a complex way: as both 'witness' and 'detached' observer. It achieves this by expanding the presentational meanings of text and image by combining them with the sounds of human response and a state of emergency. In this sense it performs a largely orientational role that is achieved through configurations of presentational and orientational meanings across the sound and image tracks. Crucial in this argument is the notion of point of audition (POA), and the enacted tension between proximity and distance (Chouliaraki, 2004b) – between 'being there' and 'distancing', which, it is argued here, is central to two news projects: of presenting 'mood reporting' in the absence of hard facts; and of distancing the event so as to gain some analytic purchase (cf. Jaworski et al, 2005).

Constructing human agency

As in the *TYFS* analysis (Chapter 6), it is argued here that sync-sound is deployed in the audio-visual construction of human agency. Again, this rests upon the foundational notion that – whether perceptually or conventionally derived (or a combination of both) – sync-sound has the potential to establish a cause and effect relation between visual events and sounds. In functional-semiotic terms, it establishes a logical, conjunctive relationship between sound and image that is of the *consequential* type: one (sound) is a consequence of the other (image). The conventions of 'mickey mousing', lip-synch, and the intuition that the actions we see occurring should make a sound (the hearing-impaired aside), form the basis of an understanding of sync-sound as 'producing' the effect of realism – that is, that what we hear is a direct consequence of what we see (see Chapter 4 for further discussion). To repeat Lastra's words:

Decades of tin-sheet thunder and coconut shell hooves prove [...] that fidelity to [visual] source is not a property of film sound, but *an effect of synchronization*. A gun firing on the screen accompanied by any brief, sudden, explosive sound *produces* the effect of source. (2000: 147)

Given sync-sound's (conventional and perceptual) potential for suggesting a basic causality between the presentational meanings of image and sound, it follows that – in this clip – the only agent of the on-the-ground emergency of the sound track is the *collision* represented in the image. In sync-sound's basic potential, no intentionality behind the act can be signified, and thus no human agent can be specified. If there *is*, as I suggest, human agency represented in this extract, it must be that it is constructed in coordination with another semiotic element. My basic thesis is that human agency is specified in the semiotic interaction of, on the one hand, the presentational meanings of the collision (image) and response (sound); and, on the other, those of the scrolling text, 'America Under Siege'. As I argue, the text extends the relationship of collision (image) and response (sound) beyond a mere cause and effect relationship, to one of *agent* (image) and *victim* (sound).

189

The presentational meanings at issue here are those of the sounds 'on-the-street' in the sound track, and those of the collision and the scrolling text in the image.
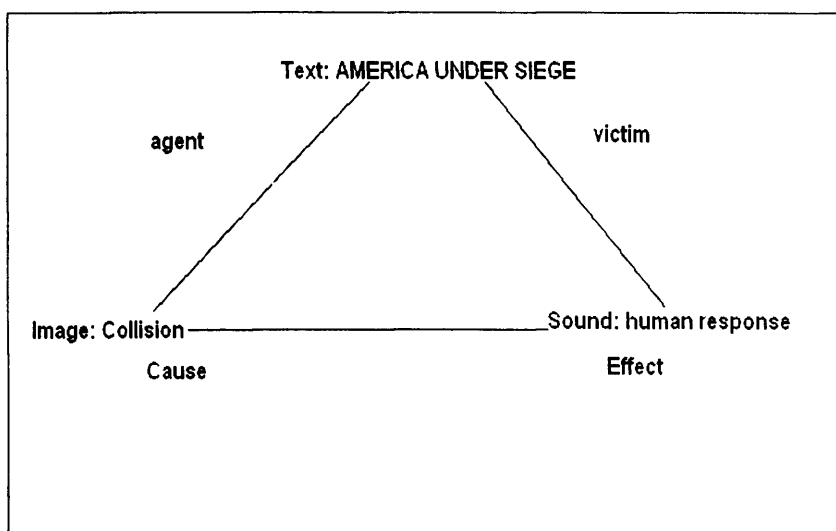
Figure 2: Contextualising relations between sound, image and text

As represented in Figure 2, the direct relation between image and sound, by means of synchronisation, is one of cause and effect. The text intervenes on this relationship, however, by modifying the presentational meaning of the collision to suggest human agency, and by modifying the meaning of human response to suggest human victimhood. The established causal relationship between the meanings 'collision' and 'response' is therefore contextualised by the text 'America Under Siege' as a relationship of 'agent' and 'victim'. These semiotic processes (which of course do not unfold temporally in the order they are analytically reconstructed here) are explained below in terms of *elaboration, enhancement* and *extension*.

In the interaction between the presentational meanings of *image and text,* and between those of *text and sound,* human agency is constructed in the *Siege* clip through relations of *extension* – precisely, *addition* relations (Van Leeuwen, 2005:

230; Halliday, 1985: 225-241; Baldry and Thibault, 2005: 235). That is, the text adds to the meaning of the image: the collision is an intentional of a human agent, not an accident. And the text adds to the meaning of the sound: the response we hear is that of a *victim*. Also operative here are relations of *enhancement*; that is, the text enhances the image meanings, and, separately, enhances the sound meanings, by qualifying them in terms of circumstance – specifically, in this clip, in terms of place. Already, here, we can see that the categories of 'agent' and 'victim' emergent in the sound and image tracks are linked, in a dependent way, to the presentational meanings of the scrolling text.

Yet the production of the meaning 'agent'/'victim' is dependent on the relations that obtain between image and sound, because the agent/victim meaning can only be made on the basis of a *causal* relationship between the collision in the image and the human response in the sound track. Involved in the construction of causality are relations of *extension* – between the presentational meanings of sound and the collision in the image. What might appear problematic here is that there is no way of determining *with certainty* which mode in a simultaneous presentation is to be understood as primary, and which as secondary, in a synchronous presentation of sound and image meanings. In Martinec and Salway's (2005) terms, this is an issue of the *status relations* between modes. How, specifically, is the relative status of sound and image in a synchronised audio-visual configuration *realised* in a given audio-visual text? One answer lies in an interpretation motivated by convention. That is, on the level of presentational meaning, the convention of synchronisation between sound and image has evolved to suggest a basic causal relationship between them, specifically positing image as the cause and sound as the effect (indeed, that might be the origin of the term 'sound effect'). There are instances when this convention is challenged – such as the visual appearance of bullet holes on a wall, synced with the sound of ricocheting bullets – but as long as a represented visual action is understood as a sound *emitter*, a closely synchronised sound is likely to be understood as an *effect* (see Chapter 4).

In positing the image as the *cause* of the meanings arising from the sound track, the sound extends the meaning of the collision. But this is not a one-way relation, because the image of the collision extends the meaning of the human

191

response in the sound: the sound is determined as an *effect* of the collision. Therefore, these aspects of presentational meaning in the sound and image are in a *dependency* relationship (Baldry and Thibault, 2005: 235) with each other; that is, their relationship as cause and effect is not 'mediated by a shared higher-order unit' (2005: 234). In Barthes' terms (1977) there is a *relay* between collision and response meanings that implicates the two in a 'complementary' cause and effect relationship (Van Leeuwen, 2005: 229).

There are two main conclusions to draw from the above discussion. The first is that it is the *combination* of meanings in the image, text and sound that serve to specify an agent of the collision ('terrorist'), and thus the ensuing devastation in the sound track. Working together, the sync-sound and the text disambiguate the polysemous significations of the image, from broadly specifying the collision as the moment at which 'emergency' occurs (an accident, perhaps), to precisely specifying a human *intention* behind the collision, thus constructing a human agent – albeit an unknown one – for the ensuing state of human emergency. Secondly, without the presence of the sync-sound, no human response is represented; thus, crucially, no human *victim* is specified. As I discuss in the next section, it is precisely such signs of human emergency that, as Chouliaraki (2004b: 158) puts it, are deployed 'to invest September 11 with certain "normative" discourses, of what it is legitimate and fair to feel and do vis a vis the event.'

Sync-sound, then, is involved in a dialogic and dynamic semiotic relationship with the scrolling text and collision in the image, and is at the core of the construction of a human agent of terror. For without the auditory representation of a human victim, the attack on 'America' remains abstract and underdetermined: a polysemous event in need of elaboration.

Constructing a subject position

In addition to the semiotic interplay of presentational meanings among the modes of sound, image and text that are mobilised to specify human agency, *orientational* meanings also play a role in the 'Siege' clip, by positioning the listener-viewer in relation to the presentational meanings of the sound and image tracks through the

mechanisms of auditory and visual perspective. Its functioning is crucial in signifying the concrete reality of Manhattanites on the ground, and therefore performs, not only a representational role, but a *representative* one (Batcho, 2005: 65), ensuring that the collision is not represented as without social consequence. In this section, I argue that sync-sound positions the listener-viewer to the event in a complex way: as both 'witness' (cf. Ellis, 2001) and 'distanced observer'. It achieves this by expanding the meanings of text and image by combining them with sounds of human response and a state of emergency. In this sense, sync-sound performs a largely orientational role that is achieved through configurations of presentational and orientational meanings across the sound and image tracks.

Proximity-Distance

The deployment of sync-sound in this clip is such that, at the moment of collision, the sound track represents not only the presence and activity of people on the streets of Manhattan, but also the auditory perspective of 'being there', a witness to the ensuing devastation and emotional response (Ellis, 2001; Chouliaraki, 2004b). A divergent perspective is opened up at the point of collision: between the disembodied, helicam long-shot POV in the image, and the close social proximity POA represented in the recording of human response at street-level. Understandably, the streets were crammed with commuters and Manhattanites attempting to rush to safety, so it is no surprise that most audio recordings of the aftermath of the attacks betray signs of bustling human activity. As argued in the analysis above, the image of the collision is specified as the intentional act of a terrorist *agent*, while the sound of response in the sound track is rendered as that of a human *victim*. Therefore, the listener-viewer is positioned as 'with' the victims, in the thick of the chaos on the ground. Simultaneously, the listener-viewer is positioned at considerable distance from the agent, disembodied and floating at the same height as the two towers.

The 'gap' opened up between sound and image meanings at the collision point is, for Eisenstein, the moment when 'art begins' (quoted in Barthes,1977: 61). That is, when 'the creaking of a boot on the sound-track occurs against a different visual

193

shot and thus gives rise to corresponding associations'. In other words, the simultaneous yet divergent orientational meanings of sound and image enact an *interpretation* of the event presented in the image; not 'art' as such, but certainly a way of reaching beyond the purported 'objective' presentational meanings of synchronised helicopter sound and helicam image to present an *attitudinal* stance towards the visualised action.

The duality of divergent perspectives gives rise to a duality of *orientations* to the presentational meanings of the depicted event. These orientations are linked by virtue of the *expansions* of the presentational meanings of sound, image and text that I have outlined in the previous analysis (on *agency*). Orientational meanings of sound and image are only joined by virtue of the logical connection between audio-visual presentational meanings. In other words, how a subject-position is constructed by means of a configuration of sound and image orientational meanings – whether those orientations are congruent or dissonant (Royce, 1998: 29) – can only be discussed if we can identify presentational redundancies (co-contextualising relations) between them. In the *Stalker* extract (original version), no logical relationship can be construed between the presentational meanings of the configuration of interior sound and exterior image, and therefore no orientational semiotic coupling is emergent, leaving the listener-viewer 'stranded' between both orientations to the scene. As argued, this engenders a highly disorienting, indeterminate subject position, perfectly suited to Tarkovsky's aims as a filmmaker. Here, however, the distinct perspectives on the event – in terms of space *and* time – are implicated in an orientational relationship that is enabled by presentational relations of *extension*: that is, the presentational meanings of the collision in the image *add to* or *vary* those of the human response in the sound track. This relationship establishes that the divergent perspectives are presentationally co-contextualising (Lim, 2004: 184), and therefore that there is some 'sense' to the linking of orientational meanings, that is grounded in the presentational logic of cause and effect.

In the *Siege* clip, the resources of sound and image are co-deployed to orient the

194

listener-viewer to the events – to suggest how they are to *relate* to those events – in such a way that a 'dialectic' of *proximity-distance* (Chouliaraki, 2004b: 158) is established. As Chouliaraki (2004b: 151) argues, 'the semiotic resources of television [are used] to invest September 11 with certain "normative" discourses, of what it is legitimate and fair to feel and do vis-à-vis the event.' Specifically, the proximity-distance duality is implemented in the audio-visual configuration of this clip so as to position the listener-viewer in a contemporaneous relationship of distance to the human *agent* of the attack and proximity to the human *victim*. This is not to argue that represented distance from human agents of terrorist acts is a representational strategy of western TV news; indeed, such an orientational stance is understandable given that the agent was (more or less) unidentified at the time when this clip was broadcast. Given an (as yet) unknown human agent, what it is fair to feel is sympathy for the victims (specifically, a shared sense of the *emergency*) and what it is fair to do is to 'stay close' and continue to engage with the people on the ground. The level of represented proximity entails that, in other words, 'our thoughts are with the victims'.

And so are our ears. The potential of sound-as-medium to situate the listener 'in communion' with the actors represented in a sound track is vast, as Van Leeuwen (1999: 197) argues, invoking de Buffon (1971: 199): '[i]t is above all through hearing that we live in connection with others'. The representation of the sound that engulfs the streets of Manhattan is undoubtedly amenable to situating the listener-viewer in a proximous relationship with the victims, however low-fi its reproduction. The 'ambient' sounds of human response, in Batcho's (2005: 65) terms, provide *access* to a news event, bringing the listener-viewer in closer proximity to the experience of the people on the streets. This is the *representative* (2005: 65) potential of ambient sound: its ability to represent the culture and society that are implicated in a mediated event.

However, as Batcho (2005: 63) argues, the TV news practice of emphasising the 'voice of authority' is often at the expense of ambient sound. In his words,

All efforts toward sound [in TV news] are aimed at enhancing the message of authority figures—and by extension, serving the commercial interests of the broadcaster—over the needs of the public. The ideological undercurrent of this is that it wraps an illusion of pluralism through a notion of objectivity that the public will accept.

If TV news objectivity is, in part, an *effect* of 'removing' the ambient sound that has the potential to include 'those who are not granted access to one another through the medium' of TV news, then the sole dependence on ambient sound in the *Siege* clip certainly does not contribute to producing 'conventional' news objectivity; that is, the objectivity produced by means of an authorial voice, around which ambient sound and images play an illustrative role. Objectivity, as O'Halloran argues (2005: 209), 'is the organisation of particular experiential and logical realms of meaning which are *accompanied by a contracted interpersonal stance*' (my emphasis). Given its status as introduction to a critically distanced analysis of the attacks, including the opportunity for speculation about causes and effects of the attacks, it is not surprising that this clip does *not* 'contract' its interpersonal stance towards the events, but rather *expands* it, by representing – audio-visually – both the acts of a human agent, and the reactions of a human victim, and positioning the listener-viewer to those meanings in a way that maintains *distance* from the agent's actions, while ensuring *proximity* to those of the victims.

For Batcho (2005), this can only be a good thing, since the *Siege* clip is entirely inclusive of those aspects of a soundscape conventionally disregarded in TV news sound practice, but that are crucial in providing *access* to the people represented in the sound track. It provides a way of mediating, of 'opening a channel' between listener-viewer and represented actor. In his words (2005: 69):

By including the ambient contextual sound in the broadcasting of our shared history as it happens, two critical goals are accomplished: (a) the television viewing public is granted the opportunity to hear for themselves the sound of an event as it occurs rather than relying on subjective descriptions by

voices of authority, and (b) the public being *represented by* television in its coverage of events is heard more openly. In both situations, the public is given access to that which has been denied by restrictive practice. It is the audible equivalent of widening the camera frame without losing focus. Ambient sound, if practiced with good intentions, gives the audience an answer to the question: "Yes, I hear what you're *describing*. But would you mind also letting me hear it for myself?"

'Like a movie': the authenticating function of sound

The proximity to the victims that is established in the *Siege* clip is in keeping with the practice of 'reporting' the emotional responses of eyewitnesses in the absence of any hard facts to report. In particular, this has been discussed elsewhere as a consequence of the emergence of 24 hour news channels, and of the problems and challenges of live 'disaster' reporting (Jaworski et al., 2005). In such cases, emotional responses increase in newsworthiness, and their prevalence and frequency subsequently comes to define the event for listener-viewers. Thus the fact that we hear human response (the people on the street, the emergency services) emerging from a POS with images of the moment of attack efficiently sums up 'all we know' (i.e. all we *can* know) about the event. All we can know, in short, is the human response on the ground, and the belief-defying images, rendered here in the only way we can comprehend images that defy belief: 'like a movie'.

As a number of authors have argued, since the 1980s, the tendency of newsmaking towards entertainment has involved a number of apparent linkages with the genre of fiction filmmaking (Postman, 1985; Tumber, 1999; Chouliaraki, 2004a). Indeed, the *Siege* clip does semiotic work that is surplus to the assumed presentational function of the TV news report, that functions to connect overtly with the action-movie (disaster movie) genre. The flash, deepened contrast and colour on the fire, and the scrolling text 'America under Siege' in the image, all serve to invoke the action-film genre as an intertext (cf. Thibault, 1991; Lemke, 1995). So, too, does the fact of a sync-sound at the point of collision – exactly where one would expect

197

to hear a sound if this was a fictional representation. As the presenter of the 'America under Siege' special, David Dimbleby, announces, just after the *Siege* introduction,

> This is the face of warfare in the twenty-first century. The events today in New York and Washington were so horrifying it's hard to find the words to describe them. Disaster movies invent acts like these as fantasy. But this was real.

The repeated statements by reporters that the events were 'like a movie' are clearly represented in the configuration of sound and image in this clip, which appears to perform a 'distancing' role from the events themselves, perfectly geared to the aims of the *Siege* special: to take time to analyse the events and their implications, and even to speculate on the identity of the agent(s).

Yet, while the image indexes clearly the action genre and its expected drama, the content of the sound track is comparatively pedestrian, unextravagant in its composition (aside from a deployment of a sync-sound effect). I argue here that the sound track is deployed in an *authenticating* role: it does not serve to readily invoke fiction film as context for its meaning, but rather functions as a 'proximity anchor' for the particular kind of 'distancing' that is established by visual means. In terms of the configuration of auditory and visual perspective, sound is used as a means to 'root' the spectacular representation of the image (flashes, scrolling coloured text, and so on), and the intertextual link to the action-movie genre, in an authentic (albeit mediated) 'experience' of the event.

But it is not only the auditory perspective that serves to specify the 'unbelievable' images as authentic. It is also the sonic character of the sound track. In contrast to the lavish visual modifications, the sound track's character, and the choice of sound to accompany the collision, do not readily invoke fiction film as context for their meaning. Instead, they attest to 'witness', an authentic experience from on the ground. The 'gritty', monaural sound track, and its lack of lower-frequencies (most

notably, at POS 1), bears signs of the technological and technical constraints that have come to define 'news reality' (Bolter and Grusin, 2000: 189-192). In this capacity, sound is performing a presentational function by indexing the very means of its production. (And in this sense, it is through a configuration of audio-visual presentational *and* orientational meanings that the *total* orientational stance of the clip is constructed.)

In addition to 'constraint signification', the authenticity that sound provides here is also understood in terms of its *content*: that is, it extends the image meanings (both orientational and presentational) by providing sound that is (apparently) sourced from recordings of the on-the-ground events of 9/11. Whether or not this is a matter of ease, in that sounds of human response and emergency would be readily available to the editor of the clip, the addition of sounds sourced from associated events (to be understood as parts of the *same* event) has become somewhat of a trend in recontextualising existing images, and can be observed in documentaries such as *9/11* (Jules and Gedeon Naudet, 2001), in which images depicting the work of the firefighters attempting to get beneath the rubble of 'Ground Zero' is accompanied by *additional* sounds of metal and glass being shifted about from the same recovery effort, yet from a different point in time. (These additional sounds, incidentally, are panned to the extremes of the stereo-field, suggesting a physical width far wider than the visual frame can convey. This is a common use of the stereo field.) The 'authenticity' here operates by a curious logic: what we hear is both semiotically connected *and* disconnected from what we see. The sound track has an indexical status beyond the space-time represented in the image. In the *Siege* clip, for instance, the sounds added are from a different point in time *and* space.

The apparently authentic sound track of the *Siege* clip answers the question that the image prompts us to ask: 'is this real?' The sound track provides a means for authenticating the depicted event, and for anchoring the extravagant image modifications in a way that does not 'overdo' the invocation of the action-film genre. As a point of interest, this 'restraint' can be exemplified by adding library music and sound effects suitable to the action genre to the clip (example provided).

The effect is startling: in comparison with the original version, the newly-modified clip displays almost no authenticity *as news*, because the deployment of sound and image overdetermine the generic link to fiction, leaving little behind of the original non-fiction context. While this modification resembles much cable TV news as much as it does fiction film, that only further buttresses the argument that the original sound track serves an authenticating role, for in much American TV news the audio-visual recontextualisation of actual events is indeed overdetermined and certainly transforms its object more completely into one of entertainment (Postman, 1985; Bolter and Grusin, 2000; Luhmann, 2000). In the *Siege* clip, then, the use of 'authentic' sound allows the image more semiotic scope, more space to perform *orientational* work such as linking up with the 'blockbuster' action-movie genre.

7.3   Commandante: 'Castro remembers'

Extract 4: *Commandante (Oliver Stone, 2003)*

7.3.1   Audiovisual Description

Stone and Castro are depicted as in dialogue, from various camera angles. In the sound track, we hear Stone prompt Castro to recall a moment from his career, that Stone has experienced only as footage from the film, *Life*. As he imitates the crowd chants of 'Fidel! Fidel!', the sound track deploys the archive sound of the event (from *Life*), introducing it by means of a brief plucked guitar figure, that then synthetically repeats as played backwards. The sound of the chanting is at a low level in the sound mix, and Stone's voice is dominant. However, the sound is not only at low level, but also muffled and almost unintelligible. Its sonic presence increases, however, very slightly, and gradually, as Stone poses his question. In the image, we cut to Castro, at a distance of medium close up, and then zoom in as he listens to Stone. In the sound track, the chanting is rendered as still 'emerging' in presence. In the image, Castro looks contemplative, and then raises his eyes. At which point, there is a visual cut to archive footage of the event. On the cut, the chanting sound becomes maximally present in the upper-mid range.

7.3.2   An intuitive reading

As Castro appears contemplative, and we hear the barely intelligible sound of the crowd chanting his name, it seems as if his very memory process is being represented audio-visually. Yet, at the same time, the process seems to speak directly to us, as if it is *our* memory that is emerging. What is going on here? Is it simply that the archive footage – and the very concept of archive footage – signifies to us that these memories are *ours*? Perhaps it is the sound that appears to anticipate the footage, luring us in by the ears. How are sound, narration and image orchestrated to produce these two distinct yet complementary perspectives on the same event?

I argue in this analysis that the deployment of sound in this extract performs two main functions. The first is that it establishes Castro as the *agent* of recollection, moments prior to a visual edit that also ascribes agency to Castro; i.e. by means of editing Castro's facial expression (looking up) and archive footage (of the event he is asked to recall) as contiguous shots. Additionally, the particular way that the sound is implemented, with evolving changes in frequency characteristics, contributes further meaning to this audio-visual configuration: specifically, its evolution as a sound event extends the meaning of the visual editing to signify an *emerging* memory rather than one that suddenly, and with certainty, 'springs to mind'.
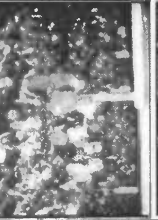
Time	Visual track	Visual Description	Sound Track	Narration
0.0		Several shaky camera shots of Castro and Stone in dialogue, in a film theatre	Plucked guitar figure leads into the low-volume, **low-filtered** sound of crowds chanting 'Fi-del, Fi-del'.	One half million people were in a square?
1.0				
2.0			Chanting gradually increases in loudness.	And shouting
3.0		Shot of Castro from behind another camera.		Fidel
4.0		Cut to Castro looking contemplative.	**[POS 1]** Chanting continues, **filtered** to admit more mid-range frequencies (presence)	Fidel, Fidel.
5.0				And I realise that you're
6.0				that you're speaking for the revolution and it's not personal but
7.0		Zoom in on Castro.		
8.0			**[POS 2]** Chanting continues, **filtered** to increase presence to maximum on the cut.	it must have been a great moment in your life.
9.0		Cut to archive footage.		
10.0				

Figure 3: Extract from *Commandante* (Oliver Stone, 2003)

202

Secondly, even while the listener-viewer is simultaneously oriented to the sound as a signifier and, importantly, an *object* of Castro's process of recollection (i.e. diegetic sound), a distinct orientational relationship between the listener-viewer and the emerging sound is maintained (i.e. the sound is non-diegetic). This orientational complex establishes that the sound is construed as an object of *mediation* between listener-viewer and Castro – that the sound is to be understood as representing both Castro's memory and the listener-viewer's own. I conclude that this complex of semiotic functions establishes a proximity to Castro that is consonant with the tenor of Stone's project, and enables a strong sense of identification with Castro.

7.3.3 Deployment of Resources

Equalisation/Filtering

Equalisation is used in this clip to dynamically transition from muffled sound to fully 'open' sound. The sound of the chanting that is introduced when Stone begins to question Castro is made fairly unintelligible by diminishing its mid-range frequencies. However, its content is recognisable because, firstly, it is introduced with Stone's imitation of the chanting; and, secondly, it has the sonic character of a crowd recorded by antiquated means, enabling us to identify it as *archive* sound. In the second case, the context it invokes (archive) is crucial to helping the listener-viewer to identify the sound, and the intertextual link with archive footage is made readily available *prior* to POS 2 by two very brief shots of a film projector that are contiguous with an extreme close up of Castro's face, and occur just moments prior to Stone's questioning. (These shots are a significant element of the text, but exceed the scope of the present analysis.) As the chanting is rendered more pronounced in the mid- to upper-mid range, it becomes more intelligible *as* chanting. When the cut to the archive footage occurs, the sound is rendered fully 'open' and intelligible.

Crucial to the deployment of equalisation here is the presentation of the *process* of altering the relevant frequencies. This is no trivial matter, because the temporal 'unfolding' of the sound is critical in being used to represent the uncovering of a

memory – as I discuss below.

7.3.4   Isolated functioning

As in the previous chapter, in order to understand the functionality of a particular resource of sound-to-picture within an audio-visual text, we first need to *isolate* that resource, and to attempt to comprehend the function it performs on its own. Here, I focus on the chanting crowd, specifically the deployment of equalisation that is central, I argue, in adding a crucial dimension of meaning to the represented process of memory recall.

Functions	*Sound*	*Image*
Presentational	Process: Narration – Stone prompting Castro	Process: Dialogue between Stone and Castro (2-shot)
	Chanting crowd, from 'distant' to 'close'.	Process: Castro appearing as if contemplating – looking up (1-shot)
	Room tone	Process: Camera roaming around the room, and 'probing' Castro.
		Circumstance: Filming of interview Participants: Castro, Stone, various crew members
Orientational	Perspective: [Figure] Distant to close chanting (muffled to clear)	Motion: Zoom to Castro
		Distance: Medium Close Up (but varying)

Table 6: Presentational and Orientational functions of the sound and image tracks.

Equalisation/Filtering

In Table 6, it can be seen that the sound of chanting in isolation is interpreted in terms of *distance*. The reason why the terms 'distant' and 'close' appear in quotation marks (at least at this point in the discussion) is that the quality of the sound, and the representation of the process, is understood as an *experiential metaphor* (Lakoff

204

and Johnson, 1980; Van Leeuwen, 1999; see Chapter 3 for discussion). In the case of 'distant' and 'close', this meaning derives from our *experience* of sound as a material event; from, for instance, hearing voices and other sounds through the separating wall of a terraced house; or hearing someone speaking closely to us, their voice exhibiting detail that we come to associate with closeness – whether as 'intimacy' or 'accessibility'. Here, I would argue that, because the sound of the crowd is initially quite abstract, the meaning of 'inaccessible' is made possible by the fact that the sound's content *is* inaccessible: its semiotic or material significance cannot be fully grasped until that content gradually reveals itself.

Terms such as 'muffled' or 'distant' bear little relation to the physical properties of the sound. Rather, they serve to entextualise that sound, so that it can be incorporated into a discussion on audio-visual semiosis. The acoustic properties of sound are only of relevance by virtue of *how they mean*; it means very little to our experience and thus our semiotic deployment of a sound that it has a particular harmonic structure. What is closer aligned with the semiotic project is knowledge of the *psychoacoustic* dimensions of sound production, since this deals with the perception of sound. For example, our perception of a sound that displays few high frequencies, but which we understand as *usually* betraying high frequencies, will tell us that the sound is *comparatively* 'muffled' or 'distant'. As Moylan explains (1992: 28),

> Two impressions lead to the perception of the distance of a sound source from a listener: (1) the ratio of direct sound to reverberant sound; and (2) the primary determinant, the loss of low amplitude (*usually high frequency*) partials from the sound's spectrum with increasing distance [...] (my emphasis)

And, of course, this works the other way around: high frequencies *increase* the closer a listener is to a sound emitter, and the fewer physical objects there are to obstruct the sonic path. Though I do not set out to discuss them in any detail in this thesis, psychoacoustic principles are *mediating* principles between the physical character of sound production and our interpretation of its semiotic potential. But

while psychoacoustics is not discussed here explicitly (aside from reference to human perception of sound) the way they are recontextualised as semantic concepts via *experiential metaphor* is central to our understanding of how a sound track functions in an audio-visual context.

In terms of presentational meaning, the increasingly intelligible sound of the chanting crowd is understood as undergoing a transformation from muffled to (relatively) intelligible. In terms of orientational meaning, the sound is posited in a relationship to the listener that is initially of close social distance but 'inaccessible', as through a thick wall (cf. Constantinou, 2002; Lastra, 2000: 141), and later fully accessible, with no apparent constraints on the listener's hearing (aside from those associated with the low-quality recording of archive sound).

The presentational meanings of the image, when considered in isolation from those of the sound track, *do* represent Castro's process of recollection, but this is a significantly different representation to that which emerges from the audio-visual integration. In consideration of the image only, Castro's recollection process is rendered as 'digital' and thus, most importantly, with more *certainty*. As for orientational meaning, the viewer is positioned at a close social distance to Castro, and the zooming, unstable camera places her in a dynamic and intimate relationship to Castro (as if 'probing' a subject). As Mercer (2002: 158) has argued, instability of camera movement, because of its evidentiary function (Nichols, 1991: 184), as proof of 'being there', is a visual signifier 'that represents the viewer/reader as if they were a participant themselves in the depicted events and constructs a more personal rather than institutional form of authority and therefore less rigid relationships between represented participants.'

7.3.5 Sound-to-picture: Contextualising relations of sound and image

Presentational meaning: Equalisation/Filtering

I argue here that the resources of sound-to-picture are deployed in such a way that a shift in presentational meaning occurs – from the sound being illustrative of Stone's questioning, to becoming the sound of Castro's imagination. By means of the material qualities of the sound (encoded as 'past'), and its frequency shaping (muffled then clearer), these in combination with the presentational and orientational meanings of the image establish Castro as *agent* of memory, and thus agent of the *sound* that signifies his memory.

As in the *TYFS* analysis of the previous chapter, and the *Siege* analysis above, demonstrating how agency is constructed audio-visually, in semiotic terms, can be accomplished by considering the sequence as without the sound track. On this view, Castro's memory process appears to be constructed by visual means: the timing of the visual cut to the archive footage is such that a relation of *extension* is actualised, specifically by cutting on the action of Castro raising his eyes (see Figure 4). Precisely, the archive shot extends the meaning of the prior shot of Castro looking contemplative to mean that he is remembering the event depicted in the archive shot. Since extension is a two-way relation, similar in function to Barthes' *relay* (1977: 45), the shot of Castro extends the meaning of the archive shot to mean that it depicts a *recollection*. Thus, despite its status as archive footage, and thus as public, the archive shot is recontextualised as a *private* memory – the contiguity of the shots establishes that it is *Castro*'s recollection that is represented by the visual editing.
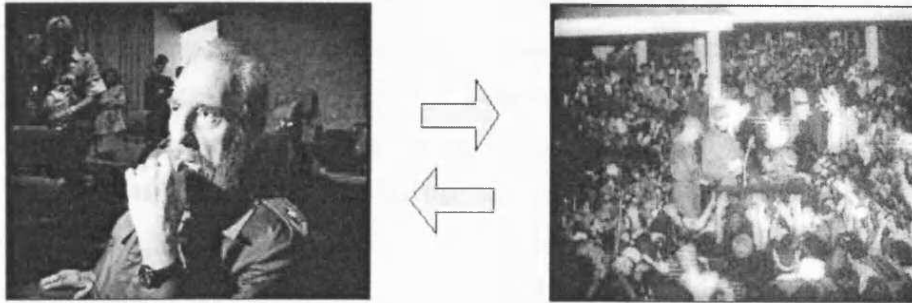
Figure 4. The cut between Castro and archive footage – an *extension* relation

However, sound plays a crucial part in this construction, rendering it more sophisticated than consideration of the visual editing alone might suggest. The memory process is actually represented as beginning several moments prior to the cut that establishes Castro as remembering the event. The sound of the crowd begins at the point that Stone is prompting Castro's memory, by imitating the chanting of the crowd. At POS 1, when we cut to Castro posed as if listening and/or thinking, a relation of *extension* is enacted between the presentational meanings of sound and image. Specifically, the cut to Castro extends the meaning of the crowd sound to mean that it is *imagined* sound – it is 'diegetically accessible' to Castro (Châteauvert, 1996: 141; Lacasse, 2000: 97). In turn, the sound extends the meaning of Castro's gestures and pose to mean 'in the process of recollecting'. (This can also be construed, however, as an elaboration relation between the presentational meanings of Castro's pose and those of the crowd sounds: Since his pose alone can be construed as 'thinking', the sound serves to *specify* that process as 'recollecting'). In Châteauvert's terms, the sound can be classed as a diegetic but 'absented voice'. In his words (1996: 144)

> The 'absented voice' corresponds to a discourse whose effective source is not, at the moment of the occurrence, present within the visualized diegesis, but which is nevertheless diegetic in that it is 'perceived' within a dream, a remembrance, or in some imaginary fashion, by the visualized character.

While this proposed status of the sound – as an object of Castro's imagination and

208

thus a logical effect of a visual source – is not reasoned on the basis of synchronisation's conventional deployment (as it is in the previous analysis) it can be reasoned on two other bases. The first is that Stone's prompting is construed as eliciting the memory from Castro. For this reason, the listener-viewer is poised for Castro's reaction to the prompting, which would expectedly involve some visible effort of recollection. The second is that non-diegetic sound is malleable: it can 'attach itself' to a source and enter the diegesis as an accessible sound. Think of the common prank in 'spoof' comedies of transforming what we assume is non-diegetic music into diegetic music: a character bumps into a record player and the music recording skips. The diegetic status of the sound is altered, and it is rendered accessible to the characters. While the present extract is not nearly as obvious, it too involves the transmutation of the sound's status from non-diegetic to diegetic (though, as I argue later, the sound is not 'fully' diegetic). I would argue that this is engendered by the *independent*, 'probing' camera movement (Mercer, 2002: 160; Bordwell and Thompson, 1993: 224) mentioned above – Castro is filmed in an 'interrogative' way, and coupled with his facial expressions (suggesting a mental process) it is as if the camera is intruding on a private moment. The sudden access that the listener-viewer is granted to this 'private moment' at POS 1 – the equivalent of the 'bump' in the example of spoof comedy above – encourages her to interpret that sound as an integral part of the diegesis, albeit an 'absented voice' (Châteauvert, 1996: 144). In Cook's terms (1998: 66), the sound 'jumps the diegetic gap', 'seeking out' the inner-thoughts of depicted characters.

Emergence: typological and topological semiosis
While it is clear that the memory process is established earlier than the visual cut to the footage, it cannot be argued that the extension relation between sound and image identified here at POS 1 is simply made 'at once'. Rather, it is the result of a dynamic relationship between each mode's presentational meanings, and central to this interaction is the *process* represented in the evolving character of the crowd sound. To explain, although the memory process is 'basically' constructed by the image – primarily by the timing of the cut and the content of the subsequent shot – the sound offers an extra semiotic dimension that enables the representation of

209

recollection as gradual and emergent. As Van Leeuwen (1999: 195) argues, quoting Wulf (1993: 10) 'while vision concentrates on the permanent and unchanging, hearing is particularly good at grasping 'the dynamics of things coming into being over time.' A distinction that Lemke (1998; 1999) makes between *typological* and *topological* semiosis is relevant here, for the sound is making meaning *by-degree* (a memory in the making) while the image determines it as meaning-by-kind (a memory is either recalled or not). Specifically, in coordination with the motion of the probing camera, sound contributes an *analogue* semiotic dimension to the digital visual construction of the process of recollection, rendering it 'emergent'. It not only 'anticipates' the moment of recollection, but constitutes that moment in an entirely different way: that is, as a process that involves uncertainty and effort.

In terms of logical relations, the relation (extension: *addition*) between the shots is typologically created (no visual dissolve). As represented in Figure 4, the process of memory *recall* is represented as a digital, not analogue action. However much Castro is represented as in the process of recollecting, the image-memory is represented as 'springing to mind'.

However, the relation between sound and image (extension: *addition*) that earlier establishes Castro as recollector is also topological: its continuous-dynamic audio-visual realisation ('developing' sound, Castro's evolving gestures, and interrogative camera) represents the process of recollection as a dynamic and *gradual* one. Precisely, the presentational meaning of the process of 'inaccessibility to accessibility' that can be construed in the sound track by means of experiential metaphor, logically extends the presentational meaning of the image (Castro's contemplation and the 'probing' camera) to signify that Castro's memory is 'surfacing', becoming accessible to him, much like the presentational meanings of the sound become accessible to the listener-viewer. Castro's process of recollection, then, is represented as emerging from uncertainty to a sudden point at which the memory is 'legitimated' and confirmed: it is a memory represented with the certainty and authority of archive footage.
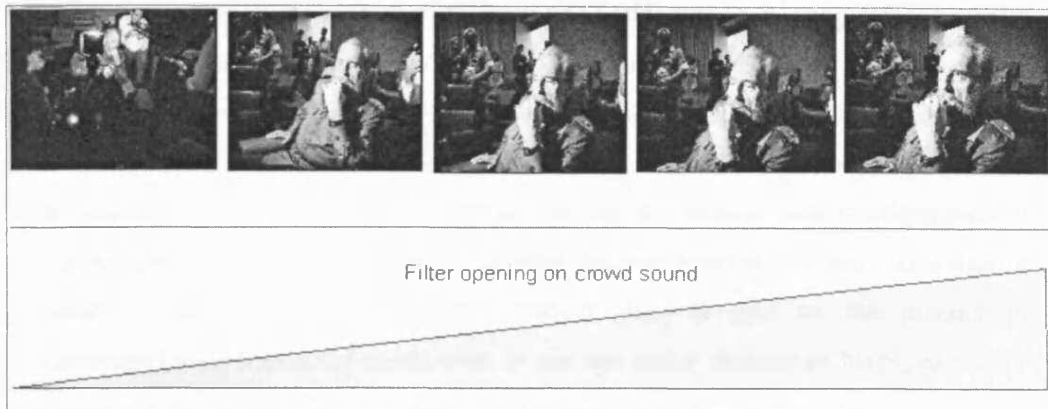
Figure 5: Topological interaction between sound and image. Audio-visual *extension*

While it is certainly true that it is not only the sound track, but also the *visual* resource of zooming, and the 'probing', medium close-ups of Castro, that represent a topological meaning (gradually remembering) the actual *moment* of recollection is visually represented as sudden, discrete and *certain*: the image is fully formed, replete in every phenomenal detail. Indeed, the sound does help to bolster this, by using the filter to discretely 'jump' a few steps to the upper-mid range on the cut, a move that redounds with the sudden appearance of the image. However, it is not the sound's presentational meanings that are suddenly 'appearing' at that point; rather, at that moment, an aspect of its represented relationship to the listener-viewer is *made certain*. The sound, then, with its low level of mid-range detail before the cut, and its gradual change from this to a more 'present' sonic character, construes the process represented in the synchronised image as one of uncertainty – the memory is coming into being by means of the 'work' that Castro is doing in his recollection.

Without this sonic addition to the total audio-visual meaning of the clip, the recollection *process* would be considerably different. As argued above, the visual construction is such that the 'memory' – in the appearance of the archive footage – is represented as 'springing to mind'. Considered with the deployment of the crowd sound's developing frequency characteristics – its topological dimension – the process is rendered as far less certain, and as far more work for Castro. The effect is similar to a visual dissolve: the image gradually 'comes into being' . The main

211

difference is that the combination of the sound and image in this clip affords a more sophisticated semiotic configuration; the memory process is represented as involving both the uncertainty of recollection *and* a definite moment of recall. While I do not want to claim here that this is a more *accurate* representation of human recollection processes, I will point out that this audio-visual configuration of topological and typological meaning affords the construction of Castro as a human *subject*. Crucially, representing his 'human' side is part of the project of *Commandante*. Indeed, the scene prior to the one under discussion involved Castro discussing his favourite movie stars, his views on comedy, and so on.

In this section, I have demonstrated that a resource of sound-to-picture is used to establish the beginning of Castro's recollection process as much earlier than with image alone. In addition, I have argued that the particular deployment of equalisation/filtering in this clip contributes a significant, *topological* dimension to the audio-visual construction, that serves to (1) represent the *process* of recollection; and (2) specify it as one of uncertainty and mental effort. In the next section, I argue that, the listener-viewer is unavoidably positioned as *complicit* in the memory process.

Orientational meaning: Equalisation/Filtering

As should be clear from the preceding discussion, the deployment of equalisation is especially attuned to making topological meanings of 'emergence' when construed in combination with the synchronised image. Also explained above is the functioning of the sound in isolation, and there we have seen that the frequency shaping of the crowd sound, by virtue of experiential metaphor, made the presentational meaning of 'distant to close'. In a functional view of semiosis, meanings are not made discretely, as, for example, *either* presentational *or* orientational ones. Rather, these terms are merely ways of typologising the social semiotic work performed in every representational act into different 'tasks': to present some state-of-affairs, to orient towards that presentation and towards potential audiences, and to organise itself into text (Lemke, 1995: 41).

The purpose of this theoretical 'reminder' is to provide some context for the present argument that the configuration of presentational and orientational meanings across sound and image, in this clip, function to establish Castro's emerging memory as *our* memory too. This is argued on the basis that the sound is not only oriented in a relationship with Castro – that is, as gradually emergent to him – but is also positioned in a dynamic relation to the listener-viewer: that is, from inaccessible to accessible *sound*.

On that view, the presentational meanings of sound and image – i.e. 'Castro recollects' – *co-occur* with the sound's orientational meanings of 'inaccessible to accessible' sound. Their co-occurrence reminds us that both presentational and orientational meanings are part of the same 'original' semiotic unity. It follows that the orientational meanings made in the sound track ('inaccessible to accessible') – when construed in synchrony with both the image's presentational ('probing') and orientational ('intimate') meanings, and indeed the sound's own presentational meanings ('emergence') – are inevitably enacted within the context of those other sound and image meanings. The consequence is that, in the context of the synchronised image, the orientational meanings of 'inaccessible to accessible sound' are attributed to Castro: the evolving sound is of his imagination and is thus developing from inaccessible to accessible sound for *him*. Because the sound is evolving for the listener-viewer too, she is oriented toward the evolving sound as if it were an *object* of Castro's emerging memory.

But even while the listener-viewer is simultaneously oriented to the sound as an *object* of Castro's process of recollection (that is, a diegetic sound), a direct/distinct orientational relationship between the listener-viewer and the emerging sound is maintained (that is, the sound is non-diegetic). This orientational complex establishes that the sound is construed as a mediating concept between listener-viewer and Castro – that the sound is to be understood as representing both Castro's memory and the 'collective memory' of the listener-viewer – thus establishing a proximity to Castro that is central to Stone's objectives in *Commandante*.

213

I propose that there are two interrelated explanations for the duality of orientations toward to sound. The first is concerned with the non-fiction genre; the second with the potential of sound as mode and medium.

(1) The sound cannot be fully integrated into the diegesis as an 'imagined' one because the context of non-fiction entails that:

> (a) Castro is understood to be an actual person, not a fictional character;

> (b) the crowd sound has an 'indexical' bond to the historical past, which is determined by both its sonic quality (which is culturally encoded as a manifestation of a shared memory resource, i.e. the archive), and by its being referenced by Stone as a historical event.

To expand, if this were fiction, and Castro a fictional 'character', we would construe the sound as that character's *private* memory, despite the sound actually being necessarily *public*. Any orientational meanings made through the sound event vis a vis the listener-viewer (e.g. 'emerging') would be snared by the overarching narrative functioning of the diegesis. The fictional context would background (automatise) the relations that the sound institutes between it and the listener-viewer, in favour of those it establishes between character and sound. Thus an imaginary social bond between listener-viewer and film character would be constructed *covertly*, based on their 'hearing' the same sound. The present clip's status as non-fiction, and the entailments of this, ensures that the sound cannot be *fully* integrated with the diegesis because the 'suspension of disbelief' on which fiction insists is not available as an interpretative resource for the listener-viewer. Instead, the sound remains ultimately non-diegetic, even when the audio-visual construction is asking us to consider it as Castro's memory. The non-fictional context of *Commandante* does not allow such 'complete' integration of the sound into the diegesis; rather, the sound's status as an object of Castro's 'memory' is only partially accorded.

(2) The direct address of sound-as-medium (cf. Kress and Van Leeuwen, 2001: 66; see Chapter 4 on 'direct address') ensures that the listener-viewer is positioned in relation to the sound *as* sound, not only as a signifier and 'object' of Castro's imagination. Because of this orientational stance toward the sound as both *distinct from* and *identical with* its constructed status as object of Castro's memory, the sound becomes a *mediating* object between listener-viewer and Castro, thus aligning Castro's 'memory' with the 'collective memory' represented by the archive.

Key to this argument is the notion that semiotic deployments of sound necessarily 'overflow' the functions of narrative, or whatever narrowly defined context they are mobilised to enact (cf. Lastra, 2000). This overflow can be considered in terms of the 'performativity' of sound-as-mode (in Austin's sense [1961]). That is, sound-as-mode's relationship to sound-as-medium is such that the latter's materiality ensures that the former cannot 'merely' *represent* its presentational and orientational content, but must necessarily also *perform* it. In Bruzzi's (2000: 154) formulation of Austin's theory (which she deploys in order to understand the recent 'performative' style of documentary representation), the 'performative' refers to 'utterances that simultaneously both describe and perform an action'. Where Bruzzi has 'description' we have 'representation', but nonetheless the concept of performativity is useful for illuminating the functional, semiotic 'overflow' of sound-to-picture: what Lastra has called its 'element of excess or [even] contradiction' (2000: 97) As he elaborates,

> Even while appearing to fulfill a particular function (such as providing intelligible, narratively important dialogue), sound may simultaneously be performing other, non-narrative or even non-representational ones. (2000: 97)

Deploying sound to represent the process of recollection involves 'enacting' the process *as sound*. For example, the narrative device of the 'absented voice' (Châteauvert, 1996: 144), in which the audience 'hears' the thoughts of a character, entails that the 'private' sound is necessarily made 'public', and thus the sound of the memory is construed as shared or common. In his discussion on sound design in theatre, Harold Burris-Meyer (1940: 348) reflects on an early deployment of a

similar semiotic device to that under consideration here, in which the audience hears the memories of a prisoner awaiting execution:

> The audience heard the auditory components of these episodes as he remembered them. Recurring through the reminiscences came snatches of the Judge's reading of the death sentence, which were progressively distorted and motivated the prisoner's finally flinging things about and shaking the bars. As the prisoner's mind distorted what he remembered, *so the audience heard it.* (my emphasis)

This 'making public' of private, 'individual' memories is a necessary part of fictional representation, but it is also, for Raphael Samuel, integral to the very act of recollection. As Silverstone (1999: 126) argues, for Samuel, memory

> is what is done in recollection, with or without tranquility, through oral testimony and shareable discourse. It is where the private threads of the past are woven into a public cloth, offering an alternative vision, an alternative reality to the official accounts of the academy or the archive.

In *Commandante*, however, Castro's represented memory *is* that of the archive. Through the orientational functioning of the audio-visual construction; the contextual determination of the sound's diegetic status; and by virtue of the direct address of the medium of sound, the 'official' account of the archive dominates, aligning listener-viewer and Castro as complicit in the memory. This mode of identification is such that we are not invited into Castro's private memories, as we would be in a similar resource in fiction film.

In this section, I have demonstrated that sound-as-mode's obviously inextricable relationship to sound-as-medium necessitates that, in this clip, sound cannot be *fully* mobilised to fulfill the audio-visual function of recontextualising the sound as Castro's memory. Rather, its materiality ensures that it maintains a *direct* relationship with the listener-viewer (Kress and Van Leeuwen, 2001: 66)

216

7.4. Conclusion

In both analyses, several issues arise that, broadly speaking, concern the semiotic functioning of sound-to-picture in the construction of: identification, subjectivity and authenticity. As I have suggested during the analyses, the contextual determination of the emergent meanings needs to be more comprehensively explored if we arè to gain deeper insight into how sound-to-picture *means* in non-fiction film. In the next chapter, I turn to examine the constraints that derive from the context of non-fiction on audio-visual semiosis, by way of a comparison of the functioning of sound across fiction and non-fiction contexts.

On initial engagement, the deployments of technical resources – sync-sound, panning, equalisation – appear to be used in both fiction and non-fiction film texts, and sometimes deployed to similar ends. For example, the audio-visual construction of Castro's memory process in *Commandante* is achieved by means of a conventional configuration of sound and image (a zooming close-up on a subject's face; a diegetic sound construed as 'internal') of the kind that might occur in fiction film. In that sense, the simple assertion can be made that 'both fiction and non-fiction film deploy audio-visual resources to construct *subjectivity*'. In itself, the presence of constructions of subjectivity in non-fiction is intriguing. As Nichols (1991: 159) argues, such meanings are, historically, rarely made in documentary film for the reason that 'the risk of subjectivity has lain its potential to colour or subvert objectivity, the prized goal for many'. But, writing in 1991, Nichols goes on to cite a number of contemporary documentary films that *do* use constructions of subjectivity as rhetorical means in making their arguments about the world. As he argues (1991: 156), the objective of 'psychological realism', so central to fiction filmmaking, need not detract from the goal of objectivity in documentary film. Instead, he claims, subjectivity and identification can be considered forms of 'emotional proof', and this evidential status renders them fully compatible with the indexical bond to historical reality that non-fiction claims.

Given Nichols' claims, it might be argued that identifying constructions of

subjectivity in documentary film, and explicating their audio-visual construction and the role of sound within it, are objectives hardly worth pursuing, for they do not reveal anything particular about the non-fiction genre with regard to the distinction from fiction film, and, by extension, the functioning of sound in that context. If this is coupled with the argument that representational techniques across fiction and non-fiction contexts are transgeneric, then no matter how 'intriguing' the issue of audio-visual constructions of subjectivity (or identification, or interior states, etc.) in non-fiction film, it is merely a trivial one. Indeed, if that were the end of the matter, all we would be asserting is that both non-fiction and fiction draw on the the same set of 'formal techniques' or 'techniques of representation' to fulfill their respective representational strategies. This is tantamount to claiming, as I have done throughout this thesis, that 'formal techniques' are transgeneric (cf. Carroll, 1996) across fiction and non-fiction filmmaking. The question arises: what would be the use in investigating the deployment of those techniques in the audio-visual construction of subjectivity in non-fiction film? The obvious response would be that there is little value in choosing non-fiction film sound-to-picture as an object of study because there is, on the basis of the arguments above, nothing *specific* about how sound is deployed in non-fiction film. (I am inclined to agree with this response, though I do recognise the value in *positive* descriptive accounts that aim to engage with the various modes of non-fiction filmmaking in terms of their deployments of sound-to-picture.)

As I have argued, the most profitable way forward in pursuing this line of investigation is to attempt to understand not merely the representational techniques of sound-to-picture that are deployed in the audio-visual semiosis of non-fiction film, but rather the different functions that such deployments are constrained to fulfill across fiction and non-fiction contexts. For example, it is to ask such questions as: what are the differences in meaning in the deployment of the 'absented voice' (Châteauvert, 1996: 141) in an audio-visual representation of recollection across fiction and non-fiction film? We may intuit that there is a remarkable semiotic difference in how such a construction is construed across contexts. The audio-visual construction of Castro's recollection process in *Commandante* is

218

recognisable to us as such precisely because such a device has evolved as a convention in *fiction* filmmaking. As such, it would appear unproblematic to claim that its deployment in *Commandante* relies on identification of the intertextual link to fiction film. But that link is not overdetermined: it does not 'pull out all the stops' by deploying reverberation on the crowd sound (as is conventionally the case in fiction film), nor by maintaining the sound as internal by *not* cutting to the archive footage. At this level of intertextual 'suggestion', the semiotic potential of the 'memory device' in non-fiction film is constrained by the overarching contextual frame of non-fiction. Understanding how that frame is established – that is, at the level of text, institution, viewer response, or a combination of these – is the task of the next chapter.

Chapter 8: Context and constraint – Sound-to-picture in non-fiction film

8.0   Introduction

In this chapter, I turn to discuss the analyses in terms of the relationship between the emergent meanings made with particular deployments of sound-to-picture and the contexts of fiction and non-fiction filmmaking. This entails a recontextualisation of the original research question; from 'What is peculiar about non-fiction film sound?' to 'How does non-fiction film *as context* constrain the meanings that can be made with sound-to-picture?' This recontextualisation of the original question shifts the theoretical focus from an attempt to understand a particular semiotic practice (non-fiction film sound) to an attempt to understand the context in which the practice of sound-to-picture operates. Such a shift represents an emergent analytic concern in the research: that differences in the roles that sound-to-picture performs in audio-visual semiosis cannot be comprehended in terms of *practice* only. Rather, if such differences can be discussed, it is on the basis of differences in contextual constraints.

In what follows, I explore the division between non-fiction and fiction in order to gain a better insight into the definitive features of non-fiction (here, documentary film and news) (Nichols, 1991; Carroll, 1997; Plantinga, 1997, 2005; Ward, 2005). These features constitute the *context of situation* of non-fiction film, in its presently delimited meaning, and as such hint at the active constraints on audio-visual semiosis and sound-to-picture in that context (cf. Halliday, 1978, 1985; Thibault, 1991; Baldry and Thibault, 2005). Then, I investigate an important emergent analytic theme of sound-to-picture – subjectivity/identification – in terms of the contextual constraints on its meaning. These constraints *emerge* from an insight into the relation between a textual instantiation (e.g. a particular use of sound-to-picture) and, what we call, 'its' context (cf. Lemke, 1995: 166). Therefore, a comparative approach to fiction and non-fiction sound-to-picture is adopted here based on the conviction that such an effort might illuminate the limits and affordances of what can be meant with non-fiction film sound. Finally, I look to the *context of culture* (Halliday, 1978; Thibault, 1991; Lemke, 1995, 1999) of non-

fiction filmmaking, in order to establish how the boundaries around audio-visual semiosis are maintained.

8.1 Non-fiction and fiction sound-to-picture: context and constraint

Although Carroll's argument concerning the transgenericity of 'formal techniques' across fiction and non-fiction has been, rather unproblematically, adopted as a heuristic in the preceding chapters, the aim of the present chapter entails that Carroll's thesis is revisited critically. There is firstly a need to define what Carroll refers to as 'formal technique'. What exactly are the limits of his definition? If a technique is a 'way of doing', then *what* is it a 'way of doing'? Here, I assume that by formal technique, Carroll means the myriad textual resources that have evolved within the general context of filmmaking: that is, techniques for *representing* the world. On that basis it is hard to disagree with Carroll, because it is extremely easy to call to mind many instances of non-fiction film that 'dramatise' the events they represent. Kevin Macdonald's *Touching the Void* and *One Day in September*, are both prime examples: the first an unqualified reconstruction, using actors to 'illustrate' the happenings; the second an unashamed deployment of archive footage, computer graphics and the recognisable voice of the Hollywood actor, Michael Douglas, to reconstruct the actions of the extreme Palestinian group 'Black September' at the Munich Olympics, 1972 as a dramatic narrative. It also takes little effort to recall the many fiction films that deploy obvious signifiers of 'documentary realism' (Nichols, 1991: 184) and 'authenticity' in telling their stories. For example, Steven Soderbergh's *Traffic* (2001) deploys shaky camera work and a *mono* sound track – a conscious choice to highlight the '"documentary" feel of the movie' (Sergi, 2004: 147). His more recent *Bubble* (2005) straddles the fiction/non-fiction divide by constructing a fictional plot that is developed from interviews with certain non-actors, and then using those very people as central actors in the film.

But while a surfeit of representational techniques is evidently common to both genres, it is argued here that formal techniques for constructing a subject position or agent are ultimately deployed to fulfill different functions across fiction and non-fiction contexts. For instance, I concur with Nichols (1991: 156) that audio-visual constructions of subjectivity and identification are 'far less frequently

221

explored in documentary than in fiction'. For Nichols, 'issues of objectivity, ethics, and ideology have become the hallmark of documentary debate as issues of subjectivity, identification and gender have of narrative fiction' (1991: 156). Yet, while this point is persuasive, it should not be forgotten, as Nichols suggests, that constructions of subjectivity and identification *do* occur in non-fiction. Indeed, much contemporary non-fiction filmmaking does not diametrically oppose the notions of 'objectivity' (the hallmark of non-fiction) and 'subjectivity' or 'identification', and this social semiotic stance – as I have demonstrated in the previous chapter, and discuss later – is represented in its practice of sound-to-picture.

As I argue below, the emergent themes in the analyses of sound-to-picture in Chapters 6 and 7 exhibit drastically different semiotic potentials across fiction and non-fiction contexts. As part of the 'constituency structure' of audio-visual film, the semiotic potential of sound in the audio-visual construction of such themes betrays indications of contextual constraint. In large part, this is inevitable: if Carroll is correct in his argument, then it suggests that decisions over what is to count as fiction and what as non-fiction must be founded on relations to some higher-order level than that of text. On that view, it follows that there are strong *contextual* constraints on the audio-visual semiosis of fiction and non-fiction film, and if we are to better understand how meanings are made with sound-to-picture in those specific contexts we must attempt to understand the points of engagement between the constraining features of those contexts and the audio-visual meanings that are made and construed in relation to such contexts.

8.1.1   The great divide?

How then is the context of non-fiction film to be described? What are its constraining features? In what ways does the genre of non-fiction film determine the meanings that can be made and construed with and within it? In this section I aim to characterise non-fiction filmmaking in terms of its salient differences to fiction filmmaking. These are at the core of how meanings are made and construed in each context, and my objective is to understand how those contexts constrain the

semiotic potential of synchronised sounds and images.

The category of non-fiction film, however, is itself in need of basic clarification. Specifically, non-fiction as a genre encompasses all films that are not considered as fictional. These include, for instance, public information films, corporate films, advertisements, news, game shows – and so on (Plantinga, 2005: 105). I do not want to pursue explication of the category of non-fiction in terms of *all* its possible members. Rather, I will attend to just two types: the documentary film and television news. From here on, I use the term non-fiction to refer exclusively to documentary film and TV news.

While broadcast news as a genre is arguably unambiguous (cf. Stephens, 2006; Tumber, 1999), what is meant by 'documentary film' is notoriously contested by theoreticians and practitioners alike. Here is Ward (2005: 6) concisely summarising the most famous attempts at defining 'documentary', and defining what he claims is the 'central issue' in all debates concerning documentary film:

> John Grierson's famous dictum, 'the creative interpretation of actuality', was one of the first [attempts at definition]. John Corner refers to the 'art of record' (1996); Brian Winston sums it up as 'claiming the real' (1995). In all of these attempts to adequately capture the meaning of documentary, there is the same dilemma: how to deal with and understand something that clearly is attempting to represent *reality* (or some part of reality), but as it does so, uses specific *aesthetic* devices. A commonsense suggestion is that the aesthetics somehow *distort* or *change* the reality being represented. This central issue has troubled documentary filmmakers and theorists (the latter more than the former, it has to be said) and has arguably had a debilitating effect on understanding documentaries.

As might be clear, this 'central issue' turns on the assumption that certain practices are proper to fiction filmmaking (aesthetic engagement) and others to non-fiction filmmaking (representing reality). Accordingly, claims abound of 'fictionalising' material through the editing process, or by 'dramatising' events through so-called *reconstructions*, in discussions concerning the 'proper' truth-value to be accorded to such moments in non-fiction texts (cf. Ward, 2005: 34-5, for discussion). As Bennet (2000) argues, reconstructions of actual events in non-fiction film, when

explicitly acknowledged as such, 'do not deceive', but rather 'they short-change us, deal in a currency inferior to truth'. As Ward (2005: 34) counters, the use of reconstruction is far more complex than Bennett suggests in her conviction that all reconstruction is 'essentially fictional'. Adopting such a position misrepresents the functional role of reconstructions in non-fiction, because of an obvious yet fundamental difference between fiction and non-fiction: whereas in fiction film 'the events and persons are precisely that – fictional or made up', in the non-fiction film 'the events and persons depicted exist (or did exist) in the real world of actuality.' (ibid.) Accordingly, a different relationship between representation and represented exists in non-fiction than in fiction film, and it is this basic fact that is central to distinguishing such films. As Nichols has put it, non-fiction requires 'awareness of an antecedent reality before it can come into being as a specific form of representation. [This] involves a tension between the representation and the represented as experienced by the viewer. Remove this tension, enter a realm of aesthetic engagement, and the specific qualities and questions of documentary no longer apply.' (1991: 232; original emphases removed)

This last statement hints at a central facet of all attempts at defining non-fiction and documentary film. Common to the arguments of writers such as Bill Nichols, Carl Plantinga, Dai Vaughan and Noël Carroll is precisely the indexical link to the historical world that documentary film and TV news claim as their distinguishing feature from the 'distracting shadow-play' of fiction filmmaking (Nichols, 1991: 4). As Nichols (1991: 5) explains, with characteristic insight:

If we consider the imaginary realm of fiction as having a metaphoric relation to history and lived experience – as a kind of carefully shaped, translucent cloud that displays contours and shapes, patterns and practices that closely resemble the ones we encounter in our own lives, we might think of documentary as a mode where this fictive cloud has settled back to earth. The elevation provided by the metaphor, the sense of remove, is drained away as special properties of photographic film and magnetic tape hold the documentary image to the exact shapes and contours, patterns and practices, of the historical world. [...] We are less engaged by fictional characters and their destiny than by social actors and destiny itself (or social praxis). We prepare ourselves not to comprehend a story but to grasp an argument. We do so in relation to sounds and images that retain a distinct

bond to the world we all share.

While this view reveals much about our *response* to the documentary film, it raises a crucial question: *how* is such a response cued? That is, what cues the expectation that a 'distinct bond' to our shared historical world obtains between the sounds and images of a film, and those we perceive in the actual world? Certainly such an expectation cannot be cued only by the 'special properties' of the sound and photographic media involved, since the same media and the same procedures are deployed in recording fictional scenes. While it is true that the phenomenal specificity of a filmed scene, particularly with synchronous sound, produces a sense of 'being there', of immediacy, that obtains whether we are experiencing a fiction or non-fiction film (cf. Plantinga, 2005), our construal of a 'distinct bond' between, on the one hand, sounds and images and, on the other, historical reality is determined by a relation to context – in this case, the 'situation' of non-fiction film.

I will return to the question of *how* the appropriate expectations and responses are generated – that is, how a text is *selectively contextualised* (Lemke, 1998b) – later in the chapter. First, though, I offer a description of the most salient dimensions of non-fiction filmmaking by way of discussing the various attempts at defining non-fiction film. This identification can then aid us in understanding the contextual constraints of non-fiction film on the semiotic potential of sound-to-picture and audio-visual semiosis.

8.1.2   Features of the non-fiction film
Indexicality

The basic tenet of all attempted explanations and descriptions of news and documentary film is that such films claim an indexical bond to the historical world. However, the theories differ with regards to how that link is to be described. An argument based on the indexicality of the photograph and moving image – which Plantinga (2005) terms 'documentary as indexical record' (DIR) – determines that link as both a material and semiotic one. In its most viable form, DIR defines

documentary as films that comprise 'moving photographical images that are indexical records or traces of the pro-filmic scene(s)' (Plantinga, 2005: 106). As Currie elaborates, the images that constitute an 'ideal' documentary 'may not represent things and events other than the things and events they are traces of' (1999: 291). By insisting on the 'identity' of material and semiotic dimensions of documentary filmic representation, Currie's formulation avoids a criticism commonly levelled against definitions of non-fiction based on photographic indexicality: that documentary cannot be defined in terms of an indexical link to the physical world because *fiction* film also claims such a link. As Nichols (1991: 28) puts it,

> The [documentary film] text presents a metonymic representation of the world as we know it (the sounds and images bear a relation of part to whole; they partake of the same order of reality as that to which they refer) rather than a metaphorical rendering (where the images and sounds operate on a separate and distinct plane of resemblance to the historical world).

Adopting a Piercean notion of indexicality – there is a causal relation between an indexical sign and its referent – DIR accords the photographic medium a privileged, evidentiary status that is then extended to the moving images of the film medium. Disregarding for the moment the obvious representational work accomplished in the procedures of framing, lighting and so on, the 'causal processes' involved in taking a photograph (or the individual shot in a given film) – that is, its 'cause-and-effect operations performed in and through a machine' – invest photographic media with 'a veracity that we do not allow for a painting' (2005: 106). And although this indexical potential has been overstated by proponents of cinema verité and certain theorists, Plantinga (ibid.) argues that 'it is nonetheless undeniable that the documentary has relied on the power of the moving photograph to "show us the world," and to do so with an authenticity that depends not only on the visual wealth and detail of the photograph, but also on the indexical, causal bond between photograph and pro-filmic scene.'

However, the status of the photograph-as-trace argument in the DIR definition rests on a rather obvious yet fundamental flaw. Whether or not we accept

226

that photographs (and individual documentary-film shots) are traces, the fact that documentary films are edited assemblies of moving images means that a given pro-filmic event undergoes interpretation ('selective contextualisation', in Lemke's [1998b] terms) and its representation 'involves intentionality in a way that indexical signs do not' (2005: 106). Just as obviously, further mediation occurs in the addition of narration, music, text and so on. On that basis, documentary film 'might be considered a trace only under conditions that very few, if any, documentaries ever meet.' (2005: 107) Furthermore, because of its insistence on limiting documentary representation to the persons and actions that can constitute a photographic trace, DIR cannot account for reenactments or reconstructions of historical events. Such representations clearly do not 'represent what they are photographs of' (2005: 109). As Plantinga points out, the same argument can be made of expository, historical documentaries concerning periods when photography was not yet invented, thus precluding documentary films about Napoleon, for instance (ibid.)

Yet the central purpose of this part of the discussion is not to decide upon a singular definition of documentary or non-fiction film, but rather to survey the various attempts in order to construct a set of 'properties' that can be used in describing non-fiction as context. For that reason, the *indexicality* of documentary and other non-fiction film, such as TV news, will be understood as a legitimate and specifiable dimension of the context of situation of non-fiction film. As Plantinga concludes in his own account of DIR, it is perhaps most usefully formulated as one component in a theory of the documentary filmmaking. On that basis, documentary film can be defined, on the DIR model, as '*a sustained discourse of narrative, categorical, rhetorical, or other form that makes use of moving or still photographic images predominantly as traces to represent what the photographic images are of.*' (2005: 107; original emphases)

Of course, absent in Plantinga's, and Currie's, formulation is any discussion on sound. It must be assumed here that the same argument concerning image-as-trace should be applicable to sound recording and editing. That is, recorded sounds are both material traces and semiotic (cf. Williams, 1980). Additionally, the qualifications of Plantinga's re-formulation concerning the editorial recontextualisation of photographic traces into the broader representation of the

documentary film (as argument, narrative, etc.) must also be applied to sound as
deployed in such filmic contexts.

Assertion

Related to the *indexicality* argument is the notion of documentary as *assertion*
(DA). In Plantinga's words, for proponents of the DA definition, *'documentaries are
moving picture texts in or through which filmmakers assert that the states of affairs
represented in the work hold in the actual world'* (2005: 108; original emphases).
The main difference between the DIR and DA attempts at a definition of non-
fiction is that, for DA, a documentary film *may* deploy photographic images and
sound recordings to directly refer to the profilmic scene, but it also may present
filmed *re-enactments* of historical events. Clearly, in the latter case, were we to
construe what we believed was an intended indexical link to the pro-filmic scene
we would *misconstrue* the relationship of the filmed events to the historical world.
Rather, the *assertion* approach to non-fiction claims that whether a re-enactment or
'original' event is represented, both representations are to be construed as asserting
that the events they audio-visually represent *actually occurred*. In other words, a
non-fiction film text takes an *assertive stance* towards its images, sounds and
narration, and the states of affairs represented by virtue of them. This is in contrast
to a *fictive* stance towards 'the world of the work' (Plantinga, 2005: 107), which
involves 'presenting the state of affairs for our delectation, edification, education,
amusement, or what have you, but not to have us believe that the state of affairs that
constitute the world of the work holds in the actual world.' (2005: 107) In Currie's
(2004: 337) words, fictions

> require us to imagine occurrences or states of affairs: things we might
> describe by means of a sentence. [With fiction] I can imagine that it is
> raining when it is not, or that 2 + 2 = 5, which is not merely false but
> impossible, or that a pipesmoking detective called 'Holmes' lives in Baker
> Street, when I know that there is no such person.

The distinction between a fiction and non-fiction film, then, is characterised in this
argument in terms of the *illocutionary* speech act of assertion. However, whereas

with DIR we can recover this assertion simply by construing the images and sounds as traces of the pro-filmic world, with DA the presence of such an assertive stance cannot be reasoned in the same (rather circular) manner. DA does, however, imply that the assertive stance of non-fiction film is constructed *within* a given focal text, by claiming that the illocutionary act is represented *through* the work, as Plantinga explains:

> When a writer or filmmaker takes an assertive stance toward the world projected through the work, he or she asserts that the state of affairs making up that projected world holds or occurs in the actual world. When filmmakers take such a stance [...] they 'make assertions about the actual world' through the work. When a text is recognised as a documentary, this 'mobilises relevant expectations on the part of the audience'. (Plantinga, 2005: 107)

What is important here is that the assertive stance is implied as being manifested through the textual forms and strategies of the 'work'. If that is true, then two questions arise. Firstly, how are the appropriate expectations cued by a given non-fiction film text? That is, what semiotic forms do such cues assume? (I address this is in more detail in section 8.1.3 below.) The second, related, question – and one I address in this section – is: how are the distinct materialities of the media of moving image and sound, that on this view serve as a conduit for the communication of a non-fiction film's assertive stance, drawn into serving that very function?

All accounts of non-fiction film that are aligned with the DA perspective 'go beyond the formal elements of films to distinguish between fiction and nonfiction on the basis of the *illocutionary act* performed through or with the work.' (2005: 108) Moreover, they all involve the intentions of the filmmaker to some degree; it is 'the *intentional* assertion of propositional content' which is at the heart of the *documentary as assertion* definition (2005: 111). Ponech (1999), for example, formulates a definition of the non-fiction film as

> one in which its makers 'openly signal their intention that viewers take the attitude of belief toward' the states of affairs presented in it. His is an intentionalist theory, one that locates the essence of nonfiction film in the intentions of the filmmaker(s). Ponech writes that those intentions are

229

discoverable in the plans the filmmaker develops in making the work, plans
that become manifest in the finished film. (Plantinga, 2005: 108)

Carroll, too, formulates a variation on the DA account – the 'intention-response'
model – that includes the filmmaker's intentions for how the audience is to respond
to the material (1997: 173-202). I mention intentionality here only because it helps
to throw into relief an aspect of non-fiction film as audio-visual text that is critical
to the present discussion. That is, regardless of the *intended* assertive stance
towards the propositional content of a film, the sounds and images that comprise it
always necessarily exceed their purported function (as claimed in the DA
definition) as evidence for an overarching argument and narrative; that is, as
semiotic potential 'harnessed to the broader argumentative strategy of the
filmmakers' (Plantinga, 2005: 110). As Plantinga argues, DA accounts of non-
fiction filmmaking are good at capturing the assertive stance that he believes is
fundamental to all non-fiction films, but they are perhaps better suited to
distinguishing prose fiction from prose non-fiction, since 'the assertion of
propositions and/or the assertive stance are well suited to linguistic discourse'. In
contrast to language, the 'peculiar nature of the photographic and sonic [...]
discourse' entails that a view of *documentary as assertion* cannot fully account for
the meanings that are made with sonic and visual media. In Nichols' (1991: xiii)
words:

> Film in general and documentary in particular does not [...] have that
> convenient level of abstraction available to spoken and written language,
> where words ('hope', 'umbrella', 'garden') leave the provision of a specific
> referent to the imagination. Film signifiers come with images attached. They
> *are* images, and sounds, and they are always concrete, material, and
> specific. What films have to say about the enduring human condition or
> about the pressing issues of the day can never be separated from *how* they
> say it, how this saying moves and affects us, how we engage with a work,
> not with a theory of it.

Thus, the sound track may make assertions that are either not consonant, or
relevant, to an overarching argument, or that cannot be *reduced to* the assertion of
propositional content that may be expressible linguistically. In short, the semiosic
functioning of the sound track (and image track) always overflows the central

230

functions of documentary film's *assertive stance*.

In the context of fiction filmmaking, the sound track is often constructed in such a way as to minimise any functional excess in respect of a film's overarching logic – that is, narrative – most commonly by adhering to the principle of narrative intelligibility, and the semiotic procedures that are conventionally undertaken to serve that principle. These include, as we have seen in Chapter 4, the spatial hierarchisation of the image by means of sound-track construction – especially the practice of creating complex points of audition through maintaining the directionality of dialogue as *towards* the listener-viewer. Such practices diminish the possibility for meanings to attach to the images and sounds that are not considered necessary to the narrative, by attempting to diminish material aspects of the pro-filmic scene, such as represented space and the subtleties of bodily motion within it, through the elimination of the spatial signification of central characters' voices. However, such a practice does not ensure that narrative excess is fully curbed. Rather, the diminishment of functional overflow is achieved in collaboration with the expectations generated through the context of fiction filmmaking. Even in fiction film sound-to-picture practice that does not value intelligibility above the phenomenal specificity of a scene – direct sound, for example – the potential excess of meaning construable in the sound track is conventionally 'automatised' (backgrounded) by the context of fiction film practice. That is, the absence of an assertive stance in fiction film, a lack which enables a metaphorical relationship between the audio-visual content and the actual world to be construed, entails that any potential narrative excess in the sound track is drawn by the centripetal pull of the narrative 'into the woof and warp of the story' (Nichols, 1991: 181). Any overflow thus becomes 'one more signifying element, more or less well motivated in relation to the plot'. (ibid.)

Saying and Showing

The above discussion suggests that the definitions of documentary film's indexicality and assertive stance would be of broader utility if they were integrated, and this is precisely what Plantinga sets out to do. I will not fully rehearse his attempt here, however, as it is only marginally relevant to this chapter's aims. It is

useful, however, to present a distinction Plantinga makes so that he may attend to the use of sounds and images both in their capacity to fulfill the assertorial functions of a non-fiction film and in their tendency to exceed these functions. Thus a distinction is made between *saying* and *showing*:

> *Saying*, in the context of a documentary, characteristically involves the assertion of specific propositional content. It is something like making an assertion or assertions about the representee, saying that it is thus and so. *Showing*, on the other hand, is something like standing in for the representee and may not involve the assertion of specific propositional content. For example, showing a person a series of snapshots taken of an event need not commit the shower to an assertion of the propositional content of the photographs. (Plantinga, 2005: 111)

Most documentaries, Plantinga goes on to argue, 'combine saying and showing and do so in different proportions depending on the type of documentary'. This view of the semiotic potential of sound and image neatly captures the complex way in which the media of sound and image are deployed to make meaning in combination. They are not merely harnessed to perform semiotic functions or 'duties' in respect of an overarching documentary assertion because not all such functions are capable of subsuming the materialities of sound and image – hence the terms 'overflow' and 'excess'. Sound, for example, can be deployed as evidence for an 'argument' (in Nichols' [1991: 111] sense that all documentary films' 'representations and propositions, tacit or explicit, aim at the historical world directly') that a documentary film presents. But sound can also make assertions independently of what is construed as the organising, argumentational logic of a non-fiction film. For instance, as Plantinga recognises, 'it may be that certain images and sounds, or sequences thereof, are meant to approximate some element of the phenomenological experience of the event, such as how it looked or sounded from a particular vantage point' (2005: 110). In this way, an assertion can be made, intentionally or not, solely with sound or image (that is, without linguistic proposition, or the assertorial potential of visual editing): an audio-visual configuration 'might be taken to be asserting that a scene shows what the pro-filmic event looked like, or approximates how the filmmakers "were appeared to"' (2005: 111). However, this is not an assertion in the same sense as it is used above,

232

because as Plantinga argues: 'what is being asserted about the propositional content of the photographs [or sounds] is underdetermined; in some cases, the propositional content is unspecified and in many cases it is unclear.' (ibid.) *Showing*, then, is more generous to the listener-viewer's interpretive actions, in that it allows for the 'apprehension of phenomenological qualities on the part of the spectator', allowing more to the "viewer's share"' (2005: 115; cf. Batcho, 2005 on *representative sound*).

We can make a (preliminary) connection here between, on the one hand, *showing* and *saying* and, on the other, *topological* and *typological* semiosis. As I have argued in the previous chapter, whereas topological semiosis affords the communication of meaning-by-degree, and thus can account for the very fine detail of a sound track or photographic image, or the fine-motor movements of a represented actor, typological semiosis 'reduces matters of degree to matters of kind' (Lemke, 2002: 322). *Saying* can be verbalised or linguistically represented. *Showing* must be apprehended in order for meaning to be construed by virtue of it. Making meaning-by-degree is clearly able to complement the making of meaning-by-kind (as demonstrated recently in O'Halloran's [2005] work on mathematical discourse). Yet, the potential for functional excess – for example, beyond the overarching function of argumentation in a non-fiction film – can also work to counter and subvert meanings that are made typologically (as Barthes argues, in relation to the 'obtuse' meaning; see Chapter 4). It is the topological propensity of early sound-to-picture that the mandate of 'narrative integration' of live sound and pre-recorded image was created to diminish (cf. Lastra, 2000; see Chapter 4). The sound track thus became structured hierarchically in terms of discrete planes of narrative 'interest' (the foreground, comprising dialogue, being the most privileged) rather than in terms of perceptual fidelity or some other more 'delicate' spatial composition that would make meanings that overflowed the functions of efficient story-telling. As Lemke (2002: 322) argues, the reduction of matters-of-degree to matters-of-kind – of topology to typology – is frequently used in the dichotomisation 'of categories (masculine/feminine, gay/straight, capitalist/communist, heroes/terrorists) through which sentiments and allegiances can be more easily manipulated.' Topological semiosis, on the other hand – and by

extension, *showing* in contrast to *saying* – 'leaves more to the "viewer's share"' (Plantinga, 2005: 115), and enables the signification of the 'shades of grey' that, to my mind, have the potential to better approximate human experience and the materiality of the world that, in typological semiosis, escapes representation. It is this aspect of topological semiosis that provides audio-visual film – fiction or non-fiction – with a *specificity* that cannot be fully subsumed by the *typological* semiosis of either a fictional narrative or an argument about the actual world. To reiterate Nichols' argument: in contrast to the abstracting potential of language, 'film signifiers come with images attached. They *are* images, and sounds, and they are always concrete, material, and specific.' (1991: xiii) The materiality of the world which is being represented audio-visually can only be *approximated* through representation – it cannot be fully incorporated or subsumed to an overarching assertive stance of a non-fiction film. In Nichols' words (1991: 110):

> The world is not only where information circulates but also matter and energy. These physical forces can be unleashed for or against us by discourse, linguistics, or even more directly, by nature. Whatever else we may say about the constructed, mediated, semiotic nature of the world in which we live, we must also say that it exceeds such representations.

Plantinga's claim that depending on the type of documentary film – and by extension TV news – saying and showing are combined 'in different proportions' is important here, as it hints at the notion of the expectations elicited by different film types. Plantinga relates his theory to two of Nichols' 'documentary modes': the expository film and the observational film. Specifically, an expository film requires less *showing* from its sounds and images, and more *saying* on behalf of its narration and visual editing. Therefore, the sounds and images used are not to serve in (solely) an evidentiary capacity – in the indexical sense that they represent only what they are traces of – but rather as illustrations of an overarching argument that is most commonly expounded through verbal and/or graphic narration. In observational films, on the other hand, sounds and images are used in *solely* an evidentiary capacity: they depict and sonically represent *indexically*. The assertive stance of observational films, then, relies far more heavily on the perceived status of its sounds and images as indexical records of a pro-filmic event than does the

234

expository documentary.

In sum, and following Plantinga, we can understand documentary films –
and I will extend the argument to TV news also – in the following terms:

> an extended treatment of a subject in one of the moving-image media, most
> often in narrative, rhetorical, categorical, or associative form, in which the
> film's makers openly signal their intention that the audience (1) take an
> attitude of belief toward relevant propositional content (the 'saying' part), (2)
> take the images, sounds, and combinations thereof as reliable sources for the
> formation of beliefs about the film's subject and, in some cases, (3) take
> relevant shots, recorded sounds, and/or scenes as phenomenological
> approximations of the look, sound, and/or some other sense or feel of the
> pro-filmic event (the 'showing' part). (2005: 114-5)

With TV news, the audience also expects that the images, sounds and propositions
represented in a given news story are in a relation of temporal proximity to the
moment of broadcast (and thus, reception). Indeed, this aspect of non-fiction film
can be characterised as *immediacy*. This is in addition to the assertive stance of
documentary film for the reason that it is not always clear or made clear in
documentary *when* a pro-filmic event has occurred. In TV news, however, the very
value and social function of news – as 'newsworthiness' – has evolved around the
notion of (relative) immediacy (see, for example, Mitchell Stephens, 2006).

8.1.3 Non-fiction film as 'situation'

In this section, non-fiction film is described as a context of situation (Halliday,
1978) in order to generate a formal-descriptive model of the relationship between
audio-visual semiosis and non-fiction film.

Documentary film and TV news, as *situation type*, may be described in terms of
field, tenor and mode values. In order to accomplish this, we must generalise the
situational values that are to be attributed to non-fiction film. These will encompass
the main types of documentary, as identified by Nichols (1991): expository,
interactive, observational and reflexive. The description of field, tenor and mode
values of documentary film, here, will be more readily applicable to *prototypical*
(cf. Plantinga, 1997) documentary films rather than marginal texts, but it does, I
believe, attend to *all* non-fiction at a fundamental level. The features of non-fiction

film expounded in the section above, and engagement with a small representative sample of non-fiction films (spanning the range of types: expository, observational, interactive and reflexive; see Appendix 1), will serve as a basis for the following description of non-fiction as situation type.

Field

The field denotes the social action or activity – that is, what is '"going on" and has recognisable meaning in the social system' (1978: 142). Subject matter is a second-order 'field of discourse' (1978: 144): it constitutes a 'type of social action that is itself defined by language'. In other words, a social action constitutes a first-order field of discourse (say, the social activity of narrating a fictional story) and the subject matter, a second-order field (that is, the social actions that unfold in the diegesis of the story).

The field values of a situation are enacted by and constrain semiosis in terms of presentational meaning. That is, the social activity type that constitutes the field of a context of situation is a key determinant of what meanings can be made in respect of describing and classifying experience, of making propositions, and so on. In Halliday's (1978: 143) words:

> The selection of options in experiential systems – that is, in transitivity, in the classes of things (objects, persons, events etc.), in quality, quantity, time, place and so on – tends to be determined by the nature of the activity: what socially recognised action the participants are engaged in, in which the exchange of meanings has a part.

In the field of non-fiction film, at a first-order level, there is the activity of *exposition*: in Nichols' (1991: 5) words, listener-viewers prepare themselves 'not to comprehend a story but to grasp an argument.' Thus the social action is not a 'narration' but a structured exposition that relates directly, not metaphorically, to the actual world. At the second-order level, the subject matter of documentary film is restricted to social actions, events, persons, relationships and so on that obtain historically. Therefore, the kinds of states of affairs that can be presented are strongly constrained by this fundamental characteristic of documentary filmmaking, to the extent that representations of phenomena that would be implausible in the

actual world – of the kind that characterise science-fiction, for instance – are considered transgressive. If they are deployed, they threaten to disrupt the 'correct' classification of a film as non-fiction.

Where the states-of-affairs proposed *are* plausibly of the same order of reality as the actual world (even those documentaries that take the future as their subject matter), the kinds of happenings, the perceived development between scenes, and the relationships of represented participants are of a different order to fiction, in which the development of character(s) within the narrative complex of plot and theme is usually central. In non-fiction, particularly in news but also in documentary, listener-viewers' engage with 'less expectation that a sustained identification with well-developed characters will follow. [...] [D]ocumentary most often draws our attention to an issue, concept, or problem that is at the center of the film's argument.' (Nichols, 1991: 29) Additionally, in fiction film, the kinds of participant types and relations represented are constrained according to a logic of narrative redundancy: for example, a variety of villains that are separately related to a given plot are extremely uncommon. In non-fiction, the participant types are represented according to a logic of argument (especially in expository film) or a criterion of circumstance (particularly, observational film): that is, a greater variety of groupings of participants, and a wider range of relations, are possible in non-fiction.

Finally, the field of non-fiction – and fiction – includes what Nichols terms 'excess'. This is inclusive of the phenomenal excess at stake in the indexical relations between sounds and images and the material, spatio-temporally specific moment that they are traces of. But, crucially, excess also denotes a fundamental feature of all non-fiction filmmaking: the excess of history (1991: 149). In fiction film, excess includes everything 'that cannot be absorbed within a theory of narrative comprehension' (Nichols, 1991: 141); in Bordwell's (1985: 53) words, 'causal lines, colours, expressions, and textures become "fellow travellers" of the story'. If we remove the 'analytic scheme' of fictional narrative, then excess is 'the noise that remains when we agree upon limits for what will pass as information' (Nichols, 1991: 141). And if what we term excess is 'that which remains', or that which is beyond fictional narrative, then in non-fiction excess is 'that which stands

beyond the reach of both narrative and exposition. [...] It stands outside the web of significance spun to capture it.' (1991: 142) Yet, the excess of history (as the referent of documentary filmmaking) and audio-visual historical evidence (as the means of representing it) constitutes a significant part of the field of non-fiction film discourse. In Nichols' words,

> Material, historical evidence exceeds all strategies of containment. Even more than fiction, where the text may motivate virtually everything that we see and hear (motivate in the sense of provide justification for its presence), documentary must constantly bear the burden of historical excess itself. It also must bear the burden, and glory, of the compelling quality of this historical evidence. Those signs or representations we see and hear of the historical world bear an indexical relation to their referent. Sound and image appear to reproduce the uniqueness of historical situations and events [...] (ibid.)

One objection here might be that since excess is that which is *beyond* representation, it cannot be identified as part of the field of the situation. In fiction filmmaking, the centripetal force of narrative renders excessive properties as automatised: the foregrounding of the essentials of narrative – plot, character and continuity – determine clearly what is to be considered 'noise' and what is not. In that domain it is accurate to characterise the aspects that are considered as 'overflow' as part of a distinctly different order of field to those that are necessary to narrative. As we saw in Chapter 4, such excess is closely aligned with the topographical mode of sound-image interaction; it is, in Barthes' terms (1977: 52-3), 'obtuse': that which overflows the denotative and connotative levels of meaning. Topographical modes of audio-visual representation threaten to disrupt fictional narrative because they can distract from, mask or subvert those filmic aspects crucial to narrative continuity. Attending to the 'grain' of each filmic moment, as Barthes' argues, 'structures the film differently.' (1977: 64) In non-fiction, however, the possibility for an indexical link between sound-image and material world to obtain as part of the representational (mimetic) plane (Hodge and Kress, 1988: 5) entails that (1) the excess of history and audio-visual historical evidence – key referents of non-fiction film – and (2) the argument and assertive stance of non-fiction, are of the *same order of field*: they are part of the same representational

plane, no matter that such excess is not necessary to the verbalisable propositional content of an argument or exposition. As such, historical excess is coded into the field of discourse of non-fiction film through its being referred to *metonymically* by both argumentation and audio-visual indexicality. To repeat Nichols' words (1991: 28), non-fiction 'sounds and images bear a relation of part to whole; they partake of the same order of reality as that to which they refer'.

Tenor

Tenor denotes the role relationships between participants in an interaction, whether concrete (a spoken interaction) or 'imagined' (the relationship of narrator and narratee; and of the social actors or fictional characters represented within a diegesis or non-fictional representation) (Halliday, 1978: 146). In filmmaking, the social roles that are immediately relevant are the roles that are engendered by means of language, and the semiotics of sound and image. Halliday terms the roles that are linguistically assigned 'discourse roles'; those that may be *nonverbally* realised are termed only as 'other types of symbolic action'. These determine the relations of power and solidarity between the film and its audience.

In fiction film, the role relationships are, firstly, concerned with that of narrator-narratee – in films in which there is no obvious narrator, we can reformulate this relationship as representation-representee; secondly, they concern the social roles constructed between characters in the diegesis. In respect of the first relationship, the listener-viewer is admitted to a world in which she can hear and see everything that there is to hear and see (i.e. that is intended to be heard or seen) and maintains a 'secret' position as *voyeur* granted privileged access to the constructed world. As argued in Chapter 4, the typical sound-track construction of Hollywood fiction places the listener-viewer in a complex and fragmented position – in both, a particular point in represented space around which other sound events are organised; and in a position of intimacy and co-presence with speaking characters from which the listener-viewer can hear everything they say with the utmost clarity, whether they are depicted as facing towards or away from the image's point-of-view. This is clearly an artificial subject position, but one that is suitable to the aims of fiction: it represents a relationship of narrator-narratee that is

built on narrative intelligibility rather than phenomenal specificity. A transparency exists between the representation and represented, and narrator and narratee, that is central to the success of, at least, classical narrative fiction, and the choices made as regards the audio-visual construction of subject position have evolved to maintain that transparency.

As regards the second role relationship – that between the characters in the diegesis – such characters are merely *projections* of types of social actor. Even in such films which are – to some degree – biographical, the actual persons, actions and relations that those films are based on are transformed by the fictional context of situation into characters, plot points and archetypal relationships that are ideally entextualised for the purposes of story-telling.

In non-fiction film, the relation between listener-viewer and the participants represented is such that, on the most general level, there is a recognised *identification* between 'representation and represented' (Nichols, 1991: 232), and thus between represented events, persons etc. and the implied listener-viewer. This identification is enabled by the indexical status of the sounds and images – what we hear-see in a documentary film or news item is construed as participating in the same order of reality as our own (1991: 230). Identification is also made possible by an assertive stance towards the represented happenings, persons and relations – that is, they are construed as actually obtaining in the historical world. This type of identification is of such a general form that it might not be immediately useful in the analysis of specific non-fiction texts, yet this generality is actually crucial to understanding the commonalities among all texts that we construe as non-fiction.

In non-fiction, the listener-viewer's relationship to both the representation and the social actors that constitute it is far more concrete than in fiction, for the reason that we understand the represented environment to be 'congruent and coterminous in quality and nature with the one in which we act' (ibid.). Where an audio-visual representation is not construed as a reconstruction of a historical event, we construe a material bond between sound-image and persons and events represented, leading us to assume a whole network of other possible shared attributes between ourselves and the participants onscreen. In fiction, this link is metaphorical, and as such leads and constrains us to identify with the social

relations that constitute a fictional representation in a somewhat 'removed' manner, in which we are encouraged to look for similarity and contrast between represented fictional relations and the relationships that obtain between us and other social actors in the actual world (cf. Currie, 2004). In non-fiction, indexicality – both material (as trace) and representational (as assertive stance) – ensures that the social roles and relations enacted onscreen are construed as specific to the roles and relations of particular historical persons.

In contrast to fiction film, we engage with non-fiction by virtue of an 'expository agent' – an often abstract, removed 'figure' that presents an argument about the world to the audience, and asserts that the represented states of affairs actually occurred. The relations that obtain between the listener-viewer and this agent are centred around rhetorical processes such as 'elucidation', 'persuasion' or 'information'. Where there is a certain formality to the expository agent/listener-viewer role relationship, for instance in examples of early British television news, and in voice-of-god expository films, there is an imbalance of power: the expository agent dominates the listener-viewer, and this is represented in the selective imparting of information. In more recent TV news, as argued in Chapter 7, this balance is skewed by a shift in focus from *informing* to attending to the emotions and reactions of participants (e.g. eyewitnesses). As Jaworski et al (2005: 140) argue,

> in the reporting of 9/11 a shift occurred from the 'sphere of legitimate controversy', which commits journalists to objectivity and balancing of opposed views, to the 'sphere of consensus', which allows journalists to re-align with their audiences presupposing shared assumptions, views and values. The tenor of talk changes from detached neutrality to a solemn, pastoral, caring one.

As Schudson (2002: 41) argues, '[m]uch reporting after September 11 turned toward a prose of solidarity rather than a prose of information'. In this 'reporting', sound is used in very much a 'representative' role (Batcho, 2005), 'opening a channel' between represented participants and representee.

The tenor values of a situation constrain the orientational meanings that can be made – or rather the range of options in orientational meaning – and the

orientational meanings that are made *enact* or *produce* the tenor dimension of the context of situation. In Halliday's words,

> the selection of interpersonal options, those in the systems of mood, modality, person, key, intensity, evaluation and comment and the like, tends to be determined by the role relationships in the situation. (1978: 144)

Mode

Mode denotes 'the symbolic structure of the situation and the specific role assigned to the text within it' (1978: 146). It concerns 'what it is that the participants are expecting language to do for them in the situation' (Halliday, 1985: 12). The mode of non-fiction film, then, is concerned with (1) the affordances and constraints of the media deployed – i.e. what can be said and done with sound, image and language – and (2) how we conventionally utilise those media to make text (i.e. meaningful discourse).

In the context of non-fiction filmmaking, two interrelated aspects of audio-visual film media are pertinent to questions of mode. These correspond to (1) and (2) above. They are also congruent with Plantinga's argument concerning *saying* and *showing*: that is, *indexicality* in the form of trace; and the deployment of audio-visual media as resources for a representation that 'aims at the historical world directly' (Nichols, 1991: 111). As deployed in non-fiction film, sound and image are indexical records of some event, both as 'trace' and in a broader evidentiary sense that allows for re-enactments of historical events. Below, I attend to the affordances of the media and the role they are conventionally called on to perform.

(1)   The mode of the non-fiction film context includes, in consideration of the *media* of discourse (Halliday, 1978), the notion of audio-visual indexicality, which determines that in non-fiction, particularly of the observational style, the relationship between sound-image and the phenomenality of the actual world is such that images and sounds may be deployed as representations *of what such images and sounds are traces of* (Plantinga, 2005: 107). The media afford a *sense* of dialogue between, or co-presence of, listener-viewer and the represented participants. However, as with written language, this is illusory – an illusion

242

enabled by the media of synchronised recorded sound and image. The audio-visual film as medium is one that does not allow for dialogue between listener-viewer and both the narrator/expository agent and the represented participants. In Luhmann's (2000: 2) words, '*no interaction among those co-present can take place between senders and receivers*' (original emphasis). The sonic dimension of audio-visual media entails that even while communication is one-way, and not 'real-time', a sense of communion that is associated with sound and hearing (de Buffon, 1971: 199) provides an illusory mode of 'access' to the unfolding representation. With moving images, too, indexicality and 'optical/acoustic harmony' involve the listener-viewer in an 'immediate' way , to the extent that it takes some effort and reflection to 'see through the replay'. (Luhmann, 2000: 39)

The same media are deployed in fiction film, of course, but in that context the impression of (at least *potential*) dialogue is conventionally constrained by the voyeuristic engagement encouraged by fictional narrative means. In stark contrast to non-fiction filmmaking, the fictional situation is disrupted if attention is given over to the indexical properties of the images and sounds deployed. As Nichols (1991: 113) argues,

> In fiction, the sense of an authoring activity or of an overt narrational process that draws our attention away from the imaginary world we have entered is normally slight and intermittent, only rarely forceful. In documentary, the sense of the filmmaker's argumentative activity or of an over expository process that directs our attention toward the historical world is often continual and highly noticeable. Without it, we would have the impression of gazing onto the world itself rather than seeing the world by means of a text, a window, and an argument.

Nichols notes that the observational film, and many instances of cinema verité, *do* generate such impressions, in which case they can be likened to fiction film on that specific dimension of similarity – that is, such films stress 'the sensation of overhearing and overlooking a world that happens to be drawn from some portion of the historical world, without making an overt argument about it.' (ibid.) Yet, as Nichols goes on to identify, the assertive stance of non-fiction – as implicit in an argument about the world – *does* manifest itself, if only as 'tacit, oblique or indirect' (ibid.). While we are invited to overhear and overlook in observational films, as we

are in fiction films, we are unable to access the thoughts of social actors that they choose to keep private.

In Soderbergh's *Full Frontal* (2003), the status of the fictional frame is continually challenged – that is, the question arises, 'Is what I am hearing-seeing fictional?'– and because of this the status of the images and sounds, as modes and media drawn into serving a fictional narrative role, is ambiguous. When the audio-visual construction of the film is abruptly divorced from any narrative function, what remains, what is foregrounded, is the material and historical excess that is conventionally automatised: the relationship between sound-image-as-trace and visual and sonic representation; and the relationship between such representation and the historically-contingent *production* of a film. Of course, *Full Frontal* would not qualify as non-fiction, on any account. Yet, neither is it prototypical fiction filmmaking, with the aims of the classical narrative film to, more or less, provide consistent narrative motivation for each sound and image. This indexicality is not a necessary condition of non-fiction film, but the point is that it is a *constraint* on what meanings can be made with sound and image in non-fiction: where an audio-visual representation is not signalled as a reconstruction, the listener-viewer will expect it to refer in an indexical way to a phenomenally specific event.

The meanings that may be made in the context of non-fiction film are dependent on a construed identification between context of situation and the historical world. Generally speaking, non-fiction depends on an implied (or automatised) exophoric and homophoric reference (Halliday and Hasan, 1976). Exophoric semiosis, common in spoken interaction, relies solely on the context of situation for meaning. In written and other texts that create their own immediate context of situation, deference to the situation must be accomplished by some semiotic means. In non-fiction, the indexicality of audio-visual representation acts in a privileged, evidentiary capacity, providing the required link between the situation and the cultural context which assigns a certain value to moving, photographic images and (often synchronised) recorded sounds.

As Halliday argues, fictional texts are 'self-sufficient': they are the '*only* form of social action by which 'situation' is defined.' (1978: 146; original emphasis)

Non-fiction, on the other hand, is not 'self-sufficient' – reference is not merely endophoric, but exophoric: non-fiction refers to the historical world via the evidentiary properties of the media and modes of audio-visual representation. Reference is also homophoric in non-fiction, in that recognition of a shared, historical reality is necessary not only to make sense of certain arguments (where certain knowledge is assumed) but also to construe the proper indexical bond to the actual world. Our understanding of this bond is a provision of culture (see section 8.3 below).

The indexicality of the media involved is an intermediary between the non-fiction situation and the context of culture. All non-fiction film is homophoric in that it depends for its import as *non-fiction* on a construed relation to the historical world. This relation is only construable through being mediated by the culturally-presumed indexicality of sound and image. For a documentary to be successfully construed as such – that is as having a privileged relation to the historical world – the role of the sounds and images used (the mode of the situation) must be construed as providing a 'port' to the shared, historical world.

In TV news, this indexicality is further constrained by the expectation, unless explicitly signalled otherwise (which implies unconventionality), that the phenomenal world indexed is coterminous (cf. Nichols, 1991: 230) with the listener-viewer's own. To quote Luhmann (2000: 39) at length,

> television has to accept a rather curious limitation when broadcasting news, which has the effect of being a credibility bonus. When filming something happening, it is tied to the *real time* of that event's unfolding. It cannot photograph what is happening (for example, a football match, a tornado, a demonstration) either before it has happened or after [...], only at the same time. Here, too, there are numerous possibilities for intervening in order to shape the material – use of several cameras and overlays during recording, choice of perspective and film clips and, of course, choice of events selected for broadcasting and choice of broadcasting time. [...] [Despite such manipulations] we are still left with evidence of something rather peculiar, which can be traced to the real-time simultaneity of filming [...] and which distinguishes it from the written fixity of texts. Television literally has 'no time' for manipulating the entire basal material.

(2)   The fundamental 'role' of sound and image – and language – in non-fiction
is rhetorical: it performs an expository, argumentational or assertorial role in the
situation. This is the case even in texts in which an 'expository agent' cannot be
readily identified. In such texts – films of the observational type, for instance – the
expository work is accomplished by means of audio-visual editing, through which
an argumentational logic is manifested (cf. Van Leeuwen, 1991). This role of the
semiosis of sound and image in the non-fiction situation is inextricable from the
primary affordances of the media as indexical trace. As Luhmann, (2000: 39),
explains, synchronised sounds and images are a powerful resource that serves the
role of the text in the non-fiction context:

> [W]hile language increasingly has to give up providing a guarantee for
> reality since everything that is said can be contradicted, the reproduction of
> reality is transferred to movable, optically/acoustically synchronised
> pictures. What one must do here is see through the replay and not mistake
> the time of the broadcast for the time of the real events; but the speed and
> optical/acoustic harmony of the series of pictures elude the contradiction
> that arises at certain points and create the impression of an order that has
> already been tested. At any rate, unlike words contradicting words, there is
> no sense in which pictures can be contradicted by pictures.

Sound and image constitute *evidence* for a given argument that is constructed by
means of audio-visual editing. Furthermore, audio-visual media act as 'reliable
sources for the formation of beliefs about the film's subject' (Plantinga, 2005: 114).

8.2   Audio-visual semiosis in non-fiction film

In this section, I turn to examine the emergent themes of the analyses across the
contexts of fiction and non-fiction filmmaking. Having attended to the documentary
'properties' that are arguably pertinent to the broad array of documentary text-types
(or 'modes'; cf. Nichols, 1991), and having described them in terms of field, tenor
and mode, these dimensions of documentary filmmaking can be utilised in
comparative analyses of non-fiction and fiction film sound-to-picture. With the
conceptual framework of 'situation', we may describe some of the emergent
differences between non-fiction and fiction film sound practices, and, so it follows,
the ways in which the semiotic potential of audio-visual semiosis and sound-to-

picture in non-fiction filmmaking is *constrained* contextually.

8.2.1 Subjectivity and Identification

In fiction filmmaking, we gain access to the diegetic world by means of a character or set of characters. It is via identification with such characters that we come to care what happens to them. In the *TYFS* analysis, a frame of identification is established between the listener-viewer and the protagonist, Nick Naylor, by means of visual direct address (a gaze to camera and the non-verbal gesture of a wink) and the orientational meanings enacted by the panning, equalisation and reverb of the sound track. It matters not that the character is also represented in terms of distaste (that is, his profession) because, this is to be understood as a variable: a quality of a character that we assume can be transformed in the course of the unfolding narrative trajectory. But even if the character clearly fulfills the role of villain, identification with that character can still occur. Alfred Hitchcock addresses this in a discussion on a common situation in 'thriller' films, in which a character is 'snooping' or rummaging secretly through another character's drawers. As Hitchcock argues, 'even if the snooper is not a likable character, the audience will still feel anxiety for him.' (Truffaut, 1993: 73)

The representation of interiority – i.e. the thoughts or, more generally, inner-life of a character – is an extremely common cinematic mode of subjective representation, and a convention for enabling identification with a given character (usually a central one). Sound plays a crucial role in this convention, as Bordwell and Thompson (1985: 194) argue:

> Often a filmmaker uses sound to represent what a character is thinking. We hear the character's voice speaking his or her thoughts even though that character's lips do not move; presumably other characters do not hear these thoughts. A character may also remember words, snatches of music, or events as represented by sound effects. This device is so common that we need to distinguish between *internal* and *external* diegetic sound. External diegetic sound is that which we as spectators take to have a physical source in the scene. Internal diegetic sound is that which comes only from the mind of a character; it is subjective.

The authors then go on to cite the example of Lawrence Olivier's *Hamlet*, in which

247

Hamlet's soliloquies are rendered as 'interior monologues' (1985: 195):

> The character registers the appropriate emotion on his face, but does not move his lips while we hear his voice saying the words, 'To be or not to be...,' and so on. Hamlet is the source of the thoughts we hear represented as speech, but the words are only in the character's mind, not in his physical space. Hence the words are simple diegetic, but internal.

Such constructions of subjectivity, as Nichols (1991: 247) argues, are as much dependant on the semiotic resources of sound and image as they are on the manifest facial expressions of an actor. In his words:

> The full weight of the cinematic apparatus can be brought to bear in the constitution of an arresting subjectivity. As Barry King [1985] notes, 'the projection of interiority becomes less and less the provenance of the actor and more and more a property emerging from directorial or editorial decision [...] [W]hile film increases the centrality of the actor in the process of signification, the formative capacity of the medium can equally confine the actor more and more to being a bearer of effects that he or she does not or cannot originate.' *These directorial and editing decisions can, of course, be taken up by documentary as well.* (my emphasis)

There are obvious similarities between a typical audio-visual representation of interiority in fiction film, and the representation of Castro's recollection process in the *Commandante* extract. Specifically, in terms of the image, Castro is depicted as registering 'the appropriate emotion on his face' (manifest signs of recollection/contemplation), and the gradual movement of the camera into a close up of his face represents a privileged intimacy between listener-viewer and subject. Additionally, in terms of sound, the chanting crowd has no visible source in the represented space on-screen. However, the particular deployment of the sound in the audio-visual representation of interiority is not typical (with respect to fiction film) in terms of the exact *resources* deployed, but may be considered a 'functional near-equivalent' (Altman, 1992: 122-5; also see Bordwell et al, 1985 on 'functional equivalence'). In fiction, internal diegetic sounds are typically represented, in part, by means of augmentation with reverb. As Lacasse (2000: 3) argues, '[i]n cinema, radio and television, who can fail to notice the presence of reverb when a character is dreaming, thinking, or remembering past events?' In the *Commandante* extract,

248

the use of equalisation to render the crowd sound as in the process of emerging is functionally near-equivalent to the use of reverb in fictional representations of interiority. Key to its functioning is the diminishment of upper-mid to high-frequencies that is effected by equalisation, and that renders the sound as distant and inaccessible.

In fiction film, reverb also fulfills other functions, such as the presentational function of classifying the sounds or voices heard as either a part of the represented physical space or separate from it. If the latter, the sound must be further classified as diegetic (for example, an off-screen space; or an internalised sound) or non-diegetic (as with most film music). In fiction films in which sounds or voices are represented as internal and diegetic, reverb helps to construct such interiority by distinguishing it as a wholly separate 'environment' from visualised space. Whether achieved by means of reverb or another resource, this classifying function is at the heart of the functional difference of fiction and non-fiction audio-visual semiosis, a difference that can be illuminated by attending to the situation of non-fiction film, and the co-patterning of audio-visual functions with field, tenor and mode values. In non-fiction film, classification of the represented space on-screen into interior and exterior 'planes' would not redound with the field values of non-fiction as situation type. In the expository activity of representing the historical world non-metaphorically (even in non-fiction of the observational or interactive kind), the indexical bond to historical reality constrains the presentational classifying 'operations', to the extent that a sound with no visualised source may be construed only as either off-screen (diegetic) or non-diegetic (cf. Chion, 1994). Accordingly, this constraint on classification of space further constrains the signification and construal of interior states of social actors.

The table below illustrates the synomorphy of situation values and the functioning of audio-visual instantiations; that is, the field, tenor and mode values co-pattern with semiosic choices made with respect to, respectively, presentational, orientational, and organisational meaning.

	Situation		Function and Realisation
Field	Activity of *exposition*. Subject matter: Personal experience of historical world	*Presentation*	Zooming camera integrated with 'archive' sound of chanting crowds. Emerging sound. Clip of relevant archive footage.
Tenor	Expository agent as inquirer. Listener-viewer as intimate, privileged observer	*Orientation*	Unsteady, 'probing' camera. Medium shot to medium close up. Sound evolves from inaccessible to accessible.
Mode	Indexical relation of media to historical world. Role of enabling access and proximity to subject	*Organisation*	Visual: Vectorised motion towards Castro. Sound: Anticipatory role. --- ---------- Sound: Sudden equalisation boost to upper-mid range on the cut. Visual editing: a cut from Castro in close-up to archive footage.

Table 7: *Commandante*: Non-fiction situation values aligned with audio-visual realisations

Situational interpretation of Commandante

As evident in Table 7, the classifying operations described above are not precluded but rather constrained by the field values of the situation. Specifically, while the subject matter (personal experience of historical event) establishes the context of subjective experience that *encourages* the construal of represented interiority, the expository activity central to non-fiction film *discourages* the making of that presentational meaning. Construed as a unified social dimension, it can be argued that the 'tension' apparent in the field values of non-fiction film strongly constrains the construction and construal of the presentational classifications typical of *fiction* film.

Consequently, any impression of Castro's recollection process that is construed – and hence any representation of his subjectivity – is subjugated to this

primarily expository frame. In other words, the construction of subjectivity primarily functions here as a form of identification that may serve as an 'emotional proof' (Nichols, 1991: 156). In Nichols' words, constructions of subjectivity offer

> a perspective not commonly found in documentary, *loosely but clearly attached* to the experience of its subjects and with which we are invited to identify. In the context of an argument about the world, rather than a story about the world, these moments [...] rejoin subjectivity to the objective: they add a perspective that runs the risk of being dismissed as fiction but that also offers the benefit of rounding out our sense of the human within the arena of history. (1991: 159; my emphasis)

Subjectivity and identification, as this suggests, are used in the *Commandante* extract to serve, not the plot- and theme-development that are central to a *fiction* context, but rather the exposition of the subjective experience of a particular social actor. In non-fiction, representations of interiority are constrained to function against the background of this primary, first-order field value.

In the *Commandante* extract, the continuous-dynamic nature of sound-as-medium is exploited in the deployment of equalisation/filtering. Specifically, the process of 'sweeping' across the frequency bandwidth enacts processual meanings and enables a shifting relationship between listener-viewer and the presentational meanings of the sound. In deploying the resource of equalisation to make *processual* meanings, the expository activity that constitutes the first-order of the field constrains the meaning of the emerging sound to construe the *process* of exposition itself. This functions in a reflexive manner, drawing attention to the expository process of the documentary, and delimiting the status of Castro's represented interiority to that of 'evidence'.

The mode of the situation, too, constrains the organisational meanings of the emerging sound. The indexicality of sound-as-medium and the role it plays in non-fiction film entail that, since the sound cannot be fully subsumed within the internal-diegetic function associated with fiction film representations of interiority, the primary semiotic role it may perform it is an anticipatory one (cf. Mansfield, 1992). That is, the deployment of the emerging sound prior to the cut to its associated image functions to anticipate the moment of revelation accomplished

251

through visual editing.

In addition to the constraints that bear upon the processual meanings of the emerging sound, and 'probing' camera, the situation values listed in Table 7 also determine that the *cut* from the close up of Castro to the archive footage is not construed as representing a fully subjective recollection. The meaning potential of the *extension* relation (see Chapter 3) realised by the cut is constrained by the expository character that pervades the situation. To repeat Nichols' (1991: 113) argument, the foregrounding of the 'filmmaker's argumentative activity' ensures the understanding that we 'are seeing the world by means of a text, a window, and an argument.' The 'transparency' essential to the fiction-film situation type, would enable the meaning of a private memory to be made. However, the 'continual and highly noticeable' expository process that maintains an indexical link to the historical world impinges on the meaning potential of the cut to the archive footage and determines that it be construed in both an evidentiary capacity (illustrative of the discussion's theme) and as representing only the *notion* of subjective recollection, not full-blown interiority.

That the sequence can function in this multifunctional capacity is due to the configuration of situational values of non-fiction: that is, the first- and second-order field values both compel and repel the construal of Castro's interiority; the indexicality of the audio-visual record – its mode value – enables its functionality as evidence for an argument about the historical world; and the tenor values of intimate access to the historical figure of Castro encourage the listener-viewer to construe the emerging crowd sound as sonically *augmenting* Castro's visible thought process, not jointly *constituting* it.

The audio-visual configuration, then, whether constrained to perform a primarily expository function, is nonetheless not *restrained* from performing other functions simultaneously. The overarching functioning of the non-fiction context of situation ensures that both *exposition and indexicality* are privileged contexts for the construal of audio-visual representations and deployments of sound-to-picture (cf. Lemke, 1995: 167). Yet the status of a film as non-fiction accords its constituent semiosic instances a certain measure of freedom to make meanings that are non-expository, or that appear to exceed the expository function – as long as

they are consonant with the overarching expository aims of the non-fiction situation type, and the filmmaker. In the *Commandante* extract, it should be clear that the representation of Castro's interiority is indeed complementary to the expository mode of representation, to which it contributes a dimension of human experience that can further serve as an emotional proof – a mode of *authentication* (Nichols, 1991: 156). It is also consonant with the interactive mode of representation (Nichols, 1991) in which the filmmaker positions himself as an obvious and integral component in the action and discourse that unfolds before the camera. In the *Siege* extract analysed in Chapter 7, the 'gap' between sound and image – akin to the political project of Eisenstein's principle of asynchronous sound (see Chapter 4) – represents an orientational stance toward the reported event, yet the 'highly noticeable' (1991: 113) expository agency of the extract (the tenor of its situation) also determines that an evidentiary, authenticating function is fulfilled by the sound in the audio-visual configuration. The POA of the street-level sound track performs ideological work, representing human emergency as a consequence of the attacks, and situating the listener-viewer at both street-level and that of the helicopter. But at the same time, the expository activity, and indexicality of the media, 'rejoins subjectivity to the objective' (1991: 159), creating a tension between the purported objective functioning of news and the attitudinal stance towards the event. The audio-visual configuration of the *Siege* extract creates a context of situation that allows scope for critique, reflection and conjecture on the events, and the invocation of this particular news context resolves any tension that might otherwise arise.

If the *Commandante* extract is contrasted with an example from a fiction film that deploys a similar formal technique for signifying subjectivity and encouraging identification, we can see more clearly how attending to situation help to delineate the manifest constraints on audio-visual semiosis and sound-to-picture. At the beginning of *Apocalypse Now* (Coppola, 1977), the protagonist, Captain Willard (played by Martin Sheen) awakes in a Saigon hotel room, and immediately looks out of the window. What we hear, in editor Walter Murch's words,

253

are the off-screen policeman's traffic whistle, the car horns, motorbikes, the little fly buzzing in the windowpane, etc. The he sits down on the bed and starts talking, in narration, about how his heart is really in the jungle and he can't stand being cooped up in this hotel room. Gradually, what happens is that all of those street sounds turn into jungle sounds: the whistle of the policeman turns into a cricket; the car horns turn into different kinds of birds; and the fly turns into a mosquito. You are watching Willard sitting in his hotel room, but what you are hearing is a very strong jungle background. One reality is exchanged for another. The thread that links them is the fact that although his body is in Saigon, his mind is in the jungle. That's what Willard really wants to get back to. By gradually making that shift you've presented the audience with a dual reality which – on the face of it – is absurd, but one which nevertheless gets at the dilemma of this particular character. (Paine, 1985: 357)

From this description we can construct the relations of semiosic functions to the fiction film context of situation. In this extract, as suggested in Murch's account, the transition in soundscapes – from city to jungle – is construed as narratively motivated only by means of identification with a character central to the narrative. The presentational meanings made in this shift in soundscape – the signification of Willard's desire to be elsewhere – are only possible by virtue of their redundancy with the (generalised) field values of the narrative fiction film. The first-order field level is constituted by the activity of narration; the second-order level, by the represented interactions of fictitious characters (as opposed to social actors) and 'imagined' states of affairs. (Of course, there is a very real historical referent to the situation's field: the Vietnam war and occupation. The situation is clearly narrative fiction, however, so the link to that historical reality is metaphorical.) Willard's fantasised states of affairs are contingent on the construal of field values appropriate to fiction film – in this case, the transparency of the activity of narration compels us to construe the actor's voice-over as an internal, yet not clearly diegetic, sound; the transition in soundscapes serves to illustrate the propositional content of the voice-over (Willard's desire to return to the jungle). Additionally, in terms of the mode of the situation, the diminished indexicality of sound and image in fiction film enables the construal of both the voice-over narration and the soundscape transition as interior to Willard. Finally, the tenor of fiction film compels us to understand representations of subjectivity or interiority as contingent on a constructed relation between listener-viewer and character, wherein she gains

254

privileged access to the inner-life of a character.

In contrast to fiction filmmaking, in which representations of subjectivity must *attach* to a particular character or group of characters, in non-fiction filmmaking, Nichols argues, it can be deployed less consistently. In non-fiction,

> subjectivity may be intermittent and less than full-blown. It may not attach to a specific character, but attempt instead to convey the feel or texture of an event or experience as does the cross-cutting between close-ups of swooning faces and medium shots of the young Paul Anka performing on stage in *Lonely Boy*. These shots, which adhere to the norm for point-of-view editing in other ways, fail to provide a particular character for us to identify with. Instead, they both allow us to observe the *folie à deux* between singer and followers, and to share the subjective space of the fan empathetically. (1991: 157-8)

In a sequence from *The Corporation* (Mark Achbar and Jennifer Abbott, 2003; included on attached DVD), an expository documentary concerning the concept of 'the corporation' and the power legally ascribed to large corporations, a camera tracks slowly backwards through a patent archive, towards an indeterminate point (as we cannot see behind the camera) and never coming to rest. In the sound track, clips and fragments of media reports concerning issues of biotechnology, specifically patenting the human genome, are presented as a montage. This is integrated with a deployment of *phase-inversion* (see Chapter 5), as well as several other aspects of the sound track, such as the hypnotic two-note bass figure of the music, and the clipped and effected sentences, resulting in a configuration of image and sound that shares many formal similarities with the fiction-film dream sequence device. There is no depicted protagonist or social actor, but the movement of the camera suggests a first-person identification with the POV. Table 8 below lists the relevant alignments of situation values and semiosic realisations.

	Situation		Function and Realisation
Field	Activity of *exposition* – specifically, providing evidence. Subject matter: Responses in the broadcast media to historical event.	*Presentation*	Sound: Edited montage of report excerpts from broadcast media. Keywords and ends of phrases 'recede' (by means of **reverb**) as each excerpt ends and overlaps with another. Image: Moving steadily backwards through patent records archive.
Tenor	Expository agent as presenter of evidence.	*Orientation*	Sound: Subtle stereo **phase-inversion** – spread image thus no objectified source. Image: Perspective aligned with image frame – first-person POV. But also: Steady motion; disembodied perspective.
Mode	Indexical relation of media to historical world. Role of elaborating the central argument.	*Organisation*	Visual: Motion that is not oriented (or vectorised; Chion, 1994) toward a definite point.

Table 8: *The Corporation*: Non-fiction situation values aligned with functional realisations of sound and image

Here, absented voices 'float free' of any character: the sounds of voices do not represent the thoughts of a protagonist or other character that we are to identify with, but rather the *evidence* of an argument that serves as the organising logic of the documentary. As in the *Commandante* extract, exposition and indexicality – as key features of the non-fiction situation – determine how the fragmented voices of the sound track, and the backward-tracking camera movement in the image, are to be construed. Specifically, the context of situation determines that the meanings made in the sound track, for instance, cannot 'attach' to a particular social actor, but instead attach, in an evidentiary capacity, to the unfolding exposition. Indeed, it might be thought that the absence of a social actor in the visual frame is a crucial determinant in construing such attachments. But it is perhaps better to argue that the very absence of a social actor is itself constrained by the situation: such a visual

presence would 'ambiguate' the meanings of the sound track so that, as with the *Commandante* extract, the signification of interiority is at risk. The film's subject is primarily *not* the historical reality of a particular individual or group, but a *shared* human reality: the bearing of the corporation-as-concept, and the acts of particular corporations, on human existence. The sequence under discussion is perfectly constructed audio-visually to cohere both with this meaning of a *shared history* and with the expository mode of representation.

The phase-inversion of the sound track creates a 'spread image' (Moylan, 1992: 50; see Chapter 5) that renders it difficult to localise the sounds' and voices' position in the stereo field. Of course, it is extremely difficult to describe this effect linguistically, except to state that the sound appears as very close to the listener's ears, and that if she changes her lateral position while facing the speakers from which the sound is emitted, the sound appears to 'follow' her movement rather than 'staying put' in the stereo field as one might expect. (The presence of stereo phase-inversion can be determined by rendering a stereo signal as mono; a mono'd phase-inverted signal will sound 'thin' and far quieter. This can be accomplished in VLC Media Player by selecting AUDIO, then AUDIO CHANNELS, and finally MONO.) Given a stereo listening environment, this sound treatment guarantees that the voices on the sound track appear as abstracted from physical reality, collapsing the distinction between listener and object. By rendering the voices as 'thoughts', so to speak, the phase-inversion emphasises the *information* conveyed by the voices and diminishes their status as individual, human voices. The effect privileges mainly the *public* aspects of the informational capacity of the voices and in doing so helps to represent those voices in such a way that they construe meanings that co-pattern with the context of situation of the expository documentary film.

Additionally, the phase-inversion redounds functionally with the notion of a shared history by disrupting the sense of auditory perspective that has become conventional: that is, a perspective that comprises a listener-viewer and a 'listened-viewed'. Conventionally in non-fiction, a clear subject position is constructed. As Lacasse (2000: 140), paraphrasing Augst (1976), argues in relation to the configuration of sound elements within a stereo configuration, 'most sound recordings are based on the principle of a fixed point in relation to which the

257

objects heard are organised [...] consequently, monocentric panning circumscribes the position of the subject.' Instead of deploying this conventional auditory perspective – which is arguably analogous to Renaissance perspective in painting, in which a single, centralised viewpoint dominates (Kress and Van Leeuwen, 1996: 130) – the phase-inversion effect creates an *abstracted* perspective. Arguably, the latter can be likened with the perspective, or lack of it, of pre-Renaissance painting such as Byzantine art. This is especially true of the iconography of Greek Orthodoxy. In orthodox icons, there is no single, circumscribed place for the viewer, in which she is rigidly positioned in relation to the figures and actions represented. Rather, due to the 'flat' perspective, the relations construed between the overall scene of an icon and a viewer, are not easily determined. They are rather *abstract* relations, in fact, entailing that the viewer is not 'fixed in place'. In fact, much like the freedom that pre-Renaissance forms allow for the viewer to determine her own preferred viewpoint, phase-inverted sound liberates the listener from the circumscription of clear POA sound.

In sum, representations of interiority can occur as long as the presentational meanings of the represented thought processes can be seen to redound with *field* in its specification of the type of social action and its determination of 'appropriate' subject matter (Halliday, 1978: 142-3). The representational techniques of fiction filmmaking conventionally deployed in audio-visual constructions of subjectivity, such as the extensive use of reverb on the imagined sounds and/or voices, and the diminishment of the sounds considered 'immediate' to the diegetic scene, would not redound with the type of social action that constitutes the 'situation' of non-fiction film. As Nichols (1991: 29) argues, in non-fiction film

> [o]ur expectation is that [...] we will view [a social actor's] speech and actions from the position of a third-person observer. Point-of-view shots, shot/reverse shots, over-the-shoulder shots, and other devices for aligning the camera with the perspective of a particular character in order to establish a first-person, more fully subjective rendering of time and space are rare [in non-fiction].

The use of reverb on 'absented' voices (Châteauvert, 1996: 141) has evolved in a

258

fictional context to become a signifier of first-person identification with a character. In the *Commandante* extract, the field values of non-fiction – specifically the interactive documentary film (Nichols, 1991: 44) which typically 'stresses images of testimony' – constrain both the type of audio-visual construction of subjectivity produced *and* what it is possible for it to mean. Instead of deploying a heavy dose of reverb on the archive crowd sound, *Commandante* draws on a less conventionally 'coded' sound resource to construct subjectivity: the dynamic manipulation of audio frequency through equalisation. The use of this less-conventional resource for representing interiority circumvents, to a degree, the problems that might arise in deploying a resource considered 'proper' to fiction film.

But contexts of situation do not separate the production of meaning and the interpretation of meaning; the constraints of context do not only determine what a sign-producer can mean, but also what meanings a sign-interpreter can construe (Kress and van Leeuwen, 2001: 8). Even if a producer did deploy an obvious amount of reverb on a sound, the expectations generated by the situation of non-fiction film in general – that the sounds and images we hear and see are of the same order of historical reality as the ones we hear and see in the physical world, and in other audio-visual representations of the historical world – constrain the degree to which the listener-viewer can accept a fully subjective rendering of an event, or the extent to which she is to believe that she can 'access' a represented social actor's thoughts and memories. We might therefore interpret the use of a convention of non-fiction sound-to-picture in a manner conventionally attributed to the aims and expectations of fiction as a *reflexive* strategy (cf. Nichols, 1991; Ward, 2005); for instance, in terms of parody or satire of the norms and conventions of documentary (see Harries, 2000: 6-9). It does not detract from the status of that film as non-fiction, but rather attests to the very constructedness of the documentary 'way of seeing' the world. *Commandante* is replete with self-referential modes of filmic construction – for example, multiple camera views that capture the goings-on of the media, often filming members of the film crew; sudden, abrupt close-ups, as if in preparation of the 'proper' shoot. In such cases, representations of subjectivity are unlikely to be convincing because the means for representing audio-visually are *deautomatised* (Thibault, 1991: 162) – they are foregrounded such that identification with ('access

to') a social actor is hindered.

In the above discussion, I have described the synomorphic 'fit' between semiosic choices made in respect of non-fiction sound-to-picture, and the dimensions of the context of situation that those choices realise (in the sense of 'realisation' defined in section 8.1.3). As argued above, attending to situation can only help to describe the constraints that emerge – it cannot explain them. In order to comprehend how the non-fiction context of situation is regulated and maintained as a *type*, we must attend to the operations that constitute the context of culture. Below, I describe those constraints that function to regulate contexts of situation. These are, specifically, *institutional* constraints.

8.3   Constraints of 'culture'

It is because there are patterns and limits to what is expected to go with what [...] that meaning becomes *possible* (Lemke, 1995: 167)

As Nichols (1991: 156) frames it, the divide between non-fiction and fiction in terms of the issues of, on the one hand, the representation of objectivity, and on the other, that of subjective experience is solely a matter of 'aesthetic convention and historical circumstance'. There is no institutional *force* precluding non-fiction filmmaking from 'incorporating moments of identification' and no mandate restraining it from 'pegging' such moments, such perspectives, to the *indexical* bonds that non-fiction has to the historical world. (1991: 156) In other words, the constraints on what can be meant with sound-to-picture in non-fiction are dynamic, and should not be reified as stable, overbearing structures. If boundaries around meaning making exist, then they are the result of people (as 'institutions') who *maintain* them (Lemke, 1995; Graham, 2001; Van Leeuwen, 2005). These are not necessarily all-powerful individuals consciously policing the perimeters of semiosis – although there are those who will try. Rather, at least in terms of the contexts of fiction and non-fiction filmmaking, certain practices have evolved to appear independent of the actions of people, and their status as fundamental principles is

260

maintained simply by practitioners continuing to practice them in ways they feel are 'proper' to the context.

While the context of situation cannot be considered as somehow separate from its semiotic realisation, precluding the possibility for explaining choices in meaning through appeal to a somehow pre-existant, pre-discursive and reified situation (Thibault, 1991: 122), we can point to aspects of the context of culture that are not immediately 'available' in a given film text. We can posit that some fundamental constraints on non-fiction sound-to-picture occur at an 'institutional' level – the array of interactions and restrictions that are interdependent with the semiotic practices that constitute non-fiction filmmaking. This view can account for the range of possible meanings made with sound-to-picture in non-fiction film in a way that the generalised features of the non-fiction situation type cannot.

8.3.1 Non-fiction as institution and activity

I do not have the scope here to comprehensively expound the ways in which constraint cascades 'down' from culture to sound-to-picture choices in non-fiction film texts. This would involve attending to a general theory of intertextuality (Lemke, 1995; Thibault, 1991) in addition to the interaction between register (configurations of field, tenor and mode values) and genre. Instead, I will limit discussion to the manifest constraints of non-fiction film as *institution*. This entails describing non-fiction as *action genre*: a repeatable and predictable 'activity formation' that enacts and maintains the meanings, and meaning configurations, germane to non-fiction film as *artefact*. Action genres, for Lemke (1995: 31-32), are

> activity formations, the typical doings of a community which are repeatable, repeated, and recognized as being of the same type from one instance or occurrence to another. A baseball game, a train ride, writing a check, making a phone call. [...] Among the special cases of action genres are speech genres and written genres, but these are clearly also definable as the products of the activities that produce them.

Non-fiction sound-to-picture is first and foremost an activity. The textual products of that activity are, most obviously, non-fiction films. These artefacts are produced

by a community of practitioners – comprising, among many others, sound practitioners and film editors – whose work proceeds in accordance with the explicit and implicit values of the institution of non-fiction film.Motivated and guided by a fundamental disposition to represent the historical world, Nichols argues, 'various organisational principles, patterns of distribution and exhibition, styles, structures, techniques, and modes will arise and contend' (1991: 16). Invoking Lyotard, Nichols describes the institution-like nature of documentary filmmaking as an 'ordinary conversation [...] with added constraints' (ibid.). To quote Lyotard (1984: 17) himself,

> An institution differs from a conversation in that it always requires supplementary constraints for statements to be declared admissible within its bounds. The constraints function to filter discursive potentials, interrupting possible connections in the communication networks: there are things that should not be said. They also privilege certain classes of statements (sometimes only one) whose predominance characterises the discourse of the particular institution: these are things that should be said, and there are ways of saying them.

While documentary filmmaking is not as stable a practice as this characterisation suggests, it can be described in institutional terms. For instance, 'there are material supports for the discourse that occurs and this apparatus of groups, organisations, festivals and conferences, companies, films, media and journalism schools, funding and sponsoring agencies, and news networks provides for a measure of regulation in the discursive traffic that takes place' (Nichols, 1991: 16). Fiction filmmaking, too, has its institutional characteristics, though in that context 'entertainment' and 'aesthetic engagement' are the dominant organising principles, and the discourses within fiction-film-as-institution that are able to occur are likely to be those that serve those principles in one way or another.

With the notion of institution comes that of constraint. At the institutional level, the constraints that are emergent in the social semiotic analysis of non-fiction texts appear as highly regulated and explicit. To quote Nichols (1991: 17), the constraints, deriving from documentary-as-institution, that bear upon the meanings that can be made within documentary filmmaking 'may gain the density of codes,

262

ethical dicta, and ritual practices'. The constraints that Nichols identifies may be presented as follows:

Expository documentary	Observational documentary
Evidentiary editing - cutting to bring together the best possible evidence in support of a point.	Continuity editing with synchronous sound.
[T]he filmmaker's responsibility to make his or her argument as accurately as possible even if it requires recontextualising the points of individual witnesses or experts.	[T]he responsibility of the filmmaker to the institution of documentary above responsibility to the film's subjects (the right of 'final cut' remains with the filmmaker not his or her subjects).
[T]he practice of intervening in what occurs before the camera by means of the interview but without showing the filmmaker or even including the filmmaker's voice.	[T]he practice of nonintervention in what occurs before the camera.

Table 9: Institutional constraints on non-fiction filmmaking according to documentary 'mode' (adapted from Nichols, 1991: 17)

On this basis, it can be seen that the constraints on semiosis that are evident in a situational analysis of non-fiction film texts redound with 'codes' of practice that obtain in the *production* of such texts. Nichols suggests that, such constructions of subjectivity in documentary filmmaking are constrained by the desired dynamics between filmmaker and subject; precisely, that it is not often possible to stop the camera in the thick of the moment, or to set up shots of the kind that we would expect to see in fiction film. As he continues,

> Stopping the action to realign the camera transforms history into a mise-en-scene; it becomes a cue that we have crossed into the realm of narrative fiction. Subjectivity, rather than enhancing the impact of a documentary, may actually jeopardise its credibility and shift the focus of attention to the fictional representation of an actual person or event. Our identification with specific social actors therefore has less of the intensity common to fiction. (1991: 29-30)

In other words, we can identify how constraints on semiosis, such as those

described in this chapter, can also be located in the production environment as constraints on *action*. Again, however, we should not consider constraints on semiosis at the 'level' of text as cascading 'down' from the cultural 'level' of production; or vice-versa. It is more accurate to understand the relation as a formal redundancy (cf. Lemke, 1995; Thibault, 1991; Stillar, 2005). It is, nevertheless, proper to characterise production as the material site of meaning making, and therefore as fundamental to the construction of texts.

8.3.2 Indexing and Response

However, the textual products of the action genre of non-fiction filmmaking are not only film texts. They are also the types of texts that are considered 'peripheral' to film. A given film is determined as non-fiction in large part by the *indexing* (Plantinga, 1997) that operates through the peripheral discourses – marketing, critical discussion, and so on – that surround the distribution and consumption of all films, whether non-fiction or fiction. Indexing, therefore, denotes a constellation of discourses that are associated with – that *contextualise* – a film text in its status as a commodity. This is a constraint of culture, and it is interrelated with the notion of *response* – a viewer-centred approach to the contextualisation of text as non-fiction. As Nichols (1991: 24) argues, the 'distinguishing mark of documentary may be less intrinsic to the text than a function of the assumptions and expectations brought to the process of viewing the text.' Vaughan has also formulated a theory of 'response' that – while flawed – is pertinent here. As he argues (1999: 58):

> [T]he term 'documentary' properly describes not a style or a method or a genre of film-making, but a mode of response to film material: a mode of response founded upon the fact that every photograph is a portrait signed by its sitter. Stated at its simplest, the documentary response is one in which the image is perceived as signifying what it appears to record; a documentary film is one which seeks, by whatever means, to elicit this response; and the documentary movement is the history of the strategies which have been adopted to this end.

The obvious problems with the separation of expectations from genre aside, Vaughan's argument – and Nichols' before him – speaks to the idea that listener-viewers bring their own constraints to bear on a given text. These are maintained

and enacted in concert with the constraints of indexing and non-fiction-as-institution.

8.3.3 Trade principles

Importantly, such maintenance of the institutional practices of non-fiction film is achieved in large part by means of *trade principles*, handed down from 'expert' to 'novice'. Practitioners operate according to generalised principles of 'good practice', and perform in ways that are sensitive and appropriate to whatever genre they are working 'in' – in Halliday's (1978: 111) terms, they are sensitive to the 'code' of their practice. For example, the 'proper' way to prepare a contemporary, Hollywood fiction-film sound track is to organise it such that the dialogue of the central characters is always intelligible, and the music and sound effects efficiently augment the explication of narrative. The *DME* paradigm serves the needs of the dialogue-oriented narrative fiction film, and as such is currently the dominant model of sound track structure. When it is discussed in the trade literature, it is not interrogated critically or even treated as a convention which, in being conventional, is thus 'open' to change. To repeat one sound designer's words from Chapter 4:

> Everything [all sound effects and music] is balanced against the dialogue. The dialogue is the key because that's where the information is – that's the story, so people have to hear that. (LoBrutto, 1994: 179)

As such, the broadly defined context of fiction film is *realised* in this practice, and it is naturalised as 'good practice' (cf. Lastra, 2000). For Halliday (1978: 111), trade principles would be considered an intermediary mechanism between 'code' (following Bernstein) and the specific practices of sound-to-picture actuated in a specific context.

To invoke the *Commandante* extract once more, while an identified semiotic function of introducing the crowd sound prior to the cut to the archive footage is to construct Castro's memory as emergent, and to construct a bond between the listener-viewer and Castro, the making of such meanings may be understood as constrained institutionally. Indeed, the deployment in *Commandante* of that particular textual strategy could be construed in terms of the convention of

265

introducing sound in anticipation of an image. As a BBC training manual (Mansfield, 1992) suggests, deploying sound to *anticipate* an image is an effective way of luring the listener-viewer in to *watch* more closely. In discussing an editing strategy for contrasting two shots – one of a horse and cart, the other of a modern tractor – the 'trainee' is advised to use sound in exactly this capacity:

> The horse and cart would be seen and heard as above, but possibly two seconds before the cut to the tractor, the noise of its motor would be faded up until quite loud, totally swamping the faint horse and cart effects. The impact on the viewer would be immediately more arresting [than simply cutting from one sound to the other]. Subconsciously the viewer would ask 'What's that unexplained noise? Horse and carts don't sound like that! Something's wrong. I must watch more closely to find out where that new sound is coming from . Oh, I see, a modern tractor'. (Mansfield, 1992: 18)

In the role that Mansfield suggests here, sound is an effective way of 'heralding' the image – of directing attention toward the image. As much as his advice suggests that certain uses of sound can be useful in augmenting the meaning-making of the image, and can therefore serve a semiotic role, the 'trainee' is taught the general principle of using sound to arrest attention. As he continues,

> The best applications of sound and picture should always make the viewer want to watch more closely. By bringing the tractor sound in early, the cut becomes motivated, welding the shots together in a much more interesting manner.

Of course, it is to be expected that in a training manual the information on offer is of a general kind. But the focus on deploying sound to 'puzzle' or 'shock', in order to arrest the attention of the 'viewer', is indicative of the trade principles expounded throughout the manual. Elsewhere, Mansfield advises that instead of

> gradually fading up the tractor over a few seconds, the noise could be crashed in at full volume over the last second of the horse and cart. Unaided by commentary, the ancient versus modern message would have been conveyed with the additional bonus of surprising the viewer and riveting attention.

Yet trade principles are not constraints as such. They are mechanisms to 'mediate'

the constraints of institution, in order to maintain a particular 'order' in terms of how non-fiction sound-to-picture is practiced. Such maintenance is dynamic, and never fully efficient, hence the blurring of boundaries between genres that continues – and will continue – to occur across many different contexts. With regards to the constraints presented in Table 9, Nichols (1991: 17) argues that '[i]ndividuals will both adhere to and inflect these constraints through their filmmaking practice, insuring that the films we call documentary retain a relative autonomy from any final or determining definition for them'. Indeed, as Nichols has argued elsewhere (1995), the very division between fiction and non-fiction filmmaking was certainly not established in the early days of cinema, and the generic divisions that now appear somehow natural were simply not in place. As he argues,

> early cinema casually blended the staged and unstaged, actors and non-actors, fact and fiction. Only as feature fiction films gained a dominant position did all other forms become relegated to a subordinate or marginal status which still did not necessarily differentiate carefully among these alternative forms.

On this view, it might seem therefore that contemporary non-fiction filmmaking – in its current multiplicity of guises – is in the process of reasserting its 'original' identity as undifferentiated from fiction filmmaking. But this is a simplistic interpretation of the current state of non-fiction film sound. It is perhaps more sensitive to the actual practices of contemporary non-fiction filmmaking to understand the current trend towards dissolving the non-fiction/fiction divide in terms of an expansion of available semiotic resources.

8.4   Conclusion

In this chapter, I have attempted to understand the functioning of sound-to-picture – specifically, the role that it plays in the construction of subjectivity and identification – in terms of its relation to non-fiction as context of situation and context of culture. Additionally, in doing so, I have outlined a procedure for

analysing sound-to-picture across contexts – here, non-fiction film (documentary and TV news) and fiction film.

The division between fiction and non-fiction that is identifiable today attests to the potential for changes in the semiotic practices, and the possibility for expansions in semiotic resources, that constitute the action genre and text type of non-fiction sound-to-picture.

Chapter 9: Conclusion

9.0 Introduction

In this chapter, I begin by evaluating the contribution of the thesis to the field of multimodal discourse analysis. This involves relating the findings of the research to the specific work reviewed in Chapter 2, and to the field more generally.

Additionally, I use this concluding chapter as a forum for reflecting on the limits of social semiotic analysis. Finally, I make some suggestions for further research.

9.1 Contribution and evaluation

The research questions set out at the beginning of the thesis were concerned with (1) investigating the semiotic functioning of sound in audio-visual semiosis; and (2) examining the semiotic role of sound in the context of the non-fiction film. In attempting a response to these questions, I have established a number of principles of audio-visual semiosis and sound-to-picture that may inform subsequent research in the MDA field. I will not repeat these here. Instead, I present what I consider to be the main contributions of the thesis.

Sound

Where much work on multimodality has tended to avoid engaging with sound, here I have attempted to establish a foundation on which other researchers may develop their own analyses and interpretations of sound and sound-to-picture. This was deemed an urgent and critical task. Where existing research, especially Baldry and Thibault (2005), has provided a crucial 'way in' for many researchers, the conceptual framework they develop does not make explicit connections to the *practice* of sound-to-picture. In this thesis, I have incorporated technical terminology into a critical audio-visual framework for analysis. This offers several

benefits. Firstly, it anchors analyses and interpretations in *practice*. This is no mere token gesture. Practices are subject to numerous constraints, technical, technological, institutional and medialogical. If these are incorporated into the core of an analytic framework, and not merely introduced *ad hoc* as 'incidental' features, the framework allows the analyst to account for certain boundaries around meaning making. Here, the technical and traditional resources I outline and deploy co-evolved in the research process with the social semiotic framework of logical relations of expansion. For this reason, their integration in analysis – for instance, the 'emerging sound' of Castro's recollection in an *extension* relationship with the 'probing' camera – while challenging, occurred with a fair amount of ease.

In implementing a framework that integrates semiotic and practice-oriented concepts, I have contributed to the existing work on multimodal integration that also deploys logical relations of expansion. I see no great difficulty, for instance, in further integrating the framework developed in this thesis with Baldry and Thibault's framework. However, such integration might be of limited value. Unless analytic categories are erected on a practice-oriented foundation they may, as I mention above, only offer 'added value' rather than enabling the original foundation to be altered so as to incorporate practice at 'ground-level', so to speak.

Non-fiction sound-to-picture

The second main contribution is to the analytic discourse on non-fiction film sound. Such work is extremely scarce. By introducing a social semiotic perspective on non-fiction film sound, I have provided a critical basis on which to conduct analyses of sound in TV news and documentary film. Chapter 7 has demonstrated that the analytic framework is equally sensitive to the objectives of TV news and documentary film, rendering it applicable to other forms of non-fiction, in particular, modern forms of propaganda, such as the increasingly dramatic and sophisticated 'Political Party Broadcast'. It has also demonstrated that the increasingly 'interpersonalising' function of broadcast news can be discussed in terms of sound-to-picture. It is particularly encouraging to see that patterns often detected in the language and images of the news media can also be identified in the

sound track. This offers a great deal of credence and import to the often made suggestion that sound is a crucial site of semiosis. I hope that the work of Chapter 7 is engaged with by social semiotic researchers interested in sound (though, of course, those aiming to engage in the moving image should also find value in the work presented there).

The model of contextual constraint developed in Chapter 8 is also a valuable contribution here, as there are no existing accounts of non-fiction film that are considered explicitly in terms of 'situation' (field, tenor, mode) nor 'culture' (institutional constraints and trade principles). Its main contribution is in the demonstration of a procedure for attending to the patterning of functional meanings of sound-to-picture and the situation values of non-fiction as situation *type*.

9.2 Suggestions for further research

As I have argued throughout, a descriptive account of non-fiction sound-to-picture that did not attend to non-fiction as 'situation' and 'culture' would neglect a crucial dimension of analysis. Yet I believe that such work holds great value for MDA, particularly now that at least *some* account of how audio-visual semiosis relates to non-fiction as context exists. I envisage such an account to attend to the variety of documentary 'modes' that Nichols (1991) proposes: expository, observational, interactive and reflexive.

I would also suggest that a diachronic study of the modes of sound-to-picture in non-fiction would be extremely useful in developing an understanding of how change in the semiosic system arises. Documentary film, for instance, is a far broader category than it used to be. Nichols' (1991) overview of the modes of documentary representation that have evolved over time reveals much about the array of texts that are now considered as documentary. At the same time, the type of text that constitutes the *ideal* documentary has certainly changed since its early period, and is perhaps in the process of changing once more as the proliferation of 'reality TV' programmes come to redefine the central social function of documentary film. In addition, the huge increase in 'feature' documentaries over the past decade, many sponsored by the Hollywood film industry and deploying its key

271

production and post-production personnel, is crucial to defining documentary film as *primarily* expository. The types of documentary films likely to receive funding is certainly constrained by the prevailing definition of the *ideal* documentary and its central purpose. One can surmise that the gaps in the meaning system (Lemke, 1995) of non-fiction film are not likely to be textured at any point in the near future. Rather, I would predict that the commercial potential of expository documentary films with 'sexy' themes and subjects will come to define the centre-field of documentary filmmaking. Consequently, the generation of genuinely new meanings to be made, new modes of representing the historical world, and thus new ways of *understanding* and *knowing* the world are at risk. An understanding of how changes have occurred in the past, as regards the clustering of text-types that are taken as representative of a culture's epistemological and aesthetic aims, will be a key step in understanding how change can be instigated. Not only that, but combining a diachronic study of the culturally prevalent documentary text-types with a 'bottom-up' approach that is concerned to describe changes in the resources of audio-visual semiosis may help to lay the foundation for a praxis-oriented approach to bringing about radical change in the meaning-making system. This is to argue that boundaries around the meaning making of non-fiction film can be positively challenged by attempts to increase the semiosic potential of its constituent modes and media.

News has also changed quite radically over the past two decades. As a context of situation, its *field* – particularly its subject matter – has broadened to include more reflective and critical engagement with the world. Undoubtedly this is related to the emergence of 24-hour news broadcasting, initiated in the US in the 1980s. At the same time that there are more *types* of subject mediated in TV news, the *tenor* of news – particularly in the UK – has evolved to become more emotive while still maintaining an operational notion of objectivity that is accomplished primarily by means of audio-visual semiosis. Additionally, the *mode* of contemporary TV news has expanded to include a wide variety of audio-visual resources, such as stereo sound (BBC News 24 even deploys surround sound in the music mix); 3D graphics; satellite 'link-ups' that enable ostensibly 'real-time' reports from the field; and interactive engagement, in the form of links to related textual or

audio-visual information.

Also worthy of further research are those films that, while typed 'institutionally' as non-fiction (cf. Plantinga, 2005 on *indexing*), are somewhat marginal. Often, such films provoke debate in the non-fiction 'community'. An understanding of how such films' sound tracks play a part in their marginality would be a first step in an interesting and exciting direction.

I should also state that I would like to continue this research into non-fiction film sound by way of comparing *parodies* of non-fiction (e.g. *Brasseye* ['current affairs'] compared with actual current affairs; *The Office* contrasted with actual observational documentary). The processes by which parody recontextualises its 'target' text (cf. Harries, 2000: 9) can be illuminated by attending to sound-to-picture practices.

Finally, while I attend to context of culture only by way of the notion of 'institution', there is a need to extend the work of Chapter 8 to include theorisation of non-fiction as *genre*. In systemic-functional accounts of 'culture', genre and intertextuality are the processes by which culture is realised. For space reasons this is not dealt with in the thesis. Rather, I actually *avoid* the term 'genre' so as not to convolute those parts of the thesis where it would be relevant. Systemic-functional researchers will be able to detect those points at which I replace 'genre' with 'situation type'.

9.3 Conclusion

This thesis has aimed to gain an insight into the semiotic functioning of a particular semiotic mode in a specific context. Analysing the use of sound in the context of non-fiction is a new enterprise. Its antecedents are mainly concerned to understand either how *visual* devices make meaning in non-fiction (cf. Vaughan, 1999, on the zoom lens in documentary); to understand how sound functions semiotically in general (Van Leeuwen, 1999); or how sound makes meaning in a general audio-visual context (Constantinou, 2002). Insofar as this thesis represents a search for a way of understanding sound-to-picture in non-fiction film, it is to be understood as a beginning. I do not claim that the research process undertaken here is the most

profitable step forward in attempting an answer to the research questions. Instead, I proffer this thesis as an opening gambit: an initial attempt to understand both the semiotics of sound-to-picture in non-fiction film *and* how to proceed in the analysis and theorisation of audio-visual semiosis.

NOTES

1. That is, to the majority of a sample of 128 undergraduate students of classical music history.

2. Representations of unreality, or altered states, such as dream sequences are not included here.

3. Another director interested in actors' physical freedom, John Cassavetes, employed a radio-microphone system to the same ends in *Opening Night* (1979).

4. Somewhat ironically, although Altman's multi-channel radio-mic system is used by him to challenge conventional representations of human dialogue, it is also perfectly suited to the traditional aims of the *DME* structure, in that dialogue is less challenged by the many acoustic variables of the recording environment. As Belton (1985: 69) argues, 'radio microphones pick up speech (and body tone) *before* it is projected – that is before it can acquire any spatial properties'. As he goes on to argue, even though such spatial properties can be simulated in post-production, the sound quality that results from the deployment of radio-microphones is vastly different to the resultant sound quality of boom microphones and their placement beyond the camera's field of view. Through avoiding subjection to the many acoustic variables that often threaten the recording of intelligible dialogue – a problem most often solved with the post-synchronised 'dubbing' of actors' voices (also termed 'looping') – vocal intelligibility is rendered practically unproblematic. [Again, the presence of choice can be identified, establishing that practice (in this case, a technical and technological system) as a resource for meaning making.]

5. Hereafter, the phrase 'silent film' will not appear within quotation marks.

6. This fact is understandable, given that most multi-shot films, in comparison to slide lectures,
travelogues and illustrated songs, were governed by an overarching narrative functioning, which a topographical spotting of sound effects would challenge. In this sense, the topographical mode of engagement was decidedly *anti*-trade.

7. Lastra's conceptual polarisation, however, tends to suggest a very clear

distinction, and not a *tension*, between those concepts as they were put to use in film-sound practices.

8. Proposed by André Gaudreault, this term refers to films comprising more than a single shot.

9. Barthes finds further support for his claim that sound is fundamental to producing genuinely 'filmic' films in the 'classical paradigm of the five senses' in which 'the third sense is hearing (first in importance in the Middle Ages' (Barthes, 1977: 53 fn.) That his concept of a *third* meaning should congrue with the classical paradigm is a 'happy coincidence' especially because at the heart of the third meaning is the notion of *listening*, as a specific outcome of the coming of sound in film.

10. Prominent sound technician of the time, Carl Dreher, estimates that approximately 80 percent of the one thousand sound technicians working in Hollywood in the early 1930s were from other sound-related industries, such as telephony and music (Dreher, 1930: 18; O'Brien, 2005: 131-132)

11. Given that stereophony had not yet been realised technically during the early-period of the sound-film, it must not be forgotten that, as Altman (1992: 55) argues, our media shape our capacities of perception, and determine the standards against which we judge what is a realistic representation and what is not. Therefore, that the perceptual fidelity model dominated theoretical discussions and actual sound practices at a time when sound cinema was entirely *monaural* does not mean that a monaural recording that followed the principles of perceptual fidelity was judged as somehow less-than-faithful *because* it was monaural, not stereophonic.

12. That this is Straub and Huilet's claim is inaccurate - it is actually the interviewer's wording. However, in the context of the interview both filmmakers are in agreement with that statement. For that reason, O'Brien's 'mistake' remains uncorrected here.

13. Eisenstein's definition of sound and image as distinct 'spheres of feeling', and

both as 'images', cannot be fully explored in this section. Suffice it to say that the perceived *commonalities* of sound and image – their shared qualities, their common existence as 'spheres of feeling' and 'images' – were believed to be at the basis of their compatibility as elements of montage (cf. Cook, 1998: 50). Yet, by definition, counterpoint involves the joining of apparently *incommensurable* effects. Therefore, what seems to be at issue in contrapuntal deployments of sound and image is that both commensurability and incommensurability of the different media are essential to producing audio-visual counterpoint.

REFERENCES

Altman, R. 1985. 'The evolution of sound technology.' In Weis, E. and J. Belton (eds) *Film Sound: Theory and Practice*. New York: Columbia University Press

_____1992. *Sound Theory /Sound Practice*. New York/London: Routledge

_____1992. 'Sound Space.' In Altman, R. (ed) *Sound Theory, Sound Practice*. New York/London: Routledge

_____1996. 'The Silence of the Silents'. *Musical Quarterly* 80: 648-718

_____2005. *Silent Film Sound*. London: Routledge

Augst, B. 1976. .The Apparatus: An Introduction.. *Camera Obscura: A Journal of Feminism and Film Theory*, 1: 97-103.

Austin, J. L. 1961. *How to Do Things with Words*. Oxford: Clarendon Press.

Baldry, A. and P. Thibault. 2005. *Multimodal Transcription and Text Analysis*. London: Equinox

Barthes, R. 1977. *Image-Music-Text*. London: Fontana

Batcho, J. D. 2005. 'The Cultural Politics of TV News Sound'. Unpublished MA thesis. San Francisco State University, California

Bateman, J., Delin, J. and R. Henschel. 2002. 'Multimodality and empiricism: methodological issues in the study of multimodal meaning making'. *Genre*

and Multimodality (GeM) project report.

Belton, J. 1985. 'Technology and aesthetics of film sound'. In E. Weis and J. Belton
(eds.) *Film Sound: Theory and Practice*. New York: Columbia University
Press

_____1992. '1950s Magnetic Sound: The Frozen Revolution'. In R. Altman
(ed) *Sound Theory/Sound Practice*. New York: Routledge

Berger, J. 1972. *Ways of Seeing*. Harmondsworth: Penguin

Bennet, C. 2000. 'Beware of cheap imitation', *Guardian*, 3 February. URL:
http://www.guardian.co.uk/shipman/Story/0,2763,192666,00.html.

Bernstein, B. 1982. 'Codes, modalities and the process of cultural re-production: A
model.' In *Cultural and Economic Reproduction in Education: Essays on
Class, Ideology and the State*, ed. Michael W. Apple, 304-55. London and
Boston: Routledge and Kegan Paul.

Blake, L. 1999. '"What's a Binky?" Larry Blake's Handy-Dandy Film Sound
Glossary.' Intertec Publishing. Available online at:
http://www.northbeachpost.com/handy_glossary.html. Retrieved
30/03/2007.

Bolter, J. and R.Grusin. 2000. *Remediation: Understanding New Media*.
Cambridge, MA: MIT Press

Bordwell, D. 1985. *Narration in the Fiction Film*. University of Wisconsin Press

Bordwell, D., Staiger, J and K. Thompson. 1985. *The Classical Hollywood
Cinema: Film Style and Mode of Production to 1960*. London: Routledge

Bordwell, D. and K.Thompson. 1993. *Film Art: An Introduction*. London: McGraw-Hill.

Brophy, P. 2000. 'Walking with Dinosaurs: Faux Acoustica'. Retrieved on August 21st, 2004 from: http://media-arts.rmit.edu.au/Phil_Brophy/PBfront.html

_____ 2004. *100 Modern Soundtracks*. London: BFI

Bruzzi, S. 2000. *New Documentary: A Critical Introduction*. London: Routledge.

Burris-Meyer, H. 1940. 'Sound in the Theater'. *The Journal of the Acoustical Society of America* 11(3): 346-351

Carroll, N. 1996. 'Nonfiction film and Postmodernist Skepticism.' In D. Bordwell and N. Carroll (eds) *Post-Theory: Reconstructing Film Studies*. Madison: University of Wisconsin Press

_____ 1997. 'Fiction, Non-Fiction, and the Film of PresumptiveAssertion: A Conceptual Analysis'. In R. Allen and M. Smith (eds) *Film Theory and Philosophy*. Oxford: Clarendon Press: 173–202.

Carter, M. 2004. 'Nazi Indoctrination'. Available online at:
http://www.utexas.edu/opa/pubs/discovery/disc1997v14n2/disc-indoctrination.html
Retrieved 30/05/2006.

Cavalcanti, A. 1985. 'Sound in Films'. In E.Weis and J.Belton (eds) *Film Sound: Theory and Practice*. New York: Columbia University Press

Châteauvert, J. 1996. *Des mots à l.image : la voix over au cinéma*. Québec: Nuit
blanche éditeur.

Cheong, Y. 2004. 'Ideational meaning in print advertisements'. In O'Halloran, K. L.
(ed.) *Multimodal Discourse Analysis: Systemic Functional Perspectives*.
London: Continuum

Chion, M. 1994. *Audiovision: Sound on Screen*. New York: Columbia University
Press

_____1999. *The Voice in Cinema*. New York: CUP

Chouliaraki, L. 2004a. 'Shock and Awe: The aestheticisation of suffering in Iraq
war footage'. Working paper, CBS MediaHub.

_____ 2004b. 'The Moral Spectator: Distant suffering in live footage of
September 11, 2001. In P. Levine and R. Scollon (eds.), *Discourse and
Technology: Multimodal Discourse Analysis*. Georgetown University Press:
DC

Comolli, J. 1980. 'Discussion'. In De Laurentis, T. (ed.) (with S. Heath) *The
Cinematic Apparatus*. New York: St. Martin's Press

Constantine, A. 2003. 'Collateral Brain Damage: The Hollywood Propaganda of
Ministry'. *High Times*, July 2003.

Constantinou, O. S. (2002) 'Sound Constructs: audiovisual modality and the social
semiotics of sound design'. Unpublished MA thesis, Cardiff University, UK.

_____ 'Multimodal Discourse Analysis: Media, modes and
technologies'. *Journal of Sociolinguistics*, 9(4): 602-618

Cook, N. 1998. *Analysing Musical Multimedia*. London: OUP

Corner, J. 1996. *The Art of Record: A Critical Introduction to Documentary*.
        Manchester: Manchester University Press

Cottle, S. 2003. 'Media Organization and Production: Mapping the Field' in
        S.Cottle (Ed.) *Media Organization and Production*. pp. 3-24. London: Sage.

_____ 2004. 'Producing Nature(s): On the Changing Production Ecology of
        Natural History TV' *Media, Culture and Society* 2(1): 81-101.

Currie, G. 1999. 'Visible Traces: Documentary and the Contents of Photographs'.
        Journal of Aesthetics and Art Criticism, 57: 285–297.

_____2004. 'Imagination and Make-Believe'. In B. Gaut and D. McIver Lopes
        (eds) *The Routledge Companion to Aesthetics*. London, Routledge.

Dancyger, K. 1996. *The Technique of Film and Video Editing*. 2nd Edition. Focal
        Press.

De Buffon, G.L.L. 1971. *De l'homme*. Paris: Duchet

Doane, M. 1985. 'Ideology and the practice of sound editing and mixing'. In Weis,
        E. and J. Belton (eds) *Film Sound: Theory and Practice*.   New York:
        Columbia University Press

Dreher, C. 1930. 'Sound Personnel and Organisation.' *American Cinematographer*,
        10 (13): 18

Eco, U. 1976. *A Theory of Semiotics*. London: Macmillan

Eisenstein, S. 1943. *The Film Sense*. London: Faber and Faber

Eisenstein, S. M., Pudovkin, V. I. and G. V. Alexandrov. 1928. 'A Statement'. In *Film Sound: Theory and Practice*. New York: Columbia University Press

Fairclough, N. 1989. *Language and Power*, London, Longman

Fischer, L. 1985. '*Applause*: The visual and acoustic landscape'. In E.Weis and J.Belton (eds) *Film Sound: Theory and Practice*. New York: Columbia University Press

Fodor, J. A., and Z. W. Pylyshyn. 1988. 'Connectionism and cognitive architecture: a critical analysis'. In S. Pinker and J. Mehler, Eds., *Connections and Symbols*. Cambridge, MA: MIT Press, pp. 3-71

Fowler, R. (1991). *Language in the News: Discourse and Ideology in the Press*. London: Routledge.

Gaudreault, A. 1988. *Du littéraire au filmique: système du récit*. Paris: Klincksieck

Gibson, J.J. 1986. *The Ecological Approach to Visual Perception*. Hillsdale, NJ and London: Lawrence Erlbaum.

Giddens, A. 1984. *The Constitution of Society: Outline of the Theory of Structuration*. University of California Press

Goodwin, A. 1993. 'Fatal distractions: MTV meets postmodern theory'. In S. Frith, A. Goodwin and L. Grossberg (eds.) *Sound and Vision: The Music Video Reader*. London: Routledge

283

Graham, P. 2001. 'Contradictions and institutional convergences: Genre as method'. *Journal of future studies*, 5 (4): 1-30

_____2006. *Hypercapitalism: New media, language, and social perceptions of value*. New York: Peter Lang

Gregory, M.J. 1967. 'Aspects of Varietal Differentiation', *Journal of Linguistics*, 3: 177-98

_____ 2002. 'Phasal analysis within communication linguistics: two contrastive discourses'. In P. Fries, M. Cummings, D. Lockwood and W. Sprueill (eds), *Relations and Function in Language and Discourse*. London: Continuum: 316-345

Gunning, T. 1986. 'The Cinema of Attractions: Early Film, Its Spectator, and the Avant-Garde,' *Wide Angle* 8 (3-4): 63-70

Halbwachs, M. 1992 *On Collective Memory*. (Trans. Lewis A. Coser) London: The University of Chicago Press, Ltd.

Hall, E. 1964. 'Silent Assumptions in Social Communication'. *Disorders of Communication* 42: 41-55

Halliday, M.A.K. 1975. 'Learning How to Mean'. London: Edward Arnold.

_____1978. *Language as social semiotic*. London: Arnold

_____ 1982. 'The De-Automization of the Grammar: From Priestley's *An Inspector Calls*'. In J.M.Anderson (ed) *Language Form and Linguistic Variation: Papers Dedicated to Angus McIntosh*: 129-59. Amsterdam: Benjamins.

_____ 1985. An *Introduction to Functional Grammar*. London: Arnold

_____ 1992. ;How do you mean?' In M. Davies and L. Ravelli (eds)
Advances in Systemic Linguistics: Theory and Practice: 20-36. London:
Pinter.

Halliday, M.A.K and R. Hasan. 1976. *Cohesion in English*. Longman, London.

Harries, D. 2000. *Film Parody*. London: BFI.

Hippler, F. 1937. 'Der Film als Waffe' (Film as a Weapon) *Unser Wille und Weg*, 7:
21-23. Available online at:
http://www.calvin.edu/academic/cas/gpa/hippler1.htm
Retrieved 16/04/2006

Hodge, R. and G. Kress. 1988. *Social Semiotics*. Polity Press

Holman, T. 1997. *Sound for Film and Television*. Focal Press.

Iedema, R. 2001 'Analysing film and television: a social semiotic account of
Hospital: An Unhealthy Business'. In T. van Leeuwen and C. Jewitt,
Handbook of Visual Analysis. London: Sage

_____2003. 'Multimodality, Resemiotization: Extending the Analysis of
Discourse as Multi-Semiotic Practice'. *Visual Communication* 2(1): 29-57

Jackson, B. 2007. 'Unravelling the sound for *Zodiac*'. *Mix Online*, March 2007.
Available at:
http://mixonline.com/sound4picture/film_tv/audio_unraveling_sound_zodiac/
Retrieved March 20th, 2007.

285

Jaworski, A, Fitzgerald, R. and O. Constantinou. 2005. 'Busy Saying Nothing New: silence in the news broadcasts of 9/11'. *Multilingua* 5 (1)

Jewitt, C. 2004. 'Multimodality and New Communication Technologies'. P. Levine and R. Scollon (eds) *Discourse and Technology: Multimodal Discourse Analysis*. Georgetown University Press: DC

Kahn, D. 1988. 'Cage and phonography'. Paper given at the *Cage at Wesleyen Symposium*, Middletown, Conn. 27th February 1988. Online at: http://www.vasulka.org/archive/Artists3/Kahn,Douglas/CagePhotography.rtf

_____1992. 'Eisenstein and Cartoon Sound', in *Essays in Sound*, edited by Shaun Davies, Annemarie Jonson, Eddy Jokovich. Sydney: Contemporary Sound Arts

_____1999. *Noise, Water, Meat: A History of Sound in the Arts* Cambridge, MA: MIT Press

King, B. 1985. 'Articulating Stardom'. *Screen* 26(5): 27-50

Kittler, F. 1999. *Gramophone, Film, Typewriter*. Stanford: Stanford University Press

Kracauer, S. 1985 [1960] 'Dialogue and Sound'. In Altman, R. 1992. *Sound Theory/Sound Practice*. New York/London: Routledge

Kress, G. 2003. 'Representation, modes and media: new communicational designs and some likely effects'. Paper presented at the *International Workshop on Multimodal Communication*, Centre for Interdisciplinary Research, Bielefeld University, January 2003

Kress, G. and R. Hodge. 1979. *Language as Ideology*. London: Routledge and

Kegan Paul

Kress, G. and T. van Leeuwen. 1996. *Reading Images*. London: Routledge

_____1997 'Discourse semiotics'. In Teun van Dijk (ed) *Discourse as Structure and Process*. London: Sage

_____ 2001. *Multimodal discourse*. London: Arnold

Lacasse, Serge. 2000. *Listen to my voice*. Unpublished doctoral thesis. University of Liverpool.

Lakoff, G. and M. Johnson. 1980. *Metaphors We Live By*. Chicago: University of Chicago Press

Lastra, J. 1992. 'Reading, Writing, and Representing Sound'. In R.Altman (ed) *Sound Theory/Sound Practice*. New York/London: Routledge

_____2000. *Sound Technology and the American Cinema*. New York: CUP

Lemke, J. 1984. *Semiotics and Education*. Monograph in Toronto Semiotic Circle Monographs Series, Victoria University, Toronto.

_____1993. 'Discourse, Dynamics, and Social Change.' *Cultural Dynamics* 6(1): 243-275. [Special issue, Language as Cultural Dynamic, M.A.K. Halliday, Issue Editor].

_____1995. *Textual Politics*. London: Taylor and Francis

_____1998a. 'Multiplying meaning: visual and verbal semiotics in scientific text'. In J. R. Martin and R. Veel, *Reading Science*. London: Routledge

_____1998b. 'Analysing Verbal Data: Principles, Methods and Problems'. In B. Fraser and K. Tobin (Eds.). International Handbook of Science Education. Dordrecht, The Netherlands: Kluwer.

_____1998c. 'Metamedia Literacy: transforming meanings and media.' In D. Reinking, L. Labbo, M. McKenna and R. Kiefer (eds.), *Handbook of Literacy and Technology: Transformations in a Post-Typographic World.* Hillsdale, NJ: Erlbaum: 283–301.

_____1999. 'Typological and topological meaning in diagnostic discourse'. *Discourse Processes* 27(2): 173-185

_____2002. 'Travels in Hypermodality'. *Visual Communication* 1(3): 299-325

_____2003. 'Beyond Literacy: Semiotics and the Civilization of Hypercomplexity'. *Semiotica* 144: 381-387

LeVine, P. and R.Scollon. 2004. *Discourse and Technology: Multimodal Discourse Analysis.* Georgetown University Press.

Levinson, J. 1984. 'Hybrid art forms'. *Journal of Aesthetic Education.* 18:5–13

Lim, V. 2004. 'Developing an integrative multi-semiotic model'. In K.O'Halloran, *Multimodal Discourse Analysis: Systemic-Functional Perspectives.* London/New York: Continuum: 11-27: 220-246

LoBrutto, V. 1994. *Sound-on-Film: Interviews with creators of film sound.* London: Praeger

Luhmann, N. 2000. *The Reality of the Mass Media.* Cambridge: Polity Press

288

Lukács, G. 1971. *History and Class Consciousness: Studies in Marxist Dialectics.* Cambridge, MA: MIT Press.

Lyotard, J. *The Postmodern Condition: A Report on Knowledge.* Minneapolis: University of Minnesota.

Mansfield, J. 1992. *Music and Sound Effects for Television.* BBC Training and Development

Martin, J. and D. Rose. 2003. *Working with Discourse: Meaning beyond the clause.* London: Continuum

Martinec, R. 1998. 'Cohesion in action'. *Semiotica* 120(1/2): 161-180

_____ and A. Salway. 2005. 'A System for Image-Text Relations in New (and Old) Media'. *Visual Communication* 4(3): 337-371

_____ and T.Van Leeuwen. 2006. *The Language of New Media Design: Theory and Practice.* London: Routledge

Maxfield, J. 1930. 'Acoustic control of recording for talking motion pictures.' *Journal of the Society of Motion Picture Engineers.* 14(1)

Mercer, P. 2002. *Television Realism: A Semiotic Approach.* Unpublished doctoral thesis. Institute of Education, University of London.

Meyer-Dinkgräfe, D. 2006. 'Thoughts on Dubbing Practice in Germany: Procedures, Aesthetic Implications and Ways Forward'. *Scope* 5 (June 2006). Online at http://www.scope.nottingham.ac.uk/article.php?issue=5&id=131

Middleton, R. 1990. *Studying Popular Music.* Milton Keynes: Open
University Press.

Mintz, P. 1985. 'Orson Welles' use of sound'. In E.Weis and J.Belton (eds) *Film
Sound: Theory and Practice.* New York: Columbia University Press

Monbiot, G. 2002. 'Both Savior And Victim: The film Black Hawk Down is helping
to create a new myth of American nationhood, which threatens everyone on
earth.' *The Guardian,* January 29th, 2002.

Moylan, W. 1992. *The Art of Recording: The Creative Resources of Music
Production and Audio.* New York: Van Nostrand Reinhold.

Nichols, B. 1991. *Representing Reality: Issues and concepts in documentary.*
Indiana University Press

_____ 1995. 'Documentary and the coming of sound'. Available online at:
http://www.yidff.jp/docbox/6/box6-1-e.html. Accessed on 14th April 2007.

Norris, S. 2004. *Analyzing Multimodal Interaction: A methodological framework.*
New York/London: Routledge

O'Brien, C. 2005. *Cinema's Conversion to Sound: Technology and Film Style in
France and the U.S.* Bloomington: IUP

O'Halloran, K. 1999. 'Interdependence, interaction and metaphor in multi-semiotic
texts.' *Social Semiotics* 9(3): 317-354

_____ 2004. *Multimodal Discourse Analysis: Systemic Functional
Perspectives.* London/New York: Continuum

_____. 2005. *Mathematical Discourse*. London/New York: Continuum

Ondaatje, M. 2002. *The Conversations: Walter Murch and the art of editing film*. Bloomsbury Publishing: London

O'Toole, M. 1994. *The Language of Displayed Art*. London: Leicester University Press

_____ 2004. 'Opera Ludentes: the Sydney Opera House at work and play'. In K.O'Halloran, *Multimodal Discourse Analysis: Systemic-Functional Perspectives*. London/New York: Continuum: 11-27

Paine, F. 1985. 'Sound Mixing and *Apocalypse Now*: An interview with Walter Murch'. In E.Weis and J.Belton (eds) *Film Sound: Theory and Practice*. New York: Columbia University Press

Pasquariello, N. 1996. *Sounds of Movies: Interviews with the creators of feature sound tracks*. California: Port Bridge Books

Peirce, C.S. 1966. *Selected Writings*. New York: Dover

Pike, K. 1967. *Language in Relation to a Unified Theory of the Structure of Human Behaviour*. The Hague: Mouton

Ponech, T. 1999. *What is Non-Fiction Cinema?: On the Very Idea of Motion Picture Communication*. Boulder: Westview Press

Postman, N. 1985. *Amusing Ourselves to Death*. London: Methuen

Plantinga, C. 1994. 'Movie pleasures and the spectator's experience: Towards a cognitive approach.' *Film and Philosophy* Volume 2. Available online at: http://www.hanover.edu/philos/film/vol_02/planting.htm

_____ 2005. 'What a documentary is, after all'. *The Journal of Aesthetics and Art Criticism*. 63 (2)

Pudovkin, V. I. 1985. 'Asynchronism as a Principle of Film Sound.' In E.Weis and J.Belton (eds) *Film Sound: Theory and Practice*. New York: Columbia University Press

Royce, T. 1998. 'Synergy on the Page: Exploring intersemiotic complementarity in page-based multimodal text.' In *JASFL Occasional Papers No.1*. Tokyo: Japan Association of Systemic Functional Linguistics.

Ruoff, J.K. 1992. 'Conventions of Sound in Documentary'. In R. Altman (ed) *Sound Theory/Sound Practice*. New York: Routledge.

Russell, B. 1921. *The Analysis of Mind*. Electronics Classics Series. Pennsylvania State University

Saussure, F. 1974 [1916]. *Course in General Linguistics*. London: Fontana/Collins

Schafer, R.M. 1969. *The New Soundscape*. Vienna: Universal Edition

_____ 1977. *The Tuning of the World*. New York: Knopf

Schreger, Charles. 1985. 'Altman, Dolby and the second sound revolution.' In Weis, E. and J. Belton (eds.) *Film Sound: Theory and Practice*.

Schudson, M. 2002. 'What's unusual about covering politics as usual'. In B.Zelizer and S.Allan (eds) *Journalism after September 11*. London and New York:

Routledge. 36–47.

Sergi, G. 2004. *The Dolby Era: Film sound in contemporary Hollywood.* Manchester University Press.

Silverstone, R. 1999. *Why Study the Media.* London: Sage

Sperling, Cass Warner and Millner, Cork. 1994. *Hollywood Be Thy Name: The Warner Brothers Story.* Prima Publishing

Stephens, M. 2006. *A History of News.* New York: OUP

Sterne, J. 1999. 'Television Under Construction: American Television and the Problem of Distribution 1926-1962." *Media, Culture and Society* 21(3): 503-530.

_____2003. *The Audible Past: Cultural Origins of Sound Reproduction.* Duke University Press

Stillar, G. 2005. 'Loops as Genre Resources.' *Folia Linguistica* 39(1/2):197-212

Straub, J.M. And D. Huilet. 1985. 'Direct Sound: An Interview.' In E.Weis and J.Belton (eds) *Film Sound: Theory and Practice.* New York: Columbia University Press

Tagg, P. 1991. *Fernando the Flute: Analysis of Musical Meaning in an Abba Mega-Hit.* Liverpool: Institute of Popular Music.

Thibault, P. 1991. *Social Semiotics as Praxis. Text, social meaning making and Nabokov's "Ada".* Minneapolis and Oxford: University of Minnesota Press

_____2000. 'Multimodal transcription of a television advertisement'. In

Baldry, A. P. (ed.), *Multimodality and Multimediality in the Distance Learning Age*. Campobasso: Pallidino Editore: 311-385

Truffaut, F. 1993. *Hitchcock/Truffaut*. France: Gallimerd.

Truppin, A. 1992. 'And then there was sound: The films of Andrei Tarkosvsky.' In R. Altman (ed) *Sound Theory/ Sound Practice*. New York/London: Routledge

Tumber, H. 1999. *News: A reader*. London: OUP

Van Dijk, T.A. 1993. 'Principles of Critical Discourse Analysis'. In *Discourse and Society*, 4(2):249-283.

Van Leeuwen, T. 1991. 'Conjunctive Structure in Documentary Film and Television.' *Continuum*, 5: 76 -115

_____ 1999. *Speech, Music, Sound*. London: Macmillan

_____ 2005. *Social Semiotics: An Introductory Textbook*. London: Routledge

Van Leeuwen, T and C. Jewitt. 2001. *Handbook of Visual Analysis*. London: Sage

Van Leeuwen, T. and R. Wodak. 1999. 'Legitimising Immigration: a discourse-historical analysis'. *Discourse Studies* 1(1): 83-119

Vaughan, D. 1992. 'The Aesthetics of Ambiguity.' In P.Crawford and D Turton (eds) *Film as Ethnography*, Manchester UP: 99-115

Ward, P. 2005. *Documentary: The Margins of Reality*. London: Wallflower

Weis, E. and J. Belton. 1985. *Film Sound: Theory and Practice*. New York: Columbia University Press

Whittock, T. 1990. *Metaphor and Film*. New York: Cambridge University Press

Williams, A. 1980. 'Is sound recording like a language?'. *Yale French Studies*, 60: 51-66

Winston, B. 1995. *Claiming the Real: The Documentary Film Revisited*. London: BFI

Wulf, C. 1993. 'Das mimetische Ohr'. *Paragrana* 2(1/2): 9-15

Wurtzler, S. 1992. 'She Sang Live, but the Microphone was Turned Off: The Live, the Recorded, and the Subject of Representation'. In R.Altman (ed) *Sound Theory/Sound Practice*. New York/London: Routledge

Yewdall, D. 2003. *The Practical Art of Motion Picture Sound*. Focal Press

Appendix 1: FILMOGRAPHY

A is for Autism (Tim Webb, 1992)

America Under Siege (BBC News 24; broadcast 9/11/2001)

An Inconvenient Truth (Davis Guggenheim, 2006)

Apocalypse Now (Francis Ford Coppola, 1979)

Brasseye (Chris Morris, 1997)

Bubble (Steven Soderbergh, 2005)

California Split (Robert Altman, 1974)

Casablanca (Michael Curtiz, 1949)

Commandante (Oliver Stone, 2003)

Countdown (Robert Altman, 1967)

Derrida (Amy Kauffman, 2003)

Faranheit 9/11 (Michael Moore, 2004)

Fight Club (David Fincher, 1999)

Full Frontal (Steven Soderbergh, 2002)

Lost Highway (David Lynch, 1997)

Mon Oncle (Jacques Tati, 1959)

Old San Francisco (Alan Crosland, 1927)

Planet Earth (BBC, 2006)

Stalker (Andrei Tarkovsky, 1979)

Supersize Me (Morgan Spurlock, 2004)

Thank You For Smoking (Jason Reitman, 2006)

The Company (Robert Altman, 2003)

The Corporation (Mark Achbar and Jennifer Abbott, 2003)

The Day Today (Armando Ianucci, 1994)

Touching the Void (Kevin Macdonald, 2004)

Traffic (Steven Soderbergh, 2000)

The Office (Ricky Gervais and Stephen Merchant, 2001, 2003)

Walking with Dinosaurs (BBC, 1999)

Zodiac (David Fincher, 2006)

9/11 (Jules and Gedeon Naudet, 2001)