Re-Sounding Images: 
sound and image in an audiovisual age

Collin Chua, BA with First Class Honours in Communication Studies.
This thesis is presented for the degree of Doctor of Philosophy of Murdoch University, 2007.
I declare that this thesis is my own account of my research and contains as its main content work which has not previously been submitted for a degree at any tertiary education institution.
Acknowledgements:

I would like to thank my supervisors, Gail Phillips and Mark Gibson.

The dissertation could not have been written without the love, support, and patience of my family.

I am also grateful for the comments provided by Greg Hainge.

This is dedicated to Keeva, Taylor, and Ava.
Abstract

This dissertation examines the evolving articulation of sound and image in contemporary culture, with particular reference to film. It argues that sound and image have undergone a historical machined separation, followed by a machined fusion or recombination. The machined fusion of sound and image has enabled the creation of *soundful images*, which are more than simply the sum of their parts. Through the infusion of sound, images are now routinely reinforced with a performed sense of presence, where they are made to *sound* more real, more powerful, more authentic. Through association with the image, sounds are reinforced to the extent of becoming ‘realer than real’. By tracing the history of sound and image from their initial machined separation to their subsequent machined fusion, it will be argued that a new relationship has been created that has shaped an influential new mode of communication and perception.
## Contents

Abstract 1

Introduction 3

Chapter 1: Sound and image in sync 25

Chapter 2: Sound and image after Edison 61

Chapter 3: Ingenious audiovisual deformations 99

Chapter 4: Secondary orality/aurality and audiovisuality 135

Chapter 5: The sound of an image 179

Chapter 6: Music, ambience, sound effects, and the flow of images 212

Chapter 7: Hyperreal soundful images 258

Chapter 8: An audiovisual imaginary 294

Conclusion 330

References 334


**Introduction**

This dissertation takes as its starting point the observation that we have come to live in a culture of audiovisuality; we are surrounded by images and sounds on an everyday level, a situation which is facilitated by the proliferation and naturalisation of media environments. Cinematic representations and cultures have had an immense impact on the ways we interact with and perceive the world. Lev Manovich calls cinema the “key cultural form of the twentieth century”.¹ Cinematic media have contributed to shaping a significant evolution of our culture’s previous modes of temporal and spatial consciousness, as well as transfiguring our embodied sense of subjectivity and existential ‘presence’. Vivian Sobchack has written: “It is obvious that cinematic and electronic technologies of representation have had enormous impact upon our means of signification during the past century”.² It can still be argued that “the cinematic still remains the sensorial dominant of this century, of a modernity defined by mass production, mass consumption, and mass destruction”.³ Television, video games, and computer-based media are now particularly influential newer audiovisual forms. However, as we live our contemporary everyday life immersed in an audiovisual and information age, cinema’s history of images and sounds should be viewed as the progenitor of audiovisual culture, and cinema’s ways of working through the relations of sound-image and concept should be acknowledged as having become particularly significant to our strategies of seeing, hearing, and saying.

---

¹ Manovich, 2001, p. 9.
Studies addressing our audiovisual culture have generally tended to focus on the potency of the visual image. It cannot be disputed that images have become an important aspect of and shaping force on our contemporary society. “Now the language of images is everywhere”, as Daniel Boorstin writes. Much less attention has been paid to what can be clearly discerned – that sound and listening have also been transformed with the emergence of new practices and technologies. Furthermore, an important, influential, and evolving articulation and interaction is taking place between images and sounds. This dissertation examines how sound and image are fusing together in contemporary culture. My argument is that sounds are increasingly associated with images, and images are becoming soundful, resulting in the emergence of soundful images. This dissertation theorises the machined fusion of sounds and images, and the implications of the emergence of what will be termed ‘soundful images’. It looks at how images are resounded in our audiovisual age, where images, fused with sounds, have become increasingly familiar as persuasive and pervasive elements in our everyday lives. The conjoining of sound and image, made possible by technological advances in the cinema, has become such a dominant mode of representation that we are now surrounded by soundful images in many different areas of representation. In theorising the machined fusion (or recombination) of sound and image, I will argue that machined sounds act to reinforce images, imbuing them with a heightened sense of authenticity and presence, so that they sound more real, more powerful, more authentic, with images appearing to be the natural source of sounds, and sounds appearing to emanate seamlessly and unproblematically from images.

---

The history of sound and image is the story of a relationship that has undergone a machined separation, followed by a machined fusion or recombination. The image track and the soundtrack, captured separately by two disparate recording mediums, are subsequently mechanically linked and exhibited simultaneously. Images and sounds interact and are fused together within the perception of the audience. The machined fusion of sound and image has enabled the creation of soundful images, which I will argue are more than simply the sum of their parts. Through the infusion of sound, images are now routinely reinforced with a performed sense of presence, where they are made to sound fuller, more alluring, and more spectacular. Through association with the image, sounds are reinforced to the extent of becoming ‘realer than real’. By tracing the history of sound and image from their initial machined separation to their subsequent machined fusion, I will contend that a new relationship has been created that has shaped an influential new mode of communication and perception.

This dissertation engages with the evolving articulation of sound and image primarily through the context of the cinematic. But, at the same time, my dissertation is not a specialised work on film. It will address the cinematic conjuncture of sound and image through the perspective of a more general cultural and social approach, with consideration of practical knowledge of the field, as well as of relevant scholarship. My argument draws in particular upon the analyses of theorists such as Baudrillard and Virilio in which the world has become a simulacrum, accessible to us only through mediated forms. The work of these theorists, by their own admission, does not consider the aspect of sound – my argument therefore extends the work of these theorists in order
to encompass the sonic realm – or, more precisely, the sonic realm in its articulations with the image.

There have been two major surges of attention paid to film sound – the first wave of (largely critical) interest emerged logically with the coming of sound in the late 1920s to early 1930s, and the second wave came about due to the innovation of Dolby sound and surround sound in the late 1970s. First, with the coming of sound in the late 1920s, came the simple debate about whether sound had a positive or negative effect on film, an art form which was endeavouring mightily to gain aesthetic autonomy from the theatre, and hence had no wish to revert to being merely ‘filmed theatre’. Thus, much of the early theorising within film sound dealt with questions of how precisely to deal with the intrusion of synchronised sound. Gradually, the conceptual transition was made, as film soundtracks began to involve notions of sound as constructed, instead of sound as captured. This led to a second surge of (more positive) interest in film sound, as Star Wars (1977) ushered in Dolby and the Surround Sound technology, and was hailed by film historians and technicians as marking a “second sound revolution”.

A number of key texts in the 1980s and 1990s initiated debates regarding film sound, opening up spaces that have enabled a better appreciation and understanding of sound in film. The groundbreaking special issue in 1980 of Yale French Studies entitled Cinema/Sound, which included work by French and American scholars, was an important catalyst in theorising film sound. Later, John Belton and Elizabeth Weis edited Film Sound: Theory and Practice (1985), a collection that presents essays by film historians

---

5 For an anthology of articles regarding this period, see Film Sound: Theory and Practice, edited by Elisabeth Weis and John Belton, 1985.
and theorists of the past sixty years, encapsulating the classical development of film sound and attitudes toward it from the late twenties to the eighties. Edited by Rick Altman, *Sound Theory Sound Practice* was published in 1992, broadening the previous field of research on sound, and renewing the debate over the importance of sound to cinema from a theoretical as well as a historical perspective. Michel Chion's *Audio-Vision: Sound on Screen* was published in France in 1990, and translated into English by Claudia Gorbman in 1994.\(^7\)

In the introduction to *Cinema/Sound*, Altman highlights the persistent assumption that film is essentially visual, in particular the fact that this perception is embedded in our habitual ways of speaking about film. Indeed, the very names for the cinema – the movies, motion pictures – emphasise the inscription of visible phenomena, intended for spectators (not auditors) who go to see (not hear) a film. Altman argues, on the contrary, that the soundtrack should not be seen merely as a secondary addition to the images on screen. Given the ‘silent’ beginnings of the cinematic experience, it is not surprising that sound and listening have often been undervalued, or even ignored. The importance of visual culture has been firmly established in film theory, with many early filmmakers being sceptical about the value of sound, a prejudice indirectly perpetuated by generations of critics who viewed the cinema as an essentially visual art. Sound was generally seen to be little more than a superfluous accompaniment. However, developments in film sound over the past few decades have resulted in a significant extension in acoustic depth, clarity, definition, and detail, granting the soundtrack a

---

\(^7\) In their influential textbook on film, David Bordwell and Kristin Thompson describe Chion as “the most prolific researcher into aesthetics of film sound”. (“Sound in the Cinema”, *Film Art: An Introduction*, 6th ed., 2001, p. 306.)
degree of heightened prominence, and gaining film sound a greater theoretical emphasis and appreciation. But even after the advent of the recorded soundtrack and synchronous sound, the overwhelming majority of film criticism and theory still remains resolutely image-bound. Today, despite the age of THX specifications, with the array of surround sound speakers in the contemporary cinema, the soundtrack is still relatively undervalued in the theoretical arena. The dialectical relation between sound and image demands the further study of the complex interactions that take place; this challenge has not consistently been met.

In 1930, the English film historian (and later documentary filmmaker) Paul Rotha lambasted “the attempted combination of speech and pictures”. He insisted that they are “two separate mediums, which appeal in two utterly different ways”. This is plainly no longer the case: what we see on the screen is, in part, determined by what we hear, and conversely, how we hear something is partly determined by what we see. In our audiovisual age, images and sounds are increasingly being fused together to form new sound-image events. In order to investigate how we apprehend sound and image together, we need to begin by looking more closely at the theorisation of sound. The theorisation of image, and of film, has on the whole generally focused on visuality, neglecting the relationship with sound, or (at best) relegating the investigation of sound to a footnote, tucked neatly away. However, the comparative critical neglect of sound, together with the

---

8 THX was developed by Tomlinson Holman for Lucasfilm. THX is the trade name of a high-fidelity sound reproduction system first developed for the cinema and subsequently for the home. The THX system is not a recording technology, and is mainly a quality assurance system. THX provides certified theatres with special equipment required for compliance; theatres become certified by meeting certain acoustic and technical requirements.


10 Walter Murch explains: “We do not see and hear a film, we hear/see it”. (‘Stretching Sound to Help the Mind See’, The New York Times, 1st October 2000.)
continued emphasis on visuality, makes for a theoretical blockage that cannot be surmounted simply by resorting to heightened attention to sound.\footnote{Altman levels this criticism against Chion and Silverman, among others; see \textit{Sound Theory Sound Practice}, 1992, p. 38.}

Mary Ann Doane has complained: “It has become a cliché to note that the soundtrack has received much less theoretical attention and analysis than the image”; indeed, it \textit{must} be acknowledged that: “In a culture within which the phrase 'to see' means to understand, the epistemological powers of the subject are clearly given as a function of the centrality of the eye”\footnote{Mary Ann Doane, 'Ideology and the Practice of Sound Editing and Mixing', in \textit{Film Sound}, edited by Weis and Belton, 1985, pp. 54-55.}. The argument that sound and hearing is both under-theorised and largely unappreciated has come about in reaction to the pre-eminence of the image and the visual; it has been successful in providing an initial stimulus for the study of sound and listening, but now there is a need for theorising on a more ambitious level. In our audiovisual age, we need to examine sound and image \textit{together}. The point I want to emphasise is that after the machined fusion of sound and image, the two senses are no longer primarily separate; however, they are still predominantly discussed in terms of autonomy and separation.

Film and sound editor Walter Murch reflects on his beginnings in film sound: “it was a revelation to me that the sound had to be recorded separately from the image and ‘cooked’ – edited and mixed – before it was finished”\footnote{In Ondaatje, \textit{The Conversations}, 2002, p. 11.}. The playback system quite literally separated the soundtrack from the image. Rick Altman has suggested that this factor has done much to shape our notions about the ideologies of sound and image on a theoretical level.\footnote{See ‘Introduction to Cinema Sound’, \textit{Yale French Studies}, 60, 1980, p. 6.} As Sean Cubitt observes, “If the relations between writing and sound...
have been fraught over the last century, the relations between sound and image, especially between the near-twin arts of recorded sound and cinematography, have verged on civil war”. The invention of technologies of inscription, writing, and recording has encouraged a separation of our senses, as seeing has become our most valued sense of perception. Visual technologies and practices have had a history and aesthetics quite separate from acoustic technologies and practices. This separation of visuality and orality/aurality created two separate sensory cultures that have existed for some time in a state of “civil war”, where seeing and visuality have been privileged over hearing and orality/aurality. This “civil war” still appears to persist on numerous levels, even as sound and image have increasingly converged to develop a new relationship.

The attempt to carve out a space for theorising sound has resulted in critical attempts to re-privilege the sphere of sound over image, with assorted claims that sound came first, or that sound offers a sense of presence, or primacy, etc. These attempts to highlight the importance of sound by disparaging image cannot ultimately be defended. One cannot just reverse what has already occurred, in terms of the privileging of the visual, the profusion of the image, the crowning of sight after the separation of the senses – it becomes clear that sound can only be spoken about in light of these historical trends, and not in terms of a revisionist history that attempts to talk about sound as the most important sense. Instead, sound has become grafted onto the image.

Nevertheless, the gap that persists in the reconciliation between sound and image gives rise to an enduring tension within the emergence of soundful images. To work through the theoretical absences displayed by the isolated, separatist theorising of sound

or image, it is necessary to theorise the critical convergence of sound and image. One of the clearest indicators of our privileging of the visual is the typically prominent situating of image as the source of sounds rather than the other way around. The separate recording of both visual and acoustic elements is generally successfully hidden by the increasing refinements of sound editing. Film sound works within an ideology of transparency: sounds are meant to emanate seamlessly and unproblematically from the image. The infusion of mediated sounds has created a more effective image, lending depth and presence to the image, creating a meaningful resonance between sound and image. As a result of the machined fusion of sound and image, sound can be said to now operate through a machined, sedimented, reified, reciprocal relationship with image.

Simply put, sound extends the image, and the image sounds more persuasive. Image and sound are forming a new synthesis; sounds have engaged in an increasingly sophisticated relationship with images. Soundful images have resulted from the pairing of sounds and images that originally had no relationship, that can now fuse together to present a new media event. The recuperation of sound and image has resulted in a new relationship between sound and image that has shaped an influential new mode of perception. This relationship, which once possessed the de-familiarising power of estrangement, is gradually becoming stabilised into a system of meaning, even as it is subject to the turbulences and debates swirling around any new relationship. Image and sound are being matched through contextual, historicised, cultural convention, naturalising the arbitrary relationship between sound and image. What kind of intertextuality or articulation takes place, then, when ‘realer than real’ sounds are fused together with images? How and why do sounds add a sense of life, presence,
dimensionality, and depth to the images seen on the screen? This notion of the *resonance* of images and sounds will be the critical impetus for this dissertation.

Daniel Boorstin has lamented the creation of a modern world “where the image, more interesting than its original, has itself become the original”.\(^{16}\) This point, of course, can be applied to recorded sound, and particularly to soundful images. What is particular about soundful images is that sound allied with the image in the late modern period is not necessarily the ‘natural’ sound that would emanate from the real object behind the image in the real world, but can be a constructed sound which is felt to be more real than reality itself. In other words, the grafting of ‘unnatural’ sound (sound not of the real source or quite simply mediated sound) creates a ‘soundful’ image in which the relation between image and sound is not natural, nor indexical, not representative, not reproductive, but rather *resonant*, which is to say that there is a natural affinity to the pairing that creates a new aesthetic – or media – (hyper)reality. The hyperreality of the simulation produces a new original that both draws on reality and re-appropriates reality. I will argue that as new sound-image juxtapositions and couplings which were once strange and defamiliarising become familiar, frequently used soundful images will become naturalised, ‘unproblematic’ invocations of ‘reality’, thereby losing their power of estrangement, and providing hyperreal models of ‘reality’.

Soundful images, resulting from the constructed conjoining of hyperreal images and sounds, have become, so to speak, more ‘real’ and alluring than any real event. In the remainder of this introduction, I will extend on and illustrate this central theme in slightly

---

\(^{16}\) Boorstin, 1992, p. 204.
more depth, touching briefly on a range of pertinent ideas, in order to provide a launching pad for the more developed discussion to follow.

Today, television is a medium comprised of recorded programs. Sound effects are rarely done live for any show. The shift from live television to recorded programs that began in the late 1950s has come to privilege the model of the carefully constructed film soundtrack. However, back in the days when live news broadcasts were the dominant format, whether for radio or television, sound effects were added to the news live; and a fundamental use of sound effects in television was for ‘sweetening’ – adding sounds to (presumably) make something better. As Mott notes, “This was (almost) never done to deceive the viewing public, but simply to give the news a more professional and realistic sound”.17 To this end, what the audience at home heard in the days of live news usually took only a few frantic minutes to arrange. Robert Mott tells the story of how the news covered the ‘important’ atom bomb tests that were taking place in Nevada during the Second World War – “although the filmed pictures of the detonation were clear and more than a little frightening, the sound was terrible. There was so much noise and hiss, the soundtrack was unusable”. The only option was to use sound effects in an attempt to present the awesome impact of a sound that appropriately captured the magnitude of an atom bomb, which of course no one had heard before. “There were no sounds in the library to compare with it and so the sound had to be created… in a hurry”. Mott confesses that the impact of the atom bomb sound (which, he adds, he would not be

---

17 Mott, 1993, pp. 61-62.
surprised to find now catalogued in sound effects libraries) was ‘sweetened’ with an obscure recording of the roar produced by the Mogambi Waterfalls, in Africa.\footnote{Quotations taken from Mott, 1993, pp. 70-71.}

This practice of sweetening can be traced back to the early days of sound film, where synthetic noises were devised to replicate real ones that were either impossible to obtain, too explosive, or somehow sounded false and inauthentic when recorded. Rifle fire was faked by breaking kitchen matches; wind through the trees was produced by the old stage technique of revolving a canvas strip over a cylinder of wooden slats; a falling body was simulated by dropping a ripe pumpkin; collapsing buildings were replicated by tearing heavy paper very near the microphone. Some of these techniques are still usefully employed, but a range of other innovative and imaginative techniques and procedures have evolved alongside the steady progress of technical capabilities. Increasingly sophisticated techniques have transformed editing and postproduction technologies, allowing new combinations of sounds and images, helping to form new models of representation.

Nowadays, as Mott observes, sound effects are “done by committee in a starship-type room lined with computers and called a post-production room”.\footnote{Ibid, p. 70.} The sound effects, all neatly catalogued in computers and stored on tape, disk, and/or hard drive, are added after the show or film has been given its final edit. The process of assembling and mixing these sounds can take months for a high-budget television program or even longer for a feature film. Sound designer Gary Rydstrom tells us that animal sounds such as growls and coyote howls were employed as sweeteners for the ravenous sounds made by the fires in the 1991 movie \textit{Backdraft} which tells the story of two fire fighting brothers who
do not get along. Rydstrom remarks that the sweeteners are not heard as animal sounds; rather, on a subconscious level, they give the fire a degree of “intelligence or a complexity it wouldn’t normally have”. Coyote howls were used to enhance the suck in of air, and many “of the fireball explosions were sweetened with monkey screams and different animal howls. Cougars make a great fire explosion sweetener. There’s a complexity to natural sounds, especially animal sounds, that is really wonderful”.20

Images have even come to speak in voices not necessarily their own. In Steven Spielberg’s well-known 1982 movie, *E.T. – the Extra-Terrestrial*, a sad and lonely little boy named Elliot (Henry Thomas) comes across an extra-terrestrial in his backyard shed. The alien, nicknamed ‘E.T.’, has been left behind by his colleagues' spaceship when government agents interfere with the peace-loving species' scouting mission. As sound designer Ben Burtt explains, “E.T. starts out making an almost alien sound, and it evolves slowly into speaking English”. The vocalised sounds made by E.T. comprises of “probably eighteen different animals or people” mixed together, with the English-speaking voice provided by a non-actor named Pat Welsh, “delivered in a kind of monotone, slowed down a little and slightly treated electronically”, with a “coating of animal sound” then being added. The actress Debra Winger also provided “some really interesting, raspy breathing”. The “animal breathing and snorts” would “offset the human element just enough”, operating to persuade the audience of the alien character of the voice.21

---

21 Ibid, pp. 146-147.
The infusion of mediated sounds has created a more effective image, as sound supplies the image with what Michel Chion would term “added value”. Sounds relate to images, opening up aspects of meaning and representation not implied by the images themselves. Sounds do not just duplicate what we see but contribute to how we perceive images; sounds draw us into images, markedly contributing to the depth and dimensionality of the images. For instance, shrill screams or the raucous blasts of an explosion, synchronised with images of their source and/or its environment, will leave an imprint on the spectator’s mind, prompting a closer engagement with the moving images. In other words, duplicative sounds might even work like an appropriate musical accompaniment, enabling a more intense interaction with the pictures. If sound can affect what we see in an image and how we interpret it, it is also well suited to showing us what is not in the image, thereby extending the scope of the image – off-screen noises and voices, for instance, suggest a field of reality beyond the framed confines of the image.

An important aspect of the increasingly sophisticated relationship between sound and image is that sound has reinforced image in a mostly invisible fashion. The ‘miracle’ of film sound is largely disregarded: a situation that, understandably, upsets and frustrates many of the people who are involved with film sound. Images speak with a multiplicity of voices! Landscapes sing with wondrous beauty and sensitivity! Objects make sounds that previously could only be dreamed about! Mystical creatures roar and move with awesome effect! Film sound is the half of the cinematic equation that exists in service to the cinematic image. Despite the fact that sound is as painstakingly recorded, constructed, edited, mixed, and structured as the image, with potentially hundreds of tracks and

---

elements comprising the filmic soundtrack, film sound is largely seamless in its effect – its constructed nature is typically unrecognised. And it is this seamlessness that conceals a great deal about the art, aesthetic, and semiotic, of film sound.

This dissertation will provide an analysis of the production and reception of soundful images. In straightforward terms, I am interested in how sound interacts with the image. I want to focus on how sound has come to be constructed as a grafted virtual presence, with authenticity arriving via a documented relationship to a particular sound. What occurs now is the *performance* of reality. And it is this persuasive ‘reality effect’ that leads to the creation of the hyperreal artefact, where a sound adds a sense of dimensionality and presence to ‘reality’ as it fuses with its accompanying image. David Toop writes that “cinema is a theatre of exaggeration in which every minute sound we hear is foregrounded with hallucinogenic clarity”. Images, paired with these hyperreal (or realer than real) sounds, have become more persuasive, more powerful, infused with more depth, life, and dimensionality.

‘Chapter 1: Sound and image in sync’ looks at the transparency and seamlessness of film sound, in terms of the generally overlooked impact and influence of sound on the image, where ‘invisible’ elements of the soundtrack – such as Foley, postproduction sound, SFX, ADR, etc. – fundamentally impact on images. The fusion of sound and image, of course, has been made possible by technological advances in the cinema, where machine perception is brought into play by cinematic technologies for seeing and hearing. In discussing the operation of soundful images, this chapter engages specifically with Michel Chion’s notions of added value and synchresis, in terms of examining the

---

‘magic’ of synchronisation, where sounds and images that were not originally associated are fused together. This chapter therefore establishes the basic premise that images can be reinforced by sounds, where sounds offers the simulation of presence (or, perhaps more accurately, a sense of ‘presentness’), where images are now more soundful and resonant, as there is an articulation between sound and image. The chapter also makes the suggestion that in order to properly appreciate soundful images, we need to provide a kind of genealogy, examining the origins, evolution, naturalisation, and circulation of hyperreal sound-image assemblages.

‘Chapter 2: Sound and image after Edison’ discusses the historical origins of soundful images. It looks at important figures such as Thomas Edison and Eadweard Muybridge, providing a historical overview of the developments of the phonograph and the kinetoscope, as well as the arrival of the cinema and the coming of synchronous sound, in order to contextualise the emergence of the soundful image. The phonograph and the kinetoscope, of course, were the twin inventions of a machine for hearing, and a machine for seeing. The phonograph enabled sound to be disembodied, for voices and sounds to be taken from bodies, and placed into a machine, while the kinetoscope first introduced the possibility for capturing moving images of reality, even if they were mute. By tracing the history of sound and image from their initial machined separation (where sounds and images are recorded separately) to their subsequent machined fusion (where sounds and images subsequently come together), we can see how a new relationship has been created that has shaped an influential new mode of communication and perception. This chapter establishes the emergent possibility for the re-embodiment of sound, for
images to be re-sounded, for voices to be re-sounded – and for these relationships to be naturalised.

‘Chapter 3: Ingenious audiovisual deformations’ examines the historical development, evolution, and context of “ingenious audiovisual deformations”, as facilitated and disseminated by the cinema, which have come to be naturalised and widely accepted as persuasive fictions of reality. By matching recorded dialogue to lip-movements, and sound effects to their apparent causes, audiences are supplied with a powerful, persuasive illusion of the ‘real’ presence of the characters, objects, and events onscreen (or even offscreen). Sound provides the guarantee of the fullness of the filmic world, persuading us of the completeness and coherence of the fiction. This chapter makes the argument that we are witnessing the continuing naturalisation of various sound-image assemblages – particular sound-image relations are naturalised and stabilised even though they are at heart always hyperreal. This premise is demonstrated by examining the evolving historical conventions around the relationship between image and sound, and situating this discussion against the discourse of fidelity. The notion of fidelity is recognised to be an idealised value, which shifts according to evolving conventions and practices – this recognition underpins a discussion on the evolving conventions around the relationship between image and sound.

‘Chapter 4: Secondary orality/aurality and audiovisuality’ develops a critical reading of Walter Ong’s notion of secondary orality. This chapter examines a variety of historical attitudes towards sound/image that have been founded on a binary opposition of hearing/sight, and suggests that for many contemporary readings of audiovisuality, image/sound and sight/hearing still have the tendency to be placed in a hierarchical
relationship, and situated in terms of a binary opposition. The theoretical work of Walter Ong and Marshall McLuhan is symptomatic of this divide, which provides enduring problems for theorising audiovisuality that need to be overcome. These issues in the work of Ong and McLuhan can be seen to be carried over (however unknowingly) in the conceptual projects of many of the modern theorists of cinema sound. Many modern critics have merely attempted to invert the hierarchy and to privilege sound over the image to put an end to the hegemony of the image. The history of the technological and critical reception of sound film ultimately shows that this binary opposition is not sensible, as these two realms are now inseparable; this chapter therefore argues that we are now faced with re-sounded images which need to be apprehended as a whole.

‘Chapter 5: The sound of an image’ engages with the question of how images and sounds are meaningful for us, examining the basis of how and why they resonate in meaningful ways. The chapter provides a reading of Christian Metz’s notion of “aural objects” – where he argues that we largely understand sounds through relating them to their sources (in other words, sound make meaning for us according to what makes the sound) – before extending to elaborate a conceptual model for understanding how sound and image resonate powerfully, shaping a media hyper(reality). The chapter builds on the point that sounds inflect the image with meaning by making particular connections. Separate images and sounds bear particular meanings, and carry distinctive denotations and connotations; sound-image events utilise these meanings through the process of combination, juxtaposition, reinforcement, and/or counterpoint. Sound can stabilise the image by allocating a space/time context, or drawing spectator attention in particular ways. Sound can also destabilise the image, by inflecting it with potentially contradictory
meanings and connections. Under the influence of sound, the image can cease to signify in a predictably visual manner, becoming interactive with the re-embodied sound. The practice of re-sounding images enables the possibility for sound to alter or shape the image’s narrative legibility, creating continuities or discontinuities not mandated by the image alone. The chapter makes the argument that an evolving subjectivity has resulted from the perceptual instabilities produced when sounds are disembodied, and when moving images find themselves speaking with persuasive voices, resounding with powerful noises. The perceptions initiated by the cinematic experience have come to produce and define a new reality, which provokes important questions regarding the production and reception of a ‘hyperreal’ aesthetic and perceptual experience. Media technologies have come to construct powerful sound-image events that stake compelling claims on our senses.

‘Chapter 6: Music, ambience, sound effects, and the flow of images’ provides an overview of the role of music in film, before proceeding to demonstrate that with the increasing importance of the role and impact of ambient sound and sound effects in the modern soundtrack, these aural elements have come to work almost like music. The chapter argues that these aspects of the soundtrack increasingly function in a similar way to film music, where film music is generally acknowledged to act as a vector, anchoring the flow of images and offering a set of viewing instructions for images that it accompanies. Sound effects and ambience have become integral features of the sound of an image. In presenting this argument, this chapter touches on a range of theoretical readings of film music, including Claudia Gorbman’s pioneering work, and Anahid Kassabian’s more recent account of the increasing popularity of what she calls compiled
scores. Also, this chapter discusses a number of historical precedents and influences for ambience and sound effects, such as such as musique concrète, the work of the composer John Cage, and animation music. The chapter builds on the work of previous chapters by arguing that the aim of SFX and ambient sound is to shape an impression of reality, a hyperreality. Images are filled in, reinforced, animated, and, in a sense, ‘completed’ by the infusion of sounds, noises, voices, and/or music.

‘Chapter 7: Hyperreal soundful images’ specifically discusses the work of two theorists: Jean Baudrillard and Paul Virilio. Baudrillard’s notions of the hyperreal and simulation are applied to soundful images, where the simulation is more alluring than the ‘original’. After all, a recurring theme is that simulation is not just confined to visual images, but can also be applied to recorded/replayed sound, and to soundful images. In addressing re-sounded images, this chapter foregrounds the point that images have come to be re-sounded in ‘unnatural’ ways, giving rise to a media hyperreality. Amidst much contemporary fascination with visual special effects, digital compositing techniques, and computer-generated imagery, it should not be forgotten that hyperreal sounds have become ubiquitous, increasingly fusing with the image, and operating as key components of soundful images. Paul Virilio conceives of the mass media apparatus itself in terms of warfare, where we are bombarded with images and sounds, or we are supplied with images and sounds through a “logistics of perception”. The chapter engages with Virilio’s critique of how mass mediated perceptions have come to substitute for immediate, embodied presence and experience, by extending upon his ideas about silence. In discussing hyperreal soundful images and their simulation of reality, this chapter will particularly refer to the contemporary form of soundful images, with regards
to digital sound-image assemblages that are increasingly alluringly spectacular, in terms of how they look and sound.

‘Chapter 8: An audiovisual imaginary’ draws on a by-now-familiar idea where the sonorization of the image implies that sound ‘fleshes out’ the image (i.e., grants images depth, presence, temporality, spatiality, and dimensionality), and the image embodies sound (i.e., claims sound as its natural property, frames itself as the source of the sound). Theorists have commonly spoken of the imperialism of the image – now, however, it might be more precise to speak of an imperialism of sound and image. Our audiovisual age’s ‘moving-image culture’ is not just about how we have come to see in different ways. Rather, we live in a dense audiovisual environment, a world filled with resounding images, re-embodied sounds, soundful images, which have affected our engagement with the world through the new language of images and sounds. The stockpiling of sounds and images in our memory has resulted in an audiovisual imaginary. As images now speak to us in voices not necessarily their own, soundful images are clearly affecting the way we see, hear, and make sense of events around us. The chapter juxtaposes the perspective of Virilio and Baudrillard with the work of Gilles Deleuze, for whom the modern image is audiovisual, founded in many respects on the power of a disjunction of sound and sight.

To conclude this introduction, I would like to situate my project within a wider conceptual context. Any act of critical inquiry should be considered as a sign of its times, and typically concerned with addressing the questions and anxieties of its particular age. We live in a time and a world of technological domination and media representations. The many debates over the so-called ‘crisis of representation’ have to do with the fear that all real meaning is mediated (distorted) by media technology. Some of our most
intense anxieties involve the increasing influence of media technologies, and include questions of commodification, simulation, and machine perception. Put simply, this dissertation is a study of specific media technologies as cultural (or social) technologies, as fundamental components in human communication and perception. Why has technology become such a pressing issue for us? Technological rationalism has become one of the primary means of social control; and media technologies have increasingly come to mediate communication and perception.

The media have come to possess a fundamental role in the production and reception of meaning, increasingly providing the means through which we relate to and understand the world, the things within it, and the people that share it with us. In our digital age, the dystopian/utopian mutterings of the electric age have persevered and grown more complex. Have we come to comprehend the world through mass-mediated (machine) perception? The aim of *Re-sounding images: sound and image in an audiovisual age* is to offer one particular perspective and answer to this pressing question, through an investigation of the implications of the machined fusion of sound and image.
Chapter 1: Sound and image in sync

In this chapter, I will examine and problematize the apparent transparency and seamlessness of film sound. This chapter establishes the basic premise that images are now more soundful and resonant – images can be reinforced by sounds, where sounds offer the simulation of presence (or, perhaps more accurately, a sense of ‘presentness’) – and sets out the argument that there is an evolving and increasingly sophisticated articulation between sound and image, which is based on the various possibilities of resounding images. My point is that it is as the language of images and sounds has increasingly permeated our everyday lives, sounds have come to be paired with images, to emanate quite naturally from images, to refer to images; the machined fusion of sounds and images establishes familiar conventions of sound-image events. In discussing the operation of soundful images, this chapter engages specifically with Michel Chion’s notions of added value and synchresis, examining the ‘magic’ of synchronisation, where sounds and images that were not originally associated are fused together.

The image, of course, has rightfully garnered a great deal of theoretical interest from a wide variety of commentators. Jean Baudrillard has made mention of “the disenchanted proliferation of screens and the profusion of images”.1 In a television interview, Jacques Derrida has commented on the ascent of the image, highlighting the success of “the image experiment we are conducting now”.2 Accelerated visual literacy enables spectators to comprehend complicated montages of images, parading at breakneck speeds across the screen. Images surround us, and screens are popping up

---

practically everywhere. Paul Virilio proclaims that: “Ours are cinematic societies”.³ We need to recognise, however, that our audiovisual age has given rise to new systems of meaning founded on the convergence of sound and image. There is an influential new language of soundful images that is being produced, circulated, and consumed, and its origins can be traced to various practices and technologies of recombining and conjoining images and sounds that emerged with the sound film.

Synchronisation technology has enabled a reliable and consistent linkage between the soundtrack and the image track, so that sound and image are played together in concert. Improvements in synchronisation techniques have been made over the passage of time. Sync systems include those that simply maintain synchronisation once it has already been established, without providing an absolute reference; and newer means, which provide both relative (correct speed) and absolute (correct start point) synchronisation. The oldest method (not including the cumbersome Vitaphone system), which ensures synchronisation via perforations in film, has been around since the start of film sound in the 1920s, and is still in widespread, if declining, use. The intermediate systems invisibly perforate tape, while the newer methods add an electronic yardstick to the length of the tape so that any position along the tape is labelled with a location or address. Digital hard disc recording technology also relies on this addressing method, called time code. These technical methods enable the machined fusion and operation of sound and image, releasing the persuasive impact resulting from the convergence of sound and image.

³ Virilio and Armitage, in *Paul Virilio: From Modernism to Hypermodernism and Beyond*, edited by Armitage, 2000, p. 27; there are claims, of course, that ours is a televsional society, or a computerized society – but it seems fair to say that these subsequent developments are founded upon the cinematic revolution of moving images accompanied by sounds.
The image track and the soundtrack, separated as they are captured by two disparate recording mediums, are mechanically linked and exhibited together. And as they are exhibited simultaneously, images and sounds interact and are fused together within the perception of the audience, presenting a coherent, mass-mediated reality. Sounds are ‘meant’ to emanate seamlessly and unproblematically from the image. Film sound generally has a powerful but invisible relationship with the image. This fusion of sound and image is usually so successful that film sound is typically taken for granted, even as sound shapes our reception of the image. Contemporary film soundtracks are thoroughly constructed, and film audiences have subconsciously come to expect images to sound larger than life. Voices are generally expected to be audible at all times, and material objects seen on the screen that make a noise should generally be distinctly heard. The sound-image convergence has evolved, becoming ‘unreal’, hyperreal, or ‘realer than real’.

The components of the soundtrack have been established through evolving conventions and practices which have gradually become established and stabilised as accepted modes of practice. James Lastra tells us that: “From 1932 on (and probably as early as 1930) film sounds were conventionally divided into hierarchically ordered tracks, with ‘dialogue’ assuming the leading role. This classification, however, is neither purely objective nor neutrally descriptive – it is determined by and attuned to Hollywood categories and imperatives”. The soundtrack can be divided into three different categories: first, immediate sounds which we consciously listen to; second, sounds that support the portrayed event or environment that are heard but do not attract direct

---

4 Lastra, 2000, p. 187.
attention; and third, background sounds that pass largely unnoticed but that provide the subconscious acoustic clues that hold together the reality of the scene. Sound effects technician Alan Beeby explains: “Faced with a bewildering medley of sounds, the problem is to select those which will illustrate the scene and the accompanying commentary or dialogue to the best advantage”. To meet this challenge, he recommends the use of a ‘three-stage plan’, where “the whole sound-scene is divided “into three main parts”. Beeby explains that “the ‘Immediate’ effect is to be listened to while the ‘Support’ and the ‘Background’ effects are merely to be heard… The ‘Support’ effect refers to sounds taking place in the immediate vicinity which have a direct bearing on the subject in hand, leaving the ‘Background’ effect to its normal job of setting the general scene”.

This method of constructing the soundtrack from various recordings has been very successful, and has become the predominant model in use today. The soundtrack therefore constructs an aural reality to accompany and extend the visual images.

Images and sounds are fused in an articulation that constructs meaning through a mutual process of juxtaposition, the twin senses of sight and hearing working in conjunction to create a holistic movie experience. Walter Murch comments that the re-association of image and sound is central to the creative contemporary cinematic experience. “Sometimes”, Murch explains, re-embodied sound, or re-sounded images, are employed “simply for convenience (walking on cornstarch, for instance, happens to record as a better footstep-in-snow than snow itself); or for necessity (the window that Gary Cooper broke in High Noon [1952] was made not of real glass but of crystallised sheeted sugar”; or even “for reasons of morality (crushing a watermelon is ethically

---

preferable to crushing a human head). In each case, our multi-million-year reflex of thinking of sound as a submissive causal shadow now works in the filmmaker's favour, and the audience is disposed to accept, within certain limits, these new juxtapositions as the truth”. ⁶

The introduction and evolution of a recorded synchronous soundtrack has contributed to a perceptual revolution that has wide-ranging consequences not only for the cinematic form, but also for contemporary perceptual practices. David Toop remarks that cinema has become “a theatre of exaggeration”, where every sound we hear “is foregrounded with hallucinogenic clarity”. For example, “the harsh scrape of a beard as Henry Fonda is shaved”, or “the hugely amplified noise of kicks, punches, even the passage of arms through the air in Chinese martial arts film; the tense creak of boat timbers, the squish of a knife into flesh, the ricochet of bullets off rocks in countless westerns”.⁷ What has become ‘realer than real’, are the soundful images that have resulted from the machined fusion of sound and image: these soundful images model augmented mass-mediated realities that refer to simulated ‘original’ events. We need to investigate the evolving relationship between a sound recording and an image that are presented in sync, that are linked in a process of machined fusion. James Lastra has remarked that “the theoretical discussion of sound within film studies has returned again and again, almost obsessively, to a single central problem. What is the relationship between a sound recording and the sound it purports to depict?”⁸ This question can be re-worked slightly, in order to emphasise the development of sound-image events: what is

---

the relationship between a sound recording and the image it is fused with? Sounds and images no longer necessarily stand alone, even if the meanings and effects generated by sync sounds are usually attributed to the image alone or the film overall.

On one hand, a film’s ‘soundtrack’ often consists of several layers independently recorded and mixed, which then overlap one another. On the other hand, the very nature of recorded sound events enables one recorded sound to be seamlessly integrated with another in editing, with the join being unnoticeable. In current practice, the art of mixing soundtracks essentially consists in smoothing rough edges by degrees of intensity. Viewers are typically quite capable of differentiating between the visual objects portrayed on film; however, listeners tend not to perceive sound in this manner, as listeners hear sound as a ‘whole’ despite the soundtrack being deliberately constructed from numerous separate elements. The majority of soundtracks are thus designed according to an ideology of transparency; and it is this presentation of sound (where separate sound elements coalesce, instead of being isolable as various different sound elements) that enables the storytelling power of sound. Sound – instead of being perceived as mere acoustic fragments – invokes a narrative flow occurring within a situation, presenting a space/time of meaning. The cinemagoer no longer notices the soundtrack, only the effect of the soundtrack. But, as Michel Chion reminds us: “if you alter or remove these sounds, the image is no longer the same”.9

Sync sound has gradually fostered the important emergence of sound effects (a category which includes background noises and Foley effects). Synchronised sound effects today play a significantly larger role in the soundtrack as compared to the early

---

years of film sound. Many films utilise sound effects (such as blows, body falls, collisions, and explosions) as key contextual elements within scenes. These synchronised noises anchor the visuals, thereby convincing the audience that everything that is heard and seen is ‘real’, providing a sense of presence and authenticity to the images. Images have been re-sounded and hence sound more convincing; sounds refer back to their images, and it is these images and not the physical source of the sound which grants these sounds their meaning and presence.

Sync sound’s hyperreal performance of an ‘authenticity to source’ is modelled on the persistent notion of reproducing an ‘original’ sound-image event. The development of sync sound has contributed new standards of ‘authenticity to source’, as sounds emanate from images with remarkable clarity, presence, and impact, as images are more soundful, dramatic, and spectacular. This transition has coincided with an evolution in auditory values, as evidenced by the essential shift that has taken place within the notion of acoustic fidelity. Previously, fidelity meant that ‘correct’ spatial characteristics should be adhered to, as sound-image perspective was rigorously monitored. However, present-day stereo sound can be deployed with complete spatial incoherence between what is seen and what is heard, as audiences have become accustomed to (and indeed expect) ‘unrealistic’ depictions of sound and image, such as clearly audible actors who are portrayed in long shot. The emergence of soundful images has irrevocably altered the standard form of perceiving cinematic representations, as new models of authenticity provide what audiences generally willingly accept as a more ‘direct’ and ‘immediate’ evocation of the ‘original’ event.
It is typically not readily apparent or obvious on viewing a film that it is very much a conscious decision on the part of the sound editors to record, edit, and mix sound effects which are carefully allocated to the objects on the screen. Tomlinson Holman explains that: “the simplest definition of sound in a Hollywood movie is ‘See a car, hear a car.’ That is, everything that we see on the screen that we expect to make a sound does make a sound”.\(^\text{10}\) For example, projected on the large screen, it is not evident that the major car chase sequences for John Frankenheimer’s 1998 action thriller, *Ronin* (starring Robert De Niro) were filmed MOS (or, without sound).\(^\text{11}\) Custom recordings were made for the myriad car sound effects, including all kinds of tire skids and slides, the distinctive rumble of driving over cobblestone streets, whining gear boxes, and the various mechanical demands. For the audience seated in the cinema, the climactic chase through Paris is seen and heard in stunning detail, with the precise audio detailing granting the action sequences a whole new dimension. Except for occasional dialogue exclamations, the audio sequence is almost entirely assembled in postproduction with sound effects and Foley alone.\(^\text{12}\)

The synchronisation of formerly separate image track and soundtrack enables the audience to be *right there* at the scene, as the cars accelerate with a roar, screech around corners, rumble over cobblestones, and reverberate with a barely restrained energy. This filmic portrayal of reality is made possible by sync sound; the images on the screen

---

\(^{10}\) Tomlinson Holman, *Sound for Film and Television (Second Edition)*, 2002, p. 45; Holman is an acknowledged authority in sound for film and television, and the creator of the THX sound system.

\(^{11}\) There is a humorous story behind this abbreviation; Stephen Handzo explains that MOS is “an archaic term dating from the German-speaking colonies of directors in early-talkie Hollywood who expedited shots not requiring sync-sound dialogue by calling for shooting ‘mit out sound’”. (‘A Narrative Glossary of Film Sound Technology’, in *Film Sound: Theory and Practice*, edited by Weis and Belton, 1985, p. 395.) However, David Yewdall reports that this story is a widespread myth that is, unfortunately, not true; the more prosaic story is that MOS instead stands for Minus Optical Sound (the abbreviation dating from the period of optically recording sound on film). (*Practical Art of Motion Picture Sound*, 1999, p. 41.)

portrayed as making a noise are carefully matched up with the appropriate sounds. The innovation of sync sound is taken for granted by the present-day moviegoer, and no longer inspires a sense of wonder. Instead, the spectator only notices when images and sounds are not synchronised appropriately. Along with this increased familiarity with the reception of sound-image events come several common misconceptions about film sound: first, that the sound is recorded along with the image; second, that the sound is unproblematically ‘captured’ and simply paired with the image; and third, that sound requires much less work than the tedious labour undertaken for the presentation of the visuals. This ‘natural taken for grantedness’ of film sound is a result of the apparent seamlessness and ‘self-containment’ of sound-image events.

In fact, much of film sound is recorded separately from the image, including a significant percentage of dialogue. Film sound is constructed and painstakingly synchronised with the cinematic visuals; and finally, the construction, recording, editing, mixing, and all-round production and postproduction of sounds calls for a great deal of work that, if successful, reinforces the apparent seamlessness of the sound-image event. “As the sound track reaches the audience via the theatre sound system”, LoBrutto explains, “it creates the illusion that the sound for a film has been captured by a single, magical microphone which records the dialogue, sound effects, and music on-set in perfect balance”. However, “just as every visual component in a film is designed and executed by the writer, director, cinematographer, and design team, each single sound in a film is carefully conceived, chosen, recorded, edited, and mixed by an array of sound artists and technicians”.13 The layers of which a soundtrack can be comprised include: the

---

film score (the music, whether diegetic or non-diegetic), the production soundtrack (sound that is actually recorded during filming of the visuals), production dialogue (dialogue that is recorded during filming of the visuals), ambient backgrounds (environment or location sounds), ADR or group Walla (individual dialogue or group vocals that are specifically recorded during post-production to match up with the visuals), Foley effects (physical movement sounds made by Foley artists that are later matched up with visuals), and sound effects (made with synthesisers, or by manipulating source recordings).

The inventiveness of Foley artists can be illustrated by looking at and listening to *The Exorcist* (1973), a film based on the 1971 novel by William Peter Blatty, which was based in turn on a real, documented case of demonic possession. The story is of a 12-year-old girl, Regan (Linda Blair), possessed by a demon who purports to be the Devil himself ("Now kindly undo these straps!" the possessed Regan snarls, in one of the more memorable lines). Regan exhibits strange symptoms, including physical convulsions, violent outbursts, demonic speech, levitation, and great strength. When all medical possibilities are exhausted, her mother is sent to a priest who is also a psychiatrist. He becomes convinced that Regan is possessed and he and a second priest experienced in exorcism try to drive the spirit from Regan before she dies. In an interview, Foley artist Ross Taylor reveals: “I did the scene in *The Exorcist* where Linda Blair throws up in bed on the exorcist. I drank a warm 7-Up, waited a few minutes, and then I had this terrible belch that they recorded, and that’s what they used".15

---

14 Named after the pioneering work of Jack Foley, a Foley artist creates sound effects to accompany the noisemaking movement of actors seen onscreen.
15 In LoBrutto, 1994, p. 63.
Rick Altman refers to this phenomenon as “the sound film’s fundamental lie: the implication that the sound is produced by the image when in fact it remained independent from it”.\textsuperscript{16} This is explained by the persistent audiovisual phenomenon which convinces the audience to believe willingly that whatever sound is synchronised to the image depicted onscreen is in fact the sound being emitted by the image.\textsuperscript{17} The mental alchemy of sound and image in sync presents a persuasive filmic reality, with images that make sounds, which are contained within sonic environments, and sounds that open up the image, which imply the presence of other images. As mentioned, the sounds that are fused to the images come from a variety of sources. Therefore, by matching a performer with a sound that he or she (or it) has not necessarily created, a dramatic performance is constructed which represents an idealised real. Film sound has played a key part in the development of a cinema that depends on the symbiotic relationship between image and sound to create stylised narratives that demand a new, naturalised audiovisual competence from its audience (which is in part required in order for them to be able to remain wilfully and happily ignorant of the stylisation). This evolution of the intricate ‘self-containment’ and referentiality of the sound-image relationship has influenced new ways to both see \textit{and} hear, through the dissemination of persuasive new cinematic modes of perception.

Synchronisation works almost like magic. It convinces both the eye and the ear. Here is a sound, and here is an image. The sound might not have any logical connection to the image. For instance, a Foley artist might squish cornstarch in a bowl with her hands. The image might be of a woman running in snow. Once the image and the sound

\textsuperscript{16} Altman, ‘The Evolution of Sound Technology’, in \textit{Film Sound}, edited by Weis and Belton, 1985, p. 46.
\textsuperscript{17} See Sonnenschein, \textit{Sound Design}, 2001, p. 35 for further reading.
are played together, they make perfect sense. The woman is heard and seen by the audience to be running in snow. Or, the visuals of the giant boulder that chased Indiana Jones in *Raiders of the Lost Ark* (1981) (incidentally not made of real stone but of plastic foam) are accompanied by a sound effect created through recording a Honda Civic station wagon rolling down a gravel slope with the engine off. These are examples of the “audiovisual contract” and naturalised semiotic system that allows the audience to believe that whatever sound is synchronised to the image is the sound being emitted by the image. Chion has coined the term “synchresis” to refer to the mental fusion between a sound and a visual when these occur at exactly the same time; he points out that this join results independently of any rational logic.

Synchresis can even work with images and sounds that strictly speaking have nothing to do with each other, shaping monstrous yet irresistible patterns in our perceptions. “Synchresis”, Chion explains, “is the spontaneous and irresistible weld produced between a particular auditory phenomenon and visual phenomenon when they occur at the same time”. Synchresis is a Pavlovian process, even if “it is not totally automatic. It is also a function of meaning, and is organised according to Gestaltist laws and contextual determinations”. Sync sound lays the groundwork for soundful images, by enabling the prosthetic pairing of sound and image that did not originally have a signal/source relationship, by deploying the semiotic force of particular sound-image relations. Michel Chion explains: “sound enriches a given image so as to create the

---

18 Ibid, p. 58.
20 Chion, 1994, p. 63; Chion’s concept of ‘added value’ refers to “the expressive and informative value with which a sound enriches a given image so as to create the definite impression… that this information or expression ‘naturally’ comes from what is seen, and is already contained in the image itself”. (Chion, 1994, p. 5.)
definite impression, in the immediate or remembered experiences one has of it, that this information or expression ‘naturally’ comes from what is seen, and is already contained in the image itself”.21

Joseph Anderson's paper at the 1992 Ohio University conference on film sound provides a solid basis in cognitive science for Chion's phenomenologically derived assertions. Experiments on perception and attention in infants indicate that ‘the processing rule’ may be this: if auditory and visual events occur at the same time, the sound and image are perceived as one event.22 As Chion observes, we see the image differently when it is paired with sound, as compared to when we see the image alone; and we also hear sound differently when it is heard in juxtaposition with the image than if the sound were heard by itself.23 Filmic reality is made up of sounds and images that are more than the sum of their parts. The interaction of recorded images and recorded sounds – manipulated and subjected to the capabilities of montage and mixing – has given rise to the construction and organisation of re-embodied sounds and re-sounded images, due to the machined fusion of image and sound.

I want to suggest another productive way to think about the relationship between image and sound, where sound can be said to be re-embodied in the image. Sound needs to be embodied in order to make meaning, for within the filmic context, sounds are understood as being a descriptive property of their source. We create meaning from listening by sourcing the sound – as listeners, we decide or discover what the source (emitter) was that created the sound (emission). If someone walks along a long corridor

21 Chion, 1994, p. 5.
22 See p. 35 in Wide Angle volume 15 number 1, January 1993.
and hears a strange noise, it might be meaningless until he or she gets closer and sees, through an office door, a cleaning woman engaged in vacuuming the office. The high-pitched whine is coming from the vacuum cleaner. When a disembodied sound reaches a listener, he or she constructs a source (or emitter) for the sound (or emission), thereby contextualising and creating a series of associations and meanings. In other words, a sound only makes meaning when attributed to a source; a sound is necessarily a sound of something. The classificatory question is posed: what is that sound? Meaning is made when it is identified as (for example) the sound of the wind blowing through an open window upstairs. A loud percussive explosion is heard as a gunshot when the source of the sound is indicated to be a gun. The sound of a loud, meaty thud is interpreted to be the impact of a punch against a man’s jaw, if the appropriate visuals are provided. Sound is separated (disembodied) from its material source, and then paired with another source, allowing sound to be re-embodied through a fusion of sound and image. The re-embodiment of sound, and the re-sounding of image, allows the sound of a sack of potatoes thrown into a mound of earth to be synchronised with the image of a man falling from the second floor window to the garden below. The audience then perceives the images and sounds that document the painful landing of a human body.

The impact of sync sound is made possible by the human proclivity for organising disparate sensory perceptions, as the brain appears to impose patterns out of chaos even if none objectively exist. For example, when the image of a boy walking on screen is accompanied by music, it will seem as though his steps are in rhythm with the music, as a result of the human tendency toward entraining sound and image. Certain combinations

---

24 In Chapter 6: ‘The Sound of an Image’, I will be elaborating on this sound-to-source model by engaging with Christian Metz’s essay ‘Aural Objects’.
between sound and image come to be expected by the audience. Once audiovisual links are established, they can be broken apart, recombined, emphasised, and contrasted. The new relationships between sound and image are disseminated to the audience, who then use this audiovisual vocabulary to navigate their way through the filmic narrative. As various sound and image combinations have been explored within the cumulative history of mass media, audiences have largely developed an accelerated comprehension of audiovisual code. This ‘code’ is based on how sound extends the narrative of the image, and vice versa; this articulation of sound and image has gradually become stabilised as a coherent system of meaning.

There is another instance of how sounds and images are organised and interpreted within human perception. Synchronisation creates a “spatial magnetisation of sound by image”.25 This refers to the perceptual process of locating a sound in the space of the image, no matter where the sound might originate from in the viewing space. For example, the audience sees a character walking across the screen, and the footsteps are balanced in the centre (as on mono TV sets), and not moving from the left to the right speakers (as in a surround sound movie theatre). The sound is perceived not only in temporal sync, but also in correct spatial placement, as human perception organises and interprets sounds and images in order to re-establish ‘normality’ to the situation. The audience will ‘hear’ the footsteps as moving from left to right. Sound-image events are founded upon the premise of a single, original, unified, audiovisual event that precedes representation. Sync sound cues up, references, and installs the audiovisual perceptual template of this ‘original’ event.

25 Chion, 1994, p. 70.
Spectators have become very good at discerning whether or not sound is in sync with the picture. The majority of viewers are able to tell if the sync is off by two frames, and many viewers are able to notice when the sound is one frame out of sync. However, sounds and images manifest a curious attraction for each other when played simultaneously, even if they are not, strictly speaking, ‘in sync’. Walter Murch provides an example: “If you have one robot walking down the hallway, the feet need to be in sync. If you have two robots walking down the hallway, the feet still need to be in sync”. However, with “three robots walking down the hallway, the mind basically says, ‘You expect me to watch every footfall and decide whether that sound goes with it? I’m not going to do it.’ Therefore you can get away with nothing being in sync”. There is a certain threshold when the mind surrenders in the attempt to track synchronisation, and simply organises sound and image together.

As mentioned, the operation of sync sound remains largely transparent; the audience mostly does not realise (or is not concerned with) precisely how the soundtrack operates in constructing filmic reality. Let me reiterate the reasons for this transparency. The ideology underlying the design of the soundtrack is that of a seamless fusion with the image, effacing the tremendous amount of work required to construct an effect of ‘realism’ and naturalness. The technical apparatus and process are carefully concealed. The sync soundtrack does not possess the visible joins and cuts that are demonstrated within the montage of the image track, and the separate recording of both visual and acoustic elements generally becomes successfully hidden by the increasing refinements

26 Holman, Sound for Film and Television (Second Edition), 2002, p. 5.
27 Interview with LoBrutto, Sound-On-Film, 1994, p. 87; the term ‘sound designer’ was originally coined for Walter Murch, who has an imposing résumé in film sound and film editing, including work on films such as Apocalypse Now, The English Patient, The Godfather Part II and Part III, etc.
of sound editing. The moving images on the screen are typically and unproblematically understood to be the natural source of sounds. The ambition of virtually all mainstream filmmakers is to conceal the separation of the soundtrack and image, in the interests of constructing and presenting a persuasive filmic reality. The result is an apparently ‘self-contained’, ‘unproblematic’ sound-image event that is, however, founded on an originary absence, and an initial separation.

Images make sounds, and sounds are made by images; this is how we perceive the persuasive filmic realities that are presented in our cinematic experiences. In order to persuade an audience to enter the ‘reality’ of the filmic experience, the sound editor frequently establishes synchronisation between sound and image within a scene as quickly as possible, typically through lip-sync. After all, the human voice still holds a remarkable attraction for an audience due to its fundamental role in communication. Once sync has been demonstrated, the picture and sound work in close correspondence and counterpoint to elaborate upon the audiovisual narrative. When the machined fusion of sound with image is successful, the impression received by the audience is that sound emanates quite naturally from the image itself, even as sound’s meaning is given form and shape by the image, just as the image shapes and gives meaning to sound.

The audience willingly believes that whatever sound is synchronised to the image depicted onscreen is in fact the sound being made by the image. If this belief is challenged, the persuasiveness and ‘self-containment’ of the filmic reality is shattered. For example, the reason why films that have been dubbed into another language can be disconcerting for viewers is because the lack of lip-sync strains the believability of the sync soundtrack. The voices heard are obviously not coming from the images seen on the
The human voice has possessed a central role in the film soundtrack ever since the beginning of sound film, a period aptly called the ‘talkies’; and the vococentric nature of cinema is clearly revealed in the practice of sync sound. The voices that are played in sync with the actors are heard distinctly, and in close perspective. In spatial terms, voices generally stay front and centre, rarely straying even if actors are moving around or are located elsewhere in the filmic diegesis. Ensuring that voices are synchronised to the lip movements of the actors remains an essential basic component for the maintenance and apparent seamlessness of the cinematic reality.

In the simplest terms, sync sound is defined as a soundtrack that is employed in synchronisation with an image track. What does sync sound enable? A simple example of the possibilities available to the juxtaposition of sound and image is demonstrated in *As Good As It Gets* (1997), starring Jack Nicholson and Helen Hunt. In the restaurant scene, Melvin (Nicholson) is courting Carol (Hunt) in his typically neurotic, obsessive compulsive fashion. As he moves towards a decisive revelatory moment, as the plot reaches a significant point of denouement, the natural sounds of the restaurant (the murmur of other diners, the clatter of silverware, the clink of the glasses, etc.) fade away, leaving a ‘perceptual vacuum’, or perceptual close-up, for the audience to observe these two characters resolving their relationship, issues, and feelings. The hushed silence serves to focus attention on the emotional event that is occurring; and the meaningful words and actions that are taking place. When the tension in the scene is subsequently resolved with a humorous shift, the sounds of the restaurant unobtrusively return to their previous levels. The climatic moment has passed.
However, more sophisticated interactions are possible between images and sounds. Sounds can animate images by supplying a sense of presence and physicality, helping to build a ‘natural’, distinctive filmic illusion. Sound can operate as punctuation and guide for the flow of images, providing acoustic environments that underpin, anchor, and impose a rhythm for the parade of images. Sound can operate in combination and in conjunction with image, shaping the meaning of image, either in support or in counterpoint. Sync sound can bridge cuts between images, enabling the smooth transition from one scene to another. A famous example can be observed in Alfred Hitchcock’s *The Thirty-Nine Steps* (1935), where the sound-image link is utilised in pioneering fashion to add a new form of narrative. The movie employs Hitchcock’s classic theme of an innocent man, framed by circumstantial evidence, who must run cross-country from police and spies alike in the frantic attempt to clear himself and find the real enemies. The protagonist (Richard Hannay, played by Robert Donat) finds he must escape from his London flat after discovering that the mysterious woman he had given refuge the night before has been murdered. Later, as his landlady discovers the murdered corpse in the flat and opens her mouth to scream, there is a visual cut to a tunnel, matching the woman’s open mouth, and a train emerges with a screaming whistle that blends over the sound of her scream. Effortlessly, we have been transferred into the next scene, as Hannay escapes on the train, en route to his adventures. Subsequently, at another point, the rhythmic noise of the train’s wheels on the track becomes synthesised into a vocal chant of “He mustn’t; he mustn’t; he mustn’t”. Hitchcock was one of the earliest filmmakers to breach the limitations of a purely causal sound and image link, instead overlapping, combining, and juxtaposing visual and acoustic elements.
Surround sound technology allows sync sound to locate images spatially, as the direction of the sound paired with the image constructs a sense of space. Sync sound offers an alluring promise of the ‘real’ by constructing and performing an acceptable version of fidelity, authenticity, and ‘reality’. Images can be re-sounded and reinforced – this has been an important transformation – as images have generally become more audible in their new prosthetic relationship with sounds. For example, audiences have become accustomed to gunshots sounding louder and larger than life. David Sonnenschein describes the process of creating a soundful image: “The sound of a gunshot will often begin with the actual recorded firearm, but then much can be added to enhance the effect of the shot itself, the impact of the bullets and the ambience reverb”. For “larger weapons like torpedoes in The Hunt for Red October [1990], the layering of animal growls and shrieks, a Ferrari engine, and a screeching screen door spring helps imbue the weapon with a vengeful purpose of its own”.28 Not only gunshots, but explosions, bodily movements, even speech – images are more ‘present’, soundful, dramatic, and spectacular, as sounds reinforce images with additional narrative layers, dimensionality, and perceptual impact. Sound ‘strengthens’ the image by animating it with source-qualities that it would not ordinarily have by itself, via the process of re-embodiment. Sound, recorded, manipulated, and played back, fuses with the image to create a re-sounded image, supplying a presence that is now part of the image, operating to supply a supplementary ‘presentness’ (subsequently concealed within and claimed by the image).29

---

29 I am making use here of Jacques Derrida’s account of the supplement: see Of Grammatology, 1976, pp. 144-145; I will provide a fuller reading of Derrida and supplementarity in Chapter 4.
Sound can create new hierarchies within the framed image; drawing spectator attention to incidental features purely because they are sources for dialogue, sound, and/or noise. Under the influence of sound, the image can cease to signify in a predictably visual manner, becoming interactive with the re-embodied sound. The practice of re-sounding images thus enables the possibility for sound to alter or shape the image’s narrative legibility, creating discontinuities or continuities not mandated by the image alone. Separate images and sounds bear particular meanings, and carry distinctive denotations and connotations; sound-image events utilise these meanings through the process of combination, juxtaposition, reinforcement, and/or counterpoint. Audiences come to expect certain relationships between sound and image; certain ‘natural’ sounds are forgotten, and new machined relationships are naturalised in a process of constructed realism. New processes of perception have developed and have become transparent, which have come to be constantly delivered and circulated by the mass media.

Audience expectation remains largely subconscious because the filmic illusion is still intended to be seamless; paying attention to the ‘constructed’ nature of the soundtrack would ruin the power of cinematic realities, and reveal the originary gap in the reintegration of sound and image. New sound-image matches have been established based on the fact that some ‘realistic’ sounds do not play nearly as dramatically as sounds created by totally different sources than those seen in the film. Images have become more audible because in the process of matching an actor with a sound that he or she had not necessarily created, or by pairing a material object with a noise that has been carefully crafted, a dramatic performance is constructed which represents an idealised real.
sync soundtrack grants film a degree of clarity and dimensionality unobtainable in the real world, making the everyday world seem almost impoverished by comparison.

Astonishingly, we then come to the realisation that reality is out of sync; the varying speed of light and sound in our physical world allows the plain observation that sound should arrive after the image of the event. For example, when watching a carpenter hammer a nail from a distance, the visuals of the hammer striking the nail should not really be accompanied simultaneously by the sound of the impact – sound arrives later. Filmmakers frequently deal with this perceptual disparity, often forcing sounds into sync that in physical reality would be temporally asynchronous. Sound travels 1130 ft/sec (or 345 meters per second) in air at room temperature. This is equal to about 47 ft. (or approximately 14 meters) of travel per frame of film at 24 frames per second. Because sound is so slow relative to light, it is conventional editing practice to pull up the sound on motion picture release prints by one extra frame, printing the sound ‘early’, and thus producing exact picture-sound sync 47 ft. (or about 14 meters) from the screen, which is usually the centre of the movie theatre. The audience’s perceptual expectations have been modified by their exposure to mass media: sound should be in sync with the image. A new mode of reality is being installed.

The notion of fidelity can thus be said to be more an effect of synchronisation. In other words, “many effects in films today are designed with only cursory thought given to achieving fidelity of diegetic sound: for example, a car door slamming shut might have the hollow metal ring of a prison door snapping into place”. Lastra has commented

---

30 Holman, Sound for Film and Television (Second Edition), 2002, p. 5.
“that any fixed or purposeful relationship between sound and image legitimately may be thought of as synchronised”. However, the predominant “view of synchronisation and, more broadly, our current theories of film sound too often derive entirely from a consideration of sounds linked to pictured sources – their presumed real world sources – as if representation were merely a matter of duplicating that world”.32 Synchronisation does not necessarily entail the simple causal linkage of sounds and images, and the sound-image model Holman calls ‘See a car, hear a car’ is not the only means of utilising the language of sound and image, which can be based on more ambitious combinations and juxtapositions. Walter Murch clarifies further: “In most films everything is ‘see it = hear it’. The sounds may be impressive, but since they come from what you’re looking at they seem to be the inevitable shadow of the thing itself”.33 This is one possible sound-image model: however, instead of merely ‘duplicating’ the world, a more complex articulation is possible between sound and image.

The mechanical linkage of the soundtrack and the image track therefore contains interesting possibilities for sound that are not directly linked to a source pictured on the screen. Cinema has developed as a form that employs judicious amounts of sync points (lip-synced voices, sound effects visually matched with their sources, etc.) which operate as perceptual foundations for the filmic reality. Sounds that are not directly linked to sources onscreen have generally been allocated a smaller role in the soundtrack (mood music, nondiegetic sounds and noises, etc.). The development of various aesthetic possibilities of these sounds, and the resultant fluid (resonant) relationships between

---

32 Lastra, Sound Technology and the American Cinema, 2001, p. 94.
sound and image, are yet another aspect of soundful images that should be thoroughly investigated. The notion is that sound and image should constitute two parallel and loosely connected tracks, neither dependent on the other. However, any “fixed or purposeful relationship between sound and image may legitimately be thought of as being synchronised”, for sounds are unavoidably implicated in a synthetic relationship with the coinciding images projected on the screen. This allows the interplay of two separate, now converging, perceptual dimensions.

Murch explains that “by choosing carefully what to eliminate, and then adding back sounds that seem at first hearing to be somewhat at odds with the accompanying image, the filmmaker can open up a perceptual vacuum into which the mind of the audience must inevitably rush”.  

35 See Ondaatje, The Conversations, 2002, p. 120.

This idea of what Murch terms “metaphorical sound” involves emphasising the visual by artificially focusing on a possibly disjointed or ‘unrealistic’ layer of sound. This enables another mode of authenticity to be performed by “the re-association of image and sound in different contexts”.  

As Murch explains, “The more you get into the emotional end of things, the more you draw upon the metaphorical use of sound. Reality can only go so far and then you have to go beyond reality, beyond the frame”.  

36 When faced with metaphoric sound, which can be said to be sound that does not directly match and thus ‘expands’ the visuals, the human mind tends to search for and organise deeper patterns of meaning, as sound-image juxtapositions refer to, overlay, connect, and subvert narratives. To explain this in terms not used by Murch, the tension produced by the metaphoric distance between sound and image serves to provide an
imposed dimensionality to the separate yet synchronised fields of sound and image, unifying them into a soundful image, suffused with presence and depth.

The convergence of sound and image therefore also refers to audiovisual ‘mismatches’, or apparent disparities between sound and image. In the course of his interview with Walter Murch, Michael Ondaatje responds: “I find it fascinating that you feel that with such ‘metaphorical sound’ you can reach a deeper truth about the atmosphere of the scene”.37 In Murch’s terms, the “metaphoric distance” between the images of a film and the accompanying sounds should be continuously changing and flexible, allowing the audience to make the connections between the disparate narratives set in motion through employing apparent disparities between sound and image. And this “deeper truth” contained in sound-image events refers to a metaphorical ‘realism’ beyond the literal ‘See a car, hear a car’ model, where sound appears to support the active narrative and only occasionally creates independent dimensions. The complexity of this type of sound-image event – where sound is freed from a literal (visual) source – therefore expands the scope of soundful images, by imparting a sense of additional dimensionality and depth, through the intertextual juxtaposition and recombinant potential of different visual and acoustic narratives. If sound can affect what we see in an image and how we interpret it, it is also capable of showing us what is not contained in the image. In other words, sound can operate not only to offer clarity, but also to create ambiguity.

These expanded soundful images offer a range of potential effects. A sense of schizophrenic experience can be elaborated, where the sound tells a particular story

---

which is not in conjunction with the image. Sound-image events can reinforce a feeling of authenticity, or establish a feeling of disconcerting arbitrariness, where sound goes one way, and the image goes another, and nothing is as it seems; where either the sound or the image could be a hallucination. Sound-image disjunctions can also invoke an unsettling conjunction of past and present, by prompting a sense of memory, where certain sounds (or images) are associated with certain sequences, images, histories, or locations. Sounds can also comment on the images (or vice versa) by calling on a particular sense of emotion, or providing certain perceptual cues. These various sound-image combinations and juxtapositions establish networks of meaning that trace and refer to glimpses and echoes of other stories, other worlds, other plots, other objects, other beings; there is also the capacity to focus attention on particular elements of the filmic reality.

Walter Murch argues that the “re-association [of sound and image] should stretch the relationship of sound to image wherever possible”. Murch advocates imparting a sense of depth and ambiguity to the cinematic experience by suggesting that there is a complete story behind the selection of material projected on the screen, with hints of buried or sedimented storylines that loom in the distance. The effect can be achieved, Murch remarks, by always suggesting through sound that something is going on off-screen, by providing a layering of sounds that suggest other worlds and other plots. This can be illustrated by looking at the opening scene of Apocalypse Now (1979), the modern epic ostensibly about the Vietnam War but concerned on a deeper level with the descent into madness. Directed by Francis Ford Coppola, and based on Joseph Conrad's book

---

Heart of Darkness, the movie follows Captain Willard on his mission into Cambodia to assassinate a renegade Green Beret who has set himself up as a God among a local tribe. “Saigon...”. Willard says – “I’m still only in Saigon... Every time I think I’m gonna wake up back in the jungle”. The jungle sounds that fill Willard's hotel room do not have an onscreen source, and the only way to resolve the “disparity between what we are seeing and hearing is to imagine that these sounds are in Willard's mind: that his body is in a hotel room in Saigon, but his mind is off in the jungle, where he dreams of returning”.40

During this examination of the articulation between sound and image, as we look at Willard restlessly cooped up in his hotel room in Saigon and hear the sounds of the hotel and the street outside the window slowly transform into jungle sounds of crickets, birds, and mosquitoes, as the scene gradually reveals the emotional dilemma at the heart of the character, the realisation comes that soundful images are able to invoke, extend, and overlap subjective/objective reality. The juxtaposition of sound and image enables virtual connections, based on the interplay of sound and image, thereby suggesting supplementary meaningful narratives. A sound-image event therefore establishes “a differential network, a fabric of traces referring endlessly to something other than itself, to other differential traces”.41 This is how soundful images work: they are founded on the links, connections, and oppositions between sounds and images, enabling and/or subverting sound-image interactions and conventions, making meaningful narratives out of different combinations and juxtapositions. As Chion likes to remind us, “Sound shows

---

40 Murch, ‘Stretching Sound to Help the Mind See’, The New York Times, 1st October 2000; Walter Murch was the sound designer for Apocalypse Now.
41 Jacques Derrida, ‘Living On’, in Deconstruction and Criticism, 1979, p. 84.
us the image differently than what the image shows alone, and the image likewise makes us hear sound differently than if the sound were ringing out in the dark”.42

Murch’s use of ‘metaphorical sound’ can be productively related to Chion’s analysis of acousmatic sound. “Acousmatic”, as Chion explains, “describes ‘sounds one hears without seeing their originating cause.’”43 For Chion, acousmatic sound is opposed to visualised sound, or sound that is accompanied by the sight of its source or cause. “In a film”, Chion says, “an acousmatic situation can develop along two different scenarios” – a sound is visualised first, and then subsequently acousmatized; or a sound is acousmatic to start with, and is then visualised afterward. In the first case, a sound is associated “with a precise image from the outset. This image can then reappear with greater or lesser distinctness in the spectator’s mind each time the sound is heard acousmatically. It will be an ‘embodied’ sound, identified” and classified with an image. For the second case, “common to moody mystery films”, the sound’s source or cause is initially kept a secret, before it is revealed in a climatic dénouement. “The acousmatic sound maintains suspense, constituting a dramatic technique in itself”.44 Chion is fascinated with how sound holds power over the image, by suggesting something (or someone) that is not seen, thereby maintaining mystery, unsettlement, or even terror.

It can be observed that acousmatic sounds derive their force from referring to (or implying) a source that is visually absent. From this broader category, Chion derives his concept of “acousmêtre”, a disembodied voice, or a form of ‘phantom’ character specific to the art of film that obtains a sense of mystery, omniscience, and omnipotence derived

44 Ibid, p. 72.
from being heard and not seen. In examining this disembodied voice that affects the visual dimension, but nonetheless remains primarily an aural force, Chion privileges a space where sound is central to the creation of the sound-image event. His concept of the acousmêtre can be read as an attempt to challenge the dominance of the visual by asserting the value and contribution of film sound, by emphasising the acoustic dimension. The acousmêtre offers a means of analysing film that privileges sound as being central to the construction of particular sound-image events. In Chion’s terms, the acousmêtre does not ignore the image; rather, the presence of the voice that is accompanied by an absence of image demands a hearing-seeing that originates with sound. Sources which are withheld (acousmatic sound; the acousmêtre) result in a sense of unsettlement, mystery, or even terror, which reaches a climax (or a sense of release) when the source is revealed (Chion’s example is Norman Bates’ mother, in Hitchcock’s * Psycho*).

Chion’s privileging of sound, however, does not sufficiently acknowledge that these re-sounded images need to be apprehended as a whole. It is the machined fusion of sounds and images that enables acousmatic sound to have the power it possesses, by establishing conventions of sound-image events that trace and refer to glimpses and echoes of other stories, other worlds, other plots, other objects, other beings. Chion does not agree with the idea of a pre-reading of the relationship between sound and image, remarking that this ‘codedness’ is subject to a system of convention that is “imprisoning us in a binary logic that has only remotely to do with how cinema works”. However, sounds obtain a sense of meaning by referring to their source, which may be object,

---

subject, or environment. The acousmêtre is founded on the gap, the lack, the originary presence of absence contained with the soundful image. Chion’s notion of the acousmêtre is a privileging of the power of sound; however, the power of the acousmêtre needs to be framed in terms of an articulation with the image, in terms of a kind of intertextuality with the image. The acousmêtre invokes the ghostly (absent, or perhaps more accurately, spectral) presence of the other; it is heard but not seen, triggering, fabricating, and referring to connections between sound-image interactions, enabling and/or subverting perceptual expectations and conventions. In a way, the acousmêtre draws upon our audiovisual imaginary, which has been stocked by our immersion in mass-mediated images and sounds. “In Psycho, the mother is first of all a voice”, Chion explains. But I want to suggest that as the language of images and sounds has increasingly permeated our everyday lives, the acousmêtre is no longer simply “first of all a voice”; as voices have come to be paired with images, to emanate quite naturally from images, to refer to images – the disembodied voice of the acousmêtre trails a set of spectral images in its wake.

It has to be said that sound-image events have not only affected the cinematic form, but have also influenced the individual form and aesthetics of sound and of image. Sync sound has enabled images to be heard more distinctly; the people and objects portrayed on the screen are more audible, more soundful, more ‘dramatic’, and generally louder than they actually are. In addition, normal, everyday sounds are becoming ‘unheard’, as sound is increasingly making meaning by referring to the appropriate visual sources, just as soundful images are gradually coming to replace mere silent images. Complaints of a lack of fidelity and authenticity, as well as an erosion of reality, have

been a direct reaction to the mass media delivery of new audiovisual practices, which are threatening the previously stable conceptual landscape of perceptual ‘realism’. The dramatic portrayals of tremendously soundful images have caused an irreversible shift in our boundaries of perception.48

Not only are re-sounded images more audible, more ‘present’, and more persuasive, but sync sound has gradually trained listeners to pay attention only to sounds that have been sufficiently extended through their prosthetic source-relationship with specific images. With the “sonorization of the image”,49 images are becoming acknowledged as a ‘natural’ source of sound; and images have become innately soundful, as sound has become a ‘natural’ property of the image. It is important to remember the sociality of the evolving relationship between sounds and images; this relationship is organised through conventions and meanings that are founded on sociality and a culturally determined language. Significantly, the ‘binary logic’ that Chion refers to can be interpreted as the fear of sound-image relations rigidifying, or solidifying, into conventionalised systems that confine and regiment the flow and singular expressive power of the cinematic art form. Any notion of a ‘codedness’ therefore succumbs to a reductionist binary structure that can only ever seek to represent a poor duplication of reality. The warning against mere representation, as opposed to the cinematic potential of an uncharted ‘realism’ that leaks and expands outside the boundaries of the frame, comes

48 It should be noted that we are the final components in the cinematic machine – we are not passive viewers, but rather active participants in constructing filmic realities. Without us, the combination and juxtaposition of images and sounds would not make sense – we bring along the necessary social, cultural, and historical networks and apply them to reading the interplay of images and sounds.

49 Virilio has employed this term in a short essay (‘Silence on Trial’, in Art and Fear, 2003, pp. 69-96) where he writes of the ‘sonorization of the image’ with regards to the death of art; there is no longer the appropriate meditative silence with which to contemplate the work of art. Although the basic premise is similar – that images are now no longer silent – I will be re-deploying this term for different purposes, and with an emphasis that differs from Virilio’s original usage.
from an attitude of caution. Hyperreal sound-image events are transforming into naturalised, institutionalised codes that represent a realer than real, that present a new real, by ‘writing over’ our perceptual reality.

The convergence of sound and image has escalated the persistent concerns regarding associated notions of authenticity and fidelity. These questions are centred on the fundamental issues of technology, mediation, and representation. Advances in recording technologies and techniques have enabled the capture of a wider range of sounds. Progressions in editing and rerecording technologies and methods have allowed the manipulation and transformation of sounds. Improving sync sound technology has facilitated the technical and institutional practice of sound montage. Heightened exhibition standards and technologies (such as stereo sound, surround sound, Dolby, and THX) have been implemented in cinema theatres, reinforcing the presentation of sounds with remarkable physical clarity and power. These sounds, expanded in frequency range, command heightened attention, and grant the persuasive sense of extended presence to the images on screen.

The machined separation of sound from source, and sound from image, has problematized the formerly stabilised, ‘natural’ relationship between sound and image. Through the process of synchronisation, a process has been set in motion that seeks to ‘recuperate’ the images and sounds that have been separated by their inscription onto separate surfaces (the celluloid image and the soundtrack). The work of the soundtrack can be explained as an attempt to ‘heal’ the disjunction between sound and image perspective through the machined fusion of sound and image, which operates by fusing a disembodied sound to a particular image. The practice of re-embodying sounds and re-
sounding images has instead inaugurated a prosthetic relationship between sound and image, marshalling a perceptual transformation that is permanently altering the relations between sound and image, sound and source. Prosthetic sound-image relations are apparently quite capable of reconstructing a new reality, assembling new contexts along the way, where normal, everyday events have been silenced by the new authenticities propagated by the emergence of soundful images.

The accusation of the effacement of the real, or of a falsification of reality, is grounded on sound film’s ‘reproduction’ of an ‘original’ sound-image event that never occurred. The ‘originary’ non-coincidence of sound and image privileges the mechanisation of perception and installs the possibility of virtual reality, which has become so pervasive and predominant that it raises the fear we are witnessing actual reality disappearing into the spectacle of the ‘unreal’, or the ‘realer than real’. The concealment or vanishing of this disappearance (as naturalised soundful images reshape the boundaries and borders of perceptual reality) has been denounced as the most insidious strategy in the substitution of human perception with machined perception.50

The presence of technological mediation has been compressed, elided, and rendered transparent. The increasing permeation of reality with mechanical equipment presents aspects of reality that appear to be free of technological mediation, much like the way the camera is not seen, and the microphone not noticed.

---

50 See Jean Baudrillard, *Simulations*, 1983, and Paul Virilio, *The Art of the Motor*, 1997, for two accounts of the mass-mediated effacement of reality; I will provide a closer reading of Baudrillard and his key concepts of simulation and the hyperreal in Chapter 7.
Now, computer-generated imaging technologies are said to be capable of producing images that accurately simulate the look of objects in the real world.\textsuperscript{51} Dolby surround sound technology has enabled intricate, clear, powerful three-dimensional soundscapes that play in support of and counterpoint with the images projected on the screen. Production capabilities, digital facilities, editing packages, the proliferation of video-cameras and microphones, and the incredible growth of media industries, have allowed voraciously voyeuristic audiences to watch practically anything, anywhere, anytime. It has been argued that cinematic/televisual illusions have taken over the world; the new instrumentality of sounds and images, delivered to the audience so persuasively and pervasively that depicted events appear realer than real, in a ceaseless supply of a portrayal of ‘reality’ augmented and reinforced by perceptual technologies. The initial term in ‘sound film’ is now superfluous – sound is taken for granted. The sound designer David Sonnenschein observes: “When sound works well with image, the impression is that sound is already contained in the image itself”.\textsuperscript{52} And it can be said that these apparently self-contained soundful images are the legacy of post-Edisonian transformations of sound and image: we no longer consume images or sounds separately, for sounds are made by images, and images \textit{sound} ‘realer than real’.

To summarise, a \textit{soundful image} results from the machined fusion (or synchronisation) of sounds and images, where images are augmented by sounds that are realer than real. Soundful images are images that are reinforced with a performed sense of authenticity and presence, where images sound more real, more powerful; more ‘authentic’. The most fascinating aspect of soundful images, or sound-image events, is

\textsuperscript{51} See my subsequent discussion in Chapter 7.
the way these events trigger and fabricate connections between the potential multiplicity of perspectives, bending time and warping space, enabling and/or subverting sound-image interactions, making narratives out of different combinations, juxtapositions, and durations. Sound does not simply serve the *sole* purpose of suturing the reality gap present in cinematic experience; cinema does not necessarily provide an ‘unproblematic’ depiction of reality, as it can construct a realer than real spectacle, and it can interrogate and provide a critique of reality, as sounds and images unfold in an intricate trace of meanings that invokes the cuts, blockages, flows, and totalities, and partialities that form the perceptual subjectivity of thought and of the body.

As new sound-image juxtapositions and couplings which were once strange and defamiliarising become familiar, frequently used soundful images become naturalised, ‘unproblematic’ invocations of ‘reality’, thereby losing their power of estrangement, and providing hyperreal models of ‘reality’. Re-embodied sounds and re-sounded images are acknowledged as ‘real’ constructions, portraying ‘natural’ sound-image events. Contemporary audiences have grown very familiar with the modes and conventions of sound-image representations, perceptions, and events. Here again we must confront the antinomy present within soundful images: it might never be a ‘seamless match’, and there might always be an originary absence at the juncture of sound-image events, but these hyperreal events – what the theorist Jean Baudrillard would term “third-order simulations”,\(^53\) that construct and refer to ‘original’ events that never happened – these soundful images are beginning to insinuate and reinforce a convincing illusion of presence, propagating what can be called ‘sincere fictions’ that we, the audience, have

---

come to willingly partake in. The creation of soundful images/re-sounded images that are blatant simulacra are actually being accepted as reality or more real than real. In order to properly appreciate the impact and pervasiveness of soundful images, we need to provide a kind of genealogy, examining the origins, evolution, construction, naturalisation, and circulation of what are now hyperreal sound-image assemblages.
Chapter 2: Sound and image after Edison

In order to explore the emergence, implications, and characteristics of soundful images, we need to examine the invention and development of synchronous sound. This chapter discusses the historical origins of soundful images, in terms of contextualising the emergence of the soundful image, arguing that sound and image have undergone a historical machined separation (where sounds and images are recorded separately), followed by a machined fusion or recombination (where sounds and images subsequently come together). The twin inventions of the motion picture camera and the phonograph just before the turn of the Twentieth Century initially appeared to establish the trend of the machined separation of sound and image, with each having its own recording technology, and with no means as yet invented to bring sound and image together. In the early days of cinema, the movies were ‘silent’, with a live musical accompaniment. Sound and image were only brought together when a consistent means was discovered to synchronise moving pictures and an accompanying soundtrack. This machined fusion of image and sound is the key to my narrative, which will examine the implications of a new language of sounds and images and soundful images as they become increasingly influential in furnishing our contemporary mediatized world. By tracing the history of sound and image from their initial machined separation to their subsequent machined fusion, we can see how a new relationship has been created that has shaped an influential new mode of communication and perception. My contention is that this history has significant implications for ideas about the authenticity and meaningfulness of images and sounds. This chapter establishes the emergent possibility for the re-embodiment of
sound, for images to be re-sounded, for voices to be re-sounded – and for these relationships to be naturalised.

Historically, the invention of the sync soundtrack was inspired by the logical desire to reproduce an event simultaneously in both the visual and the acoustic dimensions. In order to gain an appreciation for the emergence of soundful images – once astonishing, now taken for granted – it is necessary to take a trip through time. The first stop – the closing years of the 19th century, via Friday, June 5, 1998, when an attentive audience at the Coolidge Auditorium stares at the grainy images projected on the screen, whilst listening intently to the well-worn soundtrack, played over the sound system from an audiocassette. This is the first public exhibition of a historic cinematic recording that has come to be known as the *Dickson Experimental Sound Film*, which was made sometime during the mid-1890s as part of a series of collaborative experiments by the Edison laboratory staff to perfect the Kineto-phonograph.\(^1\) The Kineto-phonograph technology marked the first serious attempts to combine image and sound in a pre-recorded, synchronised, relationship. As the audience watches and listens, a crackly voice is heard: "The rest of you fellows ready? Go ahead!" This terse instruction, now laden with historical significance, has been introduced as possibly the first cinematic ‘speed’ and ‘action’ captured on wax. Projected at 30 frames per second, the 21-second film features two men dancing, as a violinist (identified to be either W.K.L. Dickson or Charles D'Almaine) plays into a large megaphone.

\(^1\) Film historian Charles Musser speculates that production probably occurred during the period of September 1894 to early April 1895 (Musser, *Edison Motion Pictures: 1809-1900: An Annotated Filmography*, 1997, p. 178).
The showing of the *Dickson Experimental Sound Film* was the climax of the 1998 Domitor² conference, held at the Library of Congress, Washington, D.C., from June 1st to 5th. This short fragment, received enthusiastically by the conference crowd, is now officially the oldest synchronous film in existence, pre-dating the previous record-holders by approximately 25 years. The 35mm nitrate print was acquired from the Edison Historical Site and preserved by the Library of Congress in 1964. The simultaneously recorded wax cylinder soundtrack had been discarded into a bin of broken Edison cylinders until recently rescued and finally put together. (The sound cylinder remains under the jurisdiction of the National Park Service.) A fortuitous connection was subsequently made between the soundtrack and the short fragment of film, which enabled the film and the soundtrack to be finally recombined and restored to sync under the supervision of sound designer Walter Murch, with assistance from Sean Cullen.³

For a contemporary audience that is aware of its historical significance, the *Dickson Experimental Sound Film* acts much like a time machine, providing a riveting viewing experience. (“With this new invention”, a newspaper would announce after the motion pictures were first unveiled, “death will no longer be absolute, final. The people we have seen on the screen will be with us, moving and alive after their deaths”.)⁴ Walter Murch describes his restoration project: “It was very moving, when the sound finally fell into sync: the scratchiness of the image and the sound dissolved away and you felt the

---

² The Domitor association was launched in 1985, with the aim of facilitating the study of early cinema (from its beginnings to 1915). The name comes from the dawn of cinema, before the turn of the nineteenth century, when the father of the Lumière brothers proposed the name ‘Domitor’ for their projector of motion pictures.

³ Sean Cullen digitally converted the film to 30 fps (frames per second; video format) from the original recorded projection speed of 40fps. The soundtrack, contained on a cracked 1890's cylinder, required patient refurbishment. Murch designated various possible sync points, with the length of the track then being adjusted accordingly.

immediate presence of these young men playing around with a fast-emerging technology”. Nevertheless, the early film experiments, no matter how ambitious and far-sighted, fell well short of contemporary standards of sync sound. Although sound and picture move at the same time, there remains the unresolved issue of whether there was actual mechanical linkage at the time of recording or reproducing. Despite its bold reintegration of sound and image, the Kineto-phonograph technology did not achieve sufficient impetus to set off the sound film revolution. The point of interest with this example, with this artefact, is that it shows us how naturalised soundful images are now, and it gives us something of a sense of how revolutionary the invention of this assemblage of sound and image actually has been.

There is some contention as to who was the ‘inventor’ of the cinema: some commentators support the Lumière brothers in Paris, December 1895, while other historians credit Thomas Edison in New York four months later. It has also been claimed that Charles Cros, an amateur scientist and inventor in the turn-of-the-century Parisian bohemian cultural scene, was the first to conceive of a practical phonograph in April 1877, although he did not produce a prototype due to a lack of funds. It has been argued that Cros also described the basic technology of motion pictures as early as 1867. Furthermore, Edison’s vision for motion picture technology appears to have been actualised by W.K.L. Dickson, an inventor in his laboratory. While Edison seems to have conceived the idea and initiated the experiments, Dickson apparently performed the bulk of the experimentation, leading most modern scholars to assign Dickson with the major

---

credit for turning the concept into a practical reality. However, for the purposes of this discussion, I have chosen to ascribe the blame (or credit) to Edison.

Things were never quite the same after Edison. The phonograph, invented by Edison in 1877, actually came before Edison’s subsequent experiments for achieving moving pictures, and provided immediate inspiration for his own, slightly later work on motion pictures. Edison’s well-known manifesto proceeded as follows:

In the year 1887 it occurred to me that it was possible to devise an instrument which would do for the eye what the phonograph does for the ear and that, by a combination of the two, all motion and sound could be recorded and reproduced simultaneously. This idea, the germ of which came from a little toy called the zoetrope and the work of Muybridge and Marey and others has now been accomplished… I believe that in coming years by my own work and that of Dickson, Muybridge, Marey and others who will doubtless enter the field, that grand opera can be given at the Metropolitan Opera house at New York… with artists and musicians long since dead. ⁷

This would lead, of course, to the birth of the motion pictures. However, Edison’s hope for the combination of sound and moving pictures would not immediately come to fruition – and when sounds and images were eventually recombined, it was not Edison who managed to achieve it.

Nevertheless, the invention of the phonograph provided the eventual means for sound to be synchronised with image. The phonograph was the instrument for a new sort of reproduction, one that extended and transformed our conception of the human senses by recording them as they occurred in real time, serving as a model for further

---

experiments in recording and reproduction. More importantly, the phonograph had in
effect separated the human senses, divorcing ear from eye in a way beyond what was
accomplished by writing and print technologies. The phonograph was a machine for
hearing, which could also ‘speak’ in myriad voices. The phonographic rebirth of hearing
and speaking altered the balance of utterance and audition, and also destabilised the ways
in which the ear and the eye perceived the world. Referring to an emanating body or
object that was no longer directly present, sounds were separated from their sources and
set free to roam in space and time. Voices were separated from their subjects and
deposited into a machine. The phonograph “represented a new day in aurality through its
ability to return virtually any sound back again and again into the sensorium and into the
historical register”. The phonograph’s “ability to hold any one sound in time and keep all
sounds in mind – produced a new status for hearing, which was energetically entered into
libraries, laboratories, literature, artistic ideas, and philosophies”.\(^8\) Captured and
preserved out of its time and space, transported and played back, sound was allowed to
become an artefact, an object for analysis, dissection, and manipulation. Sound, formerly
ephemeral and temporal, no longer had to vanish into thin air.

John Peters persuasively argues that: “In many ways the phonograph is a more
shocking emblem of modernity than the photograph”. For although people have, since
time immemorial, been able to preserve images by drawing or painting, fixing “sound
events requires an altogether different sort of inscription, namely, the ability to capture
the serial flow of time itself”.\(^9\) The fascination with disembodiment and ghostly presence,
a persistent trend in the development of media technologies, was manifested in the

---

compelling narrative of a desire for the alluring infusion of lifelike verisimilitude – where the dead are brought back to life. New media technologies enabled new perceptual experiences, which came to disturb familiar models of representation and experience, as perception has come to be extended past bodily limits, and sounds and images are disembodied, evoking traces of the absent presence of bodies within space and time. Peters wryly remarks: “Media able to capture the flow of time, such as the phonograph and cinema, seemed to vaporise personages into sounds and images”.10 Our conceptions of time and space have been altered, the world has been transmuted into sounds and images, which have become things, capable of being recombined and deployed.

It should be remembered that the changes brought about by technology seemed supernatural at first, and early photography was closely associated with death – both in the numerous images of the dead made during the early years of the medium and in the way a photograph seemed to cheat death by making at least appearance permanent. This preservation of the dead was also a theme for phonography, where the phonograph “undermined the most enduring mark of human individuality, authority, and, as Derrida has shown, presence, the voice itself, by separating it from an actually present speaker”.11 The phonograph was accused of separating the voice from the human subject and depositing it into a machine. In what was then a bewildering and astonishing turn, the phonograph had made it possible to hear the voices of those not present, and furthermore, the voices of those no longer alive. Edison was fond of boasting that it was suddenly possible to speak with people on the other side of the globe, or even those far in the future.

10 Ibid, p. 142.
The phonograph’s bewildering re-mobilisation of presence, where sound was disembodied and placed in a machine, where a speaking human body was nowhere to be seen – provoked predictable reactions from astonished audiences. A newspaper reporter in London wrote: “Mr. Edison’s invention [of the phonograph] is considered first cousin to the prince of black art. My own impression after hearing it talk in English, French, German and Hungarian, all at once, was that I had gone mad”.12 Haunted by the spectral double of a voice without a human body producing it, by the eerie possibility that these voices were produced by disembodied speakers who were no longer living, early audiences of the phonograph must have presumed that some sort of magic was involved.

In response to a letter from Edison’s assistant Edward H. Johnson that announced the invention of the phonograph, *Scientific American* (1877) published an article that proclaimed: “The possibility is simply startling. A strip of indented paper travels through a little machine, the sounds of the latter are magnified, and our great grandchildren or posterity centuries hence hear us as plainly as if we were present. Speech has become, as it were, immortal”.13 Speech no longer required a human body to produce it. The phonograph could record exactly what had been said, preserving not only the words (as in writing) but also the subtle nuances of expression. In this way, an avenue was prepared for how voices and sounds would come to emanate re-soundingly from moving images.

An 1892 issue of *Phonogram* contained an article titled ‘Being Dead, Yet He Speaketh’, which addressed the wondrous capability of the phonograph:

> We have lost Tennyson, Longfellow, Whittier, Bryant, and hundreds of other celebrities, now passed away, but Tennyson’s prayer for ‘the touch of a vanished hand and the sound

of a voice that is still,’ will be answered, for although the body may be turned to dust, the phonograph will have made and preserved an exact picture of the sounds it uttered while living.14

And in an 1896 issue of Phonoscope, an article titled ‘Voices of the Dead’ mused on the same theme: “The voice, formerly invisible and irretrievably lost as soon as uttered, can now be caught in its passage and preserved practically for ever… Death has lost some of its sting since we are able to forever retain the voices of the dead”.15 Paradoxically, the ‘death’ of recording – the petrification of the living flux of an event – ensures the ‘life’ of its spirit of sense. Recording enables an existence independent of its author or source, but furthermore, recording incarnates the very principle of life in its removal from the carnal world of time, change, and death.

The phonograph made possible the preservation and reproduction of the sonorous spectacle, which henceforth was not quite ephemeral, but now had a material presence. The phonograph would inscribe a past that was now representable and malleable. Immortality would be achieved at the cost of disassociation, decomposition, and decorporealization – as voices were removed from bodies and the living spontaneity of a sonorous event became objectified into an artefact that repeated and represented the living immediacy of a soundful event. Recorded sound can be said to cause sound to “lose its immediacy”, gaining in exchange what Adorno expresses as “the hope that, once fixed in this way, it will some day become readable as the ‘last remaining universal language since the construction of the tower’ [of Babel]”.16 Recorded sound at once invokes two interrelated figures: first, the figure of presence, of an ephemeral sound

---

14 ‘Being Dead, He Yet Speaketh’, Phonogram I, no. 11, November 1892, p. 249.
captured for eternity, summoned into being on a whim; and second, the figure of
ghostliness, of a sound that is but a spectral trace of an original, present being, now
absent, and present only via the displacing technology of a machine. Recorded sound can
be described as “acoustic photographs”, but are not like photographs, as voices are set
again in motion, invoking the absent presence of a speaking body.

The phonograph made it possible to resurrect voices as living simulacra. Recorded sound could now present the simulation of immediacy, the simulation of the living flux of an event. Henceforth, space and time, subject and object, would be fragmented and susceptible to repositioning and recombination. The event, which occurs uniquely as evanescent, situational, oracular, dialogical, agonistic, potent, could be recorded and thus rendered more like a thing than the living immediacy of an utterance – it becomes autonomous, unresponsive, fixed, detached from its source, a vast receptacle of memory, a source of abstraction. However, sound recording re-tells the story of an event, and thus operates as more than a mere mode of representation. Recorded sound is therefore a material thing that simulates an event – the living flux of an event is transmogrified into an object, a material thing. Once it becomes material and hence malleable, sound can be subject to montage and recombination (both in itself and in relation to the image). Recorded and replayed sound invokes the (absent) presence of sounding bodies, even as sound now contains the potential for moving from body to body (bodies which have become images, machines, texts).

17 The term is Adorno’s; see ‘The Form of the Phonograph Record’ (1934), in Essays on Music, 2002, p. 278.
18 It is useful to draw on and compare these reflections on recording with Walter J. Ong’s discussions of writing (which should be comprehended as a recording technology) and speech in his Orality and Literacy, 1982.
So, unlike the photograph, and unlike its younger Edisonian relative, the moving picture, the phonograph was apparently able to reproduce a recording with full dimensionality. “The phonograph presented a human voice without a human body. The human soul, the breath, had taken up residence in a machine”.20 In a world dominated by visuality, sound and hearing were reintroduced to our attention through the machining of sound and hearing – which served to remake audibility and soundfulness. Speech, whose being Hegel and other commentators had linked uniquely with temporality, no longer needed to vanish permanently into thin air, was no longer fated simply to pass away, but had been granted the opportunity to gain a re-audition. The recording and replay of sound problematized established notions of sound and presence, altering and installing new modalities of presence (which would remain as important founding components of the new model of sound). The phonograph provided a voice without a body being present to make it, a sound without its object being present to produce it. And later, when paired with the moving image, the phonograph would enable the co-presence of sound and image, which would construct an articulation of sound and image (a voice made by the image of a body, a sound produced by the image of its object). The sense of a heightened, mediated (cinematic) presence would be further reinforced by the pairing with sound technology.

Edison had returned to working on the phonograph a decade after its initial invention, and began work that would make it commercially viable in the 1890s. A year after the perfected phonograph was launched, Edison filed with the Patents Office, describing his ideas for the device that would “do for the eye what the phonograph does.

for the ear”. These ideas had gained significant momentum when Edison met Eadweard Muybridge at Edison’s laboratory in Orange, New Jersey, on Monday, February 27, 1888. A few months later, a journalist from the New York World visited Edison and wrote an article based on Edison’s reflections regarding the meeting:

Mr. Edison said that Prof. Muybridge, the instantaneous photographer, had visited him lately and had proposed to him a scheme which, if carried to completion, will afford an almost endless field of instruction and amusement. The photographer said that he was conducting a series of experiments recently and had almost perfected a photographic appliance by which he would be enabled to accurately reproduce the gestures and the facial expression of, for instance, Mr. Blain in the act of making a speech… He proposed to Mr. Edison that the phonograph should be used in conjunction with this invention, and that photographs of Edwin Booth as Hamlet, Lillian Russell in some of her songs, and other artists of note should be experimented with. Mr. Edison, he said, could produce with his instrument the tones of the voice while he would furnish the gestures and facial expression. This scheme met with the approval of Mr. Edison and he intended to perfect it at his leisure.21

Moving images can be traced back to the pioneering influence of Muybridge’s motion studies. Back in the spring of 1872, Muybridge famously photographed a horse. (The resulting photograph has unfortunately not survived – the subsequent series of photographs which were produced, however, have survived and is very well known.) This first, seemingly innocuous encounter between a camera-bearing man and a fast-moving horse marked a significant advance toward a new type of image and the establishment of an audiovisual age. The inventor of instantaneous photography,

---

Eadweard Muybridge, “had captured aspects of motion whose speed had [formerly] made them as invisible as the moons of Jupiter before the telescope”. It seemed “as though he has grasped time itself, made it stand still, and then made it run again, over and over” – and in doing so, Muybridge opened up a new world “for science, for art, for entertainment, for consciousness”, pushing the old world into further retreat. By the end of the 1870s, these experiments had led to the invention of the essentials of motion-picture technology. Muybridge’s work marked the first time photographs had captured actual motion, as opposed to previous incarnations of photography that could only capture still objects over a long period of exposure. These photographic images could capture, preserve, and exhibit a frozen piece of reality.

With the aid of high-speed electric shutters, the invention of a new kind of photography had made it possible to hold on to what had formerly always been lost to the implacable flow of the stream of time. “To represent movements in real time, machinery had to move faster than the human eye, both in capturing fragments of motion and in reassembling them”. The cinema would be born from “the speed of the projector matched to the speed of the camera. It took extraordinary speed to reproduce, say, the plume of cigarette smoke slowly uncoiling above a heroine’s face, the tail of a hero’s coat swinging as he turned around”. But then the cinema did not begin with heroines and fictions, but from “photographs of horses in motion”. Before the advent of instantaneous photography, people, objects, places, and events, had all been unique – seen only once and then slipping away due to the inexorable passage of age, light, and time. Photography practically reassembled the world it entered – or, to explain it more

22 Rebecca Solnit, *Motion Studies*, 2003, p. 3.
23 Ibid, p. 182.
accurately, photography effected a profound transformation of the way we perceive and interact with the world. The world would become a potential series of images, to be viewed at leisure, as perception became disembodied and extended in new ways.

Jean-Louis Comolli suggests that with the advent of the photograph, the human eye loses its “immemorial privilege” and is relegated in favour of “the mechanical eye of the photographic machine”, which “now sees in its place”. Comolli has remarked that: “The second half of the nineteenth century lives in a sort of frenzy of the visible”, which has resulted from “the social multiplication of images”, as well as from “a geographical extension of the field of the visible and the representable: by journeys, explorations, colonisations, the whole world becomes visible at the same time that it becomes appropriatable”. In similar fashion, Bazin comments that photography has enabled “an image of the world [which] is framed automatically, without the creative intervention of man”; where “the objective nature of photography confers on it a quality of credibility absent from all other picture making”.

The photograph freeze and preserves a moment of time, which is extracted from the irreversible momentum of the stream of time, and made into an abstract artefact, separated from the living presence of an embodied event. Photography curiously transforms the living event into a museum object. Roland Barthes writes: “Now, in the Photograph, what I posit is not only the absence of the object; it is also, by one and the same movement, on equal terms, the fact that this object has indeed existed and that it has

been there where I see it”. In this way, the photograph evokes the absent presence of the object, for “in Photography, I can never deny that the thing has been there”. In other words, photography provides a glimpse of the ‘that-has-been’: “what I see has been here, in this place… it has been here, and yet immediately separated; it has been absolutely, irrefutably present, and yet already deferred”. Movement, however, which renews the momentum of the stream of time, would reinstall a sense of presence to the photographic image.

The projection of Muybridge’s various series of instantaneous photographs in quick succession would simulate motion. Muybridge’s zoopraxiscope was comprised of a powerful projector and sixteen-inch discs containing photographic images and painted images drawn from his photographs. The images were revolved in one direction while a slotted disc revolved in the other, thus providing the intermittent visibility required by persistence-of-vision illusions. The *Photographic News* of 1882 would publish an appreciative description of one of Muybridge’s photographic exhibitions, which included his zoopraxiscope:

After Mr. Muybridge had shown his audience the quaint and (apparently) impossible positions that the horse assumes in his different gaits, he then most ingeniously combined the pictures on the screen, showing them one after another so rapidly that the audience had before them the galloping horse, the trotting horse, &etc. A new world of sights and wonders was, indeed, opened by photography, which was not less astounding because it was truth itself.”

---

28 Ibid, pp. 76-77.
The solution to the problem of recording and playing longer sequences of moving images would begin to come into focus.

After meeting with Muybridge, Edison would study the zoopraxiscope before contemplating the problem of moving pictures. After many false starts, Edison encountered the French photographer Etienne-Jules Marey. Also inspired by Muybridge’s work, Marey had devised a method of ‘shooting’ a series of rapid pictures by feeding strips of film through a single camera.

Taking this innovation into account, Edison would come to construct a miraculous device that would record and reproduce objects in motion. This invention was actually two complementary devices: one, the Kinetograph, which captured successive photos on a long strip of film, and two, the Kinetoscope, into which the perforated, developed film was fed and viewed through a peephole. In 1891, a prototype was exhibited to a convention of the National Federation of Women's Clubs. “They saw through an aperture in a pine box standing on the floor”, the magazine *Phonogram* reported, “the picture of a man… It bowed and smiled and took off its hat naturally and gracefully”.\(^{30}\) This was a machine for seeing, where the audience could be shown moving images from a different time and place, which possessed lifelike presence and verisimilitude.

The Kinetoscope also pioneered a fresh perception of the temporal dimension. By exhibiting a succession of marginally different images viewed at a rapid rate of frames per second, the eye’s persistence of vision made it possible to reproduce the movement that is such a recognisable, necessary, and natural part of the world around us – providing the illusion of reality. In 1894, an invitational showing of Edison’s Kinetoscope in Paris

would spark the Lumière brothers’ interest in motion pictures, and the brothers set out to
device a machine – which they came to call the Cinématographe – that would combine
moving pictures with front projection, overcoming the limitations and problems of
Edison’s Kinetoscope. Patented on February 13, 1895, the Cinématographe dramatically
shifted the scope and form of the nascent motion picture experience by enabling
collective viewing by an entire group, instead of the single spectators awaiting their turn
at the peephole of Edison’s Kinetoscope. On December 28, 1895, the Lumière brothers
held the world's first public film screening at the Grand Café on the Boulevard des
Capucines, in Paris. In the basement lounge, ten short films were exhibited, lasting
approximately twenty minutes in total. And one of the most impactful ‘actualities’, or
short films, shown on that historical evening was ‘The Arrival of a Train at La Ciotat
Station’. This was a filmic sequence that depicted a train pulling into the station and
which presented a cinematic experience that, according to popular accounts, had
audiences screaming and ducking for cover as they believed that the train itself was about
to plough into the theatre.31

The reaction to these short films was sensational and before long there were
twenty showings a day to meet the tremendous public demand.32 Mesguich, a cameraman
for Lumière, announced that “the Lumière Brothers had established the true domain of
the cinema in the right manner”. While the novel or the theatre “suffice for the study of

---

31 James Lastra remarks that “despite a general lack of evidence that any such violent response occurred on
that December evening in 1895”, this story has “come to embody the cinema’s primal scene, serving as a
metonymic touchstone for the history of the cinematic spectator, and suggesting a set of issues and a course
of research centred on questions of belief, immediacy and illusion” (Sound Technology and the American
Cinema, 2000, p. 64.)

32 Tom Gunning cites a long-ago Montpellier journalist, who “in 1896 described the Lumière projections as
provoking ‘an excitement bordering on terror’”, his astute observation simultaneously praising the then-
new spectacle and explaining its success. (‘An Aesthetic of Astonishment: Early Film and the
121.)
the human heart”, the cinema supplies “the dynamism of life, of nature and its manifestations, of the crowd and its eddies. All that asserts itself through movement depends on it. Its lens opens on the world”.

The Lumière brothers began to open theatres to show their films (which became known as cinemas). By 1897, the Lumière brothers were a global success, with cinemas established in London, Brussels, Belgium, and New York, accompanied by an expanding film catalogue. The irony was that the Lumière Brothers had initially characterised film as a medium without a future, believing that people would soon grow weary of cinematic images which were, after all, merely ‘doubles’ of everyday reality.

Film is held to be more ‘realistic’ than photography, for it moves beyond still photography’s ‘frozen’ reality by reproducing the movement which is such a familiar and necessary part of the natural world. Barthes writes that at first glance, “the cinema has a power which… the Photograph does not have” – for “the screen (as Bazin has remarked) is not a frame but a hideout; the man or woman who emerges from it continues living”.

(Barthes’ work on photography actually sets out to prove the opposite, namely, that the photograph, through the punctum, can institute an embodied presence that the motion picture cannot. The point here is to examine how the cinema offers the simulation of a lifelike verisimilitude.) “Like the real world, the filmic world is sustained by the presumption that, as Husserl says, ‘the experience will constantly continue to flow by in the same constitutive style’… [The filmic world] is, then, simply ‘normal’, like life”.

But because film appears to free the viewer from the constraints of both time and space, which are fundamental components of the physical world, film at the same time can take

---

33 Cited in Kracauer’s Theory of Film, 1997, p. 31.
34 Barthes, Camera Lucida, 2000, pp. 55, 57.
us ‘beyond’ reality altogether. The all-seeing camera can seemingly take the viewer anywhere and any-when, to take in the scene from an array of perspectives, travelling rapidly within space and time, pausing intently in order to gaze at a given moment and situation. Nevertheless, the flow of these filmic images and the ‘reality’ of the cinematic experience are sustained by what appears to be a lifelike verisimilitude.  

Vivian Sobchack explains that the photograph “functions to fix a being-that-has-been (a presence in the present that is always past). Paradoxically, as it objectifies and preserves in its acts of possession, the photographic has something to do with loss, with pastness, and with death, its meaning and value intimately bound within the structure and investments of nostalgia”. In contrast, the cinematic (although dependent upon the photographic), “has something more to do with life, with the accumulation – not the loss – of experience… The moving picture is a visible representation not of activity finished or past, but of activity coming-into-being”. Instead of the contemplative frozen moment of the photograph, moving images supply a persuasive infusion of life, where spectators are swept away by the deluge of images, a veritable ‘frenzy of the visible’, which is given momentum and carried along by the flow of the stream of time. Sequences of images are composed, ordered, and recombined, providing us with actual and virtual worlds seen from an array of machined perspectives. Instead of the Barthes-like ‘this-has-been’ of photography, there is the ‘this-is’ of cinema, comprised of ‘activity coming-into-being’; moving images serve to animate the ‘double’ of prosthetic perception.

---

36 A comparison can be made with this to Deleuze’s concept of an any-space-whatever; see my discussion in Chapter 8.
Christian Metz puts forward the conclusion that: “It is movement (one of the greatest differences, doubtless the greatest, between still photography and the movies) that produces the strong impression of reality”.\(^{38}\) This appeal of the image would lead the filmmaker Abel Gance to announce in his *L’Art cinématographique* (1927) that “the time of the image has arrived!” For Gance, the (then-silent) cinema would endow human beings with a new synaesthetic awareness: spectators “will hear with their eyes”.\(^{39}\)

However, as described by the writer Maxim Gorky, the reality depicted by the new art of motion pictures was “a world without sound, without colour… It is [not] life but its shadow… And all [that is shown occurs] in a strange silence where no rumble of wheels is heard, no sound of footsteps or of speech. Not a single note of the intricate symphony that always accompanies the movements of people”.\(^{40}\) Because “our real-life environment is filled with sounds”, even if “we may for long stretches be unaware of their presence… our eyes cannot register a single object without our ears participating in the process”. As a result, “the extinction of sound transforms the world into limbo”.\(^{41}\) Sound and image would continue their estranged relationship, but reconciliation, although elusive, was a goal that would be realised.

In 1895, Edison had unveiled the Kineto-phonograph, which paired together his two great inventions: the Kinetoscope and the phonograph (with music conveyed through ‘ear tubes’). The phonograph allowed the recording of sounds without images, and the Kinetoscope permitted the recording of images without sounds. The Kineto-phonograph was intended to bridge this gap, restoring sound and image to their proper relationship,

\(^{38}\) Metz, *Film Language*, 1974, p. 7.

\(^{39}\) Cited in Robert Stam, *Film Theory*, 2000, p. 35.


reproducing an event simultaneously and completely in both the visual and the acoustic dimensions. Unfortunately, although its concept was bold and visionary, the Kineto-phonograph did not manage to achieve the aim of reintegrating sound and image. At the time, no adequate amplification technology was available, and no functional method could be found to present these films to an audience of more than a handful of viewers, all clustered close to the acoustic horn on a Kineto-phonograph player. Eventually the general design was abandoned, and entirely different approaches were investigated.

Audiences' clear preference for motion pictures projected onto large screens initiated film's Silent Era. For 30 years, movies would exist as juxtapositions of film and live music and noise, making for a pioneering period of sound-image experiments. The images seen on the screen were accompanied by sounds performed live-to-picture, produced by a various array of sources: pianists, effects machines and sync-sound apparatuses, lecturers and actors speaking beside or behind the screen, small or large orchestras, etc. Music was the most popular choice to accompany the moving pictures projected on the screen. Silent moving pictures were viewed by the objective eye with a feeling of detachment, of lifeless distance. Film music was originally proposed as the solution to this, as musicians played their scores ‘live-to-picture’, situated in the orchestra pit beneath the screen, or seated at a piano just off-screen.

The historian Scott Eyman explains: “From the beginning, the cinema abhorred silence; the cinema needed some sort of sound, if only to cover up the distracting noises of the projector and the shuffling of the audience. That sound was music; by the mid-

---

1920s, movie theatres were the foremost employers of musicians in the country”. In those halcyon days when the new, still-‘silent’ cinematic art form found its feet, symphony orchestras resided in the largest downtown movie palaces to accompany the films, and a medium-sized neighbourhood theatre might carry between five and ten musicians. “Even the meanest fleapit in the sticks had a piano player”. So film music had prosaic origins, beginning “not as a result of any artistic urge, but from the dire need for something which would drown out the noise made by the projector”. Sound-absorbent walls between the projection machine and the auditorium had not yet been installed. The painful noise made by the projector “disturbed visual enjoyment to no small extent. Instinctively cinema proprietors had recourse to music, and it was the right way, using an agreeable sound to neutralise one less agreeable”. In other words, “Music was needed for its magic. To drown the whirring and the coughing, to bind the separate spectators into one audience. To hypnotise, to make a mood”. In 1936, Nicoll declared that “were we to witness only the actions”, we would develop “a restlessness dependent upon the potential alertness of our auditory nerves. We should expect sound and none would come. The musical accompaniment of silent films, therefore, simply occupied our attention and prevented the disappointment resultant upon a thwarted expectancy”.

Back at the Edison labs, sound movie experiments had continued, despite the failure of the Kineto-phonograph, until a 1914 fire finally brought the sound-image project to an end. Thomas Edison would have to wait until the 1920s to see his vision of

---

43 Eyman, 1999, p. 26; Siegfried Kracauer, however, disagrees with this, suggesting that film music’s “vital function was to adjust the spectator physiologically to the flow of images on the screen”. (Kracauer, Theory of Film, 1997, p. 134.)
44 Eyman, 1999, p. 38.
45 Kurt London, Film Music, 1936, p. 23.
46 Ivor Montagu, Film World, 1964, p. 59.
47 Allardyce Nicoll, Film & Theatre, 1936, p. 125.
synchronised sight and sound realised, and by that time Edison's laboratories would not be at the forefront of the new technologies necessary for sound movies. By the mid-1920s, recorded sound began to inform the cinematic medium as new technologies solved two basic problems; synchronisation and amplification. Although Edison’s phonograph did not manage reliably to resolve the problems of amplification and consistent synchronisation that were required by the development of sound film (not being sufficiently loud enough to fill the exhibition space, and not replaying sounds at a stable enough speed to be consistently linked to the visual images), the phonograph laid the foundations for later, subsequent innovations (provided by the Vitaphone sound-on-disc system, the De Forest Phonofilm sound-on-film system, and the Case-Sponable Movietone system), which saw the development of the ‘talkies’, and the continuing evolution of the sound film. The phonograph would make it possible for a soundtrack to exist in delicate juxtaposition with an image track, presenting separate images and sounds that play together in a synchronous relationship, and this co-presence of sound and image would reproduce persuasive illusions of reality. The ‘live-to-picture’ method would come to be replaced with the sync soundtrack; sync sound, in the form of the talking motion picture, would be launched into the popular consciousness by the invention of the Vitaphone sound-on-disc technology and the release of The Jazz Singer in 1927.

The solution to the problem of amplification was first provided by Lee De Forest in 1906. Modern electric loudspeakers had their origins in telephone and radio technology. Until the invention of electronic amplifiers, a truly loud speaker was not really possible. De Forest’s key invention was a three-filament, gas-filled tube he called

---

48 Two years earlier, De Forest had electrified the 1904 St Louis World’s Fair with his De Forest Wireless Telegraph Tower, which stood several hundred feet high and could transmit radio messages fifteen hundred miles.
the Audion (from audio and ionise). “I little realised that I held in my hand the long-sought Aladdin’s lamp of our new world”, De Forest would later write in his not entirely modest autobiography, Father of Radio, “a lamp which when rubbed by gifted engineers would swiftly summon gigantic Genii to do our bidding, and all but remake our world”.49 Western Electric would purchase the rights to De Forest’s invention in 1913, and Dr. Harold Arnold, one of their scientists, discovered that by pumping out the gas with which De Forest had filled the Audion tube and creating a near vacuum, the tube could amplify sound up to 130 times its original volume without distortion. This discovery would lead to wider commercial application, and would be a key component to the emergence of a synchronous soundtrack. The impression of reality brought about by moving pictures would soon have to deal with the infusion of recorded and synchronised sound.

This dream of sound had been present for some time, and articulated by a number of people. Back in 1918, the innovative Russian filmmaker Dziga Vertov had scribbled these hasty lines in his dairy: “And thoughts while walking: I must get a piece of equipment that won’t describe, but will record, photograph these sounds. Otherwise it’s impossible to organise, edit them. They rush past, like time. But the movie camera perhaps?”50 This piece of equipment, of course, first took the form of the phonograph. Vertov had a vision of a cinematic future where voices and sounds were coupled with the images projected on the screen, thereby producing a new co-presence of sounds and images. In 1923, Vertov would write from the perspective of the cinematic apparatus:

I, the machine, show you a world the way only I can see it. I free myself for today and forever from human immobility. I’m in constant movement. I approach and pull away

from objects… recording one movement after another in the most complex combinations. Freed from the boundaries of time and space, I coordinate any and all points of the universe, wherever I want them to be. My way leads towards the creation of a fresh perception of the world.\textsuperscript{51}

In 1925, he would optimistically predict: “In the near future, man will be able to broadcast to the entire world the visual and auditory phenomena recorded by the radio-movie camera”.\textsuperscript{52} In this vision of the cinema-to-come, the audience would be supplied with images and sounds that had been captured and manipulated by machine technologies, creating a refraction/reflection of events from a potential multiplicity of perspectives, bending time and warping space, making a narrative out of different combinations, juxtapositions, and durations.

Mikhail Kaufman, co-author of Dziga Vertov’s groundbreaking 1929 film \textit{Man with a Movie Camera}, explains:

An ordinary person finds himself in some sort of environment, gets lost amidst the zillions of phenomena, and observes these phenomena from a bad vantage point. He registers one phenomenon very well, registers a second and a third, but has no idea of where they might lead… But the man with a movie camera is infused with the particular thought that he is actually seeing the world for other people… He joins these phenomena with others, from elsewhere, which may not even have been filmed by him. Like a kind of scholar he is able to gather empirical observations in one place and then in another. And that is actually the way in which the world has come to be understood.\textsuperscript{53}

\begin{itemize}
\item \textsuperscript{51} Cited in John Berger, \textit{Ways of Seeing}, 1977, p. 17.
\item \textsuperscript{52} \textit{Kino-Eye: The Writings of Dziga Vertov}, edited by Annette Michelson, 1984, p. 56.
\item \textsuperscript{53} 1979, p. 65.
\end{itemize}
A hundred years after cinema’s birth, cinematic ways of seeing the world, of structuring time, of narrating a story, of linking one experience to the next, have become naturalised modes of understanding and perceiving the world.

It was now possible for the audience to be freed from the final constraints of time and space, the fundamental boundaries of the physical world. Cinematic images and sounds could now operate to “blast the prison of conventional reality”.54 The new form of cinematic narrative could transport the audience from one moment in time to another; and from one location to another, all in the blink of an eye. The spectator would adopt the perspective provided by an all-seeing camera, and thus the viewpoint of any of the characters in the narrative; or even take up an omniscient, voyeuristic gaze. The audience would see what the camera sees, hear what the microphone hears; the camera and the microphones become extensions of the audience’s perceptual senses, encouraging a direct participation with what is portrayed on the screen – in what can be termed machine perception.

The foundations were being laid for the arrival of the synchronous sound film. By 1926, a phonograph was an essential component of any American middle-class living room, having practically replaced the piano in the role of providing family entertainment. The explosion of radio had only increased the public’s appetite for recorded music. The microphone had permanently altered the nature of popular song and how it was presented. Recorded sound was claiming an increasing role in everyday lives. In hindsight, it seems only logical and perhaps inevitable that the musicians and orchestras

---

supplying music for the movies would come to be replaced by pre-recorded, synchronised soundtracks.

The revolution facilitated by the phonograph cannot be overstated. In an 1878 article, Edison had described the accomplishments of his phonograph: it is able to capture and permanently retain ‘fugitive’ sound, reproducing the original characteristics of the captured sound at will, without the presence or consent of the original source, and after the lapse of any period of time; it is able to transmit captive sounds for purposes of communication or commercial trade; and it is able to ensure the indefinite multiplication and preservation of sounds, without regard for the existence or non-existence of the original source.55 In other words, the phonograph enables what R. Murray Schafer terms ‘schizophrenia’, which “refers to the split between an original sound and its electro-acoustical transmission or reproduction”.56 In simple terms, the phonograph gave rise to disembodied sounds. Schafer uses this idea of ‘schizophrenia’ to highlight the importance of “the discovery of packaging and storing techniques for sound and the splitting of sounds from their original contexts”.57 In Schafer’s conception, sound has been separated from the maker of the sound: “Sounds have been torn from their natural sockets and given an amplified and independent existence”.58 With the innovation of the synchronous soundtrack, these disembodied sounds were to be shortly re-embodied, where sounds were taken from their original sources and environments and reattached (or fused) to various images; these re-sounded images were to come to create new sound-image

---

57 Ibid, p. 88.
58 Ibid, p. 90.
events, or soundful images, that would refract, echo, and draw meaning from the horizons of ‘original’ events.

The dream of a complete cinematic experience that provided sounds and images would soon come to fruition, as disembodied sounds were to be re-embodied in the form of the moving image. On April 26, 1925, under a *New York Times* headlines, “PERFECT AUTOMATIC MUSIC FOR MOVIES”, 59 Warner Bros. and Western Electric announced their partnership. The result of their relationship would be the Vitaphone system. An advertisement for the Vitaphone system was promptly placed in the July 24, 1926 issue of *Brass Tacks*, the exhibitors’ newsletter published by Warner Bros.:

A new era in motion picture presentation has arrived. It will thrill and startle the world. The marvellous Vitaphone process, which will have its first public presentation at the Warner Theatre on August 6, will revolutionise the industry… The greatest artists of the operatic and musical field can be heard in the smallest of theatres as well as the largest. Millions of people will be educated to a finer appreciation of the best music that has ever been written by the foremost composers. 60

At this point in time, the idea of watching a scene accompanied by sound that was not produced by live musicians, or actors hidden behind the screen, provoked a sense of astonishment and disbelief. “When I heard a twelve-piece orchestra on that screen at the Bell Telephone Laboratories, I could not believe my own ears”, Harry Warner would reminisce of the demonstration in 1925. “I walked in back of the screen to see if they did not have an orchestra there synchronising with the picture. They laughed at me. The whole affair was in a ten-by-twelve room. There were a lot of bulbs working and things I

---

59 In Eyman, 1999, p. 71.
60 Cited in Eyman, 1999, pp. 89-90.
knew nothing about, but there was not any concealed orchestra”. Sam Warner, discussing the possibilities, made the observation: “But don’t forget you can have actors talk too”. 61

When the moving pictures gained the element of synchronised sound, music was the natural choice. (Although Harry Warner’s flippant remark that no one would want to hear speech over music now seems exceedingly foolish.) The system would be launched at the premiere of Don Juan, in August 1926. Contained in the program were explanatory notes:

Vitaphone: An instrument that synchronises motion pictures and sound perfectly, purposes to advance the presentation of motion pictures in theatres – large or small – located anywhere. It will accomplish this by making available, on a broad scale, the music of the greatest symphony orchestras and the vocal entertainment of the most popular stars of the operatic, concert and music comedy fields. 62

Don Juan (1926) would be the first release that utilised the new Vitaphone sound-on-disc system, boasting a synchronised soundtrack that enabled music to be heard during several scenes, without requiring the presence of live musicians. In the Los Angeles Times, Edwin Schallert proclaimed: “The orchestral accompaniment to the feature was a revelation of what can be done with the Vitaphone in perfectly matching the performance of a large body of musicians with the scenes of a picture”. 63 The review of Don Juan by the New York Times gushed that “no single word, however compounded, is quite adequate to suggest the amazing triumph which man has at last achieved in making pictures talk naturally, sing enthrallingly and play all manner of instruments as skilfully

61 Scott Eyman, The Speed of Sound, 1999, p. 70; in large part due to the effect of this demonstration, the Warner Brothers would decide to produce The Jazz Singer.
63 Cited in Eyman, 1999, p. 104.
as if the living beings were present instead of their shadows… marvellous – Uncanny!”

This now-familiar innovation of automated synchronised sound enabled music, and even noises, sounds, and voices – the soundtrack – to be fused with the flow of images, independent of the foibles and inconsistencies of ‘live’ musicians.

Michel Chion has pointed out an important historical point that is often forgotten: “we are indebted to synchronous sound for having made cinema an art of time”. After all, in order to maintain sync, the playing of a soundtrack requires the stabilisation of the projection speed for the image – which had “consequences that far surpassed what anyone could have foreseen. Filmic time was no longer a flexible value, more or less transposable depending on the rhythm of projection”. Sound is fundamentally temporal. It cannot be frozen, like an image, as sound occupies time through its existence as vibration. Sound invokes time by being part of time, by occurring in time, by flowing through time. Sound creates a linear narrative through its audible beginning and ending (attack, duration and decay). Because of its transience and ephemeral nature, sound anchors us within the stream of time. An image, particularly a still image that does not give reference to the passage of time, can be coupled with sounds that imbue that scene with temporality. The moving image, by itself, is still abstracted, seemingly detached from the flow of life and time. The infusion of sound creates a sense of duration.

With The Jazz Singer (1927), recorded dialogue would be matched to filmed image. There were not merely songs but also the exhilarating moments when Al Jolson, playing the lead role, spoke, providing a tangible frisson of intimacy, spontaneity, immediacy, and presence, and garnering a large measure of popular appeal from

---

64 Cited in Scott Eyman, The Speed of Sound, 1999, p. 93; the italics are mine.
captivated audiences. With the birth of the talkies, cinematic images were granted voices. Because of the fame that has become attached to it over the years, being popularly known as the first sound film, it is perhaps surprising to discover that *The Jazz Singer* (1927) is largely a silent film, being only about 25% talkie. However, at one point in *The Jazz Singer* Al Jolson says to a nightclub audience, ‘Wait a minute… wait a minute… you ain’t heard nothin’ yet!’ That famous line, plus one other snatch of dialogue (added to an orchestral score, some Jewish music and seven songs) qualified this as the first major Hollywood film to begin to make use of a synchronised soundtrack in telling its story.

For Disney's *Steamboat Willie*, released in 1928, images were created to accompany already-recorded music. The gangster film *Lights of New York* (1928) would become the first ‘all-talking’ (or all-dialogue) feature. It was discovered that dialogue, music, and sound effects could be dubbed onto already-filmed images; and this process would eventually become the standard method of creating soundtracks. And after the talkie revolution, the sound film would develop further, moving beyond simple instances of dialogue synced with the picture, and adopting increasingly sophisticated manifestations of the soundtrack. Chion observes that today, “it is difficult to imagine the amazement inspired by the first synchronised sound films in the twenties”, as “it has become commonplace to see a moving figure on a screen as we hear movement”. We must remember that sound and image, “heard and seen like a couple of perfectly matched dancers”, was once a spectacle in itself.\(^{66}\) Early sound movies presented actors performing such innocuous actions as pouring a drink into a glass, just so the images

could be accompanied by sound. The proper sound film (instead of the talkie) took some time to develop; this new form would make fuller use of its audio-visual capacities to emphasise sounds and noises other than dialogue, and would explore asynchronous sounds, off-screen sounds, and sound montages, taking fuller advantage of the possibilities of re-embodied sounds and re-sounded images.

Adorno and Eisler’s *Composing for the Film* had presented the argument that music had been introduced to films in the era of silent pictures as a specific antidote “against the picture”, in an attempt to overcome the uncanny nature of the simultaneously “living and nonliving” silent figures flickering on the screen, which possess a “ghostly effect”. Synchronous sound, fused with the flat, framed image, reiterates a phenomenology of presence, even though sounds have been detached from sources, voices taken out of the body, and placed into a machine. Sounds and voices continue to imply the presence of a subject or an object that was present in the moment and in the flesh. Sound and listener, although physically separated, are held to share the same moment in time, thereby re-invoking/re-producing the same simultaneity that produced them. The transitory nature of sound has been subtly and profoundly transformed into a conventionalised simultaneity. “Ghostly shadows, as volatile as clouds”, with the infusion of sound, are allowed to “become trustworthy shapes”.

---

67 Much early cinema was an exhibitionist cinema, a cinema of attractions, as Tom Gunning has argued: “less a way of telling stories than a way of presenting a series of views to an audience”. (Gunning, ‘The Cinema of Attractions: Early Film, Its Spectator, and the Avant-Garde’, in *Early Cinema: Space, Frame, Narrative*, edited by Thomas Elsaesser and Adam Barker, 1990, p. 63.) Of course, this early cinema of attractions also came to present a series of voices, musical pieces, and sounds to its audience.

68 1994; the text was originally published in 1947.

69 Their conclusion was that the advent of the talkies did not manage to solve this situation: the characters seen on screen, although now voiced, remain two-dimensional effigies, still lacking the depth that the spatially powerful acoustics of film sound and music aims to provide. (All references to Adorno and Eisler’s *Composing for the Film* are found on p. 75.)

A phenomenology of presence is invoked by disembodied sound, where the presence and dimensionality of sources is called upon and set again in motion, a presence and dimensionality that can then be fused or re-embodied with moving images. Sound, then, became a fundamental component for the simulation of ‘lifelike verisimilitude’ offered by the moving image. The recording and replaying of sound, fused with and juxtaposed against the image, was to enable all manner of strange couplings, founded on the disembodiment of sound from its physical source and subsequent re-embodiment of sound in the image. There is the idea that synchronised sound brings the moving image ‘to life’. Kracauer remarks that coming across old silent pictures unaccompanied by music, or films whose sound track had suddenly failed, “is a frightening experience; shadows aspire to corporeal life, and life dissolves into intangible shadows”.71 As the film theorist Bálázs notes: “A completely soundless space… never appears quite concrete, and quite real to our perception; we feel it to be weightless and insubstantial, for what we merely see is only a vision. We accept seen space as real only when it contains sounds as well, for those give it a dimension of depth”.72

In ‘The Apparatus’, the psychoanalytic film theorist Jean-Louis Baudry writes about the apparatus of cinematic projection, arguing that the cinema constitutes the technical realisation of a perennial dream of a perfect, total simulacrum (a perspective comparable to André Bazin’s idea of the ‘myth of total cinema’). In the course of this observation, Baudry turns to look at the cinematic soundtrack, and remarks that: “one does not hear an image of the sounds, but the sounds themselves… they are reproduced

---

and not copied’. Recorded sound is not an ‘image’, or a simple imitation or representation of the ‘real’ event, machined sound is not merely a copy, but appears to reproduce the phenomenality of the ‘real’, ‘original’ event. Recorded sound reinforces the illusion of a living flux, as recorded sound seems paradoxically to be both material thing and ephemeral event. It would seem only logical that this infusion of life provided by recorded sound would be paired with the moving image (and here it is useful to keep in mind the various connotations of the word) in order to produce voices and sounds that would be attributed to the images projected on the screen, thereby attaining the dream of a perfect simulation. However, recurring anxieties were to remain fixed on the echoes of voices and sounds, re-embodied and attributed to the “shadows that aspire to corporeal life”.

With the coming of the sound film, the nature of the movie experience was permanently altered. A formerly private, active experience, in which each viewer took a creative role in the structure and perception of the silent drama, was about to become collective and more passive in the delivery and reception of both images and sounds. The

---


74 The word image is derived from the root imitari. (The word ‘image’ is somewhat ambiguous, as it can signify graphic, optical, perceptual, mental, or verbal phenomena; Jay, 1993, p. 9. Also see W.J.T. Mitchell, ‘What is an Image?’, in Iconology: Image, Text, Ideology, 1986, and also Philip Nicholas Furbank, Reflections on the Word ‘Image’, 1970.) Immediately, we find ourselves confronting the spectre of the double, the reflection of the other, in the form of the image – which is most commonly defined as a reproduction or imitation of the form of a person or thing. Images are convincing imitations, or semblances, associated with the provision of a simulated presence. More technically, the image refers to the optical counterpart of an object produced by an optical device (or an electronic device) or a likeness of an object produced on a photographic material. Images are copies of an original, a tangible or visible representation containing the possibility of an exact likeness, or a picture of something or someone not actually present. The image-as-double strives to be the same, yet is not the same. And yet as images become more clearly defined, and begin to proliferate, the form of the double takes on new life, detaching from the ‘original’, and in a way signifying the ‘death’ of the ‘original’, even as it immortalizes and resurrects and invokes the presence of its source – which then becomes visible only through the refraction of the image, the double that reflects the possibility of an ‘original’.

75 For a historical overview of this period, see Scott Eyman, The Speed of Sound – Hollywood and the Talkie Revolution, 1926-1930, 1999.
arrival of sound had an overwhelming impact on the movie industry, transforming it almost overnight. In 1928, Hollywood produced over 150 silent feature film dramas. In 1931, only one was produced – Chaplin’s *City Lights*. For the audience, the main attraction was that the people on film could speak; it took some time for people to begin to care about what was said, and even longer to care about the quality of what was heard. The impact of sound was felt in two interrelated arenas: in the sphere of production, and in the appeal to the audience. In 1927-1928, Hollywood camera crews let the soundmen know that they and their crude ‘innovation’ were unwanted interlopers. But, by the end of 1929, the balance of power had shifted. The audience had spoken and silent films now belonged to the permanent, irremediable past. In his history of the talkie revolution, Scott Eyman observes that: “Sound changed *everything*. It changed how movies were made, of course, but more importantly, it changed what movies *were*”.76 The movies were now defined by the new relationship between images and sounds.

The relationship of sound and image has evolved since those early days when audiences would search the room in bewilderment and astonishment for the physical, ‘real’ source of the sounds mysteriously emanating from the screen. Audiences accustomed to viewing silent films were astounded when images were accompanied by their sounds, constructing an even more convincing simulation of reality. We no longer search for the ‘real’ sources of the sounds that are combined with or juxtaposed against images. The co-presence of immersive sound and moving image, introduced through the Vitaphone sound-on-disc technology (based on the phonograph), would enable a soundtrack to exist in juxtaposition with an image track, presenting separate images and

sounds that play together in a synchronous relationship, reproducing persuasive illusions of reality.

The image track and the soundtrack, separated as they are captured by two disparate recording mediums, are mechanically linked and exhibited simultaneously. Images and sounds interact and are fused together within the perception of the audience. Simulation, most generally understood in terms of the copying of physical or phenomenal reality, had found its most amenable form yet in the cinema. Moving images of people could be seen and heard speaking; objects on screen could be seen and heard to make noises; and sweeping pictures of places and landscapes could be accompanied by mood music. However, while technical innovation in the cinema has been associated with the production of ‘greater realism’, it is instructive and illuminating to note how sound and image after Edison has evolved, becoming ‘unreal’ (or ‘realer than real’) whilst engaged in the quest for ‘realism’. This is the story of co-presence, where sounds and images have been separated by visual and acoustic technologies and practices, and subsequently ‘reconstituted’. Through the machined fusion of sound and image, a network of multiple echoes has been established between sound and image, and soundful images are created – when sounds are sourced by their images, when sounds are accompanied by the presence of their sources, where interstitial connections have been established between sounds and images. Perhaps, then, it is not a question of reconstitution, but of creation.

The classic film, Singin’ in the Rain (1952) provides a simple illustration (on several levels) of re-sounding images. The film is set in Hollywood at the dawn of the talkies, and addresses several of the issues arising from the coming of sound. The story

---

begins by introducing a smug but romantic silent film star and matinee idol (Don Lockwood) and his glamorous screen partner/diva (Lena Lamont) – who, in order to uphold the Hollywood star system, are expected to pretend to be romantically involved with each other. With the popular success of *The Jazz Singer*, they are pressured by the studio boss to change their planned silent romantic drama into a sound picture – the musical *The Dancing Cavalier*. However, the temperamental, narcissistic Lena has a shrill, screechy accent, and a completely horrendous singing voice. The suggestion is made that Don's aspiring actress and ingénue dancer-girlfriend (Kathy Selden) should dub in her voice behind the scenes for lip-synching Lena. The results of the plan to expose the jealous Lena and put Kathy in the limelight provide the film's expected happy resolution.

Michel Chion observes: “The plot of *Singin’ in the Rain* hinges on the notion that the natural voice of the actress played by Jean Hagen is vulgar and discordant with the star image she portrays on the silent screen”. The entrance of the talkies allows the Debbie Reynolds character to provide the voice which reconstitutes “a harmonious being on the screen”. The film is “thus based on the belief in the possibility of recreating a natural unity through dream, trick effects, or fantasy, and of finding the ‘right’ voice for the ‘right’ body. This belief disappears in the 1970s: there is no ‘natural’ voice; every voice is a construction and forms a particular composite with the body”.78 To further illustrate this point, the final credits reveal that Jean Hagen, who plays the scheming and tone-deaf Lena Lamont, actually had a much better singing voice than Debbie Reynolds, who plays Kathy Selden. In a fascinating twist, as Kathy appears to be dubbing over

Lena’s shrill screech, we actually hear Hagen’s voice (she plays Lena) coming from Kathy’s mouth. In addition, some of the tapping in Kathy’s dancing was dubbed in by Gene Kelly.

In *Singin’ in the Rain* (1952) what we see and hear are tantalising glimpses and echoes of the potential contained within post-Edisonian sound-image juxtapositions. Gunning remarks that “Edison’s original intention in pursuing motion pictures was to bring [sound and image] back together”. A valid intention, if not only because in inventing both motion pictures and sound recording, Edison can be blamed for the machined separation of image and sound. It seems only natural that Edison should also receive some credit for the subsequent machined fusion, or reintegration, of image and sound. However, the reconciliation between sound and image has not been unproblematic; rather, it has been a reintegration marked by debate, tension, and complaints of favouritism. Sound and image had quite contentedly built up separate aesthetics, distinctive identities, and different conventions due to the separation and autonomy of the senses. Two distinct channels of communication with different histories, practices, and identities cannot simply be reunited without some degree of violence. This space of tension and conflict should be a focus of any attempt to theorise the contemporary relationship between sound and image in our audiovisual age – and having looked at the independent histories of sound and image, we move now to explore the frequently tense, conflicted history of their integration via technology and the impact on both practice and theory.

---

Chapter 3: Ingenious audiovisual deformations

This chapter will examine the historical development, evolution, and context of ‘ingenious audiovisual deformations’, as facilitated and disseminated by the cinema, which have come to be naturalised and widely accepted as persuasive fictions of reality. In a lovely miniature essay entitled ‘On Dubbing’ (first published in 1945), the Argentine short story writer and essayist Jorge Luis Borges spent some time musing on the “ingenious deformations” implied by the art of dubbing new sounds with moving images: “The art of combination is not infinite in its possibilities, though those possibilities are apt to be frightening”. Borges goes on to list some of the more fantastic and mythical aspects of these frightening possibilities of combination:

The Greeks engendered the chimera, a monster with the head of a lion, the head of a dragon, and the head of a goat; the second-century theologians, the Trinity, in which the Father, the Son, and the Holy Ghost are inextricably linked; the Chinese zoologists, the ti-yiang, a bright red, supernatural bird equipped with six feet and six wings but with neither face nor eyes; nineteenth-century geometrists, the hypercube, a four-dimensional figure enclosing an infinite number of cubes and bounded by eight cubes and twenty-four squares.¹

Borges then arrives at his point: “Hollywood has just enriched this frivolous museum of teratology: by means of a perverse artifice they call dubbing, they devise monsters that combine the famous face of Greta Garbo with the voice of Aldonza Lorenzo. How can we fail to proclaim our admiration for this bleak magic, for these ingenious audio-visual

deformations?” Of course, Borges is complaining about the dubbing of voices obviously at odds (and out of sync) with the moving images they are meant to accompany. However, written at a moment in history when sound film was just beginning to take off, Borges’ contemplations could not possibly take into account how these ‘ingenious audio-visual deformations’ would further rapidly evolve in new directions; as sound-image relations morphed beyond the simple dubbing of voices onto new images; as the machined fusion of sound and image embarked upon more ambitious avenues; as what were formerly viewed as ‘ingenious audio-visual deformations’ have come to seem utterly transparent and commonplace.

A significant development occurred in the early 1920s, when a Dutch-born electrical engineer named Iwan Serrurier invented the Moviola, which provided an essential addition to Edison’s legacy. The machine was initially offered to the general public as a way of viewing home movies – hence the friendly name of Moviola, reminiscent of the then-popular Victrola record player. It very likely would have withered away as a footnote to film history if not for the technical breakthrough that occurred in 1927: sound. The Moviola ushered in the Mechanical Age of editing, becoming the editing tool of choice in the movie industry. Its main attribute – the ability to study the motion of the images, frame by frame – allowed editors to lip-read the silent frames, and thereby edit and synchronise sound and image. The ‘double-headed’ (picture and sound) Moviola was enthusiastically adopted by the movie studios, where it (and subsequent editing platforms) enabled further permutations in sound and image. In the 1930s, soundtracks were converted from the sound-on-disc system to optical sound-on-film

---

2 Ibid.
technology, allowing sound and image to be edited and synced by the same machine. Images and sounds could be dissected into different durations of time, to be ordered and juxtaposed together later. Images and sounds could be mixed together and manipulated in strange ways, creating images that spoke, re-sounded, and made music in a fascinating simulation of reality.

Of course, this mixing together of images and sounds that did not necessarily belong together provoked a series of anxieties and protestations. The well-known philosophical chestnut – about whether a tree falling in the woods makes a sound if there is no one there to hear it – is a question that can be used to muse upon perception in an audiovisual age. Schafer offers his response:

It would be unimaginative to reply that it sounds merely like a tree falling in the woods, or even that it makes no sound at all. As a matter of fact, when a tree crashes in the forest and knows that it is alone, it sounds like anything it wishes – a hurricane, a cuckoo, a wolf, the voice of Immanuel Kant or Charles Kingsley, the overture to Don Giovanni or a delicate air blown on a Maori nose-flute. Anything it wishes, from past or distant future. It is even free to produce those secret sounds which man will never hear because they belong to other worlds…

However, it might be more apt to remark that when a tree crashes in the woods and knows that it is on film, the sounds made can indeed be wondrous to hear: it can speak like a famous actor; it can sing like a glorious multi-tracked choir; it can growl like a beast; or it can produce awesome sounds akin to that of the largest object in the world falling ponderously to the ground.

---

The realist film director Jean Renoir once remarked that he regarded “dubbing, that is to say, the addition of sound after the picture has been shot, as an outrage. If we were living in the twelfth century, a period of lofty civilisation, the practitioners of dubbing would be burnt in the market-place for heresy. Dubbing is equivalent to a belief in the duality of the soul”.\(^4\) Renoir felt that a person’s voice was an expression of that person’s soul, and that to fool around with it in any way was to do the devil’s work. The devil, of course, has been frequently represented as having a voice at odds with his appearance, hence leading to his association of devilry and duality and dubbing. Renoir did not recognise that the recorded voice is, in a sense, always already a mediated construct. The point that is being missed is that sound-image relations were always already constructed, and in a sense, always already “ingenious audio-visual deformations”. These so-called deformations were to become commonplace and willingly accepted.

A range of commentators have argued that the classic model of the soundtrack exists to reinforce an “ideology of presence”:\(^5\) In this sense, the dominant practices are not held to be ‘natural’, instead operating as the embodiment of a particular ideological and aesthetic position characteristic of the dominant cinema, conforming to the desire to bury the trace of work in order to give the film an appearance of continuity and transparency. By matching recorded dialogue to lip-movements, and sound effects to their apparent causes, audiences are supplied with a powerful, persuasive illusion of the ‘real’ presence of the characters, objects, and events onscreen (or even offscreen). Sound provides the guarantee of the fullness of the filmic world, persuading us of the

---

\(^4\) Renoir, 1974, p. 106.

\(^5\) See, for example, Baudry 1975; Doane 1980a, b; Neale 1985.
completeness and coherence of the fiction. As Altman argues, sound completes the illusion begun by the vision track, their interaction hiding the fact that both are artificial constructs. However, “this tells only part of the story… The classical paradigm succeeds not simply by commitment to the illusion of presence, but by playing on the techniques that make that illusion possible”. In this sense, we can see that the language of soundful images that is being produced, circulated, and consumed, can be traced to various techniques, practices, and technologies of recombining and conjoining images and sounds that emerged with the sound film. The contemporary state of the fusion of sound and image is clearly simulacral, particularly with the use of postproduction sound, but what will become clear in this chapter is that the conjoining of sound and image is in a sense always already a simulacrum.

The arrival of the sound film coincided with the encroachment of a new model of sound and listening, founded on the electrical reproduction of sound and the accompanying new habits of listening. This new model of sound has evolved from what appeared to be the straightforward writing (or recording) of sound to the construction and assembly of hyperreal (or realer than real) sound. Initially, the invention of sync sound corresponded with the goal of accurately reproducing a pre-existing historical ‘event’. After The Jazz Singer (1927) and other early synchronous sound films, Hollywood conventions of sound recording and reproduction underwent a relatively rapid shift in emphasis. Recording film sound was no longer thought of as the documentation of a pre-existing ‘event’; instead, it was soon realised that the relative independence of the soundtrack and the image track enabled the creation and construction of a pseudo-event.

---

6 Altman, 1980.
Because the sounds and images are recorded separately, and only subsequently integrated via mechanical linkage,\(^8\) an event can be produced that has never originally occurred. Lip-sync practices are a basic example of creating a pseudo-event. With the advent of the ‘talkies’, many film stars were discovered to have voices that were deemed unsuitable. A solution was available; another voice could be recorded, and later paired with the visuals so that audiences would hear an actor speak or sing in a voice that was not his or her own. A voice could be separated from its originating body and played in careful lip-sync with the images of another person. The magic of synchronisation allows film audiences to hear the voice emanating quite naturally from the actor pictured on the screen. (This scenario is famously played out in *Singin’ in the Rain.*)

It was a logical progression that sounds other than voices could be separated from their originating sources and matched to different images. By recording the sound on a medium separate from the picture, selecting and manipulating the sound, and only subsequently combining it with the images, the soundtrack can be optimised for its task in perpetrating the cinematic event. Sound designers and sound editors have generally come to adopt the aesthetic values of a transparent construction of pseudo-events. These pseudo-events are then taken as persuasive depictions of reality by audiences. As Comolli has written: “However refined, analogy in the cinema is a deception, a lie, a fiction that must be straddled – in disavowing, knowing but not wanting to know – by the *will to believe* of the spectator, the spectator who expects to be fooled and wants to be fooled,

---

\(^8\) ‘Double system’ refers to the now-standardized practice where picture and sound are recorded on separate mediums, in order for picture and sound to be edited and mixed individually, to be subsequently synchronized during reproduction by mechanical or electronic means. Typically, double system is used until nearly the end of postproduction, when ‘married’ prints become available, with sound and picture combined on the one medium.
thus becoming the first agent of his or her own fooling”. 9 This is the classic filmic disavowal tactic, where the intermingling of presence (e.g. the moving, speaking figures on screen) and absence (who are not ‘really’ there) works to enable the spectator to believe in the reality of what is on the screen, while at the same time ‘knowing’ that it is a simulacrum. This filmic tactic of disavowal can also be applied to synchrony, the assertion of a relationship between action that is seen and heard simultaneously: although I might ‘know’ that the sound was not necessarily originally paired with the image, but all the same I believe these sounds and images belong together. The fundamental gap between sound and image has become increasingly naturalised and accepted. The system of meaning that now governs the production, transmission, and reception of sound-image events was to develop over the passage of time, with numerous theoretical, practical, and aesthetic debates and struggles taking place along the way. As these debates have been resolved, aesthetic perspectives and practices have become naturalised, thus obscuring and effacing their histories of conflict and tension. Comolli makes the point that in viewing the history of ‘realistic’ media, we can clearly discern a process where there is a “a constant trade-off of codes, a chain of substitutions producing the reality effect for audiences, rather than as a asymptotic movement toward the axis labelled ‘reality’”. 10

During the early years of the sound film, two eventually contradictory requirements of good recording battled for ascendancy. One approach prioritised the need for intelligibility of dialogue; and the other line of thought preached adherence to ‘naturalness’, or acoustic fidelity to the original rendition, which demanded the maintenance of the spatial characteristics of the original, and thus a sound/image match.

10 Manovich, 2001, p. 189.
The insistence on intelligibility at the expense of fidelity in the sound film eventually won out, necessitating the *construction* of an acoustic narrative in support of the visual scene. Dialogue (voice) was privileged in the soundtrack, placed (mixed) significantly in the foreground, and generally maintained at a consistent sound level (volume), despite the spatial placement of the speaking actor on the screen. By imposing what was considered to be ‘the proper perspective’, a new aesthetic began to be created and conventionalised, accepted by audiences as the appropriate reproduction of sound-image events.

*The Jazz Singer* (1927) is well-known for the pioneering dialogue that occurs between Jolson’s character and his mother. However, as Thompson points out, its historical significance also derives from how the soundtrack begins to present a variety of acoustic spaces, as the voice of Jolson’s character was heard in a theatre, but also in a temple, a restaurant, and his mother’s front parlour. In other words, the soundtrack could now move through space, “inhabiting the numerous and diverse places that had been long represented visually in silent films”. As soundtracks were discovered to be capable of presenting a sense of space and place, the original aim “of simply creating an accurate reproduction of ‘theatre sound’ was no longer perceived to be adequate or appropriate”.11

The ideal of recording and reproducing a ‘natural’ representation of space was soon superseded by the privileging of clarity and intelligibility, not spatial realism. Many of the sound engineers who came to work in the movies were originally trained as radio and telephone engineers. These industries emphasised clear, intelligible voice signals as the criterion for ‘good sound’; these engineers had, therefore, been trained to think of the sound they produced as a product, an aural commodity, and Rick Altman has argued that

---

the kind of sound track they ultimately constructed came to offer “sound that is made for us”, quite clearly situating the listener as a consumer of sound. As the late 1930s rolled past, sound technicians adopted fresh conventions, which reflected a changing practice that remains widely in place today: the technicians would come to focus on collecting a uniformly ‘close-up’ sound signal; the goal was to capture voices clearly and directly, and this was accomplished by following the actors closely with moving microphones suspended from booms. Altman adds: “Throughout the thirties it was for the clarity of their soundtracks that sound technicians had been praised and rewarded, rather than for their spatial realism. What had once appeared as monstrosity had now become the norm”.

Sound engineers had begun to use the tools and techniques of sound recording itself to create the effect of space as early as 1928. Thompson notes that “the technique of microphone placement constituted a powerful new means by which to create or efface the aural effect of space”. Improvements had been made in microphone technology, and the use of specific microphones or multiple microphones were becoming viable options in recording sound for film, allowing the collection and mixing of different voices and sounds. As techniques for sound mixing, editing, and dubbing (or rerecording) continued to develop, the opportunities afforded by the use of multiple microphones were increased even further. This aided technicians in achieving the new clarity of their soundtracks. Dubbing and other sound manipulation techniques “makes possible”, sound engineer Joe Coffman declared in 1930, “the improvement of voices and effects through changing their frequency content by use of the requisite filters; it permits almost any imaginable

---

13 Ibid, pp. 54-55.
acoustic trick, and the inclusion of effects which occur as afterthoughts… It is probable”,
Coffman predicted, “that within a year no original sound records will be used for the
making of release prints of feature productions of high quality”.15 Indeed, as historian
Donald Crafton has documented, by 1930 the sound track “came to be seen more as an
ensemble constructed in postproduction rather than as a record of an acoustical
performance”.16 Another film historian, John Belton also notes: “Whereas initially the
sound track was ‘recorded’, now it is ‘built’”.17 Rapidly developing sound technologies
“provided new means by which to construct the sound of space, as engineers learned to
create electrically a spatialized sound that we would call ‘virtual’. The sound of space
was now a quality that could be added electronically to any sound signal in any
proportion”. As Thompson emphasises, this resulted in a “new sound [that] bore little
resemblance to that which had been heard” previously.18

The new model of sound therefore moved away from what can be defined as a
sort of sound ‘naturalism’, which postulates that sounds and images start out in ‘natural
harmony’. During the early days of sound film, the ideal of sound ‘naturalism’ can be
clearly identified as being present in many of the technical debates over ‘correct’
methods of sound recording and playback. For example, Carl Dreher, chief sound
engineer for RKO, adopted a line similar to numerous other technicians of the period,
where he stressed the importance of maintaining a ‘natural’ proportionality between
image and sound.19 Dreher’s appeal was founded on the apparently natural relationship
that exists between the picture of a speaking person and the voice associated with it, and

17 ‘Technology of Film Sound’, in Film Sound, edited by Weis and Belton, 1985, p. 70.
was largely emblematic of the general tone and force of early arguments for the importance of achieving some type of sound/image match. In a 1930 article, RCA sound technician John L. Cass pointed out that in order to maintain intelligibility of dialogue, more and more studios were resorting to the use of multiple microphones, with a mixer choosing the best, that is, the most intelligible, sound. “When a number of microphones are used, 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”.

In other words, the prevalent practice was resulting in the constitution of a monstrous spectator, presenting sounds that were not natural, nor fit for human ears. Cass concluded: “Since it is customary among humans to attempt to maintain constant the distance between the eye and the ear, these organs should move together from one point to another in order to maintain our much mentioned illusion [of reality]”. We can read Cass’ polemic against the onset of ‘monstrous’ perception as an argument fuelled by the desire to re-embody sound and image ‘correctly’, where seeing and hearing, sound and image, are fused according to the ‘appropriate’ perspective. Cass wanted to restore what had been taken out of bodies; he wanted to reinstall a ‘natural’ perspective of sounds and images. The point that Cass overlooked, however, was that there was no ‘natural’ perspective with regards to the reconstituted totality of image and sound, only a process of naturalisation that would come to bridge the disjuncture between sound and image. Through observing the transition away from this ‘naturalist’ perspective, we can trace how perceptual transformations have gradually become naturalised, as various

---

21 Ibid.
‘monstrous’ conventions have now become accepted and transparent, and as the arguments that have previously raged over ‘fidelity’ and ‘realism’ fade away, concealed within the seamless sound-image event (always to be replaced, of course, with fresh debates over newer models of ‘fidelity’ and ‘realism’). Cinema should not be conceived of in terms of a linear march toward a single possible language, or as a progression toward perfect verisimilitude. Rather, we should view “its history as a succession of distinct and equally expressive languages, each with its own aesthetic variables, and each closing off some of the possibilities of its predecessor”.

Altman remarks that the notion of fidelity can be more precisely described as “a way of assessing a recording’s adherence to a set of evolving conventions, like the parallel standards established for such culturally important qualities as ‘realism’, ‘morality’, or ‘beauty’. The concept of fidelity is thus a strange hybrid of engineers’ aspirations and ideology, serving to mask recording’s representational nature”. Changing perceptual norms are therefore the result of technological developments in conjunction with institutionalised changes in production and reception. In other words, instead of being an objective framework established outside sound reproduction in order to describe it, the discourse of fidelity is a key construct contained within the history of sound reproduction. Jonathan Sterne notes that “after 1878 [when the term fidelity was first applied to sound], every age has its own perfect fidelity”; this demonstrates that “the idea of ‘better’ sound reproduction was itself a changing standard over time”. He concludes, “If perfect fidelity simply meant a set of social and sonic relations in which participants could have faith, it would be no wonder that we find repeated declarations of

---

perfect reproduction from the 1870s on down to the present”. Sterne’s work on sound reproduction and fidelity makes the argument that fidelity is a misleading concept because all recorded sound is always already performative, always already mediated by the technology it will be recorded by. As he explains, “When one traces recordings back to their so-called sources, one finds the intersection of cultural forces that made initial and subsequent moments of reproducibility desirable and possible. There was no ‘unified whole’ or idealised performance from which the sound in the recording was then alienated”. Sound reproduction technologies should be understood as artefacts of particular historical, social, and cultural practices and relations. As Sterne observes, “it is both easy and tempting to forget the enduring connection between technology and a larger cultural context”.

The notion of a seamless, absolute, acoustic fidelity can be traced back to its historical origins. The notion of ‘better’ (mediated) sound – made possible (and a sardonic “of course!” seems called for here) by ever-improving sound technology that at the same time seems to become ever-transparent, where mediation is rendered less and less obvious – has been in play since the very inception of the possibility of sound recording and reproduction. Advertisements at the turn of the Twentieth Century touted recordings as “lifelike”, “a true mirror of sound”, “natural”, and “the real thing”. On February 1, 1904, the opera tenor Enrico Caruso recorded his first sides for Victor

---

Records. Two years later, in 1906, the Victor Talking Machine Company introduced the Victrola, the first phonograph designed as furniture, a console in ‘piano-finished’ mahogany. In the same year Victor’s prestigious Red Seal line was headlined by Caruso and Patti. The Victor recordings by Caruso were very successful – they were often used by retailers to demonstrate Victor phonographs; and Caruso’s rich powerful low tenor voice highlighted the best range of audio fidelity of the early audio technology while being minimally affected by its limitations. In the 1910s and 1920s the Victor Talking Machine Company ran a series of ads: beneath the illustration of Caruso standing next to a Victor desk record, the ad caption proclaimed that “Both are Caruso”, and went to boast that: “The Victor record to Caruso’s voice is just as truly Caruso as Caruso himself. It actually is Caruso – his own magnificent voice, with all the wonderful power and beauty of tone that make him the greatest of all tenors”.

From 1915 to 1926, the Edison Company sponsored a series of Tone Tests – where phonographic “re-creations” of musicians by the Edison Diamond Disc Phonograph were compared directly to live performances by the same musicians. “Tone Testing reached its peak of popularity around 1920, when over two thousand recitals were presented across the nation, including one at Carnegie Hall in New York”. Curious crowds gathered in auditoriums and concert halls to engage in “a very public kind of critical listening”, as the act of listening to reproduced sounds “was implicitly accepted as culturally equivalent to the act of listening to live performers”. The Tone Testing campaign was discontinued in 1926 – listening to phonographic sounds was no longer an

---

exciting novelty, but a familiar event. 31 Edison had boasted that his Diamond Disc phonograph had no tone of its own to distort the sound of the music captured on records; however, the characteristic qualities of electroacoustic reproduction have become a desired feature, a commodity to be experienced and enjoyed. As Eisenberg points out, the acoustic phonograph did in fact “have a sound, warm and wooden and all its own. It was an instrument in its own right. Accurate or not, that sound was what people had come to expect from a phonograph”. 32 A 1927 advertisement for the Orthophonic Victrola described the new sound as “Vivid! Lifelike! As radically different as the modern motor-car in comparison to the ‘horseless carriage’. And the new Orthophonic Victor Records, recorded by microphone, have a character of tone that is pleasing beyond description. Rich. Round. Mellow”. 33 Loudspeakers did not simply amplify reproduced sound; as Thompson notes: “they also added their own characteristic to the reproduction, and people generally enjoyed this new kind of sound”. 34 What the audience hears is not the ‘natural’ voice of the performers, but their reproduction as rendered by loudspeakers. In Thompson’s words, the phonograph “taught [listeners] to distinguish between desired sound signals and unwanted sounds or noises”. 35 The fundamental impact of the phonograph was to introduce people to sounds that had been severed from their originating space and time, turning listeners into consumers – and connoisseurs, in a way – of acoustic events that were increasingly produced and packaged for attention through and by sound reproduction technologies, as an aural commodity.

31 Consumers had become “far more interested in listening to the electrically generated sound of radio”. (Citations from Thompson, 2004, p. 237.)
32 Eisenberg, 2005, p. 91.
34 Thompson, 2004, pp. 239-240.
In years to come, RCA Victor would advertise their sound machines as offering the “golden throat” sound, and “Living Stereo” – cleverly summarised by the iconic use of their trademark, Nipper, listening to his master’s voice. The well-known image of a dog attentively listening to a wind-up gramophone comes from a painting originally titled *Dog looking at and listening to a Phonograph*, which was later re-titled *His Master's Voice*. It was painted by British artist Francis Barraud around 1893. Nipper, the dog, belonged to the painter Francis Barraud’s brother Mark Henry, who died young. The original version of the painting did not show the familiar gramophone found in the trademark today, but rather a cylinder phonograph. Francis painted the picture before gramophones were commercially available, and the original sound reproduction device in the painting was therefore a phonograph (which could record on its own) and not a gramophone (which could not). The dog and phonograph were perched atop what many have speculated to be a coffin; and presumably the dog was listening to the voice of his deceased owner. In 1899, Barraud sold the rights to his picture and was told to replace the phonograph with a gramophone; the Gramophone Company first used this striking image of sound and hearing on publicity material in 1900. (At the request of Emile Berliner – the inventor of the gramophone and founder of the company – the American rights to the picture became owned by the Victor Talking Machine Company. In 1929 RCA bought Victor, resulting in the RCA Victor label.) Now, this world-famous trademark has become the property of the EMI group; and is used as the marketing identity for HMV Shops in the U.K. and Europe. The ‘His Master’s Voice’ image remains almost instantly recognisable; and I would like to propose (somewhat whimsically) that this image of

---


Nipper, listening to a recording of his master’s voice, symbolises a point in time where a new model of sound and hearing began to encroach on previously established norms.

In 1931, the physicist Dayton Miller would remark: “Acoustics is a science of the last thirty years”. The 1920s saw the rise of new industries dedicated “to a range of acoustical products and services, especially the telephone and radio”, which “offered new opportunities and resources for the study of sound”. New tools for producing, modifying, and measuring sound began to transform the scientific study of sound into a respectable discipline. Acousticians began to work with, and think about, sound in new ways. With the invention and growing familiarity of equipment and technologies such as microphones, amplifiers, and loudspeakers, sound and the study of sound were transformed – of course, the naturalisation of the electric reproduction of sound and the evolution of new ‘ways of hearing’ were not confined to acoustics and acousticians. Many of the principles that govern sound recording and reproduction today were discovered and developed in the thirties (with World War II putting an effective damper on further exploitation during the decade of the forties).

“By 1933”, the historian Emily Thompson declares, “both the nature of sound and the culture of listening were unlike anything that had come before”. There was a significant transformation in what people heard – which increasingly was a new kind of sound, the product of modern technology. The phonograph was busy making sound into a commodity, a material sign, like the photographic image. Radio was born in the early 1920s. Over the course of the 1920s, electrically amplified phonographs and radios equipped with loudspeakers became increasingly popular sources of aural entertainment

40 Thompson, 2004, p. 2.
in the home, even as public address systems and talking motion pictures transformed public spaces for listening. In the late 1920s, recorded speech and music were introduced to the motion picture theatre, where the ‘talkies’ gave birth to a whole new cinematic industry. By the mid-1930s, radios were playing continuously in hundreds of thousands of living rooms, businesses, and shops across Europe and America. In this context, sound assumed a higher cultural profile. Sounds could be stored, disseminated, circulated, and reproduced, like visual images. The naturalisation of electrically reproduced sound was accompanied by new ways of listening – people developed a new consciousness and appreciation of sound, as they “learned to listen in ways that distinguished the signals from the noise. This distinction became a basis for defining what constituted good sound: clear and controlled, direct and nonreverberant”. Notions of ‘high fidelity’ and sonic ‘realism’ originally emerged with the advent of the microphone and electrical methods of recording during the mid-1920s. However, the origins of high-quality sound reproduction can be found not in the early days of radio broadcasting and receiving but in the birth of the sound motion picture industry. The sound motion picture theatres that began to proliferate after 1927 would play an

---

41 The arrival of electrically reproduced sound can be traced to 1876, when Alexander Graham Bell invented the telephone.

42 Before the introduction of microphones in the mid-1920s, recordings were made using the acoustic (or mechanical) process. “Musicians sang or played into a recording horn, which funnelled the sound to a narrow opening covered with a flexible membrane (often of mica or glass); the diaphragm, as it was called, transferred the vibrations to a stylus, which in turn engraved a cylinder or disc. No electricity was involved”. (Katz, 2004, p. 37.) This method, of course, was cumbersome, and placed tremendous demands on the performers. The advent of electrical recording in 1925 drastically improved the quality of the recording process of disc records. There was a period of nearly five years, from 1925 to 1930, when the premier technology for home sound reproduction consisted of a combination of electrically recorded records with the specially-developed Victor Orthophonic phonograph, a spring-wound acoustic phonograph which used waveguide engineering and a folded horn to provide a reasonably flat frequency response. Electrically-powered phonographs were introduced in 1930, but crystal pickups and electronic reproduction did not become common until the late 1930s.


44 See Oliver Read and Walter L. Welch, *From Tin Foil to Stereo (2nd edition)*, 1976.
important role in shaping and disseminating the new sound, as these were spaces that
demonstrated a modern emphasis on the direct transmission and clear reception of
reproduced sound. Thompson notes: “The sound of the modern [movie] sound track only
echoed that being heard in countless other contexts”\textsuperscript{45}

When the ‘talking film’ arrived, the electronic reproduction of sound would
become commonplace, popularised by the movie theatre, even as the radio took over as
the prime sound source in the home. These sounds were created electronically and
generated by loudspeakers, which converted amplified electric signals into audible sound.
Radio broadcasting began in 1920 with the historic broadcast of KDKA, America’s first
regular broadcasting station.\textsuperscript{46} The first commercial electric loudspeakers appeared
around 1921.\textsuperscript{47} By the 1920s, more and more households could afford phonographs; and
the historical displacement of ‘live’ musicians by recorded music can be traced to the
gradual ubiquity of the phonograph and subsequent sound reproduction devices.\textsuperscript{48} By
1925, radio receivers were no longer complicated devices where solitary listeners heard
intermittently broadcast signals through their headsets. Rather, electric loudspeakers
projected the sound out into the room, enabling an entire family to listen together – and
listeners at home heard a ‘live’ signal that offered a sense of currency and connectedness
to other listeners. By 1928, people were largely accustomed to hearing radio sound

\textsuperscript{45} Thompson, 2004, p. 284.

\textsuperscript{46} Most radio historians agree on this point; see, for example, the radio time-line provided by Shingler and

\textsuperscript{47} “The earliest type of loudspeaker appeared around 1921 and consisted of a small electromagnetic
receiver, like that found in a telephone earpiece, attached to a goose necked horn. This model was soon
accompanied by the ‘cone-type’ loudspeaker, an electromagnetically driven paper diaphragm that was
capable of filling a room with sound without the assistance of any horn”. (Thompson, 2004, p. 239.)

\textsuperscript{48} By 1929, when talkies arrived, cinemas were accounting for a large proportion of all paid musical
employment – more than three quarters in England, according to union statistics; see Eisenberg, 2005, pp.
59-60. Of course, this would also change when the success of the talkies inevitably made musicians
performing live-to-picture superfluous.
emitted from loudspeakers. However, it was not until talking motion pictures became a reality in the late 1920s that speakers with better sound quality and louder sound capability were introduced, as movie audiences encountered the new Vitaphone sound motion picture system that was transforming the movie-going experience. Until the arrival of the talkies in the late 1920s, little attention was paid to the acoustics of movie theatres. By 1929, as previously silent theatres were wired for sound, “theatres across the nation were suddenly discovered to be acoustically deficient”,49 as the question of hearing suddenly became a pressing matter to consider. With the burgeoning popularity of the talkies, “virtually all of the major [movie] producers were building new studios in and around Los Angeles, and they depended on acoustical experts to ensure that these structures were both soundproof and nonreverberant”.50 This, then, was the historical, social, and cultural context for the ingenious audiovisual deformations that soon were to become naturalised.

A part of the reason as to why these audiovisual deformations were so disturbing was that sound recording has a particular relationship with the notion of fidelity, because while images are selectively shot and framed, sounds are apparently ‘captured’ in their fullness. The ideological claim of absolute reproduction rests on the belief that ever-improving audio technology reproduces sound ‘perfectly’, with no trace of technological mediation to be discerned in the quality of the reproduction. ‘Better’ sound is made possible by technology, which is transparent – even as it makes this ‘better’ sound

49 Thompson, 2004, p. 259; it was generally discovered that the problem in the old theatres was too much reverberation.
50 Ibid, p. 267; the coming of sound was to cause a number of problems. For example, the silence required of the new stages was challenged by the noise coming from the studio air-conditioning systems installed to counteract the heat from the lighting. Also, cameras were noisy, which required them to be confined in immobile soundproofed kiosks; microphones were large, fixed, and relatively insensitive; and it was difficult to edit and montage both sound and image.
possible. The ideal of ‘better’ sound (an idealised perfection) is associated with the quest for and proliferation of realer than real sounds, realer than the ‘real thing’ (and thus better?) – leading to the dissemination of hyperreal soundfulness that is increasingly becoming naturalised.

So although a discourse of realism has persistently reinforced the idea of recorded sound as the unproblematic ‘capture’ and ‘preservation’ of ‘real’, ‘natural’ sound, it is crucial to understand that recorded sound is always already mediated sound. Recording sound cannot be simply said to be the capture of sound in all its ‘original’ fullness and presence. In other words, recording does more than merely record – the impact and effect of recording technologies and techniques should not be forgotten. Sound is mediated through technologies that require users to adapt their practices and habits in a variety of ways. Recording involves a particular selection and presentation of sound. In a great majority of instances, there is no ‘original’ sonic event that a record records or reproduces. Only live recordings record an event; studio recordings (which comprise the great majority of recordings) are assembled from bits of actual events. The ‘original’ sonic event never occurred; in short, the ‘recording’ constructs, performs, and invokes the idea of an idealised original. The word ‘record’, then, can be misleading.  

Cubitt emphasises that: “Recording is no more an innocent or transparent replication of real, pro-phonographic sound than the photograph is of the pro-filmic. It has its own techniques of framing and selection, its entrenched aesthetics, its practices of

---

51 Edison, however, chose this word deliberately. His invention of sound recording technology was initially intended to be used to preserve important speeches, business transactions, and voices of family members; only later did he see the potential for popular entertainment that lay in recording and reproducing musical performances (See Thomas Edison, ‘The Phonograph and Its Future’, North American Review 126, May-June 1878, pp. 528-529; also see Edison’s ‘The Perfected Phonograph’, in North American Review, 146, May 1888, pp. 641-650). Eisenberg remarks: “The use we put [sound recording] to now might strike him as fraudulent, like doctoring the records”. (Eisenberg, 2005, p. 89.)
erasing the marks of its making”. Recording and playback technologies are “locked into an endless dialectic of the delirious pursuit of an inaudible perfection”. However, “the more perfect a recording, the less it refers to a pre-existing sound. The craft of contemporary sound engineering is rich in technique to the point at which its devotion to the revelation of reality has to be doubted”. Mediated sound has come to provide a ‘perfect’, hyperreal, simulation. Jacques Attali has explained that in a certain sense, reproduction “is the death of the original, the triumph of the copy, and the forgetting of the repressed foundation”. Mediated sounds are no longer simple copies but sophisticated constructions, no longer mere imitations of ‘original’ sounds, but are new, perfect models that are more compelling, more defined, more present, in turn becoming new ‘originals’. There is a new model of meaning around sound, which is built on the notion of hyperreal sound and a new sound-image synthesis. Contemporary instances of mediated sounds present aural images that are augmented to sound ‘realer than real’, which are then transparently presented as perfectly ‘natural’ sounds in themselves. Mediated sounds are not mere imitative models, but transparently ‘real’. Mediated sounds are constructed representations – aural ‘images’ – that at the same time do not appear to be mere imitations. In fact, mediated sounds have come to be naturalised, and their ‘realer than real’ status is accepted as (almost) transparent, or conventionalised to such a degree that these mediated sounds are taken for granted as natural parts of our everyday soundscapes. Machined sound is no longer a copy, but a new real, which stands for itself.

In other words, machines did not simply ‘capture’ sounds that already existed in the world. Making sounds for the machines was always different than performing for a

52 Cubitt, 1997, p. 102.
53 Attali, 1992, p. 89.
live audience, and even in so-called live situations, the machine required a certain amount of attention, care, and technique. Sometimes, actual spontaneity would interfere with the recorded appearance of spontaneity. Recording did not simply capture reality as it was; it aimed to capture reality suitable for reproduction. Spontaneity was spontaneous only through artifice. Of course, it did eventually become possible to reproduce events as they happened for a listening audience. But, even here, the auditory reproduction of the actual event is highly contrived; the audience hears not so much the event itself as a performance concurrent with the event.

The ‘copy’ constructs the ‘original’ – sounds are carefully framed and selected for recording, and subsequently manipulated and combined in postproduction to provide the perception of an ‘original’ event that has been captured in its fullness and presence. The idea of an original is constructed through the possibility of repetition, imitation, and copying. The representation and interpretation of sound (through recording and playback technologies) depends on the idea of an original sound event – referencing the context of this original sound event. The original can only be verified as the original when there are copies or imitations, when it is compared to a repeatable model – the original is valorised, or makes meaning, when it is compared to a copy or imitation. When the replayed sound is heard, it makes meaning when we bring our knowledge, expectations, associations, and listening conventions to the sound in order to interpret it, in order to make it meaningful. The copy enables the original – it allows us to imagine an idealised, fully present original event that we then partake of, albeit in a partial, perspectival manner, forgetting that this original was always already partial and multiple. Multiple in that the original event
required context, an auditor, etc. – multiple in that the original sound is always formed of many sounds.

A recording is commonly charged with reproducing the original event. However, it is more accurate to say that a recording offers a particular perspective of a sound event: this practical limitation can be problematic or unacceptable due to our preconceptions around the ‘purity’, presence, and originality of a sound event. Recording has been critiqued on two different levels. First, the recording can be blamed for not being able to reproduce the ‘original’ event in all its fullness; and second (more seriously), the recording can be blamed for doing violence to the ‘original’ event. The fundamental theoretical problem at the heart of recording sound can therefore be summarised as follows – no practical system can be said to completely ‘capture’ the sound of most real sources in all their spatial complexity. Recorded sound always carries some record of the recording process, superimposed on the sound event itself. And then there is the playback process (and technology), not to mention any traces of editing, mixing, and mastering.

However, even before the mediation of recording and other technological processes, we need to understand sound as always already mediated by human hearing, and by various naturalised, taken for granted social, historical, and cultural practices of listening. This can be clearly illustrated by the contributions made by psychoacoustics. Psychoacoustics is the second major branch of acoustics; and it is concerned with the interaction of sound with the human hearing system. Psychoacoustics deals with subjective responses to objectively measured phenomena, as compared to physical acoustics, which is the scientific discipline that deals with measurable objective parameters of sound and its behaviour in any given medium. Most of the innovative
acoustic breakthroughs that occurred in the 1930s were the direct result of research at the Bell Telephone Laboratories under Harvey Fletcher and at the RCA research laboratories under Harry Olson. Fletcher and his associate Wilden A. Munson discovered the nonlinear behaviour of the human hearing mechanism in about 1933 (now known as the Fletcher-Munson effect), which marked a significant point in the beginning of psychoacoustic research.

In their research conducted in the early 1930s, Fletcher and Munson undertook to measure the sensitivity of the human hearing mechanism at different frequencies. They first generated sounds at very low levels through many frequencies over the audible range to determine the threshold of hearing, or the softest sound that could be heard. They made a plot of these levels versus frequency, and found that the curve is not uniform, but varies drastically with frequency. The Fletcher-Munson effect acts to illustrate the realisation that ‘pure’ sound-in-itself is an ideal because for us sound is always already mediated by the parameters of human hearing, as well as by situational, environmental, and contextual factors. Our human mode of hearing, our psychoacoustic interpretation of sound, means that sound is always mediated, always organised and perceived according to our human perceptual selectivity and limitation, even as we stand in the direct, immediate presence of sound being made by its ‘natural’ source, in its ‘natural’ environment.54

Sound is typically treated as “a natural phenomenon exterior to people, but its very definition is anthropocentric”. As the physiologist Johannes Muller pointed out over 150 years ago: “without the organ of hearing with its vital endowments, there would be no such thing as sound in the world, but merely vibrations”. Muller emphasised that our

54 As the phenomenologist Maurice Merleau-Ponty argues, perception is always an embodied perception, which always occurs within a specific context or situation. Perception in-itself does not exist. See his *Phenomenology of Perception*, 1962.
other senses can also perceive vibration. “Sound is a very particular perception of vibrations. You can take the sound out of the human, but you can take the human out of the sound only through an exercise in imagination”. Sterne explains, “As part of a larger physical phenomenon of vibration, sound is a product of the human senses and not a thing in the world apart from humans. Sound is a little piece of the vibrating world”. As such, “It is not possible to sample the acoustic world, to audit an event, without participating in it”. Holman concurs; it is clear that: “Any sound undergoes two principal processes for us to ‘hear’ it. First, it is generated and the objective world acts on it, and then that world is represented inside our minds by the processes of hearing and perception”.

Furthermore, we don’t hear in the dispassionate manner that the microphone collects sound; rather, our hearing undergoes a process of selectivity, as we focus on certain sounds that are important to us, and ignore others, which is exemplified by the psychoacoustic example of what is called the cocktail party effect – simply put, at a noisy cocktail party, when we engage in conversation, we focus on the sound of the other person speaking to us, while blocking out or selectively ignoring extraneous noises, such as the babble of other conversations and the clink of glasses. A microphone, however, placed in the same general listening situation, would indiscriminately pick up all the sounds in the vicinity. The process of selective auditory attention is far more difficult when we are listening to recorded material. As Altman puts it, while “live sound provides

55 Sterne, 2003, p. 11; the citation from Muller is also drawn from this source.
56 Ibid, p. 11.
57 Ibid, p. 284.
59 The term can be traced to Colin Cherry, a British researcher working at MIT, who reported research on what he called the ‘cocktail party problem’ in 1953.
an extraordinary number of variables, each permitting and promoting selective attention, recorded sound folds most of these variables into a single, undifferentiated source. In a live situation, we easily differentiate among the various sound sources surrounding us, but with recorded sound no such clear distinctions are possible".60 The solution would be to close-mike the conversation – an acoustic variation of the visual close-up – before mixing in the background noises, sounds, and voices, while maintaining the *emphasis* on the primary conversation. Increasingly, the sounds we hear have been *selected* for our attention; microphones hear on our behalf, as we lend our memories and perceptual expectations to machines – sounds are collected by particular microphones, mixed judiciously, subjected to manipulation, and replayed for our audition. We depend on the machined soundtrack to present us with a selected array of sounds.

R. Murray Schafer writes that: “Machines listen differently than men do. They have exceptionally wide hearing range, fine sensitivity, and no listening preferences whatsoever”.61 It is true that the microphone does not have a brain, nor a memory, and hence no listening preferences, or perceptual expectations; it cannot tell the difference between the sound that you want to capture, and extraneous noise. However, every recording application has a requirement; a right microphone exists for a particular recording situation.62 As Rick Altman has written:

To record is thus to recall to mind, as the dictionary would have it, but like most mnemonic devices, sound recordings must heighten some aspects of the original phenomenon at the expense of others. So-called recordings are thus always

62 See Yewdall, *Practical Art of Motion Picture Sound*, 1999, p. 44.
representations, interpretations, partial narratives that must nevertheless serve as our only access to the sounds of the past.\textsuperscript{63}

However, this selectivity is always present, and not just a side-effect of the recording process. Even as we listen to an ‘original’ event (without technological mediation), the psychoacoustic characteristics of our human hearing are the subjective ‘filters’ that enable us to interpret the sounds that seemingly are ‘fully present’.

So the prime goal of any sound recording is the creation of an illusion that creates an ‘honest’ sense of place. This may seem like a contradiction, but it should be realised that the recording will never be the same as what is heard in the living soundscape. Sound engineers working for the cinema began to realise that “the goal of cinematic representation is to create the ‘illusion of occurrence’, that is, the represented effect of an event that seems to exist independent of the act of recording – an event that seems simply to ‘occur’ and to be captured”.\textsuperscript{64} These constructed sonic worlds, however compelling, “presume no necessary link to the actualities of either real objects or real acts of recording”.\textsuperscript{65} These ‘original’ copies therefore model an original that never occurred; an imaginary original that serves as semiotic template; a contextless original that can be recontextualised and thereby made meaningful. Such idealised sounds are becoming increasingly prevalent; where the background ‘noise’ of the original context is becoming increasingly effaced. Just like the hyperreal image has come to define how we look at the world, hyperreal sound has come to define how we hear the world, because it is designed to be ‘realer than real’; sounds have been selected, mixed, and shaped for our perception.

\textsuperscript{64} Lastra, \textit{Sound Technology and the American Cinema}, 2000, p. 178.
\textsuperscript{65} Ibid, p. 74.
With sounds infusing moving images, providing a sense of space, place, time, occurrence, and presence, film (the audiovisual event) can be said to present a ‘complete’ depiction of events.

As Holman points out, “production sound recordists, boom operators, and recording engineers” have “become highly proficient at choosing one, or a few, microphone positions that instead represent the sound of the source, without making a valid claim to actually reproduce the source completely”.66 After all, if sound recording devices are placed in different locations at a concert hall – among the boxes, the stalls, the dress and the upper circles, on and behind the stage, and also in the orchestra between the seats of the players – the various recordings of the music will all offer different perspectives of the performance. Differences of emphasis will be easily distinguishable, with no two recordings being alike – variations of clarity and power in the reproduction of the music will be evident, according to the location of the recording machines.67 The fundamentally important question we must raise here is: which is the ‘original’ sound event? As Schafer notes, for human listeners “the perfectly pure and mathematically defined sound exists as a theoretical concept only”. After all, “the moment a sound is produced”, distortion occurs, “for the sounding object first has to overcome its own inertia to be set in motion, and in doing this little imperfections creep into the transmitted sound”. This is also true for the human ear. “For the ear to begin vibrating, it too has first to overcome its own inertia, and accordingly it too introduces more distortions. All the sounds we hear are imperfect”.68 Realism and fidelity must be understood to be performative values; they serve as regulatory ideals, as semiotic templates. Sound

67 This observation was made by Richard Semon in his book The Mneme, 1921, p. 125.
technicians have largely reconciled themselves “to a mode of production based not on the faithful collection of real events, but rather the construction of carefully hierarchized events whose realism is a function of their plausibility and their compatibility with conventions of narrative realism rather than literal, perceptual duplication”. Fidelity is judged by evolving conventions, where each successive generation claims perfect fidelity, only for subsequent generations to discern the evidence of constructedness and mediation as new conventions of perception, production, and reception are established and rendered semi-transparent in turn. The fidelity model is a discursive ideal that an ‘ever-improving’ sound technology appeals to and claims at each stage of its development. Fidelity can also be seen to be a by-product of synchronisation, as fidelity is installed by effective film sound, where sound matches up to and is ‘faithful’ to the image. In other words, “determinations of sound quality usually prove to be creatures of the historical moment, not of some timeless measure of sonic realism”.

The transition, then, from mono, to stereo, to surround sound in the cinema – a transition which can be viewed as an evolution towards hyperreal sound – can be read in terms of how these formats have been adopted, naturalised, and subsequently taken for granted. Stereo (short for stereophonic) had its origins as three-channel sound, but actually is a term that refers to a wide range of multi-channel recordings – although stereo, for the general public, has come to typically refer to two-channel sound. In the early 1950s, stereo magnetic sound was introduced as a new method of putting sound on

69 Lastra, 2000, pp. 177-178.
70 Kahn, 1999, p. 140; as Sean Cubitt remarks, a common misunderstanding is to view a recording as “always a communication without a content, pure mediation, always an effect, never the thing itself. If we have learnt to listen habitually, and habitually to identify sounds by their origins, it is because we have forgotten how to hear”. (Cubitt, 1998, p. 102, my italics.) In simple terms, a recording should be acknowledged as a ‘thing itself’.
film, a revolution which was forecast to herald a new era in sound motion pictures. Magnetic sound offered a significant technical advancement, and provided much improved fidelity over the conventional optical soundtrack. It also enabled the first multi-channel sound reproduction, dubbed ‘stereophonic sound’, to be heard by the public.71 Sound spread from one central speaker behind the screen to additional behind-the-screen speakers on the left and on the right, and sounds emanating from a source appearing to the left, centre, or right of the picture (as well as off-screen sounds) could be heard from as many as five behind-the-screen speakers and from dozens of ‘surround’ speakers mounted on the side and rear walls of the theatre auditorium. Music took on a new dimension of ‘realism’, and sound effects could emanate from the rear or sides of the cinema.

In terms of film exhibition, magnetic sound was widely adopted in the 1950s. By the 1960s, however, when the film industry experienced an overall decline, the expense of magnetic release prints, their comparatively short life compared to optical prints, and the high cost of maintaining the playback equipment led to a massive reduction in the number of magnetic releases and cinemas capable of playing them. Stereo sound failed to overthrow the mono format, and magnetic sound came to be reserved for only a handful of first-run engagements of ‘big’ releases each year. Moviegoers were again hearing low fidelity, mono optical releases most of the time, with only an occasional multi-track stereo magnetic release. However, the initial relative failure of stereo was not just due to technical difficulties. Film audiences were simply not ready for the stereophonic revolution.

71 This coincided with the wide-screen revolution that transformed the nature of the traditional motion picture experience in the mid-1950s.
From the first days of the cinematic transition to sound in 1927, theatre loudspeakers were placed to the side or below the screen. Early Vitaphone films drew upon silent film conventions, playing back orchestral scores through speakers placed in the area of the former orchestra pit. In 1927, Earl Sponable developed a porous screen material, which facilitated the placing of loudspeakers behind the screen. John Belton explains that:

This encouraged the illusion of the homogeneity of sound and image, which was achieved quite literally through their physical imposition. Over the years, this location became a rigid convention – sound came from the centre of the image. For over twenty-five years, dialogue had been played back to audiences from central speakers located behind the screen.72

As such, it was not surprising that stereo sound was initially perceived to be unnatural and artificial. The representation of events provided by mono sound had come to be associated by audiences with realistic representation.

During the first years of multi-track sound, attempts at ‘real’ spatialization were made; sound was located on the left side of the screen if its source was shown there. Although spatialized sound contributed an apparent realism to the scenes, providing close correspondence between sound and image, it also functioned as an ostentatious display of what multi-track stereo magnetic sound could do. While undeniably realistic, the newfound stereophonic practices nonetheless drew attention to the technical apparatus, violating timeworn conventions of stylistic transparency. Having been conditioned by the ‘second nature’ of mono sound, audiences could not avoid perceiving the events which

unfolded before them as being mediated (constructed) by the stereo sound technology. The introduction of stereo added a spatial, and hence a visual, aspect that at once clarified and confused. It was not simply a technical improvement. Stereo projected sound from a particular location, drawing the eye, and creating a virtual sense of space – which listeners felt uncomfortable with, being used to the ‘natural’ projection of mono sound.

The relative failure of magnetic sound would deprive motion picture spectators of high quality, state-of-the-art sound in the theatre until 1975, when a stereo sound system began to gain wide acceptance through the efforts of Dolby Labs to market a 4-track optical stereophonic sound-on-film system. The breakthrough was the development of a highly practical stereo optical release print format originally identified as Dolby Stereo. In the space allotted to the conventional mono optical soundtrack are two soundtracks that not only carry left and right information as in home stereo sound, but are also encoded with a third centre-screen channel and – most notably – a fourth surround channel for ambient sound and special effects. This format not only enabled stereo sound from optical soundtracks, but higher quality sound as well. Various techniques were applied to the soundtrack during both recording and playback to improve sound quality. Foremost among these was Dolby noise reduction to lower the hissing and popping associated with optical soundtracks, and loudspeaker equalisation to adjust the cinema sound system to a standard response curve. Dolby’s involvement with film sound first achieved wide recognition with the spectacular audio effects of such films as Star Wars (1977).

Surround sound is therefore the latest incarnation of stereophonic sound – where sounds come from ‘all directions’. Dolby Stereo also pioneered an advanced noise-
reduction process, which improved sound quality further. Today, Dolby Stereo is the analogue sound standard, thanks to its superior sound quality and relatively simple installation. The heightened sound quality of Dolby Stereo led moviemakers to make more extensive use of the surround channel. (The real innovation of Dolby Stereo is how so much audio information is squeezed into a small space on the film.) In recent years, new and enhanced standards of film sound exhibition have been introduced. These new formats provided a growing range of independent surround channels, extraordinary dynamic capability, wide frequency range, and low distortion; these improvements enabled improved ‘realism’ and more precise sound placement. Digital surround sound has become the standard, normal delivery system.

Present-day film sound demonstrates a system that employs the assemblage of a soundtrack rather than the direct recording of already constructed sounds. This constructed soundtrack enables the provision of a clean, clear, continuous sound record, oblivious to image scale but attuned to dialogue intelligibility, story continuity, and freedom of action. The current fixation on sound quality over ‘correct’ spatial alignment with the image marks the current trend within the passage of different perceptual conventions. One of the primary contemporary conventions for assessing sound-image fidelity is based on judging the reproduction of sound quality; in this way, images are allowed to be clearly audible, possessing maximum impact. The power of post-recording transformation has become apparent, with numerous opportunities for enhancing what is to be heard by the audience. The sound must be pristine, rich in frequency detail, possessing a low signal-to-noise ratio, presenting good spatial characteristics, and providing a strong sense of presence. Sounds are understood to emanate
unproblematically from the images seen on the screen, aided by a mediating technological apparatus that strives to be invisible and inaudible.

Audiences have developed a finely honed audiovisual competency, or literacy, which we use to interpret or ‘read’ sound-image events; this competency is also appealed to through current discourses of fidelity and realism. Images are viewed as an ‘unproblematic’ source of sounds, and sounds act naturally in conjunction with images. Sound-image events have been conventionalised and naturalised through the passage of time; the development of soundful images has progressed through various aesthetic and ideological stages that have seen the (incipiently contested) evolution of different conventions, which thereafter become naturalised and transparent. Furthermore, images or sounds that stand alone, without being extended by the sound-image relation, now possess the potential to refer to these sound-image events: silent images seem to resound with the echoes of sound, and disembodied sounds seem to refer to images as their sources.

Lastra makes the comment that given the “long tradition of practice to the contrary, it is somewhat surprising how the academic literature about sound in film is so thoroughly permeated by the assumption that sonic representation amounts to nothing more than absolutely duplicating some event that is otherwise fully independent of the act of mechanical inscription. This state of affairs has resulted in a situation where the constructed nature of the sound of an image then typically comes to be measured against the ‘pure’ and ‘natural’ sound of the world, even as the notion of ‘pure’ sound (separated from source, and evaluated simply as emission) can ironically be in a way seen as a

---

73 Lastra, 2000, p. 90.
product of sound recording technologies. After all, sounds were tied to their sources, until the advent of sound recording devices permitted them to be disembodied. With the establishment of new techniques and practices, people developed a new expertise and perspective, hearing and judging sound as emissions disembodied and separated from their sources. Now, however, sounds are more and more referring to images, associated with images, converging with images, invoking images as their source – sound has become more visual and increasingly image-friendly. An important, influential, and evolving articulation and interaction is taking place between images and sounds; sound and image are increasingly converging, and are creating events that operate as models of cinematic (and perceptual) reality. However, in order to properly comprehend the implications of the contemporary articulation and interaction between images and sounds, it is necessary to interrogate a range of reductive misconceptions that have impacted upon how this relationship has been understood.
Chapter 4: Secondary orality/aurality and audiovisuality

This chapter will explore a variety of attitudes towards sound/image and hearing/sight to illustrate how they have been problematically founded on a binary opposition of sound/hearing and image/sight. Many modern critics have merely attempted to invert the hierarchy and to privilege sound over the image to put an end to the hegemony of the image. However, what the technological and critical reception history ultimately shows is that these two realms are now inseparable, as what we now have are re-sounded images which need to be apprehended as a whole. In presenting this argument, this chapter will begin with the critical reception of sound during the early years of sound film, moving on to discuss the influence of Walter Ong’s and Marshall McLuhan’s ideas about secondary orality on a range of modern ideas about sound, before concluding with a discussion of how re-sounded images need to be apprehended as a whole.

In the easy familiarity of our contemporary cinematic cultures, it might seem odd that any sort of a negative reaction would have been attached to the infusion of the sync soundtrack. (Even as elements of this historical bias still persist in the continuing classification of cinema as a visual art.) Even though the advent of the talkies quickly garnered popular appeal, not everyone was entranced by what sound technology had brought to the art of motion pictures. There were numerous technical issues to deal with during the introduction of sound. Also, the re-introduction of the voice was vociferously debated across various theoretical, conceptual, and aesthetic arenas. As Christian Metz observes, “The advent of the talking movies, which should have changed not only the films but the theories about them, in fact modified the latter in no way, at least during the
first several years. Films talked, and yet one spoke about them as if they were silent”.¹ Metz is somewhat overstating the situation – when sound came to the cinema, a number of theorists and filmmakers did seek to conceptualise the possibilities and complications introduced by this new aesthetic dimension, but the point that is made here is that films were still largely discussed in terms of a binary opposition between image and sound, where film was held to be and discussed as being a visual art, and the coming of sound was generally received as a problematic, unwanted intrusion. There were fears of an unnecessary verism, wherein film would merely become a poor imitation (or double) of reality, instead of an authentic art form – which was similar to the Lumière brothers’ original fear. Ironically,² the closer film got to ‘reality’, the greater the success in simulating immediacy and presence, the more film was criticised for being a pale, lifeless imitation (or representation) of reality, instead of being an artistic, authentic, original form of expression.

Historically, the early critical, conceptual, and theoretical debates over film sound can be separated into several categories. One initial argument was that film sound defaced an art built on silence, as film had achieved its independence due to a mode of representation constructed solely on the moving picture. This perspective gave rise to several supporting arguments: that sound was secondary to image; that sound was an unnecessary addition; that sound would cause the cinema to regress to nothing more than ‘filmed theatre’. This resistance to sound also existed in another more clearly articulated form: where the problem was not sound per se, but the ‘ancient human bondage’ of

² And here we can see shadows of Metz’s argument that the movies have a greater sense of reality than its predecessor, the theatre, because it has less of a direct presence (See Metz, 1974, pp. 9-11).
speech,\(^3\) which was the main cause for the relapse to ‘filmed theatre’.\(^4\) This later, more nuanced argument, gave rise to critiques of just how sound was to be employed: for instance, sound had to be subject to the same techniques of montage as the image, asynchronous sound was the way to go, etc.

A prevalent accusation was that: “The only distinction between [theatre] and cinema – silence – has been removed by Edison with his latest invention”.\(^5\) And a pseudonymous columnist in *American Cinematographer* of August 1928 would observe (in somewhat colloquial terms):

> Go to any talking picture, shut your eyes, and if you don’t get exactly the same effects as in a theatre, there’s something wrong that somebody’s gonna get heck for. For the ideal they’re working for in speech from the screen is to duplicate as far as possible the stage voice quality, the stage spacing of words, the stage delivery of speech, and the stage construction of sentences. All of which is the rankest kind of hooey – for pictures. Now reverse yourself and stop your ears, and see if you don’t see the punkest kind of movie action you’ve seen in all this world.\(^6\)

Filmmaking, which had already largely adopted certain habits of practice, all of a sudden had to take sound technology (together with its distinctive technical conventions and aesthetic practices for recording sound) into account when shooting sound films. And the theatre loomed as an older model and tradition, however regressive, with which to

---

\(^3\) A comment made by Rene Clair; see Kracauer’s *Theory of Film*, 1997, p. 102, and ‘The Art of Sound’, by Rene Clair, in *Film Sound*, edited by Weis and Belton, 1985, pp. 92-95.


\(^6\) Cited in Eyman, 1999, p. 194.
attempt to deal with the sudden influx of voices, and the unsure chorus of actors freed from silence.

Harry Carr, journalist and screenwriter, remarked quite matter-of-factly in a column: “It is quite possible that some new kind of talking picture might be developed”. However, he declared that the present style of filmmaking “will never become talkative. Instead of making the movies more real, it makes them less real. The voice accentuates a fact we sometimes forget – that movie characters are flat shadows on a wall”. He concluded that “the danger of disillusion” would be why pictures would never come to successfully talk. Sounds and images were happily separated, and should not be reunited. Rudolf Arnheim was of the opinion “that speech cannot be attached to the immobile image (painting, photography); but it is equally ill-suited for the silent film... It was precisely the absence of speech that made the silent film develop a style of its own, capable of condensing the dramatic situation”. For Arnheim, the cinema’s defects (for example, the lack of a third dimension) were a means toward an authentic form of artistic expression. The advent of sound therefore derailed what would have been the normal train-like progress toward a consciously artificial cinema by tempting filmmakers to submit to ‘the inartistic’ demand for a superficial ‘naturalness’.

In his autobiography, the director Josef von Sternberg (who worked in the days of early sound film) made the complaint of the ‘redundancy’ of sound, arguing that: “To be correctly and effectively used, sound had to bring to the image a quality other than what

---

7 Cited in Eyman, 1999, p. 105.
9 Arnheim, *Film as Art*, 1957; originally published in German in 1932. However, what Arnheim saw as something to be transcended – film’s mechanical reproduction of phenomenality – was seen by André Bazin to be the key to achieving cinema’s potential. For Bazin, the valorisation of realism was encapsulated in the realization of what he called the ‘myth of total cinema’. I will discuss this idea in more depth in a subsequent chapter.
the lens included, a quality out of the range of the image… [But] sound became a blatant addition, a saccharine, charmless frosting.”¹⁰ The critique of the ‘redundancy’ of sound was based on the argument that the talkies would then become a mere representation, imitation, or simple double of reality. The notion of the redundancy of sound is thus based on the idea that sound possesses too much of a ‘reality quotient’, thereby detracting from the abstraction of the image that is required for art. In this way, sounds are an unnecessary addition to images. The ‘added dimensionality’ of the soundtrack was scorned as being superfluous.¹¹

In the famous ‘Statement on Sound’ (1928)¹², signed by the Russian filmmakers Sergei Eisenstein, Vsevolod Pudovkin, and Grigori Alexandrov, the writers correctly predicted a tendency for the content of many films to be dominated by verbal rather than visual considerations. These authors were particularly distressed at the way synchronised sound cinema was reduced to the less ambitious ‘talking picture’, with its presumption that dialogue would play the leading role. “To use sound in this way will destroy the culture of montage, for every ADHESION of sound to a visual montage piece increases its inertia as a montage piece, and increases the independence of its meaning – and this will undoubtedly be to the detriment of montage, operating in the first place not on the montage pieces but on their JUXTAPOSITION”.¹³ For Eisenstein, Pudovkin, and Alexandrov, montage, as a principle based on the dialectical model of clashes between

---

¹⁰ Cited in Eyman, 1999, p. 270.
¹¹ ‘Redundancy’ is when images are prosaically accompanied by their straightforward sounds, but we need to consider the possibilities for soundful images, where sounds expand images, and images re-embryo sounds. The separately recorded image track and soundtrack are now quite capable of being fused together to create a new sound-image event.
shots, had “become the indisputable axiom on which the worldwide culture of the cinema has been built”\textsuperscript{14}.

Lev Kuleshov, recognised by many as the founding father of Soviet cinema, was the first person to use the term ‘montage’ in relation to film images. In an article in 1917, Kuleshov wrote that “to make a picture the director must compose the separate filmed fragments, disordered and disjointed, into a single whole and juxtapose these separate fragments into a more advantageous, integral and rhythmical sequence, just as a child constructs a whole word or phrase from separate scattered blocks of letters”\textsuperscript{15}. Montage, or editing, “is the key twentieth-century technology for creating fake realities”\textsuperscript{16}. In a basic sense, montage can be defined as editing together separate, often discontinuous shots taken at different times (and perhaps different places) to form a juxtaposed narrative taking place within consecutive moments in time. Montage defines the cinematic language as we know it.

Kuleshov devised an experiment, now famously called the Kuleshov effect, where viewers would see a close-up of Ivan Mozzhukhin, a famous Russian actor, projected immediately after three different images: first, a bowl of soup; then, a woman lying dead in a coffin; and finally, a girl playing with her toy bear. Audiences praised Mozzhukhin’s range – showing pensive hunger when he saw the soup, grief when gazing at the woman, and joy in seeing the child playing. Viewers interpreted the actor's expressions as being different in each shot and relating to the images, but in fact the three shots of the actor, one after each image, were exactly the same. The meaning of a shot therefore depends on

\textsuperscript{14} Sergei Eisenstein, \textit{Film Form and Film Sense}, 1957, p. 257.
\textsuperscript{16} Manovich, 2001, p. 148.
what precedes it in the movie, and this succession of scenes creates a new reality which is not merely the sum of its parts.

The authors of the ‘Statement on Sound’ believed this dynamic model of montage was threatened by the coming of sound for two reasons. First, sound threatened to smooth over the dynamic interplay required of montage pieces by dictating a scene naturalistically at the slow pace set by the synchronisation of speech emanating from bodies and sound from objects and actions. And second, the shots would acquire unwanted independence from the overarching theme of the film expressed in the montage. The ‘Statement’ contained the warning that editing, the cinema’s fundamental tool for the production of meaning, faced the danger of degenerating into merely being a device for facilitating the visible forms of sounds or speech. Calling for a depth of experimentation that sought an enrichment of cinematic meaning through the employment of asynchronous sound, the writers argued that an image tied to its natural sound became prosaic (or redundant), and was thereby prevented from having independent meaning. The creative montage of sounds, noises, and images was recommended; and the voice, plainly reattached to the image of a speaking body, was held to be an impediment to the development of an authentic cinema. Their ‘Statement on Sound’ proposed that sound montage should be developed along the lines of visual montage and that the two should maintain an asynchronous relation to one another. The rationale behind asynchronous sound is that sound and image should constitute two parallel and loosely connected tracks, neither dependent on the other. Eisenstein had earlier theorised that if film were to be its own art, it would need its own artistic raw
material. The “contrapuntal use of sound” would enable “the creation of a new orchestral counterpoint” between sound and image.\(^{17}\)

In contrast, Eisenstein’s contemporary Dziga Vertov wrote in 1925: “If, with respect to vision, our kinok-observers have recorded visible life phenomena with cameras, we must now talk about recording audible facts”.\(^{18}\) Vertov actively championed location sound, creating complex audio tracks for his first sound film *Enthusiasm*,\(^{19}\) while dismissing claims that filmmakers and radio producers should concentrate on recording sounds in soundproofed studios because the world was not ‘audiogenic’. Vertov also disagreed with the ‘Statement on Sound’ that Eisenstein had co-authored in 1928, particularly with the contrapuntal, asynchronous aesthetic proposed there, instead insisting that “neither synchronisation nor asynchronisation of the visible with the audible is at all obligatory”.\(^{20}\) Workers speaking with their own voices in their own audio environments were integral to his documentary concept of film. Vertov’s perspective was that film could overcome its indexical nature through montage, by presenting a viewer with objects and events that never existed in reality, and that sounds and images could be combined in a variety of different ways.

Vertov’s insight was that the camera eye would see on our behalf, and provide a machine assemblage of images. Sounds too, could be recorded and organised, and presented to the audience to listen to. In a sense, what Vertov can be said to have glimpsed is that film editing – which could just as easily be called ‘film construction’ – identifies and exploits underlying patterns of sound and image that are not obvious on the

---


\(^{18}\) Vertov, 1984, p. 56.

\(^{19}\) Fischer, ‘*Enthusiasm*: From Kino-Eye to Radio-Eye’, in Weis and Belton [Eds.], 1985.

\(^{20}\) Vertov, 1984, p. 111.
surface. Putting a film together is, in an ideal sense, the orchestrating of all those patterns; as such, voices need not be discounted. What Eisenstein and his cohorts fought for was for a creative practice and montage of sounds, and of images and sounds. Despite their limitations, these perspectives were early insights into the potential interplay between two separate, now converging, perceptual dimensions, in terms of an emerging ability to create a new reality which was not merely the sum of its parts.

Nevertheless, the British documentary maker Paul Rotha also insisted that the addition of sound and dialogue to the visual image on the screen would emphasise its isolated significance, as sound and dialogue take longer to apprehend than the visual image. Dialogue, he pointed out, represents real time and not the filmic time of the visual image, as a result of its very realism; and dialogue had imposed such restrictions on the director that all forms of cutting and cross cutting were now impossible.\(^{21}\) Despite this view of dialogue, Rotha would effectively foresee the possibilities made available by the montage and manipulation of sounds and noises, placed in juxtaposition with images:

> In the same way as an effect is built out of the pieces of film by the act of montage, so will little portions of sound be built up into new and strange noises. The process of short-cutting in visual images will be paralleled in the mixing of sounds. Even as visual images mix and dissolve one into another so will sound images mix and dissolve, according to the nature of the scene and as indicated by the scenario montage… Contrast of sound will be used in the form of the relationship of sound volumes.\(^{22}\) Sounds could be cut up, manipulated, and recombined – just like images, and \textit{together} with images.

\(^{22}\) Ibid, p. 310.
As Christian Metz noted, “Noise and music were accepted but not speech, which, of all the sounds in the world, in whose midst it nevertheless exists, was kept – in theory – under a mysterious and specific taboo”. However, Rotha (and many of those commentators who critiqued dialogue and yet supported asynchronous sound and the creative deployment of sounds and noises) did not realise that this recombination and manipulation of sounds could also apply to the voice – which was still viewed as the guarantor of an unproblematic presence and immediacy. The Hungarian film theorist Bela Bálázs, who initially complained that the sound film had undermined the expressiveness of film acting, later changed his position, examining the dramatic possibilities of silence and the “intimacy of sound” which makes us perceive sounds which are usually drowned out by the accustomed din of everyday life. In this way, he was another commentator who began to acknowledge the potential in new sound-image assemblages. He also pointed out that the anti-sound critics never objected to sound per se in the cinema, but only to dialogue as the real enemy. These opponents to dialogue and the voice did not see the potential in new sound-image assemblages, where voices could be re-sounded and even re-embodied in new images. Voices could be matched to images to which they did not originally belong. Voices could be manipulated, edited, and recombined. Sounds, noises, and voices could be selected, transformed, recombined, juxtaposed, together with the stream of images – creating new sound-image events that then operate as cinematic representations (or ‘doubles’) that present (imitate) an ‘original’ that never occurred. These monstrously inhuman simulations would model a ‘real’ that never occurred, but which still possessed the illusion of living flux, of lifelike

23 Metz, 1974, pp. 51-52.
24 Theory of the Film: Character and Growth of a New Art, 1972, pp. 210, 221
presence and persuasive flow. ‘Unnatural’ perspectives of sounds and images have now become naturalised, as new cinematic conventions have become transparent and familiar to contemporary eyes and ears.

In essence, numerous early commentators on the coming of sound to the cinema had the tendency to privilege film as a visual art and reinforce a particular hierarchy of image and soundtrack. However, Rick Altman points to a related problem that can be discerned in the work of the defenders of sound. In “Composing for the Films, Theodor Adorno and Hans Eisler attribute to hearing a privileged relation to pre-individualistic collective times; music thus has a pre-capitalistic nature, being more direct and more closely connected to the unconscious”.25 His critique can be extended. Adorno and Eisler make the comment that the sense of hearing has lagged behind in an age dominated by the sense of sight as the principle path to knowing. They observe that the ear is generally regarded as a passive organ when contrasted to the “swift, actively selective eye”.26 For them, there is more to the difference between the two senses: while vision is distancing, hearing possesses an enveloping and communal aspect. This sensory gap allows sound and hearing a special role: the spatiality and dimensionality of sound enables a corporeality and immediacy that cannot be achieved in the two-dimensionality of the screen image.27 In their Composing for the Films, Adorno and Eisler advise their readers that “listening is subject to historical change”28 – advice that they themselves do not appear to take fully into consideration. Altman points out that this approach to hearing

---

26 Theodor Adorno and Hans Eisler, Composing for the Films, 1994, p. 20; this text was originally published under Eisler’s name in 1947.
and sound is a problematic, idealised and transhistorical ontological claim which does not acknowledge the contextual, historical, social, and cultural conditions of the role and function of hearing and sound. As I have contended, the comparative critical neglect of sound, together with the continued emphasis on visuality, makes for a theoretical blockage that cannot be surmounted simply by resorting to heightened attention to sound. With the machining of sound and hearing, and the machined fusion of sound and image, both seeing and listening in our contemporary age must be acknowledged to have been subject to historical change and transformation.

A second surge of (more positive) of interest in cinema sound can be traced to the innovation of Dolby sound and surround sound in the late 1970s, heralded by *Star Wars* in 1977. This was hailed by film historians and technicians as marking a “second sound revolution”.29 Film theorist Robert Stam makes mention of “the ‘lateness’ of the study of sound in the cinema”; indeed, he claims that “the theoretical ramifications and aesthetic possibilities of sound have only begun to be explored”.30 Even with a recent increase in critical attention, recognition of the significance of sound and hearing has been slow in coming. The discussion of the acoustic dimension has largely been quiet in comparison with that of visuality. This second wave of critical interest in sound can clearly be seen to be, in many ways, an attempt to acknowledge and recuperate the importance of the filmic soundtrack, by critiquing many of the prejudicial presumptions of the early critics of the sound revolution.31 These later theorists generally argued that sound should not be seen merely as a secondary addition to the images on screen. With exciting developments in film sound that resulted in a significant extension in terms of acoustic depth, clarity,

---

31 See, for instance, Altman, Chion, Silverman, Gorbman, Doane, Lastra, etc.
definition, and detail, this later generation of theorists granted the soundtrack a degree of heightened prominence, and gave film sound a greater theoretical emphasis and appreciation.

But even after the advent of the recorded soundtrack and synchronous sound, the overwhelming majority of film criticism and theory still remains resolutely image-bound. This should not be surprising. Commentators and writers interested in visual media have for some time supplied various sophisticated conceptualisations of visual culture. However, a parallel construct – acoustic culture, or sound studies – has been comparatively slow to develop in terms of providing sustained perspectives and analyses on orality, aurality, listening, and hearing.32 There are numerous accounts of the image and vision.33 There are far fewer theoretical accounts of recorded sound and listening. While sound is considered a unified intellectual problem in some science and engineering fields, it is far less developed as an integrated and coherent subject in the social and cultural disciplines.

Modernity is admittedly distinctly ocularcentric, even hypervisual; it has been said to be marked, as the theologian and philosopher Jacques Ellul has complained, by “the unconditional victory of the visual and images”.34 And it seems that “the history of diminished hearing” has been the unavoidable counterpart to “the history of increasing ocularcentrism”.35 The benefit of the emergence of sound-reproduction technology in the

---

32 The back cover of The Auditory Culture Reader (edited by Michael Bull and Les Back, 2003) announces: “Providing a definitive overview of an emerging field, this pioneering reader is the first to redress a glaring imbalance, by investigating how auditory culture subtly and profoundly impacts on our everyday lives”.


34 The Humiliation of the Word, 1985, p. 2.

Nineteenth and Twentieth Centuries has been to provoke scholars to begin to acknowledge and contemplate the historicity of hearing. Kahn explains:

> Sound inhabits its own time and dissipates quickly. Its life is too brief and ephemeral to attract much attention, let alone occupy the tangible duration favoured by methods of research. Only recently in historical terms have there existed the conceptual and technological techniques available to sustain a full range of sounds outside the unstable environs of their own time.\(^{36}\)

It needs to be acknowledged that “both sound and listening [in the Twentieth Century] have been and continue to be transformed through the cultural elaboration of technology”.\(^{37}\) As a result of the work done by various theorists of sound and aurality, the acoustic dimension is no longer quite as neglected in relation to the sphere of visuality. Nevertheless, in theorising audiovisuality, the spheres of sight and hearing, visuality and orality/aurality, image and sound, still have the tendency to be placed in a hierarchical relationship, and/or situated in terms of a binary opposition. I want to turn now to look at the theoretical work of Walter Ong and Marshall McLuhan, which illustrates this divide, and which provides enduring knotty problems for theorising audiovisuality that need to be overcome. These troublesome issues in the work of Ong and McLuhan can be seen to be contained within (however unknowingly) and symptomatic of the conceptual projects of many of the modern theorists of cinematic sound.

The contribution of Ong and McLuhan has been to show how sound and hearing have increased in attention and importance with the advent and proliferation of acoustic technologies, and with the convergence of these acoustic technologies with visual

---

\(^{36}\) Kahn, 1999, p. 5.

technologies and the image. Ong and McLuhan attempt to attribute a recuperative role to sound, based on the ‘secondary orality’ of “present-day high-technology culture”, in which “a new orality is sustained by telephone, radio, television, and other electronic devices that depend for their existence and functioning on writing and print”\textsuperscript{38} – which can supposedly restore a sense of presence, immediacy, and transparency to social relations. The sound film is a significant marker of this secondary orality. However, I will argue that this so-called revival of oral/aural culture in our own electric age must be acknowledged to exist in a close relation with the regimes of visuality. We are now concerned with not just image, and not just sound; we must theorise the \textit{convergence} and transformation of the relationship between sound and image, between listening and seeing, that has occurred through the cultural and social elaboration of technology.

Walter Ong proposes that we live in an age of secondary orality, where sounds are recorded, ‘written’,\textsuperscript{39} and inscribed; manipulated and transformed, transmitted and played back. Our era of secondary orality is said to have resulted in the restoration of a sense of spontaneity, immediacy, community, and presence. The ideas introduced by Walter J. Ong with regards to orality and secondary orality can be clearly seen to have been influential in shaping approaches to acoustic culture and the nascent field of sound studies. The concept of secondary orality was introduced in Walter Ong’s well-known book \textit{Orality and Literacy} (1982), where he studied the differences between oral and literate cultures. He defined our age of secondary orality as a new mode of orality based

\textsuperscript{38} Ong, \textit{Orality and Literacy}, 1982, p. 11.
\textsuperscript{39} An apt visual representation of the ‘writing’ of sound is provided by EMI’s famous early trademark of the Recording Angel (1898-1909), where a winged cherub is sprawled atop a recording and is using a quill to inscribe (or write on) it. The phonograph performed a new kind of writing of sound, which appeared to reproduce the sound-event in its full, immediate, complete presence. The phonograph can be described as perhaps the landmark acoustic technology that has encouraged us (or provoked us) to think about sound again in a world organized by print and visuality.
on technologies of inscription. Secondary oral cultures are literate cultures, such as our own, that have been rendered significantly oral once again by the appearance of dominant new electronic communication media. Ong remarks: “This new orality has striking resemblances to the old in its participatory mystique, its fostering of a communal sense, its concentration on the present moment, and even its use of formulas”.40 Ong’s work can be explicitly linked with that of his mentor and influence Marshall McLuhan, the media theorist and ‘High Priest of Popcult and Metaphysician of Media’,41 whose “cardinal gnomic saying, ‘The medium is the message,’ [registers the] acute awareness of the importance of the shift from orality through literacy and print to electronic media”.42 As McLuhan explains, “the personal and social consequences of any medium – that is, of any extension of ourselves – result from the new scale that is introduced into our affairs by each extension of ourselves, or by any new technology”.43 Simply put, McLuhan’s point is that: “Each medium of expression profoundly modifies human sensibility in mainly unconscious and unpredictable ways”.44

McLuhan argues that (the technologies of) writing and print were primary causes for the separation of the senses and the privileging of seeing and visuality, particularly once “print stepped up the visual component in Western experience to extreme intensity”.45 The development of the printing press and perspective painting in the West at the time of the Renaissance are key events that have signalled the transition from the

41 As he was flamboyantly billed in his interview with Playboy magazine; McLuhan, ‘Playboy Interview: Marshall McLuhan – A Candid Conversation with the High Priest of Popcult and Metaphysician of Media’ (March 1969), in The Essential McLuhan, edited by McLuhan and Zingrone, 1995.
42 Ong, 1982, p. 29.
ear to the eye as the most important gatherer of information. The dominance of the written and printed word over the spoken word resulted in “the breaking apart of the magical world of the ear and the neutral world of the eye”. This ‘neutral world of the eye’ has henceforth laid claim to a predominant position in the hierarchy of our perceptual senses. The hegemony of the visual has coincided with the valorisation of values such as scientific observation, the abstract power of the gaze, the objectivity of sight, the clear demarcation between the viewer and the viewed, the clarity made possible by the dissecting force and isolative capabilities of the eye, etc. Both Ong and McLuhan attribute the opposition between ear and eye, hearing and seeing, to the transition (and subsequent opposition) between the spoken word and the written word, which laid the groundwork for the separation of the senses and the predominance of vision.

The spoken word has traditionally been linked with full presence and immediacy; whereas writing (which can be seen as the first sound recording technology) has been associated with mediation, imitation, mimesis – being, as Ong calls it, “a ‘secondary modelling system’, dependent on a prior primary system, spoken language”. The classic text that first framed the opposition between the spoken word and writing is Plato’s *Phaedrus* (360 BC). In the *Phaedrus*, Plato recounts an Egyptian fable that warns against writing (the practice of which causes people to neglect the exercise of memory and to

---

48 In his book *The Third Ear*, the German author and jazz critic Joachim-Ernst Berendt insisted that the dominance of ‘The Eye’ limits our imagination – instead, he makes the suggestion for ‘a democracy of the senses’ (1985, p. 32).
49 In his book, *The Audible Past: Cultural Origins of Sound Reproduction*, 2003, Jonathan Sterne describes what he terms ‘the audiovisual litany’, or common assertions about the difference between hearing and seeing, explaining that: “Outlining the differences between sight and hearing begs the prior question of what we mean when we talk about their nature”. (p. 15.) He concludes with the astute observation that “the audiovisual litany carries with it a good deal of ideological baggage”. (p. 16.)
depend on symbols), and makes the observation that books are like the painted figures that seem to be alive, but do not answer a word to the questions they are asked. It is at the end of the *Phaedrus* that Socrates (ironically acting as a character in this book written by Plato) embarks on his famous critique of the written word, complaining that writing is inhuman, is impersonal, diminishes memory, lacks interiority, parodies live presence, destroys authentic dialogue, disseminates promiscuously and indiscriminately, is incapable of acknowledging the individuality of its interlocutors, and disembodies speakers and hearers. Writing is a distortion or travesty, which masks the presence of an immediate, authentic meaning, in the ‘pure’ form of the spoken word.\(^5\) Derrida points out that in Plato’s *Phaedrus*, the central metaphor is that of speech as divine or natural truth, as opposed to writing as a fallen, human, laborious, finite and artificial inscription.\(^5\) The key here is that Ong and McLuhan must be understood to share a conception of speech and orality/aurality as being immediate, spontaneous, caught-up-in-the-moment, implying ‘truth’, presence, and life. In our age of secondary orality, then, their argument is that speech (disseminated through the electronic mass media) is again allowed to be a primary mode of communication, serving to re-invoke these values around the spoken word, offering a welcome corrective to the privileging of visuality and sight that has occurred largely as a result of the shift from the spoken word to the written (and subsequently printed) word (which, of course, marked the beginning of the

\(^5\) Writing can be seen as the first sound recording technology – as sounds in the form of spoken words are recorded onto the page in the form of the written word. For the ancients, the written word was initially nothing more than a substitute for the spoken word; books are read aloud, and the written words are henceforth ‘naturally’ returned to sound, to the spoken word. As Lisa Gitelman puts it, there was “the common but intricately held belief that written words were the graphic representation of speech”. (*Scripts, Grooves, and Writing Machines*, 1999, p. 26.) The advent of the (revolutionary!) practice of silent reading – reading without saying the words, omitting the return of sounds – would inaugurate a new status for books, and a new hierarchy between seeing and hearing, visuality and aurality.

\(^5\) *Of Grammatology*, 1976, p. 15.
hegemony of the visual). This perspective, of course, encourages a positive reading of the coming of sound to the cinema, and the emergence of characters on film that could speak and engage in dialogue.

As Ong and McLuhan have argued, where sight has created and supported various notions of detachment, objectivity, representation, mimesis, etc., sound is in many ways its opposite. Ong writes:

Sight isolates, sound incorporates. Whereas sight situates the observer outside what he views, at a distance, sound pours into the hearer. Vision dissects... Vision comes to a human being from one direction at a time: to look at a room or a landscape, I must move my eyes around from one part to another. When I hear, however, I gather sound simultaneously from every direction at once: I am at the centre of my auditory world, which envelops me, establishing me at a kind of core of sensation and existence.... You can immerse yourself in hearing, in sound. There is no way to immerse yourself similarly in sight.53

In this sense, sound operates to invoke a sense of reality for us in a way that differs from sight. In his book Ocean of Sound, David Toop provides a formulation that is strikingly similar to Ong’s:

Sound places us in the real universe. Looking ahead, I can see a plane enlivened by visually represented objects. I can touch within a limited radius. I can smell a body, a glass of beer, burning dust. But sound comes from everywhere, unbidden. My brain seeks it out, sorts it, makes me feel the immensity of the universe even when I have no wish to look or absorb.54

While sight is associated with being a sharp and delineated sense of perception, sound seemingly envelops the body, as if it were an atmosphere to be experienced rather than an object to be dissected.

In addition, Ong comments that: “There is no way to stop sound and have sound... All sensation takes place in time, but sound has a special relationship to time unlike that of the other fields that register in human sensation. Sound exists only when it is going out of existence. It is not simply perishable but evanescent, and it is sensed as evanescent”.55 Ong celebrates the second coming of orality by arguing that secondary orality (made possible by the new media) enabled a return to the utopian qualities of sound: sound is immersive, unified, spontaneous, present, immediate, harmonious, interiorising, centralising, incorporating, participatory, communal, actualising. After all, sound is multiple: sound is typically comprised of many sounds, jostling together within a time and space. Sound is ephemeral; it occurs within time and space, as sound begins, lasts for a time, fills a space, and then fades away. Sound – largely because of this distinctive relationship to time – possesses enduring associations with the notion of presence and immediacy. Sounds generally make meaning for us by being a characteristic of an object or person or environment making those particular sounds – sound is embodied, or contextual. Sound is dimensional, filling and occupying space and time, unbidden – as Cage famously remarked, there is no such thing as silence (and as R. Murray Schafer has pointed out, we have no earlids) – and thus sound connects us to the world; sound signals life; sound immerses us in a situation, within a context, in an environment, within a space and time, in a world filled with and occupied by movement, life, things and people,
nature and culture, all producing sound. Sound therefore tends to signify the presence of an object, person, or environment. Recorded sound, unlike the recorded image, appears to retain its dimensionality, still fills and occupies space and time (unlike the image, which is two-dimensional, appearing flat and lifeless without sound), and evokes a disembodied (or absent) presence – as replayed sound acts by presenting this prosthetic presence of an object, person, or environment within a new context, while recurring within a new time and space, enabling a particular perspective of a past event, time and space.

McLuhan made the important observation that audiovisual and acoustic technologies have returned the field of hearing (subsequently elaborated via Ong’s account of secondary orality) and recuperated the separation of the senses, reinstating the dimensionality and immediacy of sound. Ong’s account of secondary orality argues that the electric reproduction and transmission of sound works to heal the isolation, detachment, and fragmentation of visuality and enables the “retribalization” of humanity; restoring a sense of community in “a seamless web of kinship and interdependence”.56 (These perspectives would be adopted enthusiastically by sound theorists, who attempt to challenge the perceptual focus of the many media theorists who remained completely enthralled by the image.) McLuhan believed that with the image coming to displace writing and print in a position of prominence, sound (secondary orality) has the opportunity to return, reinstalling such values as dynamism, immediacy, presence, spontaneity, community, etc. “The new media – the new languages – which have increasingly supplemented writing and print, have begun to reassemble the multiple sensuousness of integral speech. Touch, taste, kinaesthesia, sight and sound are all

56 These phrases are borrowed from McLuhan.
recreating that acoustic space which had been abolished by phonetic writing”.\textsuperscript{57} For McLuhan, this convergence (and return of the ear) results in a utopian vision of retribalized man, and the founding of a global village.\textsuperscript{58}

McLuhan strives to challenge the visual (writing and print) bias by invoking the ideal of presence, in the form of the celebration of the ear, the dimensionality and simultaneity of sound, and the values of oral cultures. McLuhan observes that: “A basic feature of acoustic space is its inclusiveness. Visual space is inclusive. As our world recreates acoustic and oral culture simply by pushing on with devices of instantaneity and simultaneity, we need not fear the suppression of visual and written culture”.\textsuperscript{59} Instead, in McLuhan’s terms, visual and written/print culture will benefit from the return of sound. McLuhan has commented on the continued influence the model of literacy and writing has had on our perception. “Literacy gives people the power to focus a little way in front of an image so that we take in the whole image or picture at a glance”.\textsuperscript{60} McLuhan adds: “many of the conventions of literacy are built into even non-verbal forms like film”.\textsuperscript{61} However, he does not sufficiently emphasise how this culture of visuality and literacy has also shaped sound and hearing – the ‘return of the ear’ is founded on inscription, visuality, and literacy. I want to argue that secondary orality, built on writing and the culture of visuality that developed around writing, has in many ways fused with the visual, as sound increasingly becomes ‘written’ (or resonant) with the narrative of the

\textsuperscript{58} For other, more pessimistic commentators, this retribalization and establishment of a global village has seen the administering of reification, commodification, and an instrumentalized interplay between (pseudo) difference and repetition. Baudrillard, for instance, argues that the impact of the electronic media is geared towards global homogenization and control, influencing a systemic detribalization instead of the retribalization espoused by McLuhan.
\textsuperscript{60} McLuhan, \textit{The Gutenberg Galaxy}, in McLuhan and Zingrone [Eds.], 1995, p. 132.
\textsuperscript{61} Ibid, p. 132.
image. Instead of the nagging persistence of this binary opposition between sound and image, sight and hearing, we need to examine sound and image together. As the image has come to gradually displace writing, so sound has moved past simple ‘writing’ (storage and dissemination) and established an evolving relationship with the image (particularly with regards to the possibilities around the re-embodiment of sound, and the re-sounding of images). Even as the technological parameters around sound have subtly shifted; even as our ideas around sound have changed (but paradoxically stayed the same) – sound has come to make meaning in new ways (even as it is founded on established models of meaning). Ong and McLuhan have claimed that the new communications media have apparently returned us to a fluid, shifting, evanescent, and open-ended communication. Is this truly the case, or can this utopian vision (a phrase rendered ironic by their oral/aural emphasis) espoused by Ong and McLuhan be profitably critiqued? After all, in this age of secondary orality, sounds too are now recorded, manipulated, and replayed – in a way, ‘fixed’ in material form (like writing). Sound recording devices appear to reproduce sounds in their fullness, instead of the two-dimensional representation that images provide. However (like the written and printed word), mediated sounds are subject to processes of perspective, selectivity, manipulation, construction, artifice, all while offering and furnishing a persuasive depiction of reality. (Many of the persistent debates and analyses around the recording of sound can be traced to the idea that recording is comparable to the writing of sound.) Mediated sounds provide aural representations and/or reproductions that are selected, constructed, manipulated, and which offer particular perspectives on reality.
Ong’s work can be said to be the impetus for sound studies’ generally reactive stance in defending the priority and transcendence of sound and hearing by denigrating sight and visuality. This is a perspective that needs to be critiqued as being reductive and equally as one-sided as the assorted studies of visuality that neglect sound and hearing. As Sterne comments, sound has typically “been treated as a non-historical or transhistorical substance, immune to human action and human practice, with a fixed predetermined relevance for historiography – in short, a constant”;\textsuperscript{62} and this critique applies to Ong’s theoretical project. Because of these failings that are still found in many theoretical and conceptual approaches to sound and audiovisuality, it is essential to address sound and hearing as social phenomena and to trace the historical, social, and cultural grounds of oral/aural experience.

It is true that the proliferation of various acoustic technologies has also led to a resurgence of auditory cultures. It is evident that sounds (and not just speech) have returned to a position of perceptual importance – we are surrounded by machined sounds, which increasingly shape the way we perceive our acoustic environment. Sound and hearing and their associated technologies and practices have increased in attention and importance.\textsuperscript{63} However, Ong’s discussion of secondary orality can be seen as just another variation of the privileging of sound and hearing, and thus inherently limited. With the convergence of acoustic technologies and sounds with visual technologies and the image, my argument is that the increasing dissemination of machined sound and hearing has come to support the image, being comprehended and theorised through already established visual models of understanding.


\textsuperscript{63}
The work of the philosopher Jacques Derrida can be employed to problematize the value system espoused by McLuhan and Ong – a possibility which Ong himself has addressed. Derrida provides his well-known critique of what he terms the ‘metaphysics of presence’ to dismantle the connections among speech, sound, voice, and presence in Western thought. Derrida famously deconstructs and excoriates the metaphysics of presence, styling it ‘logocentrism’ and diagnosing it as deriving from ‘phonocentrism’, that is, from taking the logos or sounded words as primary, and thereby debasing writing in comparison with oral speech (which is what Ong and McLuhan can clearly be accused of). Writing problematizes this logocentric model, as Derrida demonstrates that writing has an economy of its own so that it cannot simply transmit unchanged what it receives from speech. And, as Ong himself points out, it is also evident that the model of logocentrism can be problematized even earlier, by examining spoken words, “which do not themselves transmit an extra-mental world of presence as through transparent glass”.64 When Derrida raises the notion of presence, he is referring to the fiction of direct reference, or the ideal of the transcendental signified, and concealed behind the foundational myth of presence lie various sedimented discourses, normalised associations, taken for granted rules, and possibilities for play – which, though structured into the very constitution of textuality, are almost never acknowledged. Within the Western metaphysical tradition, there has been an obsession with a sense of plenitude, with the theme of presence, with the privileging of origins, with an original state of purity and simplicity. Derrida’s critique of presence culminates in the point that originary presence is fundamentally multiple, contextual, and always mediated. The observation he

64 Ong, 1982, p. 163.
makes is that speech cannot be privileged over writing; after all, the spoken word is fully implicated within a system of signs that pre-exists the speaker, and the ‘natural’ and ‘unmediated’ immediacy of the spoken word is necessarily an illusion.

The speech/writing opposition is central to McLuhan’s understanding of the transition from what he has called the Gutenberg galaxy to the new secondary orality. Derrida has criticised McLuhan for his simplistic opposition between hearing and seeing, speech and writing, which privileges orality by viewing writing as a secondary mediation of the immediacy and presence of speech, thereby idealistically appealing to (what Derrida would call) a logocentric notion of presence – all of which propelled McLuhan’s utopian claim that the return of the acoustic dimension would take over and resolve the problems brought about by the predominant culture of writing, print, and visuality, recuperating and reinstalling the ideals of immediacy, spontaneity, and community (the perspective which influenced and was adopted by Walter Ong). Derrida argues for the originary logic of supplementarity, and remarks: “We are not witnessing an end of writing which, to follow McLuhan’s ideological representation, would restore a transparency or immediacy of social relations”. Instead, Derrida emphasises “a more and more powerful historical unfolding of a general writing of which the system of speech, consciousness, meaning, presence, truth, etc., would only be an effect, to be analysed as such”. Derrida rejects McLuhan’s vision of the psychological and social wholeness of pre- and post-literate cultures. McLuhan viewed writing as a supplement to speech; Derrida points out that writing is sandwiched between two oralities, the first originary and the second neo-originary, whose unity it interrupted. For McLuhan, of course,

writing separates and specialises and undoes the “tribal web”\textsuperscript{66} by granting the individual emotional freedom; it is also civilising, intensifying, visual and uniform.

In summary, McLuhan posits that writing is exterior to the speech whose place it takes and keeps, and this belief situates McLuhan firmly in the Western metaphysical tradition that Derrida has consistently taken aim at. McLuhan suggests that the electronic media manage to restore a space of plenitude and presence; however, for Derrida, McLuhan’s transition from writing to speech is logocentric in its attempt to invoke transparency, immediacy, nowness, and presence within the global village. As Derrida famously argues in \textit{Of Grammatology},\textsuperscript{67} writing should be thought of as an originary supplement that takes place \textit{before} and \textit{within} speech. Secondary orality is an \textit{inscribed} orality, and thus McLuhan’s unbridled celebration of an audile-tactile retribalization, and his conceptualisation of writing as a supplement to speech, needs to be carefully examined. As Gary Genosko has insightfully commented, McLuhan’s media theology “had a specific faith underlying it: salvation from the fall of literacy might be found in electric technology”.\textsuperscript{68} (Ong also adheres to a version of this ‘faith’, which I will address shortly.)

The situationist Guy Debord has also rather unkindly caricatured McLuhan as “the spectacle’s first apologist, who had seemed to be the most convinced imbecile of the century”.\textsuperscript{69} Continuing his acerbic comments, Debord has characterised McLuhan as a Global Village idiot, remarking that even McLuhan had come to eventually realise that the mass media could not deliver on promises of freedom and accessibility. McLuhan’s

\footnotesize{\textsuperscript{66} McLuhan, \textit{Understanding Media}, 1964, pp. 82-84. \\
\textsuperscript{67} \textit{Of Grammatology}, 1976, pp. 313, 315. \\
\textsuperscript{68} Genosko, \textit{McLuhan and Baudrillard}, 1999, p. 13. \\
\textsuperscript{69} \textit{Comments on the Society of the Spectacle}, 1990, p. 33.}
focus on technology – the medium is the message – and his relative neglect of the social, led to a one-dimensional focus on the technological production and transformation of speech (secondary orality). So this account of secondary orality is flawed, and McLuhan’s hyperbolic claims should be mitigated – sound and hearing are still comprehended through visuality, and media technologies do not simply determine social transformations (the medium is the message!); rather, there is a dialectical relationship between sociality and media technology. I want to suggest that McLuhan’s account of secondary orality can be made productive through a supplementary account of secondary aurality.

I use the term secondary aurality in order to emphasise that attention also needs to be paid to the reception of technologized sounds and transformation of ways of hearing. Over time, certain practical understandings have become sedimented around the processes and practices of sound reproduction and the resultant relationships of original and copy. People have come to form particular relationships with sounds made by machines, and the various discursive narratives have become firmly established and taken for granted. I want to focus on the logic of supplementarity (to utilise a Derridean formulation) that informs sound, and the relationship between sound and image; I will then turn to examine how this founding absence becomes solidified into conventionality and codedness; I will also look at the sociality of sound, and the relationship between sound and image, and how this has given rise to a new language of sound and image.

In simple terms, orality can be defined as the voice producing sound; and aurality can be described as the ear hearing sound. I want to include the important notion of secondary aurality (as well as orality) because the development and increasing
naturalisation of acoustic technologies and practices have affected not only the production of sound (orality) but also the reception of sound (aurality). Secondary orality is a useful idea, but the machined production of sound (secondary orality) has coincided with the machined reception of sound (secondary aurality), where new modes and ways of listening have developed. The point of secondary orality/aurality is to acknowledge that listening and hearing have also been transformed through technological means. Secondary orality/aurality – based on the machining and technologization of sound and hearing – is founded on prior, established cultural elaborations of media and communications technologies, which bear witness to histories of writing, inscription, literacy, and visuality. These two overlapping categories, with their nearly inaudible difference (orality/aurality), mark the technologically-mediated evolution of the production, manipulation, and reception of sounds, not just confined to the privileged sphere of speech.

We have looked at secondary orality/aurality – and the implications therefore – which can (almost flippantly) be called the narrative of sound once-removed (and a key point here is, removed by technology? Or transformed by new technologies of sound and practices of hearing?). I want to focus now on how this secondary orality/aurality is intertwined with visuality and the image: as audiovisual technologies become more prevalent, the auditory has been revitalised, but in conjunction and convergence with the visual image. Contrary to what McLuhan argues, it is becoming evident that it is the reinforcement of the visual which is the most remarkable cultural effect of the mass media. And contained in McLuhan’s work is the admission: “The revival of oral culture in our own electric age now exists in a… fecund relation with the still powerful written
and visual culture”. Even so, Walter Ong maintains the faith with McLuhan’s utopian perspective on oral/auditory culture by claiming that: “Writing can never dispense with orality”, because “in all the wonderful worlds that writing opens, the spoken word still resides and lives. Written texts all have to be related somehow, directly or indirectly, to the world of sound, the natural habitat of language, to yield their meanings”. Ong nostalgically advocates a utopian return to past forms of social organisation founded on the idea of face-to-face conversation embodied in mutual spoken dialogue.

The Jesuit priest and scholar Ong hopes for some of the mystery of orality to return so that people can once again hear the Word of God; and this ideal is placed in opposition against large-scale societies founded on writing, print, literacy, and visuality. The opposition between visuality and orality/aurality contains what can be classed as deeply conservative ruminations about the shape of human societies. When discussing the purportedly unique temporal and spatial characteristics of aural phenomenology, Ong argues that sound is more real or existential than other sense objects, despite (or because of) the fact that it is also more evanescent. Sound itself, Ong continues, is related to the actual present rather than to past or future – sounds exist only as they go out of existence. But, as Sterne points out: “To say that ephemerality is a special quality of sound, rather than a quality endemic to any form of perceptible motion or event in time, is to engage in a very selective form of nominalism”. Strictly speaking, Ong’s claim is true for any event, any process that we can possibly experience – and should not be seen as a quality that is special or unique to sound. In other words, the persistent binary notion that sound

---

71 Ong, 1982, p. 8.
draws us into the world while vision separates us from the world rests upon an essentialist metaphysical assumption that nevertheless (no matter how powerfully and convincingly it is deconstructed) has remained extremely influential in demarcating the way we view and engage with the world around us. Sterne summarises: “The attempt to describe sound or the act of hearing in itself – as if the sonic dimension of human life inhabited a space prior to or outside history – strives for a false transcendence”.73 This, in essence, is the critique that Derrida’s work enables us to level at Ong and McLuhan’s privileging of orality and auditory culture.

The ideal acoustic community championed by the soundscape theorist R. Murray Schafer can be profitably compared to Ong’s notion of secondary orality and McLuhan’s call for “the return of the ear”. Schafer grounds his ideal sound culture on what he calls a human scale – in other words, bounded by the natural spatiality of the unamplified human voice. This conservationist definition of our acoustic culture and environments – traceable to Plato’s privileging of face-to-face dialogue – equates cacophony with social disorder and celebrates authentic, ‘pure’, natural sounds over machined, mediated sounds which are framed as unnatural, disturbing, and noisy. In his seminal text *The Soundscape* (1977), R. Murray Schafer introduces his concept of “schizophonia”, which “refers to the split between an original sound and its electro-acoustical transmission or reproduction”.74 Schafer uses this idea to highlight the traumatic impact of “the discovery of packaging and storing techniques for sound and the splitting of sounds from their original contexts”.75 In Schafer’s conception, sound has been separated from the maker of the sound: “Sounds have been torn from their natural sockets and given an amplified and

75 Ibid, p. 88.
independent existence”\(^7^6\). Schafer’s account of this notion of disembodied sound is used to reinforce his arguments for ‘pure’ sound, as he laments “the fact that precisely at the time hi-fi was being engineered, the world soundscape was slipping into an all-time lo-fi condition”\(^7^7\). Schafer’s critique clearly contains the argument that secondary orality is based on the ‘writing’ (recording and reproduction) of sound and speech, as he implies that recording is (as Plato remarked about writing) an artificial and misguided process that rips sounds out of bodies.

Schafer’s idealistic project of protecting “natural soundscapes” from the challenge of these brash disembodied sounds leads him to observe that: “Originally all sounds were originals. They occurred at one time in one place only. Sounds were then indissolubly tied to the mechanisms that produced them. The human voice travelled only as far as one could shout. Every sound was uncounterfeitable, unique”\(^7^8\). This idea of schizophonia is useful, if flawed: Schafer’s notions of “pure” sounds that “have been torn from their natural sockets” and his overall perspective needs to be expanded and altered, employed as a launching pad, in order to examine issues around re-embodied sound, inscribed sound, sound-as-trace, presentness, as well as to criticise. After all, sounds are always embodied, as we understand sounds in the world as being packaged with, or embodied with, meanings. Schizophonia – sounds ‘detached’ from their ‘natural’ sources but still carrying the supplementary ‘traces’ of these sources – is only the initial step in the process (where the machined separation of sound and image has led to the machine fusion of sound and image), for sounds are then re-embodied, and attached to ‘prosthetic’ sources; and this re-embodiment is becoming increasingly naturalised. As a result of the

---

\(^7^6\) Ibid, p. 90.
\(^7^7\) Ibid, p. 88.
\(^7^8\) Ibid, p. 90.
machined fusion of sound and image, images have increasingly come to be seen as the ‘natural’ sources of sounds.

The composer John Cage – who should be read as a major philosopher of sound, noise, and silence – has worked towards “discovering means to let sound[s] be themselves, rather than vehicles for man-made theories or expressions of human sentiments”\(^7\). However, as Maurice Merleau-Ponty was fond of remarking, we are ‘condemned to meaning’ – applied here, this means that as we engage with and make sense of the meaningful sounds contained within our world, we can problematize the reification of meaning into a rigid code, but this problematization must admit that sound is never ‘pure’. We can deconstruct sound, but at the same time allow that sound will revert and solidify into codes of meaning – for we are not largely concerned with the ‘pure’ materiality of sound. The codedness of sound does not necessarily remain static; the articulation and flow of this code are prone to moments of stability, and instances of rigidity – but the code contains an originary fluidity, in terms of what Derrida would call \textit{différance}, and we can map the connections, blockages, assemblages, articulations, and re-routing of the flow of this code.

\(^7\) Cage, 1994, cited in Cubitt, 1998, p. 96; “Sound is vibration”, Cage tells a reporter from \textit{Newsweek} in 1967, “everything is vibrating. So there is no earthly reason why we can’t hear everything”. (Cited in Revill, \textit{The Roaring Silence}, 1992, p. 222.) This celebration of the vibration of the world – an essentialist, almost spiritual notion that all is comprised of sound – affords a privileged ontological status to sound. However, Cage’s concept of sound also fails to take into account the \textit{sociality} of sound, within which meaningfulness is elaborated within the discourses and codes of daily life or artistic practice. Sound requires a situation, and a context, as well as perceivers, in order for sound to be \textit{meaningful for us}. The notion of sound-in-itself makes reference to an ideality, thereby leading to the fallacy of ‘pure’ sound. There is no such thing as ‘pure sound’ for us. Sound inevitably becomes semanticized within a system of interpretation; we can problematize the reification of meaning into a rigid code by re-emphasizing the sheer materiality of sound (where sound is felt and experienced as a physical/material phenomenon, not so much as a signifying phenomenon), but this problematization must admit that sound is never ‘pure’.
These conceptual realisations enable a further clarification to be made. The harsh rhetoric targeted at modern sound-reproduction technologies is based on their disembodying capacity to separate sound from its ‘source’ (and Schafer’s schizophrenic definition of sound reproduction remains a sterling example); this rhetoric maintains a questionable set of prior assumptions about the fundamental nature of sound, communication, and experience. Human experience and the human body are assumed to be categories outside history, as face-to-face communication and bodily presence are valorised as the yardsticks by which to measure communicative activity. Sound reproduction can therefore be defined negatively, in terms of modifying or doing violence to a ‘natural’ co-presence of sound and soundful body, or ‘original’ sound. The claim that sound reproduction has disembodied and alienated the voice from the body implies that the voice and the body once existed in some prior holistic, unalienated, and self-present relation. There is the assumption that sound-reproduction technologies are ontologically separate from a ‘source’ that exists prior to and outside its affiliation with the technology. The resulting distinctions between ‘sources’ and ‘copies’ diverts our attention from the historical, social, and cultural nature of the technological process and practice.

If interpersonal interaction is taken to be the primary or ‘authentic’ mode of communication, this then explains why sound reproduction has been denigrated as inauthentic, disorienting, and possibly even dangerous by virtue of its capacity to ‘decontextualize’ (or disembody) sound from its ‘proper’ interpersonal (or embodied) context. However, my point is that we should be talking about the recontextualizing capacity of sound reproduction, instead of its decontextualizing function. The ‘thing itself’ as we imagine it from the recording was never there at the moment of the
recording; the performance itself was selected, framed, crafted, and manipulated in order to be reproduced. In other words, there is no ‘originary’ presence with regards to sound and speech, and therefore recording technology does not deface or perform violence on this ‘originary’ presence, but rather invokes it, transforms it, colonises it, refers to it, uses it as a template, decontextualizes and recontextualizes it. And subsequently, this new ‘mediated’ presence becomes naturalised in turn.

The new model of sound and listening that now defines our contemporary audiovisual age has installed sedimented habits and practices of perception, and supplied us with pre-selected, reinforced sounds. A recording-as-‘text’ should not be viewed as just a static container of ‘knowledge’ about the world but rather, as creator, preserver, energiser, or mediator. A ‘text’ simultaneously dissolves, recovers, rejects, transposes, and projects various discourses of associations and meanings. To borrow the words of Edward Said: “The point is that texts have ways of existing that even in their most rarefied form are always enmeshed in circumstance, time, place, and society – in short, they are in the world, and hence worldly”. Derrida’s textualist and deconstructionist approach has been beneficial in theorising “the interplay between textuality and twentieth-century acoustical technologies”, which has remained relatively “undocumented and under-theorised”. (For example, Derrida’s critique of presence and his ideas about difference and iterability have facilitated the observation that each replaying of mediated sound produces a new audition of that sound, even as mediated sounds appear to be reproduced in all their original fullness and presence. In this way, it can be said that these reproduced sounds are not quite the same sound for us as these

---

sounds subtly shift in potential meaning and acoustic character through each shift in
performative context, situation, and environment.) The strategy of employing Derrida’s
productive point regarding the originary supplementarity of writing to invert and
decompose Ong’s formulations has been embraced (or at least noted) by numerous sound
thorists, to further various conceptual agendas.

Derrida’s work on writing and speech, and the primacy of writing, is very useful
for the study of sound and how it has come to be dominated by recording, editing, and
playback – possibilities that enable the inscription of sound, and which act as a mode of
writing. The inscription of sound has introduced a new quality to a formerly ephemeral
sense of perception. I have consistently noted the importance of what can be called
supplementarity or sedimentation – where media technologies open up new networks,
avenues, and discourses of meaning that have to be navigated. Derrida’s oeuvre can be
said to be built around the critique of presence, which he associates with speech, the
spoken word, and which I have linked more generally to sound – where sound operates as
immediacy, as pure presence, as an ideal of authentic embodiment. (I would like to
suggest that this can also be traced from his influences, in terms of the Heideggerian idea
of being-in-the-world.) One wonders, however, why Derrida does not turn specifically to
examine the more recent developments of media technology, applying his ideas of trace,
supplementarity, and *différance*. In my discussion and theorization of sound and soundful
images, then, I am explicitly doing what Derrida does not, in terms of applying and
employing a range of productive conceptual tools.

Sound, recorded, manipulated, and played back, fused with the image, operates as
*trace*, as supplement, invoking an absent presence, operating by supplying a
supplementary presentness. Sound is slowly but surely being written into the tapestry of
the image. Sound does not exist in its ‘pure’ form even as we stand in the direct,
immediate presence of sound being made by its ‘natural’ source, in its ‘natural’
environment. Our human mode of hearing, our psychoacoustic interpretation of sound,
means that sound is always mediated, always organised and perceived according to our
human perceptual selectivity and limitation. Derrida writes, “There is not a single
signified that escapes, even if recaptured, the play of signifying references that constitute
language”.82 This is to say that, with every decontextualization, there must be a
subsequent recontextualization; with every deterritorialization, there must be an
accompanying reterritorialization. Machined sound produces a different, deferred event,
which refers to an originary presence, that is fundamentally multiple, contextual, and
always mediated. This is the story of co-presence, when sounds are accompanied by the
presence of their sources, when sounds are sourced by their images. But even when
sounds are not directly sourced by their images, an implied link has been established
between sounds and images. We can say that there is a network of multiple echoes
between sound and image, and a soundful image is created as a result, which is a
meaningful sound-image event in itself.

The soundful image is a sign that is composed and makes meaning due to the
articulation of sound and image – its referentiality is founded on the multiple echoes of
sound and image, which are contained within and bound by the specific soundful image.
As alluded to by Ong, Derrida’s project can be characterised as being focused on a sort of
limitless intertextuality, whereas I am interested in considering the space where the

signifier (and the ‘text’) pause to rest – a place rarely ‘innocent’, dimensionless, or without the affirmative authority of discursive discipline – in other words, I am concerned more with investigating the practical force by which the signifier and the ‘text’ occupy a place, a situation, a context. To qualify Derrida’s contribution of a critique against phonocentrism and logocentrism and the privileging of writing – I argue that sound has power, sound possesses presence – not because sound contains these transcendent essential qualities, but because we have idealised sound, we engage with sound, we have granted and allocated sound these sedimented social, cultural, and perceptual meanings within our being-in-the-world. The stories and histories sedimented around sound-for-us remain fundamental in determining how these sounds signify, how these sounds make meaning – even as these sedimented stories, histories, conventions, practices, and aesthetics become semi-transparent and taken for granted.

Maurice Merleau-Ponty writes that: “The world of perception, or in other words the world which is revealed to us by our senses and in everyday life, seems at first sight to be the one we know best of all”. Of course, “this is a delusion”. Merleau-Ponty suggests that “one of the great achievements of modern art and philosophy… has been to allow us to rediscover the world in which we live, yet which we are always prone to forget”.83 Merleau-Ponty observes that the ‘natural attitude’ of common sense leads us to overlook the phenomenon of the perceived world. As Merleau-Ponty explains, the senses in perception cover their tracks as they organise experience in such a way as to present a world of things arrayed before us in a three-dimensional objective space – within which we are located as just another object. In other words, “as we get on with our life we do

---

not notice the role of the senses in organising experience and ‘constituting’ the physical world; it is precisely their business to make this role invisible to us”.  

This means that mediation is transparent; our perspective on the world is concealed, as objects appear to be entities entirely independent from our particular perspective of the perceived world. We have come to consume signs (messages, images, mediated sounds) rather than material commodities. The rise of the empire of signs has installed a system of habits, behaviour, and perception.

Merleau-Ponty notes: “It is commonplace to say that we have five senses, and it would seem, at first glance, that each one of them is like a world out of touch with the others”. Nevertheless, he goes on to say: “My perception is… not a sum of visual, tactile, and audible givens: I perceive in a total way with my whole being; I grasp a unique structure of the thing, a unique way of being, which speaks to all my senses at once”. What you see is, in part, determined by what you hear; and vice versa, how you hear something is partly determined by what you see. (This reciprocal relationship can be explained by tracing the linkage of sound to source.) So, the mutual influence of sound and image is inextricable. Merleau-Ponty has written: “We have been considering sight and sound by turns, but in reality they way they are put together makes another new whole, which cannot be reduced to its component parts”. I have argued that the infusion of mediated sounds has created the most effective/affective image. The meaning of sound is deeply influenced by visuality and image. What we hear is shaped by what we see – sounds are embodied, and made by soundful sources. When we hear sounds, we tend to

---

86 I will discuss the sound-to-source model of meaning in further detail in the next chapter.
look for the source, in order to help us to interpret the sound, to make the sound semantically meaningful.

It can be said that Derrida has tended to focus predominantly on examining textual constructs, to the neglect of perceptual experience. “I don’t know what perception is and I don’t believe that anything like perception exists”, he has remarked in an interview. However, perception is inevitably intertwined with language. Soundful images are subject to discursive pressures – how are soundful images meaningful and consumed within particular discursive contexts? Soundful images are caught up within a concatenation of production, reception, appropriation, and reproduction that are located within particular performative circumstances and situations. Image is associated with semblance and appearance; sound is associated with presence and immediacy. These two dimensions have begun to engage in a close relationship. Images are easily absorbed by the audience, instead of requiring the obvious acts of reading, interpretation and mediation, like written text. Sounds, too, are typically processed without awareness of their material acoustic dimensions and the social conventions that are present. The contemporary model of sound has become more spectacular, and concerned with a representation of sound that is no longer modelled simply on the original/copy distinction. Sounds have never looked so good, and images are sounding ever more present, immediate, and impactful.

The articulations between sound and image are stabilising. We now live in an age when every mediated sound comes practically (or not so practically) equipped with its own stereotypical image. Sounds, too, like images, have become commodified and reified.

---

– sounds have also become fused to images. As a result of the commodification of sound (which is not synonymous with the codification of sound), fetishized sound is valued in its commodified form, which then becomes formulaic, instrumentalized, and reified. Meaning is layered, disseminated, and solidifies into widely accepted discourse, as enabled by mass media. Social meaning becomes increasingly reified, sedimented, and codified – effacing the material aspect of being, and falling into place as components of the productive and receptive systems of social meaning.

Lastra rightly points out that vision has not been “the only sense to be transformed nor the only one to act as an agent of transformation”; instead, in a “quiet but no less powerful story of aural modernity and modernisation”, hearing has been “just as surely dislocated, ‘mobilised’, restructured, and mechanised. The ‘annihilations’ of space and time affected hearing as much as seeing, and acoustic experience was as thoroughly commodified as its optical counterpart”.89 Perhaps Ong and McLuhan have overlooked the real significance of this point, in their idealistic espousal of the return of the ear, the mediated renewal of the spoken word, the reclamation of the immersive power of sound. The notions of secondary orality/aurality and the machining of sound and hearing do not refer only to the production and reception of mediated sounds, but also address how developing technologies and listening practices are affecting the way we hear and listen to our changing acoustic environments. In the modern age, sound and hearing have been transformed. “Simply put”, Douglas Kahn has written, “there were more sounds, and people could hear them more quickly”.90 In other words, modernity has “entailed more sounds and produced a greater emphasis on listening to things, to different things, and to

89 Lastra, Sound Technology and the American Cinema, 2000, pp. 3-4.
90 Kahn, 1999, p. 12.
more of them and on listening differently”. Rather than people waking up one day and discovering everything suddenly different, the transformation of soundscapes and listening and hearing practices has occurred gradually over a period of time. Attali has described “the stockpiling of music”; however, the scope of this arresting phrase can be broadened to speak of the stockpiling of voices, sounds, and noises. Not only have there been more sounds, but mediated sounds have been specifically selected, constructed, and disseminated for our attention and consumption. Mediated sound and machined hearing/listening have been successfully installed as new and now familiar soundscapes and events that have become naturalised components of our world.

Shaped by the advent of sound reproduction and dissemination technologies and practices, people have come to share acoustic experiences that are both collective and yet individualised. Transformations in sound, hearing, and listening have been part of massive shifts in the landscapes of social and cultural life of the last three centuries. As Theodor Adorno, Walter Benjamin, and other writers have pointed out, the issue of mechanical reproduction is fundamental in understanding the changing shape of communication in the late Nineteenth and early Twentieth Centuries. And the compelling problematic of sound’s reproducibility, like the reproduction of images, is centred on its apparent abstraction and disembodiment from the social world. However, mediated sound, disembodied sound, sound seemingly immortalised and perfectly reproduced – must also be acknowledged to be caught up within worldliness, framed by aesthetic, perceptual practices and conventions that have become semi-transparent.

---

New, naturalised recording practices, aesthetics, and technologies have produced machined sounds that are increasingly foregrounded, possessing a hallucinogenic clarity, becoming hyperreal (or realer than real), where the evidence of mediation appears increasingly transparent, and this transparency becomes naturalised, as machined sounds are constructed that refer to their own models of originality, authenticity, and reality. Sounds are being increasingly comprehended through the sedimented history of machined sound and hearing; where socialised meanings have been attached to particular sounds, and sounds are heard according to these codified conventions – machined sounds present particular acoustic pictures of the world.

Michel Chion has described the onset of an important transformation, where we are now in our everyday lives totally immersed in “mediated acoustical reality”, where sound is relayed to us by amplifiers and loudspeakers. This new and naturalised sound reality (which we can call secondary orality/aurality) has become a familiar and standard form of listening, “supplanting unmediated acoustical reality in strength, presence, and impact”. This form of listening “is no longer perceived as a reproduction… but as a more direct and immediate contact with the event. When an image has more presence than reality it tends to substitute for it, even as it denies its status of image”.

Mediated sound increasingly seems to offer “a more direct and immediate contact with the event”. Mediated sounds are coming to possess more ‘presence’ than “the reality it tends to substitute for” – even as this new model of sound (in a hyperreal turn) “denies its status as image”, as mere appearance, or semblance, or representation, or likeness.

---

93 Chion, 1994, p. 103.
With the history of ocularcentrism, this increasing dissemination of machined sound and hearing has come to support the image, being comprehended and theorised through already established visual models of understanding. Sound is being linked to the more developed narrative of the image. Sounds are increasingly associated with images; images are becoming soundful. The machined fusion of sounds and images is allowing sounds to be sourced by images (where images have come to be understood as the ‘natural’ and ‘unproblematic’ source of sounds), enabling the multiple echoes of sounds and images, where images are extended by sounds, where the force and impact of sounds penetrate and territorialize images. As cinema has contributed to a new model of sound and hearing, the relationship between mediated sounds and moving images has grown more natural and familiar. In the next chapter Chion’s ideas will provide a stepping off point as we proceed to investigate how this naturalised, familiar interplay between mediated sounds and moving images should be understood in terms of a resonant relationship between sound and image.
Chapter 5: The sound of an image

Despite his many contributions to cinema sound, Michel Chion should be situated as belonging to the tradition of privileging sound by inverting the hierarchy of image and sound. In making the provocative comment that a “scar” marks the “reconstituted totality” of image and sound,1 Chion is, of course, rehearsing a psychoanalytic reading of lack, which clearly encourages a hierarchical opposition between sight and hearing. This chapter begins by problematizing Chion’s remark that sound and image do not stick together. My interest, however, is in how sound and image fuse together, how sound inscribes and infuses a sense of presence on and to the image, how this articulation facilitates particular meanings and representations, how this interaction or machined fusion produces soundful images which are, in a certain sense, always already hyperreal (to use Baudrillard’s term) – always already realer than real. This chapter, then, makes the argument that sound and image need to be recognised as sticking together, where the “reconstituted totality” of sound and image, now composed of re-embodied sounds and re-sounded images, has the potential to resonate powerfully. It will explore how sounds and images, recorded separately, are subsequently conjoined, which results in a resonant relationship – a kind of intertextuality, where there are multiple echoes between image and sound, based on the ideal that sound should not explain or merely accompany the image in prosaic fashion. In so doing, sounds inflect the image with meaning by making particular connections. In arguing that sound and image now stick together in a resonant relationship, this chapter provides an extension of Christian Metz’s essay ‘Aural

---

Objects’, in order to elaborate on a conceptual model for understanding how sound and image resonate powerfully, shaping a media hyper(reality).

It is evident that a persistent gap remains within post-Edisonian sound and image; the remnant and result of the historical separation and autonomy of the senses. At stake in the machined fusion of sound and image is the mutual reinforcement between channels of communication that cannot forget or put aside their distinct histories, practices and identities. “But isn’t the talking picture precisely a form that reunites and reassembles, more than it cuts in two?” Michel Chion asks. “If the talking cinema has shown anything by restoring voices to bodies, it’s precisely that it doesn’t hang together; it’s decidedly not a seamless match”.2 This is the antinomy still at work within the emergence of soundful images, which can be traced to the machined separation and fusion of sound and image: to the fundamental gap between sound and image that has to be acknowledged; to the lack of sound that was originally embraced as a defining characteristic of a new, distinctive art form; to the gap between sound and image that has been bridged, but persists as a ghostly presence (or awareness) of absence (it will never be “a seamless match”).

Chion focuses our attention on the “scar” underlying the “reconstituted totality” of the cinematic experience of sounds and images. He writes: “The physical nature of film necessarily makes an incision or cut between the body and the voice. Then the cinema does its best to re-stitch the two together at the seam”.3 Chion writes about the cinematic operation of “grafting the non-localised voice onto a particular body that is assigned symbolically to the voice as its source”. In Chion’s opinion, this operation

---

“leaves a scar, and the talking film marks the place of that scar, since by presenting itself as a reconstituted totality, it places all the greater emphasis on the original non-coincidence”. Cinema, of course, was to seek to “reunify the body and voice that have been disassociated by their inscription onto separate surfaces (the celluloid image and the soundtrack)” via the process of synchronisation. “But the more you think about synchronisation, the more aware you can become”, Chion comments, “of the arbitrariness of this convention, which tries to present as a unity something that from the outset doesn’t stick together”.4

Chion is pointing out the constructed nature of the reconstituted totality of sound and image. From a historical perspective, it is clear that the critical and theoretical responses to this reconstituted totality of sound and image has been fraught with tension, conflict, and debate, due to this reunification and recombination of image and sound which, according to Chion, never manages to quite stick together. I have argued that this recombination is in a sense always a simulacrum, and becomes increasingly so with the postproduction sound, as highlighted by the use of Foley, ‘wrong’ sound-image perspectives (where we can hear sounds clearly and closely and in synchronisation even though events are at a distance), and other techniques. But what I want to argue now is that image and sound have come to “stick together”! Soundful images are now produced, which have become accepted as natural, and indeed, in many cases, as being realer than real. Chion’s protestation that this “reconstituted unity” is merely arbitrary and does not stick together offers one particular reading of the separation and reintegration of sound and image – however, I am more interested in how the ‘arbitrariness’ of connections and

---

linkages soon becomes naturalised, and ‘inhuman’ or ‘monstrous’ perception (or sounds and images that fuse to form a unity that “does not stick together”) soon seem to be ‘whole’ and ‘natural’. Indeed, we are presently confronted with “the problems of matching sound and image in an age when every sound comes equipped with its own stereotypical image”. The appearance of ‘wholeness’, then, where the image track and soundtrack appear seamlessly matched ‘together’, is indeed initially composed of two separate narratives (sound and image) that are recombined and juxtaposed in a kind of intertextuality. However, once images are re-sounded, once sounds are re-embodied in the image, these two recombined narratives are no longer separate, but rather result in the emergence of a resonant soundful image.

Rick Altman points to a problem that can be discerned in the work of the defenders of sound. For instance, in Composing for the Films, “Theodor Adorno and Hans Eisler attribute to hearing a privileged relation to pre-individualistic collective times; music thus has a pre-capitalistic nature, being more direct and more closely connected to the unconscious”. His critique can be extended. Adorno and Eisler make the comment that the sense of hearing has lagged behind in an age dominated by the sense of sight as the principle path to knowing. They observe that the ear is generally regarded as a passive organ when contrasted to the “swift, actively selective eye”. For them, there is more to the difference between the two senses: while vision is distancing, hearing possesses an enveloping and communal aspect. This sensory gap allows sound and hearing a special role: the spatiality and dimensionality of sound enables a

7 Theodor Adorno and Hans Eisler, Composing for the Films, 1994, p. 20; this text was originally published under Eisler’s name in 1947.
corporeality and immediacy that cannot be achieved in the two-dimensionality of the screen image. Altman points out that this approach to hearing and sound is a problematic, idealised and transhistorical ontological claim which does not acknowledge the contextual, historical, social, and cultural conditions of the role and function of hearing and sound. The comparative critical neglect of sound, together with the continued emphasis on visuality, makes for a theoretical blockage that cannot be surmounted simply by resorting to heightened attention to sound.

Altman has made the point that the work of Michel Chion, as well as that of a number of other psychoanalytic theorists such as Mary Ann Doane, Kaja Silverman, and Claudia Gorbman, are in danger of attributing hearing with a privileged relation, by attempting to challenge the regime of visuality and the image through resorting to heightened attention to sound and inverting their hierarchical position. Altman writes that Chion (and these other theorists) tend to characterise “the voice as archaic, based on the notion that we hear the soothing voice of the mother from the womb long before we are able to see”. For instance, Chion argues that from “the very dawn of time, voices have presented images, made order of things in the world, brought things to life and named them”. In this sense, the voice, sound, and hearing are held to be primary, closer to the Real.

Chion is interested in particular examples of how “a strong link has been created between the voice and the body, the voice and face. It’s in this relationship that something can emerge, and not in the voice taken in isolation”. Chion provides the

---

8 See Thomas Levin, ‘The Acoustic Dimension: Notes on Film Sound’, Screen 25.3 (May-June 1984), pp. 60-64.
10 Chion, 1999, p. 49.
example of *The Exorcist*, which “uses the very ordinary process of dubbing to make multiple voices – old woman, monster – come out of the ‘possessed’ girl”. Other effects are also employed, “such as ‘backwards’ sound and ‘slow motion’. The whole thing takes on its horrific effect because of the relationship, the comparison the viewer makes between the visible body and the voice”.\(^{11}\) For Chion, the “general realisation that the voice is radically other than the body that adopts it (or that it adopts) for the duration of a film” is taken to “be one of the most significant phenomena in the recent development of the cinema, television, and audiovisual media in general”.\(^{12}\)

Referring to Marguerite Duras, Chion writes about “the idea that the contemporary cinema stringently requires voices to be *nailed down* to bodies”. This “nailing-down via rigorous post-synching”, Chion argues, is present “to mask the fact that whatever lengths we go to, restoring voices to bodies is always *jerry-rigging* to one extent or another”\(^{13}\). Chion’s important concept of the acousmêtre, the voice without a body, or the voice that is made more powerful through its lack of physical presence, a voice that is present and yet mysteriously absent, is an exemplary illustration of his privileging of the voice. I want to suggest that the acousmêtre can be more profitably explained through the resonance of sound and image, which is founded on the separation and subsequent fusion of sound and image. This resonance between sound and image – made possible by re-sounding images and re-embodied sounds, by the articulation of sound and image, which results in the multiple echoes of sound and image – can also be used to re-examine Chion’s discussion of synchresis, where sounds and images, separated

\(^{11}\) Chion, 1999, pp. 170-171.
\(^{13}\) Chion, *The Voice in Film*, 1999, p. 130-131.
through the forms of soundtrack and image track, are unified (synchronised) in a reconstituted totality and wholeness, which although bearing the traces of this original separation, gains its power and potential of resonance from this reconstituted totality, from “sticking together”.

Chion makes the provocative claim that “there is no soundtrack”. By this Chion means that dialogue, sound effects, and music are merely the “end-to-end aggregation of all sounds in a film” which, in the absence of the image, do not operate to articulate “an internally coherent entity on equal footing with the image track”.14 Chion argues against (whereas Rick Altman argues for) the three elements of the soundtrack (dialogue, music, and effects) constituting an integral, parallel track to the images. James Buhler comments that: “sound… remains supplemental to the image in Chion’s account, which merely accepts uncritically the suspect, though admittedly pervasive notion that sound film is essentially a visual medium. Ultimately Chion… reduces film to image plus soundtrack”.15 This reductionism is despite Chion’s own important observation that sound affects our impression of an image every bit as much as the image affects our impression of a sound. After all, Chion himself points out that “the creators of a film’s sound… know that if you alter or remove these sounds, the image is no longer the same”.16 Perhaps he does not take his own injunction far enough, for particularly with the contemporary ubiquitous profusion of images, images are no longer simply ‘absent’ – increasingly, sounds conjure up images, which have become their sources, operating to

16 Chion, 1999, p. 4.
structure our reception of sounds, and images are re-sounded, claiming properties of sounds as their own.

In *The Acoustic Mirror* (1988), Kaja Silverman’s important feminist work on sound in cinema, Silverman focuses on the sound of the voice, especially the female voice, and provides a critique of Chion’s notion of the acousmêtre. Silverman gives a psychoanalytic interpretation of the voice in cinema, arguing that dominant cinema ‘contains’ the female voice in order to facilitate the male viewer’s own disavowal of his own insufficiency. Silverman argues that classic cinema “has the potential to reactivate the trauma of symbolic castration in the viewer, and that it puts sexual difference in place as a partial defence against that trauma”. Silverman says *partial* because the protection is only for the male viewer, not for the female. Silverman demonstrates how in most classic Hollywood films this psychoanalytic trauma is connected to women’s embodied and discursive powerless positions. The female voice then becomes an acoustic mirror for male anxiety. So, also in psychoanalytic interpretation, sound – especially the female voice – is considered an acoustic mirror for the image. Non-discursive physical and embodied sounds like screams, cries, and pure sounds are considered negative but fascinating at the same time. Silverman makes reference to Chion and argues that the body in relation to the voice often functions as a prison of sexual difference. She demonstrates that Chion’s work, despite its descriptive ‘neutral’ appearance, is heavily inspired by castration anxiety, especially when it comes to the voice. A point that strikes Silverman in the work of Chion is his statement that cinema is “a machine made in order

---

17 Silverman, 1988, p. 310.
to deliver a cry from the female voice”. 19 What is demanded from women, Silverman concludes, is an involuntary sound: the scream, the cry, or possibly a (non-discursive) melody. Silverman points out that the voice of Chion’s acousmêtre is almost never designated to a woman. In classic cinema, the all-knowing voice-over is reserved mostly for male characters. 20 If there is a female voice-over, it is an embodied one, the voice of an actress whom we can see in the image, while she is speaking in voice-over. Silverman demonstrates, still following Chion’s arguments, that the only disembodied female voice in the cinema is a terrifying voice, like the voice of ‘the mother’ (Mrs. Bates) in Psycho. The power of the acousmatic voice, then, is related to the omnipresent maternal voice. Silverman therefore makes the “critical claim that the acousmêtre, the threatening, all-powerful, free-floating voice which cannot be attached to any diegetic personality, is inherently masculine, i.e. the male voice of the master controlling the hysterical woman”. 21 (Žižek refutes this claim, and points out several exceptions, of which, he remarks, Silverman is well aware. Žižek argues that the voice, especially the female scream, is an impossible object of desire, for the voice as such (without speaking comprehensible language) makes a connection to the trauma of the Real. According to Žižek, the voice can be uncanny, “because the voice can threaten the established order. For that reason, it must be brought under control, subordinated to the rational word of meaningful language”. 22 The voice, then, can be liberatory, opening up new connections and affects.) In providing a feminist reading of the voice and performing a critique of Chion’s work, Silverman nevertheless is still in danger of a similar criticism that has been

---

19 Silverman, 1988, p. 78.
20 Ibid, p. 50.
21 Žižek, 2001, p. 42.
levelled at Chion. By privileging sound over the image, Silverman glosses over the productive insights that would be possible in thinking about the resonance of sound and image. Silverman, then, does not fully appreciate that sounds and voices gain power from this resonant relationship with particular images, where certain sound-image relationships have come to be naturalised, and sounds and images are thus caught up in a signifying network constrained by symbolic power, where certain soundful images, such as a screaming woman, come to be accepted as ‘natural’.

Sounds and images that do not necessarily start out together are fused together, creating a meaningful resonance between sound and image. Sound engages in a close relationship with memory, due to its existence in time. Sound refers to a time and space, evoking memories of a particular time, a particular space. Sound evokes the (acousmatic) spectre of a sounding body. These connotations can be utilised to provide presence to images, invoking the (onscreen or off-screen) presence of a source within a particular space and time, or grounding (situating) the images within a particular space and time. Moving images can claim the sounds as properties, which are made (sourced) by the images, giving depth, dimensionality, and life to the images, and using these sounds to penetrate the abstract space and time around the images, drawing further attention to the images. The machined fusion of sounds and images thus has the potential to create soundful images that possess a reinforced, augmented presence, inaugurating new models of meaning and perception.

The potential for sounds and images to resonate powerfully is founded on the ideal that sound should not explain or merely accompany the image in prosaic fashion. “Ever since the silent film gave way to the talkie”, the film director Kurosawa has said,
“sound has interfered with the image – and at the same time this flood of sound has become largely meaningless. That is why the director must be very careful – because a motion picture must be the most effective combination of both image and sound”.

Kurosawa reiterates that cinematic sound should never function as mere accompaniment, as the soundtrack should not simply be made from “what the sound machine caught while you took the scene. Real sound does not merely add to the image, it multiplies it”.23 With the creative deployment of the multiple echoes of sound and image, sound multiples the image, building layers of narrative that surround and expand the screen. However, it should be noted this articulation of sound and image is also subject to the processes of naturalisation and codification, as sound-image relationships rapidly become fixed in place – and the effect of the articulation of sound and image then becomes prosaic, as we enter an age where images lay claim to their ‘natural’ sounds, and sounds come equipped with stereotypical images.

To summarise briefly: the disembodiment of sound, enabled by recording technology, allows sound to be decontextualized and subsequently recontextualized (and I have been arguing that an important strand of this recontextualization is the re-embodiment of sound in the image, or the resonance of sounds and images arising from the machined fusion of sounds and images). Images can be re-sounded. In the wake of sound recording technology, “the voice is projected as a partial, disembodied object, no longer the sign of subjective interiority, but now a free-floating entity that can become an integral component of other systems and machines”.24 Voices and sounds have become exterior objects, removed from the interiority of a sounding body, but still invoking the

---

(absent) presence or the traces of a sounding body – disembodiment is then followed by an integration with “other systems and machines”. Re-embodiment occurs through the “reconstituted totality” of sounds and images, comprised of separate objects that are fused together (and marked by Chion’s ‘scar’, which I have read as the desire to point out the constructedness of sound-image linkages); re-embodiment is thus constructed by recombinations and dislocations, while presenting the naturalised appearance of a resoundingly ‘whole’ body.

It can be said that there are two basic types of sound-image events. First, where the sound-image ‘seam’ is not evident, or has become bridged (and this bridge has become naturalised); second, where the ‘seam’ is evident, or the bridge has not yet become naturalised. In both cases, the resonance of sound and image operates to construct a new sound-image event, which then possesses its own connotations and denotations of meaning, founded on the multiple echoes of sound and image that this new event is constructed out of. Several aspects of these multiple echoes (or articulations) are foregrounded (or highlighted), to form the new characteristics of this new sound-image event. And the miracle of recording is that, although it appears to offer a simple double or reflection of the world, it actually offers a particular perspective, interpretation, or refraction of the world, with the potential to precede and determine the world, making it (and us) a reflection of its appearance. I argue that soundful images have come to be valued as potentially new objects to be contemplated for their own meaning, not seen merely as reproductions of prior events – even as their meaning is still linked to this persistent narrative of imitation and reproduction.
So in examining the fading, often scarcely visible “scar” of the “reconstituted totality” of sound and image, we can observe the development and operation of the machined fusion of sound and image – evidenced in the multiple echoes of sound and image, the re-embodiment of sound in the image, the naturalisation of ‘arbitrary’ links between sounds and images, and the emergence of soundful images. Perhaps it might be more accurate to speak of ‘presentness’ being invoked by fusing sound to the image, instead of a logocentric notion of presence. This ‘presentness’ is shot through with supplementarity and trace, and is founded on the separation and subsequent fusion of sound and image. Instead of being a mere imitation, the new model of presence (the cinema re-appropriates presence) comes to inaugurate a new model of meaning (founded on and referring to existing models of meaning).

Sound fuses with the image to create a re-embodied event. Recorded sound appears to be a straightforward reproduction, but is actually carefully constructed, selected, and crafted. Machined sounds and images are increasingly being fused together, and are creating events that operate as models of cinematic (and perceptual) reality. And with the acceleration of media technologies, we can observe the continuing force of “this power of the image that eclipses reality”25 – even as images now increasingly become soundful, and resonant in new ways. Sounds are increasingly referring or linking themselves to images.

We now live in a world filled with re-embodied sounds, re-sounding images, and soundful images, which have affected our “cognitive mapping” of the world – machined images and sounds have been fused together, transforming the way we both see and hear.

I have described the encroachment of a new model of sound – where ‘copies’ are
becoming their own, new model of the ‘real’. I have argued that there has been a gradual
naturalisation of the way mediated sounds are increasingly being sourced by acoustic
machines, and heard as extensions of hearing – even as these mediated sounds become
increasingly hyperreal. In this chapter, I will elaborate on how these realer than real
sounds are increasingly associated and recombined with images. Mediated sounds,
sourced by acoustic machines, are increasingly being made meaningful through being
fused with images, through being ‘made’ by images. In arguing that sound and image
now stick together in a resonant relationship, this chapter provides an extension of
Christian Metz’s essay ‘Aural Objects’, in order to elaborate on a conceptual model for
understanding how sound and image resonate powerfully, shaping a media hyper(reality),
setting in motion new systems of meaning made possible through the articulation of
image and sound.

Some preliminary groundwork has to be done, however, before turning to look at
Metz’s essay. John Belton has commented that the sound track does not duplicate the
world set before it; instead, “it realises an imaginary world, endowing the space and
objects within the story space with another dimension that complements their temporal
and spatial existence as representations. What the sound track seeks to duplicate is the
sound of an image, not of the world”. This “sound of an image” is, of course, made
possible by the naturalised, semi-transparent practices around the production and
reception of mediated sounds that accompany the evolution of sound technology. Belton
continues: “The evolution of sound technology and, again, that of studio recording,
editing, and mixing practice illustrate, to some degree, the quest for a sound track that
captures an idealised reality, a world carefully filtered to eliminate sounds that fall outside of understanding or significance; every sound must signify”. For Belton, this process means that “the goal of sound technology in reproducing sound is to eliminate any noise that interferes with the transmission of meaningful sound”. The key element I want to extract and deploy from this passage is contained in the statement that “every sound must signify”, which establishes and emphasises a focus on the transmission of “meaningful sound”. My argument is that sounds, once evanescent, transitory, have been recorded and replayed by machines, and are fusing with images, being made by images – sounds now are signifying and making meaning for us through this articulation with images. I will refer to this development as “the sound of an image”.

The sound of an image refers to the sounds that are heard in accompaniment with images, that appear to be made by images, that invoke images; and this sound of an image is experienced, for instance, as the audience watches a film. As mentioned in Chapter 1, contemporary film audiences are generally unaware of the work involved in constructing the “idealised reality” supplied by the film soundtrack, which is comprised of various different intricate layers of sound. Practically all contemporary movie soundtracks are created in a mixing studio. Sound editors and mixers take a number of different audio recordings – dialogue recorded on the movie set, sound effects recorded in a dubbing studio or created on a computer, a musical score – and decide which audio channel or channels to put them on. This aural mosaic, a literal polyphony of sounds, is then mixed together to form the complete, ‘whole’ soundtrack – subsequently resulting in images furnished with the presence of sound.

---

26 Belton, in *Film Sound*, edited by Weis and Belton, 1985, p. 66.
I want to turn now to dissect the emergence and naturalisation of the sound of an image, via the arguments provoked by the film theorist Christian Metz’s seminal study of ‘Aural Objects’. Metz defined cinema’s “matter of expression” as being composed of “five tracks” – image, dialogue, noise, music, and written materials. Three of Metz’s ‘tracks’ are aural elements; and his work has certainly stimulated research into sound by undermining the formulaic view of the cinema as an “essentially visual” medium. Metz made the observation that “unless it is a question of the sounds of spoken language, sound has been studied far less than the visual, our civilisation greatly privileging the latter. Caught between the two, ‘sound’ is often left aside”. In the course of looking more closely at film sound, Metz examines “what auditory perception and language have in common”. He explains: “In this study, I wanted to show that the perceptual object is a constructed unity, socially constructed, and also (to some extent) a linguistic unity”. In this essay, Metz argues that sound is experienced not as a concrete object or thing (or, to phrase it in a different way, not as sound-in-itself) but rather as an attribute or characteristic of what he terms an “aural object”. He posits that filmic sound exists “as an autonomous aural object, the pertinent traits of its acoustic signifier corresponding to those of a linguistic signified”. In other words, sound signifies as a socially constructed object, and should not be framed as a purely material phenomenon.

Metz makes his key point that “auditory aspects, provided that the recording is well done, undergo no appreciable loss in relation to the corresponding sound in the real world: in principle, nothing distinguishes a gunshot heard in a film from a gunshot heard

27 Metz, in Weis and Belton [Eds.], 1985, p. 154.
29 Ibid, p. 159.
on the street”. Metz can be read as making the problematic and controversial suggestion that the sound of an image and the sound of the world are interchangeable. (Very likely influenced by Christian Metz, Gerald Mast has similarly asserted that: “there is no ontological difference between hearing a violin in a concert hall and hearing it on a sound track in a movie theatre”. This perspective, of course, is clearly opposed to the argument of this dissertation, which is that, in effect, all recorded sound, even if not dubbed or constructed, is hyperreal and therefore not the sound of some originary, stable, authentic reality.) Metz argues that sounds are usually classified according to the aural objects that transmit them rather than by their characteristics (whereas with visual objects the reverse is true). “Aural objects” are the embodiment of the sound source; for instance, an aural object would be the gun that has just fired (the source of the sound), while the ‘bang’ would be termed as the sound’s ‘characteristic’ – and the final identification of ‘gunshot’ is dependent on the acknowledgement of the aural object. In other words, “the recognition of a sound leads directly to the question: ‘A sound of what?’ At first glance, this seems paradoxical, since the… initial identification (‘whistling’, ‘hissing’, ‘rubbing’, etc.) corresponds to strictly aural profiles, while… the final identification (the wind, the river)… [involves naming] the source of the sound rather than the sound itself”. (Metz also examined the idea of off-screen sound, pointing out that sound cannot evade the screen area since it is either heard or not, despite the fact that sound is more difficult than a visual object to locate spatially.) Sounds are meaningful through reference to the source of the sound rather than the sound itself; sounds are meaningful with reference to and by

33 Metz, in Weis and Belton [Eds.], 1985, p. 155.
invoking their “aural objects”. In other words, to obtain meaning from sound, listeners typically look for or imagine its source. Generally speaking, any familiar sound or noise evokes mental images of its source as well as associations of activities, modes of behaviour, etc., which are either customarily connected with that sound or at least related to it in the listener’s recollection. (As Kracauer observed, “All the notions a noise is apt to evoke more or less revolve around its source”.34) Localisable sounds and noises therefore often carry familiar symbolic meanings, and can be classified as material units which, like verbal statements, serve as components of mental processes. Sound becomes codified as a unit of meaning.

For Metz, comprehending “a perceptual event is not to describe it exhaustively but to be able to classify and categorise it: to designate the object of which it is an example. Therefore, sounds are more often classified according to the objects which transmit them rather than by their own characteristics”.35 Metz examines “the conception of sound as an attribute, as a non-object, and therefore the tendency to neglect its own characteristics in favour of those of its corresponding ‘subject’, which is in this case the visible object, which has emitted the sound”.36 I will describe this as the sound-to-source model, where sounds are meaningful in terms of what has made the sound. Language has largely conceptualised sound in terms of the image, which has come to predominate in nearly all forms of communication. As Metz remarks: “The situation is clear: the language used by technicians and studios, without realising it, conceptualises sound in a way that makes sense only for the image. We claim that we are talking about sound, but

34 Kracauer, Theory of Film, 1997, p. 128.
35 Metz, in Weis and Belton [Eds.], 1985, p. 156.
36 Ibid, p. 158.
we are actually thinking of the visual image of the sound’s source"). Metz’s study and conclusions about ‘Aural Objects’ have been based on the understanding that the perceptual object is a socially constructed unity; and this fundamental point is essential in any discussion regarding the validity of Metz’s observations.

Metz can be critiqued (and indeed, has been critiqued) as offering a simplified account of sounds as mere components within a filmic narrative structure. According to his account, sounds are largely interchangeable, since they perform clearly delineated functions within the filmic narrative. Cinematic gunshots, although sounding nothing like real gunshots, are simply interchangeable commodities when situated within the filmic narrative. Thomas Levin disagrees with Metz’s position, remarking that a gunshot does not sound the same in the different acoustic spaces of the street and the inside of the cinema. “If differences remain unnoticed, this is a function of a socially constructed auditory practice which emphasises the similarity of such sounds in order that they be understood (i.e. linked to a common source) by the hearer”. However, Metz’s account remains attractive as it provides a persuasive and practical reading of how film sound works in terms of locating the audience within a framework of socially constructed meanings. The logical question then would be: has there been a reification of sounds and images, where particular sounds automatically refer to a certain set of images? Has this reification resulted in a ‘closed universe’ of meaning?

Metz remains prone to unsympathetic interpretation. For example, in his discussion of Metz, Sean Cubitt begins by bringing up “sound’s metacoding in language”, which “gives sound perception a socially constructed meaning”.

37 Ibid, p. 158.
Collin Chua  
Chapter Five

There is a pressure, in listening, not to hear the sound but the name of the sound, to infer from it its source, or to impute to it a certain semantic function. So instead of hearing what it is in itself, we ‘recognise’ the sound, after which we can identify what is making it and say to ourselves, ‘Now I understand.’

(Here Cubitt can be seen to be essentially summarising Metz’s position.) He then continues: “this coded listening emphasises the opposition between subject and object, recognising the historical nature of hearing as an acquired skill, framed in the determinations of a specific culture. Metz emphasises a formal, if not originary, division of hearing and sight from one another”. Cubitt notes that: “perhaps because his argument depends on the gap between subject and object constructed by recording sound, severing hearer from heard, Metz’s semantic relation of coded hearing suggests an instrumentalization of listening”. Cubitt says that Metz frames hearing and listening as a “sociological phenomenon, as something given in the contemporary cultural context; we should be wary of mistaking it for an inherent quality of all listening”.

Metz, however, is not saying that cinematic sounds are merely sounds that reproduce a simple coded meaning. As Lastra points out, it can be said that, for Metz, sound is always already a social/semiotic phenomenon, not a purely sensual or phenomenal one. The process of identifying something “as a sound, as a particular type of sound, necessarily involves some social or cultural dimension”; the very possibility of a sound having a particular identity presupposes its social character. In other words, Metz’s argument about sound should not be understood as a phenomenological account,

---

40 Cubitt, 1998, p. 94; Cubitt concludes by remarking that: “indeed, there is a tendency in contemporary acoustic arts specifically opposed to the semanticization of sound in any form”.

198
but rather “a semiotic or structural one holding that sound in the cinema is a socially produced phenomenon”.⁴¹ We do not hear the sound-in-itself – the sociality of sound means that we hear sound as meaningful sound, rarely as the pure materiality of sound. As Martin Heidegger points out:

> We never… originally and really perceive a throng of sensations, e.g., tones and noises, in the appearance of things… rather, we hear the storm whistling in the chimney; we hear the three-engine aeroplane; we hear the Mercedes in immediate distinction from the Volkswagen. Much closer to us than any sensations are the things themselves. We hear the door slam in the house, and never hear acoustic sensations or mere sounds.⁴²

This important observation allows us to return to Sean Cubitt’s discussion, where it becomes clear that Cubitt is pointing to long-standing habits of imagining that sounds transcend or escape meaning or that sounds elude sociality despite the fact they are made, heard, imagined, and thought by humans.

Admittedly, Metz offers a theory based on one particular model of hearing, and one that does not necessarily render the most sensitive discrimination between sounds. There is no valid reason for suggesting that one socially constructed practice of sound production and reception should ground the discussion and evaluation of all others. For someone like Metz who is more concerned with function and signification than with acoustics, there is no ‘loss’ between original and copy because even the act of listening to the original sound is a mediated and directed experience – at the very least, directed

---

toward identifying the source. It is perfectly obvious that experience of hearing a gunshot in the street does not resemble hearing a gunshot at the movies, as Metz claims. One cannot even say that the source of the sounds is the same, since the sound of a movie gunshot is taken from different sources, resulting in a sound more effective than the real thing. For Metz, what matters is our ability to identify a sound’s source, and thereby its meaning. In other words, as Lastra puts it:

Metz understands sound events to be inherently legible, or recognisable, across a series of different contexts as relatively stable signifying units. Against the definitions of sound as an essentially unrepeatable event, Metz describes sound as an eminently repeatable and intelligible structure… Metz thereby emphasises less its perceptual uniqueness than its capacity to generate meaning in a particular context... For Metz, it is not a question of there being no literal difference between street and theatre, but rather there being no difference in meaning.43

Sound recordings are constructed in order to ensure maximum legibility so that they may take their place as clearly defined elements in a narrative structure. Given these conditions, the sound of an image and the sound of the world can be recognised as being interchangeable in that they perform a particular narrative and semiotic function, and are held to derive from the same source.

Dissenters can be said to belong to another category of theorists, who have generally argued that instead of the sound of an image and the sound of the world being interchangeable (or identical), the sound of an image is nonidentical, or different from the sound of the world. Alan Williams is an example of this category of sound theorist,

---

43 Lastra, 2000, p. 126.
arguing that: “it is never the literal, original ‘sound’ that is reproduced in the recording, but one perspective on it, a sample, a reading of it”. What is this original sound? Is this original sound an idealised sound, complete in all its fullness, heard by an idealised auditor? For all ‘original’ sounds are necessarily heard from a particular perspective by the auditor, mediated by human hearing, and contained in a specific spatio-temporal context. Recordings are thus more than mere copies; as theorists like Baudrillard and Deleuze would doubtless argue – replayed recordings are new events that refer to the notion of an ‘original’, becoming virtual models that precede the real, that model the real. Thus, it can be said that this perspective of the ‘original’ becomes a new original, installs a new model – this accusatory ‘doubling’ of sounds does not necessarily imply a fatal gap or divide between the sound of an image and the sound of the world.

Lastra calls critics such as Williams non-identity theorists, and he comments that we do not hear sound like that. He writes:

Positing an unbridgeable gulf between original sound and copy, non-identity theory is always and in every instance sensitive to the most minor acoustic variations, allowing us to make finely nuanced distinctions between any two sounds, recorded or not. It correspondingly tends to regard all such differences as inherently meaningful. Like the exquisitely sensitive symphony listener, who might truly feel his concert had been ruined because he chose to move his seat two rows in the wrong direction, the auditor implied by non-identity theory is – or should be – constantly aware of the absolute uniqueness of his experience, and its

---

absolute difference from all other possible experiences. But… such sensitivity is not characteristic of the way we engage with most sounds.45

The non-identity of reproduced sound means that sound can be re-associated within different networks of meaning; sound can be read in different contexts, and from different perspectives. However, these different sounds are then categorised, ordered, and made meaningful according to what I like to term a semiotic template – a template which does not remain static, but evolves according to changing conventions and usages, much like a language. This is where Metz’s account can be given a sympathetic interpretation.

As mentioned, Metz’s notion of “aural objects” is usually also critiqued for placing sound in a subservient role to the image. (For instance, how can we explain the powerful role of speech and the voice via this hierarchy?) Contrary to Metz’s account, it should be evident that there is not a straightforward causal linkage between sound and image. I am arguing, however, that Metz’s account can be profitably read as an interpretation of how sounds can be used to reinforce images, and how images can be infused with sounds. Sound theory’s desire to re-order the dominance of image and the subservience of sound is understandable, but I want to theorise the machined fusion of sound and image – a relationship which does not necessarily imply a straightforward causal linkage between sound and image – through describing what I have variously called the resonance of sound and image, the articulation of sound and image, and the multiple echoes of sound and image. Metz’s model has facilitated the understanding that sounds largely make meaning by referring to their source – “aural objects”. And I am taking this point to its logical extension, in arguing that as our world is increasingly being

---

45 Lastra, 2000, pp. 131-132.
populated by mediated sounds and a profusion of images, this model of meaning – where sound is meaningful for us through a connection with its source – necessarily implies that sounds are increasingly being sourced by images, that there has been the emergence and naturalisation of soundful images, that the machined fusion of sounds and images is continuing on its implacable path. Metz’s point can be productively extended to suggest that the source that sounds point toward in a (technologically or socially) mediated environment is not so much a thing in nature but an image of simulacra – which means, of course, that constructed and manipulated sounds (such as Foley effects) are no less ‘real’, relatively speaking, than the ‘natural’ referents.

A relatively peaceful resolution to the controversies instigated by Metz’s account of aural objects can be found in the acknowledgement that there are different models of hearing/listening; varying degrees of attention; and different categories of auditory focus, in terms of what and how we perceive auditory stimuli. At the simplest level, we can differentiate between hearing and listening – where hearing is passive and listening is active. While hearing involves receiving auditory information through the ears, listening relies on the capacity to filter, selectively focus, remember, and respond to sound. Michel Chion has divided these modes of listening into three categories. Chion’s three modes of listening provide important models with which to comprehend how sound makes meaning. Chion defines causal listening as listening for the purpose of gaining information about the sound’s source;\(^{46}\) this mode can be directly linked to Metz’s argument. Semantic listening involves listening for the purpose of gaining information

about what is communicated in the sound (this usually refers to language);\textsuperscript{47} in my opinion, this mode can also be connected to causal listening (where information that is communicated in the sound is contextualised by referring to whom or what makes the sound; after all, the same spoken words can be given a radically different interpretive shading if they are attached to or re-embodied with a different speaking source). And reduced listening entails listening for the purpose of focusing on the quality of the sound itself (pitch, timbre, etc.) independent of its source or meaning.\textsuperscript{48}

Chion borrows the concept of reduced listening from Pierre Schaeffer; according to this mode of listening, sound is heard in its materiality, rather than listening for the sound’s cause or meaning. Reduced listening is recommended as an enterprise “that disrupts lazy habits and opens up a world of previously unimagined questions for those who try it”.\textsuperscript{49} Chion remarks that “reduced listening has the enormous advantage of opening up our ears and sharpening our power of listening”.\textsuperscript{50} In our contemporary age, where technologically mediated sounds are flooding our environments, Pierre Schaeffer’s category of reduced listening seems to offer a utopian ideal that has been gladly adopted as a rallying point by various artists and commentators. Cubitt, for instance, postulates that “by renewing the confrontation of auditor and vibration”, it is possible to return “to the act of listening a materiality which it had lost”. This then supports “the \textit{aesthetics of pure hearing}, a devotion to the \textit{autonomy of sound}”, which “will lead us towards the necessity and radical difficulty of assembling truly audiovisual media”.\textsuperscript{51} This is clearly an idealistic and utopian perspective, aimed at addressing our audiovisual age where an

\textsuperscript{47} Ibid, p. 28.
\textsuperscript{48} Ibid, pp. 29-33.
\textsuperscript{49} Ibid, p. 30.
\textsuperscript{50} Ibid, p. 31.
\textsuperscript{51} Cubitt, 1998, p. 92; my italics.
acoustic (or aural) spectacular has arisen to accompany visual spectacle; where sounds are louder, more clearly defined, and dramatically powerful; where there is an implacable reinforced materiality of sound as the physical shape of sound is buttressed by technology; and where this reinforced sound has become naturalised and oh-so-familiar to our ears.

I am proposing that sound can be divided into two categories – ‘natural’ (materiality of sound) and ‘cultural’ (sociality of sound). This goes back (again) to the old question: if a tree falls in a forest and there is no one to hear it, does it make a sound? Does sound require an auditor? Should sound be meaningful, or should it be defined as a physical, material phenomenon, that does not necessarily require an auditor? Should sound be meaningful? How should sound be meaningful? Should we focus on sound in a ‘natural’ or a ‘cultural’ way? How can we trace the dialectical relationship between these two categories? Cubitt represents the perspective, of course, of those who are anxious about the ongoing commodification, reification, and instrumentalization of sound, where in an ongoing process of sedimentation, meaning is layered, disseminated, and solidifies into widely accepted discourse, as enabled by the mass media. “Recording”, as Cubitt remarks, is “a technology formed in the traces of Metz’s metacode, a way of controlling the soundscape so that it will always conform to an already catalogued expectation, magically confirming our desire to hear with the desired sound”.52 Continuing this polemic, Cubitt insists that: “The invention of electro-mechanical recording, and its revolution into the digital, has doubled the quantity of sounds in the world. They too are part of the sound worlds we inhabit, along with the river and the pneumatic drills, the

clattering keyboards and the refrigerator’s hum”. Sound recording is no longer viewed with “the magical intensity that surrounds any new technology with glamour”; rather, sound recording has come to be integrated into our everyday lives and worlds, “even as it maintains its distinctive qualities. It has made the alien familiar”.

However, the ‘naturalist’ conception of sound, which privileges an idealistic purity of sound as opposed to the ‘alien’ sounds of machines, must be examined in light of the sociality of sound. Do we have to learn how to interpret sound (thereby subscribing to a culturalist or cognitivist argument) – or is there a primary level to comprehending sounds that precedes learned responses and complex mental cognition? The answer is that these two categories are not mutually exclusive – in other words, sound works through codified associations and conventions, operating as a learned trigger for representative ideas, but is also simultaneously able to be an emotional primary effect. Every sonic event, even those that do not appear to need interpretation, is subject to a process of codified and contextualised listening. For instance, Douglas Kahn writes that a scream “is thought to be an irrepressible expression, instantaneously understood through unmediated communication”. Of course, “screams in their natural habitat usually demand and receive a direct response. However, the literary, theatrical, musical, or cinematic habitats in which modernist screams reverberated are very different. Does anyone rush to the stage to lend assistance?” Even reduced listening, which attempts to hear sound without the imposition of codification, cannot completely escape the grasp of socialised meaning.

When we hear foreign languages, we hear the materiality of the sounds – but we always attempt to make meaning of sounds. When we are inundated by the recontextualizing

---

53 Ibid.
54 Kahn, 1999, p. 4.
force of avant-garde noise music, we are meant to immerse ourselves in the ocean of sound, and float on its surface – but we are again teased with the (perhaps inadvertent or involuntary) search for meaning. Reduced listening should be understood as an ideal form of audition, providing a curative that can be used to challenge the inevitable sedimentation and instrumentality of perception.

There is an increasing and inevitable reification and codification of the sociality of sound; and it is not difficult to observe that sound is increasingly commodified and instrumentalized – but this does not mean that sounds are completely confined within the code. Reification is never complete (as commentators such as Adorno have helpfully pointed out), and sounds can subvert and escape reification, through lines of flight, and by re-assembling their mode of organisation. (Avant-garde music, for instance, is a constant re-organisation and expression of sound.) In other words, the language of sound is never completely stable; like any other language, the codes and meanings of sound and listening are constantly being negotiated and re-negotiated, according to different listening situations and discourses. The materiality of sound is made meaningful when we interpret it through the unavoidable frames of language and culture and situation; sound has to make meaning for us; otherwise it is meaningless, noise, or silence. For us as listeners, sound is typically always subject to a constant process of organisation into a system of meaning. We seldom pay attention to the materiality of sound, merely how sounds are (socially) organised into coherent structures of meaning. Sounds can become so familiar that no one really hears what is going on anymore. The codedness of sound means that the organisation of sounds and noises becomes reified and naturalised. The essential point that needs to be emphasised here, of course, is that with the continuing
evolution of the language of sound and listening (which I have discussed in terms of a new model of sound), sound increasingly makes meaning through an image-sound articulation and resonance. Sounds have come to be fused with images, in more ways than one – the influence of the cinema has enabled the effective re-stitching of sound and image together, where sounds now naturally and unproblematically produce meaning by referring to images as their aural objects.

In other words, in our audiovisual age, where there has been a machined fusion of sound and image, an important transformation has occurred to the modes of listening. First, the causal mode has been increasingly sourced by images; second, the semantic mode has become shaped by new mediatized contexts; and third, reduced listening has increasingly been seduced by the allure of the acoustic spectacular, which has also resulted in the idealisation of the notion of a natural acoustic experience. As sound becomes more and more reified, codified, and socialised, I want to suggest that Metz’s account of aural objects becomes increasingly persuasive and relevant. Metz has been critiqued as being a reductionist, who is concerned solely with sound as a socialised object (causal listening), and as being an identity theorist who argues that recorded sound is analogous (or functions identically) to ‘real’ sound. But Metz’s argument describes how we tend to make sense of sound, and my argument is that aural objects as images are transforming (or influencing) all the various modes of listening. What I am saying is that the machined fusion of sounds and images is allowing sounds to be sourced by images, enabling the multiple echoes of sounds and images, where images are extended by sounds, where the force and impact of sounds penetrate, invoke, and territorialize images – resulting in the emergence of soundful images. I have discussed the persistent
theme of sound and presence, where sound refers to the immediate presence of an embodied source. In this way, sounds call out for their sources to complete them. My argument is that in this contemporary age, sources are increasingly becoming images. Sounds *call out for their sources* in order to make meaning; sounds need to be sourced (by their images).

Reproduction has not necessarily meant the death of the original and the triumph of the copy. The copy (in the form of recorded sound) appears to have triumphed over the original, but at the same time, reproduction has shaped a new idealisation of an original, as the proliferation of the copy has enabled the valorisation of authenticity and the original. The original is imaginary, created, and functions as a semiotic template to provide echoes of meaning. To quote Chion:

Curiously, the more unmediated [not technologized] acoustical reality loses its value as real experience and the less it is the lived standard to which we compare what we hear, the more it becomes the abstract reference we call on conceptually [as an *idealised* conception of sound] – for example, regarding the notion of acoustic fidelity that the cinema demands. The more we use recorded and/or transmitted sound, the more we mythify its contrary: a natural acoustical experience...

It can be said that the more sound is codified, and mediated, and the more the constructedness of mediated sound is naturalised, the more we come to idealise and appeal to the ‘real’ materiality of sound. This wistful idealism occurs against the

---

backdrop of a new hyperreal model of sound, and the machined fusion of sound and image.

Images now need to be comprehended as aural objects. Fidelity is now defined by the relationship of sound to its source-image. Sound has come to make meaning in conjunction and association with the image; and in a neo-Metzian formulation, the logical and inevitable conclusion would seem to be that *the sound of an image is becoming the sound of the world*. Recording and playback technology transforms sound into a repeatable object, possessing a prosthetic presence; sound becomes codified, and able to be arranged and organised within a system of meaning. Sound is still ephemeral – but this ephemeral nature has become reproducible. Sounds can become so familiar that no one really hears what is going on anymore. It has become increasingly natural for sounds to be fused with images (where mediated sound presents aural images of the world), just as it has become familiar for images to be soundful (where images are reinforced with the infused presence of sound). The contemporary transformation of sound and hearing centres on new aural representations that operate as extensions of hearing, and new auditory practices that are founded on increasingly mediated, augmented sounds, which are increasingly sourced or re-embodied by images. This is not to propose some rigid system or code of meaning, where one sound links unproblematically to a particular image – rather, it is to observe the ongoing production of systems of meaning, where sounds refer to images, in an ongoing articulation, and this referral is a fluid system that we organise according to contexts and our perspectival requirements. We understand and expect sounds to meaningfully and ‘naturally’ emanate from sources – these sources are contextual, and *can* be reified to form a rigid system. The re-association of sound – the
deployment of metaphoric sound – can estrange our perception, and create extended systems of meaning – but this re-association can be rigidified into convention, thus losing this power of estrangement.\textsuperscript{56} The links and connections between sound and image continue to evolve, and thus can break down, become blocked up, or re-directed, within varying systems of meaning. We need to rethink any proposal of pursuing a devotion to the autonomy of sound and an adherence to the aesthetics of pure hearing, in order to achieve the necessary goal of assembling truly audiovisual media. Instead of this idealistic stance, a more practical and creative perspective should be adopted: in order to achieve a creative use of film sound, sound and image should add up to make a larger, more nuanced whole, evoking multiple echoes of sound and image, instead of sound and image simply and prosaically each supporting the other. As Cubitt writes, “Images comment on and expand the sound; sounds interpret and open up the funnel of the image”.\textsuperscript{57} This is how best to explore the creative and communicative potential of the evolving articulation between sounds and images, as produced, developed, and circulated by our contemporary mass media. As part of this evolving articulation between sounds and images, the next chapter will begin by looking at the historical relationship between music and moving images, before turning to look at how sound effects and ambience have come to help redefine the contemporary form of the soundful image.

\textsuperscript{56} Cubitt, 1998, p. 119.
\textsuperscript{57} Take, for example, a stereotypical movie scene involving a terrified woman alone in a house on a stormy night, where various auditory narratives clues can be employed – loud claps of thunder, a ticking clock, thunder and rain beating against the window, the howling wind, the shutters banging against the side of the house, and the sound of steps coming slowly up the creaking stairs. The dark, flashing images, of course, are also underscored by the requisite creepy music, set in minor key, which plays incessantly upon the sustained tension of the scene. This particular interaction between images and sounds has become intensely familiar and somewhat clichéd.
Chapter 6: Music, ambience, sound effects, and the flow of images

Film, of course, is composed of a melange of sounds, not just the voice. In this chapter we look at the contribution other types of sound can make to the composition of the soundful image. The chapter begins with an overview of the role of music in film, before proceeding to analyse the role and impact of ambient sound and sound effects in the modern soundtrack. The chapter argues that these aspects of the soundtrack increasingly function in a similar way to film music, where film music is generally acknowledged to act as a vector, anchoring the flow of images and offering a set of viewing instructions for images that it accompanies. Sound effects and ambience are now integral features of the sound of an image. In presenting this argument, this chapter touches on a range of theoretical readings of film music, including Claudia Gorbman’s pioneering work, and Anahid Kassabian’s more recent account of the increasing popularity of what she calls compiled scores. It also examines a number of historical precedents and influences for ambience and sound effects, such as such as musique concrète, the work of the composer John Cage, and animation music. Building on the work of previous chapters, it is argued that the aim of SFX and ambient sound is to shape an impression of reality, a hyperreality. Images are filled in, reinforced, animated, and, in a sense, ‘completed’ by the infusion of sounds, noises, voices, and/or music.

Music and images have a history of being paired together. In 1932, Beaton would write that the combination of the visual sense being “engaged in following the action on the screen” as “the aural sense was employed in listening to music” is “perfect from a physiological stand-point”. After all, “when the visual sense is occupied exclusively, it tires more readily than when another sense, the aural sense, is occupied in sympathy with
Some decades later, Siegfried Kracauer explained that: “Music is not just sound; it is rhythmical and melodious movement – a meaningful continuity in time”. In other words, music intervenes so that “we perceive structural patterns where there were none before. Confused shifts of positions reveal themselves to be comprehensible gestures; scattered visual data coalesce and follow a definite course. Music makes the silent images partake of its continuity”. The thread I want to pursue in this chapter is how this sense of continuity has been increasingly provided by ambient noises and sound effects. But before discussing the role of ambience and SFX, it is necessary to set the stage by looking at the function of film music according to a range of perspectives.

In her book on film music, Kassabian elaborates, “Music draws filmgoers into a film’s world, measure by measure. It is... at least as significant as the visual and narrative components that have dominated film studies. It conditions identification processes, the encounters between film texts and filmgoers’ psyches”. Music, of course, should be understood as having been transformed through the advent of soundful images. For instance, the “stockpiling of music” has begun to involve the packaging of music with images; as people who are conditioned through a dominantly visual culture, we tend to remember certain auditory and musical themes alongside certain movie scenes.

Music for the movies is a strange phenomenon. It provides a cinematic bath of sounds that are akin to viewing instructions: you are meant to feel sad here, or happy here, or think that this landscape is magnificent. Joseph Lanza noted that in 1932, “Audiences, still new to the sound film, were known to deride a phantom symphony of

---

1 Welford Beaton, *Know Your Movies*, 1932, p. 82.
violins, harps, and horns coming out of thin air whenever characters were in distress or succumbing to a kiss”. These auditory viewing instructions, however, would come to be accepted, conventionalised, and naturalised as part of the movie-going experience, and audiences now take it for granted that music accompanies movie scenes; our viewing of the supposed ‘realism’ of the visual image is accompanied by carefully selected melodies, lyrics, and rhythms. Russell Lack reminds us that we should not forget how bizarre film music really is. “When placed alongside the supposed realism of the photographed image our experience of film music is radically at odds with our experience of the world”. For “the film image resembles reality in a way that the soundtrack does not”; after all, everyday life “is not as a rule serenaded by alien melodies”.

Nevertheless, this relationship between music and images has gradually become stabilised, codified, organised, and reified into a coherent system of meaning. Music enables us to navigate the flow of images; and music is tied to, or made meaningful, by the flow of images.

According to Deutsch, “Narrative support music accounts for about 80 per cent of all moving-picture music. It is effective and if done well, almost inaudible to the conscious ear”. Narrative music is composed in order to support the linearity of the film. It is constructed in this way to provide a vector (as Michel Chion might define it) for the particular scene it accompanies, guiding us along a path, which crescendos or dissipates at a pre-determined time. Images and music mutually anchor and reinforce each other. Music carries the spectator over the rough spots of the diegesis, keeping the audience’s

---

attention from the film’s process of fabrication. Stam elaborates on this (in somewhat melodramatic fashion):

The musical scores of Hollywood dramatic film lubricate the spectator’s psyche and oil the wheels of narrative continuity; music goes for the emotional jugular. Rather like aesthetic traffic cops, film music directs our emotional responses; it regulates our sympathies, extracts our tears, excites our glands, relaxes our pulses, and triggers our fears, usually in strict conjunction with the image.8

Heard in isolation, film scores often don’t make much musical sense; the music is deliberately written to enhance the mood of a scene and to underscore the action, not as a foreground activity, but as a background one. The function of the music is to ‘tell’ the audience how to feel, from moment to moment: soaring strings mean one thing, a single snare drum another. As Eicher writes, “Music and film-making have a close relationship because of rhythm, inhalation and exhalation, colours and layers of movement – horizontal movement, vertical movement... And very often the sound may be an inspiration for a film-maker to conceive an image”.9

Royal Brown argues that music, as a non-iconic medium, when accompanying the other tracks of film, can have a generalising function, encouraging the spectator to receive the scene on the level of myth, while also triggering a ‘field of association’ likely to foster emotional identification.10 Sound designer David Sonnenschein explains: “In a kind of trance with music, our lowered threshold of belief is essential to watching films

---

8 Stam, 2000, p. 221.
9 Eicher, in Sider, Freeman, and Sider [Eds.], 2003, p. 156.
10 Royal S. Brown, Overtones and Undertones: Reading Film Music, 1994; more generally, Brown notes that (non-diegetic) film music functions on at least three levels: as “wallpaper soporific” to allay fears of the dark or silence and to hide noises elsewhere, as an “aesthetic counterbalance to the iconic/representational nature of the cinematic signs”, and as a “co-generator or narrative affect”. (p. 32.) Brown adds that we cannot ignore the manipulative aspect of film music. (p. 110.)
and being immersed inside the story”. The dream factory of the cinema appears more real than real not because of the validity of the claims about cinema’s uniquely mimetic power, but rather because cinema enacts a representation of an idealised, fantasised, whole, imaginary reality, facilitated by the presence of music and sound. The cinema sets in motion a mode of subjectivity that allows us to absorb these sounds and images as being idealised (dream) forms that impose themselves on and submerge the audience. These sounds and images become part of a unified reality that we are immersed in – our subjectivity becomes part of the filmic reality. As such, we are not detached observers of the cinematic experience – we are within the sounds and images, and they become our reality.

Music has power to affect the visual field and the imagination; music carries particular connotations and associations, and contains sedimented meanings that emanate from various historical, social, and cultural contexts. In Donnelly’s words, “Music can suggest, or even lead directly to an elsewhere, like a footnote”. According to Annabel Cohen, images unaccompanied by music “seem prosaic, mundane, even lifeless; with music, however, the world of film comes alive”. Music operates to load our sensorium with sympathetic energies; music functions to bring the pictures into focus. Music also increasingly draws meaning from the images it is paired with. “Screen music”, Donnelly observes, “constitutes a way of looking at the world”. Music enables us to commune more intensely with the picture, thus operating to ensure “our ‘proper’ identification

---

within the scene, determining in a significant way how we comprehend it”. The modern audience has grown knowledgeable enough to be conscious of how musical cues operate in the movies, as the recognition value of popularly successful music like the *Jaws* and *Psycho* themes allows them to be parodied. Composers can also use contrapuntal music that plays against the textual theme, creating a sound-image event that is more than the sum of sounds and images.

Music and images are bound in a close relationship, resulting in an audiovisual language or code that we increasingly take for granted. Music, together with sounds, noises, dialogue – the other elements on the soundtrack – operates to form a continuum that grants an affective, emotional depth to images, and enables an extension beyond what is seen in the picture. We hear music, and it moulds images for us, it conjures up images; even as images absorb and claim music, voices, sounds, in order to extend themselves, to obtain a sense of presence, emotion, immediacy, and dimensionality. George Antheil suggests that Hollywood music can be understood as a type of “public communication, like radio. If you are a music fan (and who isn’t?) you may sit in a movie theatre three times a week listening to the symphonic background scores which Hollywood composers concoct”. As a result, “musical tastes become moulded by these scores, heard without knowing it. You see love, and you hear it. Simultaneously. It makes sense. Music suddenly becomes a language for you, without your knowing it”.

In our world of sounds and images, we are constantly garnering and honing an accelerated comprehension of audiovisual code, even if we are not aware of this process.

---

Since our normal process of listening is largely about making sense of what we are hearing, we form hypotheses or expectations about what is coming up based on what we have heard so far. Lack says: “Music organises and dredges memory, invoking something akin to a feedback system. The repetition of musical experience creates a residual psychic structure which becomes archetypical”. Film music indexes the perpetual present of film time, since it develops and defines our sense of visual tempo. Film, like music, unfolds in time. The subjective perception of the unfolding of changing filmic events, on both a plot level and in terms of the film passing through the projector, is predicated on an awareness of rhythm – the rhythm of change, the internal circles, repetitions, deviations and changes that forms the experience of watching a film or listening to music. As Sonnenschein argues, we tend to impose “order on our perceptions, as the brain will pull out patterns from chaos even if none objectively exist. Watching a person walk on screen and putting music to that image, it will seem that their steps are in sync with the music, just because we tend toward entraining sound and image”. However, film music is not an isolated system, in that it cannot be compared to a fixed code or language (as per Metz’s well-known attempt to theorise film as a language). No given system of semiology can manage completely to interrelate and/or classify the many different kinds of signifier on offer in the film. Film music therefore serves varying semiotic functions in individual instances of cinema, and analysis of film music thereby involves a nuanced decoding of the ways in which the score has been deployed as an isolated case. This is not to say that film music does not operate to perform a specific

---

range of functions within the filmic narrative, in its articulation and relationship with the flow of images.

In their *Composing for the Films*, Adorno and Eisler presented a well-known critique of the standardisation of film music, claiming that the majority of film music acts to “numb the listener”\(^20\). Indeed, in the early period of sound cinema, music became quickly rooted in a kind of mimesis. Music tended to track dominant physical action or melodramatic display of emotion. It wasn’t particularly subtle. For example: “Mountain peaks invariably invoke string tremolos punctuated by a signal-horn motif”\(^21\). Any moonlight scene is typically accompanied by the *Moonlight Sonata*. In opposition to this, they argued that the cinematic images should be accompanied by sensuous musical scores that reinforce the narrative or visual shock, instead of musical scores that serve to anaesthetise audiences with formulaic musical clichés. Unlike Eisenstein’s work on asynchronous sound, Adorno and Eisler did not espouse a lack of relation between sound and image. Instead, they critiqued the underlying attitude that music should serve as handmaiden to the potency of the screen image. According to Adorno and Eisler, the vast majority of film music served the purpose of suturing the ‘reality gap’ that is present in cinematic experience, by instilling a sense of passive receptivity in the audience. In their view, the image-score relation needed instead “to act in a unified accord to critique the socio-cultural here and now, and against the ‘natural’ tendencies of the visual medium itself”. They supported the technique of montage between image and score, in terms of exploiting the gap between the two, performing “a kind of visual-acoustical back-and-forth that acknowledges, rather than attempts to cover up, the difference”. Adorno and

---


\(^{21}\) Ibid, pp. 13, 15.
Eisler advocated “a denaturalising of the experience of viewership. Instead of music’s serving as an over-determination to the suturing proclivities of film in relation to the viewer, it should draw attention to the reality gap itself. The music, in other words, ought to de-familiarise, distance, or estrange the viewer”.22

They explained that film music’s “social function is that of a cement, which holds together elements that otherwise would oppose each other unrelated – the mechanical product and the spectators”.23 In ‘The Acoustic Dimension’, Levin points out24 that the spatiality of musical sound enables a corporeality – even when recorded, instead of live – that cannot be achieved in the two-dimensionality of the screen image. However, Adorno and Eisler concluded that the advent of the talkies did not manage to solve this situation: the characters seen on screen, although now voiced, remain two-dimensional effigies, still lacking the depth that the spatially powerful acoustics of film sound and music aims to provide.25 Eisler and Adorno professed that: “music is supposed to bring out the spontaneous, essentially human element in its listeners and in virtually all human relations. As the abstract art *par excellence*, and as the art farthest removed from the world of practical things, it is predestined to perform this function”.26 However, Adorno and Eisler’s contention that film music conceals alienation does not seem to be a charge that can be levelled at the film music itself, but rather at the finished film. Film music cannot be proved to be quantifiably any more responsible for lulling an audience into passivity than sensuous lighting, or ‘realistic’ acting. Also, their resolute claim that the flood of sound has become largely meaningless serves as an overly blunt opposition to

---

24 ‘The Acoustic Dimension’, *Screen* 25.3 (May-June 1984), pp. 60-64.
26 Ibid, p. 20.
the realisation that film music and cinematic sound still possess the potency and potential of moving past the role of mere accompaniment, instead functioning to add to the image, working to multiply the dimensionality and depth of the image.

The ability of film music to open up the image (contra Adorno and Horkheimer’s pessimistic outlook) can be illustrated with an experiment recommended by Michel Chion. This experiment, involving a forced marriage between sound and image, can be favourably compared to Kuleshov’s better-known experiment with visual montage. First, the instructions:

Take a sequence of a film and also gather together a selection of diverse kinds of music that will serve as accompaniment. Taking care to cut out the original sound (which your participants must not hear at first or know from prior experience), show them the sequence several times, accompanied by these various musical pieces played over the images in an aleatory manner.27

Chion concludes: “in ten or so versions there will always be a few that create amazing points of synchronisation and moving or comical juxtapositions, which always come as a surprise”.28 In addition, what he defines as anempathetic and empathetic sound work as productive elaborations of audiovisual montage. Empathetic sound refers to music or sound effects whose mood or rhythm matches and/or reinforces the mood or rhythm of the action on-screen; whereas anempathetic sound refers to sound – usually diegetic music – that seems to exhibit conspicuous indifference to what is going on in the film's plot, creating a strong sense of the tragic or ironic.29 For example, when the director Stanley Kubrick chooses to introduce the song ‘Try a Little Tenderness’ to accompany

---

27 Chion, 1994, p. 188.
28 Ibid.
two planes refuelling in midair at the beginning of Dr. Strangelove, he makes a disquieting aural pun that, in its contrast with the dark implications of the rest of the film, infuses his images with particular connotations. In other words, via the link between sound and image, music has the potential to strengthen the image by animating it with source-qualities that it would not ordinarily have by itself, via the process of re-embodiment.

In her seminal book *Unheard Melodies* (1987), Claudia Gorbman has pointed out the erroneous assumption that “the image is autonomous”. She offers instead the idea of “mutual implication” between the image and film music. As Stam has argued, “Conventional film music has always worked to efface the instruments of production of the cinematic illusion, channelling and directing the audience’s emotional response. Although in the sound film these two functions have been integrated into the filmic text and rendered more subtle, their purpose has not changed”. Drawing on narrative theory, Gorbman imported two terms and applied them to music: diegetic film music, which comes from a source that we actually see on screen or understand to be part of the diegetic world, and non-diegetic film music, whose presence is not ‘explained’ within the film’s world. (Gorbman also classifies meta-diegetic sound, which she explains as sound imagined, or perhaps hallucinated, by a character. She states that when non-diegetic music is played from any character's point of view and its subjectivity is strongly emphasised, it may be considered as meta-diegetic.)

---

30 Gorbman, 1987, p. 15.  
31 Stam, 2000, p. 220.  
33 Ibid, pp. 22-23, 83.
To elaborate further, diegetic music is perceived as coming from within the world of the narrative portrayed on film – for example, if a choir is depicted as performing an aria, singing voices are heard. (It is important to note here that music that comes from a source outside the cinematic frame can still come from the story – for example, if a radio is playing off-screen, and a song is heard.) Nondiegetic music is perceived as coming from outside the filmic narrative – for example, patriotic music is heard as heroic soldiers achieve a stirring victory. Gorbman identifies the co-ordination of music and action as a Hollywood convention allowing music to remain unobtrusive.34 Whereas nondiegetic music is clearly selected to set the tone of a scene, diegetic music suggests spontaneity. It is introduced as music that ‘happens to be there’, as a soundtrack to the characters’ lives. This seemingly casual relationship between the events and the music permits the establishment of further narrative connections, without disrupting the impression of realism. Gorbman observes the discursive possibilities this offers: as diegetic music is played, it comments on a scene while appearing ‘indifferent’ to the dramatic situation.35

However, the problem with these two categories of diegetic and nondiegetic is that they tend to function as mutually exclusive categories; whereas the actual boundaries between diegetic and nondiegetic typically blur, as usually what seems to be nondiegetic music quite often can turn into diegetic music, or vice versa. As mentioned, film music is not an isolated system, and should not be reduced to a fixed code or language. The binary code of diegetic/nondiegetic sound is useful, but does not manage to completely interrelate and/or classify the many different kinds of signifier on offer in the film. Rather, it remains as a tool utilised to dissect the relationship between sounds and

34 Ibid, p. 78.
images; where sounds call upon embodied histories and narratives and associations, which are then echoed and reverberate both within and outside the frame of the cinematic image.

Gorbman explains that music behaves as an *ancrage* (or binding-in technique) for the listener, as film music anchors the instability of visual signification. The French term *ancrage* (commonly translated as anchorage), was initially employed by Roland Barthes to describe the process by which the meaning of an image is ‘anchored’ by an exterior text. In other words, anchorage describes the semiotic articulation between two different systems of expression. For Barthes, linguistic elements can serve to 'anchor' (or constrain) the preferred readings of an image, operating “to fix the floating chain of signifieds”. Barthes introduced this concept of textual anchorage primarily in relation to advertisements. For example, photograph captions typically present themselves as neutral labels whilst actually operating to define the terms of reference and point-of-view from which the image is to be interpreted. For instance, “It is a very common practice for the captions to news photographs to tell us, in words, exactly how the subject's expression ought to be read”. Film music, then, anchors the flow of images by subliminally (or even blatantly) informing the audience just how the visual narrative ought to be interpreted and received. (Although this is not always the case, as film music can also act with disquieting, disruptive effect.) This concept of textual anchorage can therefore be applied to other genres (as Gorbman has done); its influence can also be discerned in my discussions of the articulation and intertextuality of sound and image.

---

Barthes, however, also introduced the idea of relay, which describes a reciprocal and complementary relation between text and image, in that each contributes its own part of the overall message. Relay also operates in terms of relating a sequence of pictures to each other (here we can profitably insert and explain the concept of montage). Barthes points out that “this relay-text becomes very important in film, where dialogue functions not simply as elucidation, but really does advance the action by setting out, in the sequence of messages, meanings that are not to be found in the image itself”. Barthes explains that most semiotic systems are actually a combination of anchorage and relay and "the dominance of the one or the other is of consequence for the general economy of a work". These notions of relay and anchorage can be usefully applied to a point made by the writer and film composer George S. Burt, who declared that film and music can interact contrapuntally, with each medium perceived independently, but combining to “make a statement that is larger than each of the component parts”.

Claudia Gorbman’s *Unheard Melodies* has attained an established position of authority, but some of its conclusions need to be re-addressed, particularly in the light of contemporary developments with the soundtrack. Gorbman’s claim is that much (orchestral) film music acts to smooth over the artificiality and ‘constructedness’ of film. Due to her influence, the orchestral score has typically come to be understood as always working in this way; and this type of orchestral score has also been situated as a ‘proper’ avenue for research. However, a theoretical discussion of, for example, a soundtrack built on ready-made pop music, is still largely unexplored territory. After all,
there has recently been a discernable shift in the soundscape. A whole new aesthetic has come to be associated with the way that people use music as a soundtrack to their lives, as popular music has become a dominant feature of contemporary life.

Up until the 1970s, the vast majority of film music was symphonic, wholly tonal and derived directly or indirectly from the conventional practices that had grown up with cinema, which in turn had derived from practices common in nineteenth-century theatre. However, the dominance of the symphonic score has been challenged by the increasing use of pop music which has been deliberately and selectively compiled for use in film. As the basic blueprint of film music established by the Hollywood studio system, the classical film score has become a heavily dated language. Pop music speaks in a very current and direct emotional language, allowing films to register an emotional impact with audiences accustomed to pop music rather than orchestral art music. Pop songs are ready-mades, carrying pre-packaged associations and images that can be utilised, repurposed, redirected, and recontextualised. In her study, Hearing Film (2001), Anahid Kassabian distinguishes between two main approaches to film music in contemporary Hollywood. First, the composed (or symphonic) score, which gives rise to a body of musical material composed specifically for the film in question; and second, the compiled score, which is built of songs that often (but not always) pre-existed the film. For Kassabian, composed scores, which are most often associated with classical Hollywood scoring traditions, condition what she calls assimilating identifications. These composed scores “are structured to draw perceivers into socially and historically unfamiliar positions, as do larger scale processes of assimilation”. However, compiled scores operate quite differently. “With their range of complete songs used just as they are heard
on the radio, they bring the immediate threat of history”. As compared to composed scores, compiled scores offer what Kassabian terms *affiliating identifications*, where these compiled songs construct narrative ties that “depend on histories forged outside the film scene, and… allow for a fair bit of mobility within it. If offers of assimilating identifications try to narrow the psychic field, then offers of affiliating identifications open it wide”.44 ‘As Time Goes By’, a pop song within a largely symphonic score, can be seen as an early example. ‘As Time Goes By’ has come to be indelibly associated with the images of *Casablanca*, invoking Rick sitting at a table after hours, demanding Sam to play: “You know what I want to hear. You played it for her. You can play it for me. If she can stand it, I can. Play it!”45

The surge of popularity of compiled scores can be largely traced to the mid-1970s, when pop music and its originary metaphor, the radio, become an all-conquering global phenomenon. The pop song as tacit commentator has almost become a genre of film music in its own right and can be said (however cynically) to have reinvented the value of many a record company’s back catalogue. Over the last few decades, the use of more and more pre-recorded music (whether classical or pop) in film soundtracks has resulted in a transformation of our sense of musical content. Quentin Tarantino has claimed that certain pop songs have been used so successfully in certain movies that the movie comes to lay claim to the song, blotting out other prior associations. He seems to have a good point – songs used in his films (such as *Reservoir Dogs* and *Pulp Fiction*)

45 There is an interesting history with regards to this song. Strangely enough, Max Steiner, the film’s musical director, didn’t want the song. The song, written in 1931, was only there because Murray Bennett, the teacher who wrote the unproduced play (with Joan Alison) on which the film is based, liked it. Steiner wanted to drop it, and might have succeeded – except that Ingrid Bergman had already cropped her hair for her next film, ‘For Whom the Bell Tolls’, and retakes were impossible. So it stayed.
have become integral parts of these films, fusing quite naturally with the images of these films. Pop music has become increasingly visual – that is, pop music has increasingly become image-friendly, actively seeking out links and associations with images, in terms of a closer symbiosis and articulation with images.\footnote{The arrival of MTV has played an important role. The music video became permanently established as a media form with the debut of MTV in the USA on 1 August 1981, as a channel dedicated to the broadcast of music videos. Appropriately, the first video screened was The Buggles’ ‘Video Killed the Radio Star’. This provocative title insinuates that the music video is a case of the image working to reinforce its hegemony over sound (where image has ‘killed’ sound). While various commentators have suggested that MTV does not privilege sound, but the images that are held to produce the music, Andrew Goodwin proposes that the opposite is true – that the video illustrates the music, that the images embody the music. His analysis attempts to show “exactly how we might invert the position of film studies, by demonstrating exactly how the visuals support the sound track”. (Goodwin, 1992, p. 70.) This ongoing debate, of course, again demonstrates the tension and opposition between sound and image.} An example from Reservoir Dogs (1992) provides a good illustration.

Joe Egan and Gerry Rafferty were a duo known as Stealers Wheel when they recorded a Dylanesque pop hit, ‘Stuck in the Middle with You’, in April of 1974. The single reached number five on the charts – improbably, eighteen years later it would become a cult favourite. In 1992, Quentin Tarantino, then a little-known writer/director, took the Cannes film festival and the world by surprise with his motion picture Reservoir Dogs, a movie about the difficulties that occur when five ‘master’ criminals are hired by a crime kingpin named Joe to pull off the biggest diamond heist of the century. Stuck squarely in the middle of the film, the Egan/Rafferty hit is played as an introduction and accompaniment to the infamous and bloodily violent torture scene, a centrepiece of the movie. Tarantino has said:

Personally, I don't know if Jerry Rafferty necessarily appreciated the connotations that I brought to ‘Stuck in the Middle with You’. There's a good chance he didn't. But that's one of the things about using music in movies that's so cool: the fact that if you do it right, it's about as cinematic a thing as you can do. You're really doing what movies do better than...
any other art form. It works in this visceral, emotional, cinematic way that's special. And when you do it right and you hit it right, then you can never really hear that song again without thinking about that image from the movie.47

Through techniques like this pop music has become increasingly visual – that is, pop music has increasingly become image-friendly, in terms of actively seeking out links and associations with images, in terms of a closer symbiosis and articulation with images. (And as films become marketed more and more as events rather than simply as films it seems likely that this process of exploitation (or ‘marketing’) by entertainment companies will eventuate in compiled scores being employed more frequently. After all, as Donnelly points out, “film music has had a massive audience, in fact probably the largest audience for music ever”.48)

To reiterate, the functions of film music can be placed in two main categories.49 First, film music provides a wide range of information to the audience. Film music operates to enhance viewer perceptions of historical verisimilitude, as music carries associations with the historical period suggested by the visual images, thereby capitalising on cultural nostalgia and audience knowledge. For instance, Gorbman wryly quips that: “If one hears Strauss-like waltzes in the strings, it must be turn-of-the-century Vienna”.50 Film music also tells you about character, about story, plot, mood, and, by the

---

49 In an oft-cited newspaper article, the respected concert-hall composer Aaron Copland explained that film music performs a useful function in terms of “tying together a visual medium which is, by its very nature, continually in danger of falling apart”. (Aaron Copland, ‘Tip to Moviegoers: Take off Those Ear-Muffs’, *The New York Times Magazine*, November 6, 1949.) Copland posited the following five categories of film music function: ‘creating atmosphere, highlighting the psychological states of characters; providing neutral background filler, building a course of continuity; sustaining tension and then rounding it off with a sense of closure’. Here, however, I have chosen to classify the function of film music into two broad categories.
use of motifs that recur, it creates connections, either subliminally or consciously, for the audience. In the *Star Wars* movies, the appearance of the villains onscreen is accompanied by deep or military sounds. In a particularly familiar illustration, the movie *Jaws* uses a sinister but very simple double bass which begins in long, heavy notes which gradually speed up in tempo and attack, creating a feeling of ominous tension, operating extremely effectively/affectively in introducing the entrance of the killer shark.

Second, film music serves to generate affect in the audience; in other words, the score is aimed towards bathing the audience member in feeling, or emotion – “a lot of film-makers put music in their films just for that reason; they want the audience to dissolve away in a bath of feeling”.

Director John Carpenter, who creates and performs the music for almost all of his own films, agrees that the soundtrack should be implicit: “you shouldn't be aware of what I'm doing… It should be working on you. ... I don't want you to be aware of the technique. I just want you to feel it”. Film music works to cultivate receptive and particularly emotive moods for the presentation of information, as music is intended to guide viewers toward feeling a certain way about the events being depicted. Film music serves to correspond with ‘invisible’ or continuity editing, enabling a transition between scenes, developing and sustaining dramatic tension, and/or providing a contrast with periods of no music. Film music can function in a subtle and transparent fashion; or its role can be more overt. Film music therefore provides information and comfort; without the guidance of the music, audiences are released into a position of

---

uncertainty and discomfort, and this is a strong reason as to why some filmmakers have made the particular aesthetic choice of not having music in their movies.

Walter Murch suggests that: “Music functions as an emulsifier that allows you to dissolve a certain emotion and take it in a certain direction. When there’s no music, the filmmakers are standing back saying, simply, Look at this. Without appearing to comment”.54 Director Robert Zemeckis has employed a deliberate use of silence on Contact, the 1997 film starring Jodie Foster. In the film's opening scene, the camera pulls back from earth and the viewer sees entire galaxies flashing by while old radio waves fill the air. Eventually, the earliest broadcast has been passed by, and the sound of static ensues. Then there is nothing. “There's silence for about 45 seconds”, says sound designer Randy Thom. “The temptation is to cover [the scene] with music in order to convey to people what they should be feeling from moment to moment. Zemeckis, however, was brave enough to allow it to be silent, and as a result I think it's a very powerful sequence”.55

Murch explains that when a musical score is heard, “you recognise that you are now being manipulated musically – and you reposition yourself slightly. A new voice has entered. Whereas source music [and sound effects] can enter a scene without being perceived as a new voice”.56 They are part of the diegetic world, coming from sources seen on the screen. A marvellous example is provided by The Birds (1963), a Hitchcock film in which sound plays an essential role. The Birds is “Hitchcock's most stylised sound track – it is composed from a constant interplay of natural sounds and computer-

55 ‘Making Contact’, interview with Randy Thom and Tom Johnson by Richard Buskin; article found online at http://www.filmsound.org/randythom/contact.htm, accessed on July 28, 2005.
generated bird noises\textsuperscript{,57} and entirely eschews the use of background music. Hitchcock told François Truffaut in his discussion of \textit{The Birds}, "Until now we've worked with natural sounds, but now, thanks to electronic sound, I'm not only going to indicate the sound we want but also the style and the nature of each sound".\textsuperscript{58} In the film, ominous bird screeches become “even more important than visual techniques for terrorising the audience during attacks. Indeed, bird sounds sometimes replace visuals altogether”. 

Oftentimes, as Elisabeth Weis points out, “Hitchcock carefully manipulates the sound track so that the birds can convey terror even when they are silent or just making an occasional caw or flutter”.\textsuperscript{59} Truffaut observes: “The bird sounds are worked out like a real musical score”.\textsuperscript{60} In the film, bird sounds function like music; as orchestrated sound effects replace the use of orchestrated instruments. Instead of music being played under the opening titles, bird sounds are heard. And once the menace of the birds has been established, Hitchcock “controls suspense simply by manipulating the sounds of flapping and bird cries that recur quite unrandomly for the rest of the film. At any point in the film a bird noise can be introduced naturally”, and this allows for “a means of controlling tension even more effective and less obtrusive than musical cues”.\textsuperscript{61}

Donnelly has noted that: “‘Sound design’ – often conceived in a ‘musical’ manner – has largely replaced the ‘classical film score’ as the musical paradigm for contemporary cinema”.\textsuperscript{62} My contention is that these traditional functions of film music are being filled by ambient sound-beds and sound effects – and with the evolution of modern

\textsuperscript{57} Elisabeth Weis, \textit{The Silent Scream: Alfred Hitchcock’s Soundtrack}, 1982, p. 136.
\textsuperscript{58} François Truffaut, \textit{Hitchcock}, 1967, p. 224.
\textsuperscript{59} Weis, 1982, p. 138.
\textsuperscript{60} Truffaut, \textit{Hitchcock}, 1967, p. 223.
\textsuperscript{61} Weis, 1982, p. 139.
soundtracks, marked by the rise of sound effects and ambient sound-beds in the aural hierarchy, it is important to look at the role, function, and interrelationship of these elements of the soundtrack. Anahid Kassabian has complained that “no significant body of criticism actively includes analysis of film music; it has rarely been pursued by the semiotic, narratological, or psychoanalytic theorists of (what they call) ‘film’. The film of these theorists and their theories, seen but never heard, was more silent than any silent film”.

A case should also be made for the neighbouring categories of ambient noise and sound effects, which, having grown in importance and role, and being organised in comparable fashion to the more established sphere of music, have also come to exert a powerful influence on the reception and perception of soundful images. Contemporary audiences and viewers are increasingly being bombarded with ever-deepening visual information; and in order to manage the production and reception of escalating visual spectacle, improved theatre speaker systems enable filmmakers to supply audiences with heightened musical backgrounds, ambient noises, sound effects, and voices, in order to anchor and manage the relentless flow of images. Sound-beds, sound effects, as well as music, are increasingly depended upon to provide the necessary bath of information, characterisation, affect, emotion, and feeling. As Mancini notes, “there is no better way to authenticate chimeric worlds than through familiar noises” – and there is no better way to destabilise and estrange familiar worlds than through surreal, discomfiting noises, voices, and sounds.

---

63 Kassabian, 2001, p. 5.
64 Mancini, in Weis and Belton [Eds.], 1985, p. 361.
The continued evolution of the synchronous soundtrack has led to new models of sound-image assemblages. Film music has transited from an assortment of various musicians playing ‘live’ to picture, to the recorded synchronised soundtrack. And as the soundtrack has grown in depth, definition, and complexity, the music contained on the sync soundtrack has come to comprise more than the organised sounds made by traditional musicians, orchestras, symphonies, and instruments. As early as 1924, Ezra Pound professed that “music is the art most fit to express the fine qualities of machines. Machines are now a part of life”. In 1939, the musician Edgard Varèse declared: “Personally, for my conceptions, I need an entirely new medium of expression: a sound-producing machine (not a sound-reproducing one)”. In 1977, the soundscape theorist Schafer made the announcement that: “the amplifier [has] replaced the orchestra as the ultimate weapon for dominating acoustic space”. What is clear is that the engagement with new acoustic technologies changed the face of music. New sounds were made; sounds could be prefabricated, recorded, reproduced, manipulated; new modes of authenticity and performance were gradually naturalised; and a type of ‘sonic tourism’ was made possible, as reproducible and portable sounds invoked soundful contexts and worlds for the sedentary-yet-sophisticated listener. Listeners were also granted an unprecedented ability to control their listening and viewing practices, creating new practices of consumption, production, and reproduction. Due to the changes made possible by new electronic technologies and practices, film music has become more of a

---

hybrid, in that there is an increasing fluidity between sound effects, ambient noises, dialogue, and music.

Among other influences, the design of film sound – with sound effects and ambient noises garnering growing attention and laying claim to more important roles – has been shaped by tape music, musique concrète, sound effects for radio plays, and the work of John Cage. In tracing this genealogy, the radio play might be a good point of departure, where sounds without images tell a story, where sounds invoke images in a narrative form that has frequently been billed as the theatre of the mind. Radio sound effects are built around sustaining illusion; after all, it is an art of painting pictures for the imagination.

It’s taking an ordinary bowl of cooked spaghetti and convincing the listening audience they are hearing a giant worm hungrily devouring people in their sleep. It’s eating an ear of corn in such a way as to conjure a picture in the audience’s mind of a horde of terrified rats trying to escape from drowning by gnawing their way through the wooden walls of a lighthouse. Or using glass wind chimes for the sound of sunlight – or two moist rubber gloves twisted and stretched for the sound of a human body being turned inside out...68

The creative impact of these sounds, emanating wonderfully from the radio, depends on the listener’s imagination. It is a familiar proposition; sound allows the listener to ‘see’ an event in their mind’s eye.

The first play written for radio is generally agreed to be A Comedy of Danger by Richard Hughes, which was transmitted on 15 January 1924. The genre’s affective power was to be demonstrated on the Halloween night of 1938, when the Mercury Theatre on the Air performed a radio adaptation of H.G. Well’s War of the Worlds, which had been

---

transformed into a faithful representation of an American entertainment program interrupted by urgent news bulletins. The panic created by this radio play is famous in broadcast folklore. The play’s mastermind, Orson Welles, remembers: “Six minutes after we’d gone on the air, the switchboards in radio stations right across the country were lighting up like Christmas trees. Houses were emptying, churches were filling up; from Nashville to Minneapolis there was wailing in the street and the rending of garments”.69 That landmark performance illustrates the power of radio, where the ‘authenticity’ of the performance fooled an entire country. Sound’s potential to persuade the audience results from the subtle manipulation of listening conventions, enabling the creation of rich (and in this case, terrifying) mental images within the theatre of the mind.

Orson Welles came from a highly developed theatre and radio background, where he had explored various possibilities contained by the deployment of sound. Welles’ use of radio sound effects was extremely innovative, and he brought the techniques he had learned in radio to films. When he began to make films, in the early forties, he utilised the assorted techniques that he had pioneered with which to simulate space with sound. This contribution was significant because Hollywood had already settled into a pattern of visual representation, where a spatial reality was represented by simply turning the camera on a scene. In radio, however, mental images are evoked through the narrative use of sound. As Murch points out, Welles’ appreciation of sound was developed “by the need, in radio, to tell the whole story through audio. All of his interest in sound effects, in ‘realistic’ dialogue overlaps, in the manipulation of acoustic space, came from innovations that he developed for his own radio programs, and then imported wholesale

69 Peter Bogdanovich, Orson Welles and Peter Bogdanovich, 1993, p. 18.
into the films that he made”. Welles found that his radio techniques could be profitably transposed to the film form, and that he could creatively combine the aesthetics of the radio play and the cinema. (These innovations were an important contribution of his first film, *Citizen Kane*.) In the aftermath of Welles’ work, film soundtracks were no longer quite so bound by previous reductive conventions.

The origins of film sound effects have been traced to the discoveries made through certain fortuitous ‘accidents of recording’ resulting from the primitive quality of early microphones used for film sound – where something would record better than the real object. In his history of early film sound, Scott Eyman writes:

Someone quietly shelling peanuts next to an open mike contributed a sound that resembled sharp crashes of thunder and lightning. A woman wearing silk stockings and crossing her legs produced something akin to the noise of a Kansas cyclone. Often, the technicians had to devise synthetic noises to replicate real ones that were either too explosive or somehow sounded false when recorded. Rifle fire was faked by breaking kitchen matches; wind through the trees was produced by the old stage technique of revolving a canvas strip over a cylinder of wooden slats; a falling body was simulated by dropping a ripe pumpkin; collapsing buildings were replicated by tearing heavy paper very near the microphone. As for cyclones, well, the technicians took advantage of serendipity and rubbed silk stockings together.

The constructed nature of film sound results in an apparently whole soundtrack that is actually made up of a mosaic of various sounds, as sound is always multiple. These created soundscapes have come to eventually model an original event that never occurred.

70 Murch, in Sider, Freeman, and Sider [Eds.], 2003, p. 87.
The philosophy and aesthetics of the avant-garde composer John Cage have been an enduring influence on people working in film sound. Walter Murch, for example, has shared his thoughts about the impact of Cage on his sound aesthetic: “I was… moved by the idea of what he was doing – that by taking humble sounds out of their normal context you could make people pay attention and discover the musical elements in them”.\footnote{In Michael Ondaatje, \textit{The Conversations}, 2002, p. 9.} Cage’s influence has been in encouraging attention to the rhythm of interesting sounds, as they occur and interrelate – everyday sounds can be employed and juxtaposed, as in a musical composition.\footnote{I will address the musicality of sound effects and ambient sounds in more detail in Chapter 7.} There is a fascinating story behind Cage’s proposed soundtrack for the radio play \textit{The City Wears a Slouch Hat}. In 1941, Cage was commissioned to provide the sound score for the Columbia Workshop production of Kenneth Patchen’s script \textit{The City Wears a Slouch Hat}. Cage espoused a vision of radio music which took into careful consideration “the possible environmental sounds of the play itself; so that, if it was a play that took place in the country, it would be natural to have the sounds of birds and crickets and frogs and so forth. But, if it were a play that took place in the city, it would be natural to have the sounds of traffic”. What he wanted was “to elevate the sound effect to the level of musical instruments”.\footnote{Conversing with Cage, edited by Richard Kostelanetz, 1988, p. 158; this aim has subsequently been achieved – see my discussion in Chapter 7.} In other words, his aim “was to explore the raising of sound effects to the level of music”.\footnote{Revill, 1992, p. 77.} When told by the sound effects engineer that "anything was possible", Cage let his imagination loose, returning “to the studio with a 250-page schema for an hour’s music, for which he had allowed just under a week for recording. The effects engineer told him that while the score was
technically feasible it was practically impossible, because of the expense”. 76 No doubt the engineer, in telling him that anything was possible, was not counting on the fantastic imagination of a composer devoted to musical experimentation. Cage’s original intent to write a score exclusively for sound effects (using them not as effects, but as sounds, as musical instruments) thus did not come into fruition. However, his goal – “to elevate the sound effect to the level of musical instruments” – while not being achieved at that time, has come to be evident as a sort of guiding principle for the deployment of ambient sound-beds and sound effects in the construction of modern soundtracks.

Another idea of Cage’s was described in his book *Silence*, where he recommended bringing about “a music which is like furniture [a reference to Satie’s *Furniture Music*] – a music, that is, which will be part of the noises of the environment, will take them into consideration”. 77 Although usually solely attributed to Erik Satie, *musique d’ameublement* (furniture music or furnishing music) was a collaboration with Darius Milhaud, which first took place in 1920 at an art gallery to act as an interlude for a play by Max Jacob. The introduction was read by Pierre Bertin:

> We present for the first time, under the supervision of MM. Erik Satie and Darius Milhaud and directed by M. Delgrange, ‘furnishing music’ to be played during the entr’actes. We beg you to take no notice of it and to behave during the entr’actes as if the music did not exist. This music… claims to make its contribution to life in the same way as a private conversation, a picture, or the chair on which you may or may not be seated. 78

---

76 Ibid, p. 78.
This not-to-be-listened-to music provided an unobtrusive ambient musical background, which operated to ‘furnish’ the space with sounds. The filmic ambient soundtrack was to take this ideal to a different, subterranean level. Douglas Kahn writes: “since film music must as a rule never overwhelm the images, action, or speech, it is relegated to a music heard but not to be listened to”.  

Cage also had a brief flirtation with film sound. David Revill chronicles Cage’s meeting with the experimental animation director Oscar Fischinger in the early days of the talkies. “Fischinger made abstract films to follow music; until meeting Cage, he had been using popular classics such as the Hungarian Dances by Brahms”. It seemed only logical for Fischinger to give Cage a commission to write for one of his films. Whilst occupied by this project, “Fischinger suggested to Cage, ‘Everything in the world has its own spirit, and this spirit becomes audible by setting it into vibration.’ ‘That set me on fire,’ Cage recalls, ‘He started me on a path of exploration of the world around me which has never stopped – of hitting and stretching and scraping and rubbing everything.’” This comment by Fischinger would influence Cage “toward an almost mystical respect for all sounds”. This combinatory potential of all sounds was also a focus of the development of musique concrète and tape music.

Pierre Schaeffer founded musique concrète in 1948. Together with Pierre Henry, Schaeffer and other composers of the Groupe de Recherche de Musique Concrète began experimenting with the process of piecing recorded sounds together in order to make music. For his early work, Schaeffer used phonographic recording equipment, before later moving to use tape recorders. He rejected his very first composition Étude aux

---

chemins de fer (1948) shortly after completion on the rationale that the train station sounds remained too recognisable. Thereafter, Schaeffer took care to employ “a variety of manipulation techniques that would more assuredly diminish or entirely eradicate any associative properties a sound might have”.81 Once sounds had been detached from their direct association, Schaeffer believed that music was inevitable: “From the moment you accumulate sounds and noises, deprived of their dramatic connotations, you cannot help but make music”.82

These so-called concrète sounds weren’t simply recorded – they were played backwards, at different speeds, with their attack transients spliced off, scrambled up and processed by various electronic devices, including filters, amplifiers and echo chambers. R. Murray Schafer clarifies that:

In the practices of musique concrète it became possible to insert any sound from the environment into a composition with tape, while in electronic music the hard-edge sound of the tone generator may be indistinguishable from the police siren or the electric egg-beater. This blurring of the edges between music and environmental sounds may eventually prove to be the most striking feature of all twentieth-century music.83

The advent and development of electronics has changed the whole basis of composition, blurring it with performance. The appearance of various sound production technologies has opened up the possibility of creating a range of new sounds. As Lack observes, “Any aspect of recorded music can now be edited down to the microsecond at any stage in the life of the recording to be redeployed in a completely new musical context”.84 Speaking

81 Kahn, 1999, p. 110.
82 See Diliberto, 1986, p. 54-59, 72.
83 Ibid, p. 111.
of *musique concrète*, Walter Murch has said: “I could hear a real similarity with what I had been doing – taking ordinary sounds and arranging them rhythmically, creating a kind of music on tape”. He goes on to remark: “It was an early, technically primitive form of sampling”.85

The structuralist Levi-Strauss has written a well-known commentary on *musique concrète*, which can be usefully related to the use of fragments of sound that are then assembled into a soundtrack. He remarks that in *musique concrète*, the collected noises are carefully treated and made unrecognisable, “so that the listener cannot yield to the natural tendency to relate them to sense images: the breaking of china, a train whistle, a fit of coughing, or the snapping off of a tree branch”. This process of estrangement is intended to remove “a first level of articulation”. Levi-Strauss suggests that wiping out this initial level of meaningfulness is not difficult, “since man is poor at perceiving and distinguishing noises, perhaps because of the overriding importance for him of a privileged category of noises: those of articulate speech”.86 Once sounds are disembodied, and replayed from a machine, it is typically fairly difficult to relate them to concrete “sense images”; in other words, it is quite hard to figure out just what made the sounds.87 (This, of course, as Levi-Strauss points out, does not apply to articulate speech.) This “first level of articulation”, as Levi-Strauss terms it, provides an important initial level of meaningfulness; sounds are meaningful for us as we automatically comprehend sound through what made the sound (the breaking of china, or a train whistle.) Once sounds are removed from their sources and stored in a machine, to be subsequently

---

87 This can be simply illustrated by thinking of the various ‘mystery sound’ contests that have been in vogue with radio shows, where a ‘mystery sound’ is played, and the audience is invited to phone in and guess what made the sound, in order to gain a prize.
replayed, any two sounds that sound similar or alike can be comprehended as the same sound, even if they come from wildly different sources. (Such as faking rifle fire by breaking kitchen matches.) After all, in the habit of tracing sound to its source, we find nothing but a sound machine. However, disembodied sound, freed from this “first level of articulation”, can be re-embodied with images (as sources), and hence rendered meaningful again in a new way, forming another mode of articulation, creating fresh “sense images” that are naturalised in turn. (When, for example, the sound of matches breaking is meaningfully articulated with the images of a man firing a rifle.)

More generally speaking, in the latter half of the past century, modern music has increasingly undergone an evolution of musical colour, of pitch, of musical texture. The move into the domain of electronic music has resulted in the manipulation of ‘blocks’ of sound rather than individual melody lines. The development of electronic music can be seen advancing down two largely separate avenues of experimentation: first, the emancipation of natural sounds through electronic manipulation, and second, the fabrication of an entirely new range of sounds through the use of electronics. The electronic manipulation of the human voice is perhaps the most direct metaphor of the machine age. In particular, the symbolic loss and disembodiment of the human voice by the machine has provoked all kinds of dystopian nightmares, even as the re-embodiment of voices has shaped the emergence of new sound-image interactions. Lack remarks that the dividing line between music and sound effects (and even dialogue) has been blurred by electronic film music. Film music’s function was typically to serve “to integrate other elements on the soundtrack, but with the incorporation of the synthesised score (and in the 1980s the MIDI-based sampler technologies) the soundtrack became seamless. Music
tracked not only significant dialogue or action but smaller minutiae that most audiences were not even aware of.” 88 Consequently, the diverse textures of electronic music have moulded “an ambiguity of musical meaning which undermines many of film music’s basic conventions. In isolation such music or sound flows can create ambiguous feelings in the audience since it can’t be coded in terms of simple pitch or rhythmic structure”. 89

The montage of sound – the combination and juxtaposition of various sound elements, the organisation of noises, voices, and sounds – has enabled new musical or sonic flows.

Brian Eno, musician and innovator of what has come to be termed as ‘ambient music’, can also be said to have had an influence on the shape of the contemporary soundtrack. Eno crafted his ambient ideal in 1975 during months of lying in a hospital bed recovering from a car accident, forced to listen to too-quiet 18th century harp music that his body cast prevented him from turning up. 90 This alerted him to the way that recorded sound can effectively merge with the environment in which it is played, appealing to many levels of listening attention without enforcing one in particular. He aimed to create a cocoon for thought and reflection through a music that could be used with utilitarian purpose. He has described the process as a painter taking the human figure out of a landscape. In music, this figure took the form of his own voice, a cohesive melody, and other evidence of human intervention – by eliminating these, he created a sense of space where there was once an object. Sounds can be tuned and sculpted into a soundscape that emphasises their immersive properties. Film music and background sound have already engaged with this aesthetic, as judicious deployment of music,

90 On the cover of his Ambient 4: On Land (1982), Brian Eno noted: “cluster all disparate sounds into one aural frame: they become music’. This echoes Erik Satie’s desire for musique d’ameublement, or furniture music.
sounds, noises, voices, combine to conjure up environments, oceans of sound, which we find ourselves immersed in, as audiences are bathed in affect.

In relation to this analysis of the relationship between score/soundtrack and visual images in the cinema, animation provides a particularly interesting and complex example of the tight fusion (yet not necessarily complementarity) of image and sound. Animation was one of the first film genres to take advantage of the potential of the articulation between music and images. In 1985, the British filmmaker John Grierson singled out the precedence of sound as the basis for Walt Disney’s success:

Out of the possibilities of sound synchronisation a world of sound must be created, as refined in abstraction as the old silent art, if great figures like Chaplin are to come again. It is no accident that of all the comedy workers of the new regime the most attractive, by far, is the cartoonist Disney. The nature of his material forced upon him something like the right solution. Making his sound strip first and working his animated figures in distortion and counterpoint to the beat of the sound, he has begun to discover these ingenious combinations which will carry on the true tradition of film comedy.91

The animated feature-length ‘concert’ film Fantasia, produced by Disney in 1940, can be appreciatively viewed as his pièce de résistance, in which enchanting animated figures such as dancing mushrooms and hippopotami were imaginatively choreographed and integrated with eight well-known pieces of classical orchestral music. The individual pieces were conceptualised within a wider framework, bringing together the various themes of prehistoric times, the four seasons, nature, hell/heaven, the themes of light vs. darkness and chaos versus order, dancing animals, classical mythology, and legend. This Disney production was an ambitious attempt to popularise classical music by

accompanying it with animation. Originally, the film was to consist of only The Sorcerer’s Apprentice segment, but it was expanded to include the full anthology of shorts.

When the Russian filmmaker and theorist Sergei Eisenstein first investigated Disney’s methods, it was the innovative use of sound that caught his attention. In 1928, Eisenstein had issued a ‘Statement on Sound’ together with Vsevolod Pudovkin and Grigori Alexandrov, where they argued that sound was the most significant development for the cinema. Sound – as Adorno and Eisler would also stress in Composing for the Films (1947) – was to be used contrapuntally, in “sharp discord with the visual images”.\(^92\) For Eisenstein, Disney created a captivating current of interflowing images, imaginatively shaped by the influence of sound – “Disney was so interesting because his metier was sound and moving drawing, together, apart, in dialogue and in conflict”.\(^93\) Effusively, Eisenstein would write that: “Nobody else has managed to make the movement of a drawing’s outline conform to the melody. In this Disney is inimitable”.\(^94\)

In a lecture given at the Sorbonne in February 1930, Eisenstein remarked that that future belonged to the sound film, “Particularly Mickey Mouse films. The interesting thing about these films is that sound is not used as a naturalistic element. They look for the sound equivalent of a gesture or a plastic scene, i.e., not the sound that accompanies it in reality but the equivalent of this optical fact in the acoustic domain”.\(^95\) Perhaps what Eisenstein found so intriguing was that animated characters are two-dimensional creatures that make no sound at all unless the illusion is created through sound out of

\(^{93}\) Leslie, 2002, p. 244.
\(^{94}\) From The Eisenstein Reader, edited by Richard Taylor, 1999, p. 171.
\(^{95}\) Eisenstein, 1988, p200
context: sound from one reality transposed onto another. In essence, this adhered to the model of asynchronous sound espoused by Eisenstein. When Eisenstein gave into his fascination and finally wrote about Disney Company cartoons,\(^96\) the main concern was the “poly-formic capabilities of an object”,\(^97\) the mouldable forms and the elasticity of bodies that accompanied the coming of sound to Disney’s animated universe, which he called *plasmaticness*. Eisenstein was deeply intrigued by the proto-plasmatic qualities of Disney films: “It is *beyond* any image, *without* any image, *beyond* tangibility – like a pure sensation… [It] is music!”\(^98\) For Eisenstein, Disney’s images had a musical quality and a molecular fluidity. The cartoon dazzled with brilliant tones and mutable forms, and its flexibility was shared and shaped by music which was subject to constant change.\(^99\)

Cartoons presented the viewer with beings that behaved like the “primal protoplasm”, not possessing yet a stable form, but dashing up and down the “rungs of the evolutionary ladder”.\(^100\) The elastic force of sound and music worked powerfully on animated bodies, just as sound stretched across the cinematic cut. As Kahn argues, “In terms of cinematic montage, sound did not resemble a suture, which as a figure is too inscriptive; it resembled a gum or a glue, an adhesion that could stretch”.\(^101\)

In cartoons, music functioned to shape and structure the visual movement of gestures and actions, with voices, sounds, and music also being supplied to and spread out over the bodies of both characters and objects, such as a squeaking robot elbow joint, fly foot steps, or a rib-cage xylophone enthusiastically played with left-over bones. As

\(^{96}\) See *Eisenstein on Disney*, edited by Jay Leyda, 1988, p. 69.
\(^{97}\) Ibid, p. 41.
\(^{98}\) *Eisenstein on Disney*, p. 46.
\(^{100}\) Ibid, p. 21.
\(^{101}\) Kahn, 1999, p. 150.
Kahn puts it, “The exaggeratedly tight coordination of sound and image in the novel content of sound cinema meant that the visual experience of animated cartoons was itself animated by sound”. It is essential to note that in animation, what is seen and what is heard commonly share common elements of structure. A musical tune, for example, is often characterised by a simplicity that is not unlike the simplicity of personality typically recognised in cartoon characters. In many animated films, musical tempo is related to the visual ‘beats’ of the action taking place on the screen. The power of synchronous sound was discovered quite early by animators, and ‘Mickey Mousing’ is still the term used to describe the effect when a character walks along with his footsteps perfectly in sync with a musical beat, or when actions (blows, falls, doors closing) are punctuated by instrumental flourishes (horn blasts, cymbal crashes, a flurry of notes, etc.). Chion provides a good explanation: “Owing to the eye’s relative inertia and laziness compared to the ear’s agility in identifying moving figures, sound helps to imprint rapid visual sensations into memory”. Sound can even be said to play “a more important role in this capacity of aiding the apprehension of visual movements than in focusing on its own substance and aural density”.

In his guide to animation, Kit Laybourne contends that we “are pattern-seeking creatures”. It comes naturally for us to “listen for the completion of various patterns fashioned from individual notes, intervals, instrumentation, themes, and rhythms”. When faced with a familiar melody, we automatically anticipate how it will play itself out. “If the composition follows the pattern we are expecting, there is a satisfying sense of

102 Ibid, p. 149.
104 Chion, 1994, pp. 121-122.
closure and ‘release’. If our anticipation is thwarted, we are disturbed and our attention increases, and if we experience a similar pattern that is embellished in a new and unexpected way, we are delighted”.¹⁰⁵ Curtis asks the question: “But how do we think about the relationship between sound and image when the entire shape of a cartoon is complemented and determined by the music?” He continues by pointing out that the ‘diegetic/nondiegetic’ distinction which functions so centrally in most discussions of film sound is clearly inadequate. “The term ‘diegetic’ refers to that which is accessible to the characters of a film, and yet buildings in cartoons sway to music that has no source in the diegesis. Even when there is a source, animators play with the notion of diegesis and the audiences’ acceptance of the convention”.¹⁰⁶ In other words, the music is absorbed and claimed by the images. This flexibility in animation music has also spilled over into sound effects, where the combination of fast moving animated visuals and ‘unrealistic’ sound effects create audiovisual metaphors that often have humorous aspects. Curtis makes the point that: “The relationship between the sound effect and its visual representation is not one of fidelity, but of analogy”.¹⁰⁷ In animation, synchronised sound effects are usually employed to provide a source of kinetic energy. In the Classic Hollywood Studio Cartoon, the soundtrack can be heard to be mostly constructed from boinks, petangs, pings, and kerthuds – these sounds, organised like music, are used to energise a scene. Taking a more contemporary example, in Disney’s 1998 animated feature Mulan, visuals of the cricket jumping around a piece of parchment writing a letter are synched to the sound of a typewriter busily clacking away.

It is important to realise that the various ways in which film music interrelates with the flow of images has continued to evolve and diversify. Mickey Mousing, once a forceful innovative technique, has come to be generally disdained as a painfully obvious method by many established film music composers, where music blatantly and simplistically matches the action. Music in the cinema is no longer obliged to correspond to the visual image. It is no longer forced to provide either a narrative or an illustrative function, since so much of this function can now be handled by other elements on the soundtrack. The articulations between music, voice, sound effects, and image are no longer necessarily bound by the requirements of a plain, straightforward match, but can be freed to seek new avenues of expression, that extend and bring new meanings to sounds and images, in creating a new sound-image event.

In many ways, Eisenstein’s comments in particular on the proto-plasmatic qualities of Disney animation can now be applied to how sound effects and ambient sound-beds tend to be used with contemporary films (this pertains to the live-action genre, not merely to animation), particularly in terms of the fluidity of sound-image linkages. In many ways, the contemporary moving image is animated by the creative use of music, ambient noises, sound effects, and voices. There is a sort of plasmaticness to the way music, sound effects, sound-beds, and even re-sounded voices operate to flesh out and expand bodies, things, and places, which then seem almost to vibrate and throb to these aspects of the soundtrack. In terms of contemporary cinema, Manovich defines digital cinema as ‘a particular case of animation that uses live footage as one of its many elements’.  

With digital filmmaking, Manovich notes, special effects have become
predominant, and shot footage is no longer the final point, but now merely the raw material to be manipulated on a computer, where the real construction of a scene will take place. With the fluidity and malleability offered by Computer Generated Imagery (CGI) and digital compositing techniques in contemporary digital cinema, this plasmaticness of the sound-image link can be said to have even more potential than ever before.

Ambient backgrounds operate as unheard melodies that can be likened to the subterranean foundations within film, lacking the grandiose, sweeping ambitious of symphonic scores, and the familiar choruses of compiled pop songs. Soon after synchronised sound made its entrance into the movies, it was quickly realised that when background sound continued over changes of location and behind dialogue, the illusion of a continuity of time and space was reinforced in an extremely tangible way. “Recordings of actuality music of sound atmospheres – crowds, traffic, wind, rain and thunder – could be laid behind the spoken lines, to produce a feeling of continuity and presence seemingly inseparable from the experience of reality itself”.

Most basically, ambient sound has come to be used in a manner similar to the way conventional cinema has come to use music: to bridge the cut. Because aural reality is perceived as a constant unbroken flow, introducing and sustaining quiet background music or sound effects from an apparently identifiable source helps to make a scene much more persuasively real. It marks out and establishes a section of time as believable because the ear confirms it, and the spectator is then drawn into the constructed reality of the scene more readily and more pleasurably. Ambient sound, functioning like unobtrusive background music, works

---

to ‘furnish’ a filmic space, and to flesh out the continuity (or discontinuity) of particular scenes.

Ambience is sound that produces a space for the film to exist in. Ambience supplies artificial ‘presence’ in the sense that it provides a ‘space’ that wasn’t there during shooting. Ambience most typically consists of more or less continuous sound, often with a low frequency emphasis we associate with background noise of spaces. Ambience plays a significant role in scene continuity. If ambience stays constant across a picture cut, it says subliminally to the audience that while we may have changed our point of view, we are still in the same space. An ambience track can provide a unified space for the action to play in that may not have been there at all during the shooting. Ambience may be used to smooth over small changes in presence that may otherwise draw attention to the artificiality of building the scene from different shots. When there is a continuation of the sound across a picture cut, we sense the illusion of continuity in the image as well. Conversely, if there is an ambience change at a picture change, it says the opposite – we are in a new scene. Sound can introduce or terminate a scene with an abrupt shift in such parameters as rhythm, volume, or timbre. To aid transitions, sound can overlap from one scene to another, either lingering or anticipating. Ambience may even be overlapped across certain scene transitions, either to create an effect of the former scene lingering into a new one or to anticipate a cut to a new scene. From a spatial point of view, ambience is one of the most interesting categories within the soundtracks, using stereophony to achieve its goals.

Sound designer Gary Rydstrom points out that a lot can be done quite subtly via soundtracks: “It doesn’t have to be in-your-face, traditional, big sound effects. You can
especially say a lot about the film with ambiences – the sounds for things you don’t see. You can say a lot about where they are geographically, what time of day it is, what part of the city they’re in, what kind of country they’re in, the season it is”.\textsuperscript{110} The rerecording mixer Richard Portman elaborates that the soundtrack should not just supply a cacophony of sound: if “it’s an establishing shot here in New York City and no one’s talking, you want to simulate the overpowering city that just rushes on you when you go outside – \textit{whoosh!} You give a second for that to establish and then go to a general traffic with selective sounds”. Constant sirens can then be placed between the dialogue, as city sounds that would be familiar to a New Yorker — but these should be played, Portman concludes, “as in a musical score”.\textsuperscript{111} This illustrates the growing trend towards the musicalizing of sound effects and ambience.

Ambience has become a ubiquitous form within film. Ambience can be described as a sort of unobtrusive film music, another layer to the soundtrack, made up of found sounds, environmental soundscapes, background noises and voices, as well as sound effects. A noticeable contemporary trend has been the rise of sound effects (SFX) to the top of the aural hierarchy. With the continued development of sophisticated recording, editing, and mixing techniques, sounds can be taken from their original sources, or built from raw and manipulated sources, be treated, combined, and/or manipulated, to produce all manner of ambience, SFX, and music. Even actors’ voices are frequently treated. Subsequently, these sounds, noises, voices, are fused with images with which they have no originary relationship, in order to create a new, realer than real, soundful image. Sound effects are commonly used to enhance the cinematic spectacle, by layering the

\textsuperscript{110} LoBrutto, 1994, p. 229.
\textsuperscript{111} Ibid, p. 47.
ambience track with convincing noises, through reinforcing images with persuasive sounds that sound realer than real. The traditional struggle between dialogue and music for supremacy can be said to be superseded by the triumph of SFX – as sound effects increasingly infuse dialogue, music, and ambience. New modalities are constantly being created and installed, in terms of performing new forms of fidelity to source, disseminating new types of sound-image engagements, naturalising new re-embodied sounds and soundful images.

During a 1991 lecture at the Sibelius Academy, Helsinki, Finland, sound designer David Yewdall declared: “Music is nothing more than organised sound effects, and sound effects are nothing more than disorganised music”.112 It seems “that Rene Clair’s 1929 prediction that ‘the interpretation [rather than the mere imitation] of noises may have more of a future’ has finally borne fruit, and it is the modern sound designer who has taken charge of the harvest”.113 Gary Rydstrom remarks: “You can use it [SFX] just like music. I always think of film sound in terms of rhythm. There’s sound happening over time, so it’s the rhythm of the sound effects over time that people remember”.114 Acknowledging the influence of Cage, Rydstrom continues: “John Cage talked about listening to the sounds around him wherever he was, the natural sounds or the sounds of the city, and hearing them as music, as interrelationships between rhythms and sounds. He made sound effects into music, and that happens in real life. Interesting sounds occur and they interrelate”. Rydstrom concludes that “we should do this in movies”.115 Speaking of musique concrète, Walter Murch – the man who first laid claim to the title of

113 Mancini, in Weis and Belton [Eds.], 1985, pp. 361-362.
114 LoBrutto, 1994, p. 236.
sound designer – has said: “I could hear a real similarity with what I had been doing – taking ordinary sounds and arranging them rhythmically, creating a kind of music on tape”.\textsuperscript{116} Murch also comments on John Cage: “I was more moved by the idea of what he was doing – that by taking humble sounds out of their normal context you could make people pay attention and discover the musical elements in them”.\textsuperscript{117}

In another example, the sound editor Richard Anderson tells a fascinating and amusing anecdote of the time he obtained a pig carcass from an auction for the purposes of creating a range of sound effects. “I want to emphasise that this pig was dead and didn’t suffer in any way”, he first takes care to announce. The pig was brought onto a Foley stage, where it was then hit with everything imaginable – “hands, hands with gloves… Then we got into hardware – two-by-fours, hammers; we actually had a sledgehammer. We also did stab wounds with a large knife. At the same session… we crushed and split open a variety of vegetables – watermelons, green peppers, you name it. We also broke wood”. The end result was a variety of short impact sounds. The resulting sound effects, made for the hits and impacts in a carefully choreographed fight scene, must be meticulously \textit{orchestrated} and organised for maximum effect/affect. “The important thing was to make a symphony of sounds. So if a guy got hit in the arm three times, you don’t want a \textit{boom, boom, boom}, you want \textit{boom, BOOM, boom} – more musical”.\textsuperscript{118} And, like music, sound effects can also become standardised in terms of a particular sound-image relation. The first recordings of dubbed-in punches began in the 1930s, with the early emergence of sync sound films. Ben Burtt makes the dry

\textsuperscript{116} In Ondaatje, 2002, pp. 7, 8.
\textsuperscript{117} In Ondaatje, 2002, p. 9.
\textsuperscript{118} LoBrutto, 1994, p. 165.
observation: “Obviously people weren’t really hitting each other on the set, so one had to produce a sound”. When caught up in the dynamic images of a good fight scene, it does not become immediately evident that the sound of a face punch has become intimately familiar to many audiences because they’ve heard it hundreds of times in a film.

Like film music, the purpose of sound effects is not to achieve ‘realism’, but for its magic, to create a mood, to present information, to invoke and liberate the sense of imagination, to make the audience feel a certain way, to hypnotise the audience by placing them in a particular situation, to augment the image by providing a sense of immediacy and presence, to reinforce the impression of cinematic reality. Rydstrom says, "Your first thought when you see a lot of special effects is that sound's job is to not only do something as fantastical as the visual, but also to make it real. It's not competing with the special visual effect, because people perceive the visual and the sound differently. Walter Murch had a way of putting it: The eyes are the front door, and the ears are the back door". Propelled by the musical ethos of a meaningful organisation of sound, film sound thereby installs a sense of presence upon being fused with the flow of images. When sound is fused to the image, it is re-embodied, re-attached to an originating space/time/source. Images are filled in and, in a sense, completed by the infusion of sounds, noises, voices, and/or music. In other words, we must investigate the sight (and site) of sound: just how sound links to (and is situated in and/or extends) the image. The importance of soundful images, enabled by conventionalised systems of sound and image

---

119 Ibid, p. 140.
in sync, lies in the networks of meaning that are set in motion by the articulation of sound
and image. Theorists such as Jean Baudrillard and Paul Virilio have elaborated a similar
critique of the simulation and substitution of reality performed by mass mediated images;
the focus of the following chapter will be to apply their critique to soundful images,
where images, reinforced by voices, sound effects, noises, and ambience, have come to
resound in hyperreal fashion.
Chapter 7: Hyperreal soundful images

In discussing hyperreal soundful images and their simulation of reality, this chapter will particularly refer to the contemporary form of what I have termed soundful images, with regards to digital sound-image assemblages that are increasingly alluringly spectacular in how they look and sound. The chapter specifically investigates the argument that cinema has contributed to a contemporary model of hyperreal soundful images. After all, as has already been argued, hyperreal sounds have become ubiquitous, increasingly fusing with the image, and operating as key components of soundful images. To conduct this investigation, this chapter engages with the work of two influential thinkers who have developed related critiques of the simulation and substitution of reality performed by mass mediated images. Jean Baudrillard originated the concept of the hyperreal, and is well-known (or infamous) for his writings on simulation. Paul Virilio has been prominent for his interrogation of (particularly media) technology, in relation to how the accelerated speeds of transmission and communication enabled by modern technologies has resulted in media representations increasingly coming to substitute for immediate presence and lived embodied experience. This chapter applies Baudrillard’s conceptual tools to the contemporary transformation of sound and hearing in terms of the naturalisation of hyperreal sound (particularly the present-day digital model of sound), and the chapter will conclude by applying a reading of Paul Virilio’s work to the relationship between sound-image assemblages and various conceptions of silence.

*The Matrix* (1999) is often used as a useful entry point into Baudrillard, simulation, and the hyperreal. Despite (or perhaps because of) being a somewhat misleading point of entry, it is nonetheless productive in revealing some of the
misconceptions involved with comprehending Baudrillard’s work, as well as providing a sense of the seductive allure of his theories. In the film, a computer hacker named Neo (Keanu Reeves) lives a relatively ordinary life, until he is contacted by the enigmatic Morpheus (Laurence Fishburne) who reveals a terrible reality – the world has been laid waste and taken over by advanced artificial intelligence machines. Computers have created a false, virtual version of 20th-century life – the ‘Matrix’ – to contain and placate the sleeping humans, whom the AI machines are using as biological power sources. Neo is hailed as ‘The One’ who will lead a small band of human revolutionaries to overthrow the machines and reclaim the Earth. Early in the film, before the revelation of the ‘Matrix’, Neo is visited by his friend Choi, who has arranged to buy some software. Neo picks up a book from the floor beside his bed; the book has been hollowed out, and contains several computer disks. The book is Jean Baudrillard’s *Simulacra and Simulation*. This reference pays homage to the influence of Baudrillard’s ideas on reality, simulation, and the virtual. Baudrillard’s reading of simulation and the simulacrum, of course, is far more radical than what is espoused by the film – which, as Baudrillard says, deals with “the growing blurring between the real and the virtual” rather “crudely”.¹ Ultimately, *The Matrix* provides a safe and conservative conclusion, by holding out the hope that beneath the simulacra of everyday life the real survives and can be rediscovered. William Merrin points out that for Baudrillard, “simulacra are not unreal media productions covering a real that can be rediscovered, as *The Matrix* suggests; rather they are as efficacious as the real, representing our everyday experience of the real’s own volatisation. Cinema itself provides an example of this, Baudrillard argues, in

invading and imploding with our lives”. For Baudrillard, then, simulation is not simply a layer to be peeled away, to reveal the real reality hidden underneath, as *The Matrix* suggests. Simulation is the reality constructed around us.

Baudrillard’s theoretical project has been to examine our saturated media environments, addressing the profound changes to society influenced by the growing role of the media. Rex Butler says that Baudrillard’s work “has always been about the essential paradox of representation, which is that the copy cannot get too close to the original without it no longer resembling it at all”. After all, “insofar as the copy completely resembles the original, it is no longer a copy but another original”. This focus explains the particular applicability and relevance of Baudrillard’s work in our digital age; after all, as Pierson remarks, one of the most powerful discourses around computer-generated imaging technologies is “the possibility that this technology might one day produce images that so perfectly simulate the look of objects in the real world that it will be impossible to tell that they have been computer generated”. The

---

2 Merrin, 2005, pp. 127-128.
3 Baudrillard’s theorizing of the hyperreal and simulation can be usefully supplemented by referring further to the philosopher Gilles Deleuze, who has contributed to a more developed understanding of simulation and simulacrum by pointing out that the simulacrum is not a copy but a phenomenon of a different nature altogether; the simulacrum operates by referring to and undermining the very distinction between copy and model (Deleuze, ‘Plato and the Simulacrum’, *October*, No. 27, Winter 1983, pp. 52-53). Deleuze’s notion of the simulacra rejects the implication that there was once a time when life was ‘more real’. For Deleuze, simulation is actually a process that *produces* the real. Simulation does not destroy the real, but rather transforms and reshapes the real by producing new models and originals that are different and singular. The point is that what we understand as reality is always already virtual. (Also see Brian Massumi, ‘Realer than real: The Simulacrum According to Deleuze and Guattari’, *Copyright* No.1, 1987; available online at: [http://www.anu.edu.au/HRC/first_and_last/works/realer.htm](http://www.anu.edu.au/HRC/first_and_last/works/realer.htm), accessed on May 9, 2005.) Deleuze will be addressed in more depth in the following chapter, as a critical counterpoint to the perspectives of Baudrillard and Virilio.
5 Butler, *Jean Baudrillard*, 1999, p. 14; Butler notes that this paradox was first raised by Plato in his dialogue *Cratylus* (1875, p. 257) and has been addressed by Derrida in his essay ‘Plato’s Pharmacy’ (1981, p. 139) – see Butler, 1999, p. 21, n. 5.
6 Pierson, 2002, p. 120.
perceptions initiated by the cinematic experience have come to produce and define a new (hyper)reality, which provokes important questions regarding the production and reception of a hyperreal aesthetic and perceptual experience. Media technologies have come to construct powerful sound-image events that stake compelling claims on our senses. What happens when we find moving images speaking with persuasive voices, resounding with powerful noises?

In engaging with what often seems to be the extreme perspective espoused by Baudrillard, it is important to keep in mind that his critical project should be read as an attempt to provoke a symbolic violence, a “theoretical violence – speculation to the death, whose only method is the radicalisation of hypotheses”.7 Baudrillard argues that theory should not be true; instead, it should constitute a symbolic challenge, provoking an opposition, defying the world to escalate to its own position.8 As Baudrillard says, “theory is… both simulation and challenge”.9 This strategy has often been overlooked or ignored. As a result, Baudrillard has been a controversial and provocative figure, and his work has been criticised as being nihilistic, fetishizing the spread of media simulations, and celebrating the “murder of the real” – Douglas Kellner, for instance, has not entirely kindly described Baudrillard’s writing as “perhaps that first self-consciously produced science fiction social theory to project futuristic anticipations of the world to come, the world right around the corner”.10 Baudrillard’s response to this acerbic observation can be seen in his comment that simulation is now “our absolute banality, our everyday

8 Baudrillard Live, 1993, pp. 56, 122-123.
10 Kellner, 1989, p. 84.
obscenity”¹¹ – no longer simply dismissible as science fiction, the complex matrix of mass-mediated simulations has become a familiar and even desirable aspect of our contemporary lives. Baudrillard observes that contemporary filmmaking has become only “a spectacular demonstration of what one can do with the cinema”,¹² where the cinema has become “a performance game” – one that displays “a supreme contempt for the image itself, which is prostituted to any special effect whatsoever”, as well as for the viewer, who has become an ‘impotent voyeur of this prostitution of images”.¹³ As Sean Cubitt quips about The Matrix: “Although the narrative wants us to puncture the illusion, it is illusion we came to witness”.¹⁴ We willingly consume and are consumed by simulation and hyperreality.

Baudrillard argues that the mass media inundate us with images.¹⁵ The moving image has become an important aspect of and shaping force on our society, just as the written and printed word had significant impacts upon past societies and civilisations. As Ron Burnett has written, images “have become central to every activity that connects humans to each other and to technology”¹⁶. The contemporary state of our mass-mediated world, where images are proliferating and becoming ever-more persuasive, means that images can no longer be seen as mere copies, as simple imitations, but rather as models that are communicating their own real, their own reality. Visually spectacular images now contain associations, connotations, and meanings that frequently can be more compelling and/or appealing than ‘reality’. In this way, the once seemingly

¹⁵ These images are “all-too-visible” (The Ecstasy of Communication, 1988, p. 22), providing us with an “absolute proximity of the thing seen” and a “hypervision in close-up” (Fatal Strategies, 1990, pp. 59-60).
uncomplicated copy-original relation has become problematized. Images have come to be their own entities; and in this way, new ‘originals’ have come into play.

For Baudrillard, the electronic mass media transforms “the lived, eventual, character of the world” into a finished consumable product that offers “the simulacrum of the world”, which thereafter “assumes the force of reality”.17 Baudrillard describes simulation as “the generation by models of a real without origin or reality: a hyperreal”.18 According to Baudrillard, we live in saturated media environments where signs proliferate at an immense speed.19 Hyperreal images are “substituting the signs of the real for the real”, where these signs are now their own pure simulation.20 Simulation, then, “is precisely this irresistible unfolding, this sequencing of things as though they had a meaning, when they are governed only by artificial montage and non-meaning”.21 For Baudrillard, the hyperreal is not the unreal but the more than real, the realer than real, as “the very definition of the real has become that of which it is possible to give an equivalent reproduction”, and the hyperreal is “that which is always already reproduced”.22 Baudrillard says that whereas “in earlier times an event was something that happened – now it is something designed to happen. It occurs, therefore, as a virtual artefact, as a reflection of pre-existing media-defined forms”.23 The hyperreal presents a model of a real without origin or reality; this hyperreal then shapes and transforms reality in its wake, generating or enabling meanings and real events on its own terms.

19 Baudrillard writes of “the murderous power of images, murderers of the real; murderers of their own model”. (Ibid, p. 5.)
23 The Transparency of Evil, 1993, p. 41.
Collin Chua  
Chapter Seven  

Cinematic assemblages of sound-image events have, in a sense, always already been hyperreal simulations; but with more recent developments in technology and practice, it would seem that hyperreal soundful images have reached a new, alluringly spectacular zenith of simulation – where we voraciously and willingly consume sound-image assemblages as they are produced, reproduced, exchanged, and circulated in our everyday lives at an unprecedented rate, and as these soundful images inextricably become part of the fabric of our everyday realities.

The hyperreal, then, refers to an event or representation that is ‘realer than real’. The hyperreal describes an event or representation that is presented with a hallucinogenic clarity. On an initial level, this ‘realer than real’, or idealised, event or representation draws attention to itself by being different from the ‘real’, more persuasive than the ‘real’, more conspicuous and/or enthralling than the ‘real’ – subsequently, it inaugurates new values, understandings, and meanings that come to define a transformation of the ‘real’, that come to (re)shape or re-appropriate the outlines of our reality with reference to this hyperreal. For example, contemporary digital sound recordings are hyperreal, offering a ‘pristine’, ‘perfect’, more persuasive depiction of acoustic reality. This hyperreal sound therefore works to normalise this new standard, as well as (paradoxically) to idealise analogue sound recordings as somehow being more ‘authentic’, more ‘real’.

Simulation, then, appears to be a hyperreal (or realer than real) copy or representation of an original event, but in actuality is a different event. This hyperreal simulation produces new ‘originals’ and new meanings. This original, hyperreal simulation is positioned as resembling or being a copy of a pre-existing event, therefore
producing or modelling an ‘original’ event that, in actuality, is new and different. In other words, simulation works by erasing the difference between original and copy; there is no original which can be reconstituted, as all we have are simulacra which precede (and thus erase) the possible constitution of an originary point. For example, soundful images are hyperreal, as they are constructed from disparate sound and image elements that are carefully fused together, forming a persuasive, impactful new sound-image event that models or appears to be a reproduction of an original event that in actuality did not occur. The ‘original’ event is constructed by the simulation that is the soundful image; the hyperreality of the simulation produces a new original that both draws on reality and re-appropriates reality.

Baudrillard’s ideas can be productively elaborated by cross-referencing to Daniel Boorstin’s 1962 book *The Image*. Boorstin describes the creation of a modern world “where the image, more interesting than its original, has itself become the original. The shadow has become the substance”.24 The image becomes reality: “Fact or fantasy, the image becomes the thing”.25 Boorstin suggests that the overwhelming impact of the mass media has given rise to the pseudo-event.26 These pseudo-events transform actual lived experience into reassuring intelligible and digestible images,27 which are “more vivid, more attractive, more impressive, and more persuasive than reality itself”.28 The consequences are, Boorstin says, “new categories of experience… no longer classifiable

---

24 Boorstin, 1992, p. 204.
26 Ibid, pp. 7-44; I have described how soundtracks have come to be constructed to depict a pseudo-event.
27 Ibid, pp. 185-194.
28 Ibid, p. 36; Boorstin has been criticized by Debord, who pointed to Boorstin’s failure to recognize the economic and political basis of our spectacular world, even as he liberally borrowed Boorstin’s notion of the ‘pseudo’ world and critique of the media for his 1967 book *The Society of the Spectacle* (1994, theses 198-200). Marshall McLuhan dismissed Boorstin as a ‘cultural reactionary’, who failed to appreciate the new reality formed in the electronic world (*Understanding Media*, 1964, p. 199). Unfortunately, I do not have the scope available here to engage with these critiques.
by the old common sense tests of true or false”, and a “reshaping of our very concept of truth”.  

Boorstin concludes that “the multiplication of images has had a revolutionary effect on all our imaginations, on our concept of verisimilitude, on what passes for truth in common experience”. Boorstin’s notion of the pseudo-event leads to a logical conclusion: we are threatened by a new menace, he writes. “It is not the menace of class war, of ideology, of poverty, of disease, of illiteracy, of demagoguery, or of tyranny, though these now plague most of the world. It is the menace of unreality”. It was perhaps inevitable that people would get worried. Sound and image after Edison have supplied a constantly developing, evolving, unreal (or realer than real) simulation of reality, a new reality, which does not necessarily coincide with the rules and laws of the real world (the physical reality in which we are living). We see things, and hear things, as if via an extension of our senses – and yet these sights and sounds excerpted from and/or modelled on the phenomenal world have been constructed, manipulated, and selected for us, on our behalf.

One wonders what Boorstin, writing before the digital revolution, would have made of computer-generated special effects imagery, which would seem to bear out his worst anxieties. No need to wonder with Baudrillard, who suggests that “the useless perfection of images” leads to the hyperreal modelling of an alluringly coherent spectacle, whose offer is of a coherence that is impossible in the contemporary world.

---

29 Boorstin, 1992, pp. 211, 205.
30 Boorstin, 1992, p. 212
33 This brings to mind Baudrillard’s well-known example of Disneyland, which “is presented as imaginary in order to make us believe that the rest is real, whereas all of Los Angeles and the America that surrounds
Baudrillard points out that what threatens us is our very “ultra-reality”, as media representations have become realer than real.\textsuperscript{34} He argues: “the more closely the real is pursued with colour, depth and one technical improvement after another, the greater does the real absence from the world grow”.\textsuperscript{35} In our time, he suggests, “there is an inverse negative relation between the cinema and reality”, where the real is now “absorbed in cinematographic (or televised) hyperreality”.\textsuperscript{36} Baudrillard’s project is an attempt to look at the escalation of mass media representation and mechanical reproduction, raised to the power of simulation. His work presents simulation as establishing its own frame of reference; simulation creates the real. In a remark curiously adjacent to Baudrillard’s position, Cubitt observes that: “In return for the simple sacrifice of reality, the digital cinema of the new Hollywood offers the assurance of a whole and satisfying if circumscribed existence”.\textsuperscript{37} As we consume mass mediated consumable products, the overexposed, realised simulacral model is given material, persuasive, force; we are utterly absorbed by the soundful images produced for our rapt attention.

The point I wish to make is that simulation is not confined to hyperreal visual images; sounds can also be hyperreal, or realer than real – and further, sounds can facilitate the construction of a hyperreal image by granting it soundfulness. Baudrillard does not specifically address sound in any sustained detail. Indeed, he has candidly confessed in an interview that “sound, the sphere of sound, the acoustic sphere, audio, is really more alien to me than the visual... sound is less familiar to me. I have less

\textsuperscript{34} The Perfect Crime, 1996, p. 66.
\textsuperscript{36} The Evil Demon of Images, 1987, p. 33-34.
\textsuperscript{37} Cubitt, 2004, pp. 268-269.
perception, less analytic perception, of this aspect”.

While Baudrillard might feel that he is not quite capable of extending his thought to address the sonic and audiovisual realm, I would suggest that his ideas are extremely productive and can provide useful avenues and conceptual tools to think about the contemporary state of soundful images in an age where digital technologies and practices have come to the fore. Mediated sounds should no longer be conceived of as a simple copy or representation of ‘reality’. Developing technologies and listening practices have indisputably affected the way we hear and interact with the world around us. Mediated sounds now are no longer merely reproductions, but are now productions of sounds with reinforced clarity and apparent fidelity – leading to a new model of sound that acts to produce its own real, its own reality, that is consumed according to naturalised ‘ways of hearing’. What’s more, sounds have come to be understood in conjunction with images, resulting in persuasive hyperreal sound-image simulations. The advent of cinema did not only mark a transformation of vision and visuality – with the arrival of sound film, sound and hearing have also been significantly restructured and impacted, and sounds have become as thoroughly commodified as images. Sound technology has also been fetishized, as evidenced by the pre-movie trailers for aurally spectacular various sound reproduction formats such as Dolby and THX. The contemporary soundtrack has become increasingly spectacular and hyperreal. The cinema can be located as an important site for an increased production and reception of hyperreal soundful images, where sound-image assemblages are increasingly “more vivid, more attractive, more impressive, and more persuasive than reality itself”.

38 Caroline Bayard and Graham Knight, Vivisecting the Nineties: An Interview with Jean Baudrillard, Part 1, *Ctheory 18*, nos. 1-2, August 1995, article 24a.
39 Boorstin, 1992, p. 36.
The digital revolution in the cinema has come to offer its viewers the latest development in hyperreal soundful images at the present moment. How did digital soundful images develop? In coming to the forefront in the early 1990s, the spectacular computer-generated imagery (CGI) featured in Hollywood genre films “became an object of intense fascination, curiosity, and scrutiny in the popular and mass media”.

The introduction of digital editing systems arguably marks the most significant change ever to have occurred in film-editing practice, in that the materiality of the process is almost entirely excluded. Non-linear editing involves film being recorded digitally into the computer memory in such a way that is can be easily recalled and identified for instant access, so that sections from any part of it can be assembled in any order and rearranged and adjusted at will. The system provides a degree of flexibility unavailable to traditional linear editing tools, with instant access to all the material and virtually no cumbersome handling of physical film.

The adoption of digital non-linear editing systems has changed the way film images are viewed and assembled into sequences, influencing new kinds of assemblages, new perspectives on montage, and faster rhythms.

In 1991, *Terminator 2* showcased visually stunning, groundbreaking computer-generated special effects, featuring the now well-known spectacular morphing effects of

---

40 Pierson, 2002, p. 3.
the T-1000 cyborg.\textsuperscript{43} Pierson observes – “Sound and image tracks work together to clear the way for the presentation of one of the film’s key CGI effects sequences: the revelation of the T-1000 in its liquid metal state”. This signature scene is “preceded by a visually and aurally frenetic chase sequence”, which “ends with the truck being driven by the T-1000 exploding into a spectacular fireball”. After the hyperkinetic action, dizzying visual montage, and explosive sound effects, “there is a moment of stillness and quiet, when the low roar and crackle of the flames that now fill the frame are the only sounds to be heard. A horn blast announces a burning tire as it rolls out from the fire… swiftly followed by a familiar drumbeat”. With these anticipatory cues, “a luminous, chrome-like figure begins to emerge from the flames, its shimmering, reflective form morphing into the familiar figure of the police officer as it moves out of the fire”.\textsuperscript{44}

In the December 1995 issue of \textit{Wired}, Parisi wrote: “A decade ago, only an intrepid few, led by George Lucas’ Industrial Light and Magic, were doing high-quality digital work. Now computer imaging is considered an indispensable production tool for all films, from the smallest drama to the largest visual extravaganza”.\textsuperscript{45} That year, in 1995, \textit{Toy Story} was released, as the first completely computer-generated feature film. A year later, in 1996, \textit{The English Patient} was one of the first films to be completely digitally edited in both image and sound, winning its editor – Walter Murch – two Academy Awards for film and sound editing. In 1999, \textit{Star Wars: Episode 1 – The Phantom Menace} was the first major motion picture theatrically exhibited as digital

\textsuperscript{43} For more about morphing effects in the cinema, see \textit{Meta-Morphing}, edited by Sobchack, 2000.
\textsuperscript{44} Pierson, 2002, p. 125.
cinema with digital projectors. These milestones in digital filmmaking and exhibition have coincided with aesthetic developments in sound-image assemblages, which have come to resonate and impact in different ways, some subtly different, and some resoundingly so.

The composition of the moving image itself has changed. Special effects have become predominant in digital filmmaking, when it was once historically relegated to cinema’s periphery. Manovich has defined digital cinema as “a particular case of animation that uses live footage as one of its many elements”. As Manovich observes, shot footage is no longer the final point, but now merely comprises the raw material to be manipulated on a computer, where the real construction of a scene will take place. “In short, production becomes just the first stage of postproduction”. For Star Wars: Episode 1 – The Phantom Menace (1999), traditional on-set filming took only sixty-five days. However, postproduction “stretched over two years, since ninety-five percent of the film (approximately two thousand shots out of the total 2,200) was constructed on a computer”. The soundtrack, too, gained increasing postproduction attention. Co-supervising sound editor Tom Bellfort estimates that 60% to 70% of the finished dialogue was looped (i.e., recorded separately from and after the visuals during post-production, and subsequently matched up with the image track). In addition, sound effects were painstakingly crafted for various characters, machines, and objects, as images have come to be re-sounded in far more detail and with more meticulousness. Digital technology has enabled a formerly unimaginable degree of attention to the soundtrack, where for

instance, types of sounds could be assigned to individual sound effects editors to work on.50

In 1999, The Matrix was to signal another apotheosis in the evolution of visual special effects, with its signature ‘bullet-time’ sequences. Bullet-time allows the filmic illusion whereby the passage of time is dramatically slowed or frozen in order to allow a viewer to observe imperceptibly fast events (such as flying bullets). In a well-known sequence, as time slows to a crawl, we can observe Neo arc his body in seemingly impossible positions to dodge bullets being fired at him, which come towards him in tantalising slow-motion, trailing slipstreams in their wake. As Pierson describes, it is “a technique used for creating the illusion of movement around objects in the foreground. In defiance of the natural laws of cinematography, elements of the filming could be made to appear to be moving at different speeds, with higher-speed camera moves miraculously tracking slow-motion action scenes”.51 The soundtrack, too, is intricately designed. The Matrix sound designer Dane Davis has talked about the sound design for the climatic scene “toward the end where Trinity kisses Neo while a battle is raging around them”. The romantic intensity had to be maintained “without losing the dramatic tension of the background conflict”. The solution was to transition “the full-on attack into a surreal, deep metallic booming like cannons in the distance while occasionally bringing some mid-range frequencies back in when Trinity pauses in the kiss. The scene was built very carefully in terms of where all these resonating metal hits are positioned throughout the action and it let the intimacy of that crucial kiss build while the battle continued”.52

50 Yewdall, 1999, p. 162.
51 Pierson, 2002, p. 163.
we can see is a development, popularisation, and adoption of a digital aesthetic for both image track and soundtrack.\(^{53}\)

In relation to the visual image, it has become “possible to generate photorealistic scenes entirely on a computer using 3-D computer animation; modify individual scenes or whole scenes with the help of a digital paint program; cut, bend, stretch, and stitch digitised film images into something with perfect photographic credibility, even though it was never actually filmed”.\(^{54}\) The “mutability of digital data” has redefined moving-image culture by problematizing the value of cinema recordings as documents of reality; “cinematic realism is being displaced from the dominant mode to merely one option among many”.\(^{55}\) The predominance of digital image production has enabled virtually any image to become possible, and hence the connection of images to solid reality has become tenuous. As simulacral images acquire a hyperreal life and dynamism, we become more sceptical about the image’s truth value (even as we increasingly consume these hyperreal commodities) as new possibilities emerge for (soundful) images to be created and constructed.

Manovich asserts that due to computer culture, traditional montage – a time-based mosaic of different shots – “is no longer the dominant aesthetic, as it was throughout the twentieth century, from the avant-garde of the 1920s up until the postmodernism of the 1980s”. The aesthetic of montage, which “aims to create visual, stylistic, semantic, and emotional dissonance between different elements”, has been superseded by the aesthetic of compositing, which “aims to blend [different elements] into a seamless whole, a single

\(^{53}\) Of course, this should not be taken to apply to all genres and models of cinema.

\(^{54}\) Manovich, 2001, p. 295.

gestalt”. These “elements are not juxtaposed but blended, their boundaries erased rather than foregrounded”. Manovich goes on to add that digital compositing, “in which different spaces are combined into a single seamless virtual space, is a good example of the alternative aesthetics of continuity”. Manovich remarks that “this new key technique redefines our concept of a moving image”. For example, Cubitt comments that filmic space has been restructured, where the “classical practice of holding an open line of sight into a scene into which a new character can enter becomes, in the digital, the structuring of space as a wholly navigable virtual map or virtual architecture, in which no space is empty that is not already filled (with smoke and digital ‘fog’ to create depth and concretise light, and with real and virtual objects and their trajectories) or destined to be filled”. The computer era has introduced a new paradigm for the image and montage, where “the problem is no longer how to generate convincing individual images but how to blend them together”. This problem has increasingly been overcome in terms of successfully generating synthetic images – and sound needs to be recognized as playing a key role in terms of anchoring the flow of these digital images.

Manovich comments: “common opinion still holds that synthetic 3-D images generated by computer graphics are not yet (or perhaps will never be) as ‘realistic’ in rendering visual reality as images obtained through a photographic lens”. He suggests that “this common opinion is mistaken. Such synthetic photographs are already more ‘realistic’ than traditional photographs. In fact, they are too real”. The problem with typical images produced with 3-D computer graphics is that they “still appear unnaturally

---

60 Manovich, 2001, p. 199.
clean, sharp, and geometric looking. Their limitations especially stand out when juxtaposed with a normal photograph”. The issue, then, with computer graphics is that they are too perfect; and Manovich points out that “one of the landmark achievements of Jurassic Park was the seamless integration of film footage of real scenes with computer-simulated objects”. In order to accomplish this integration, “computer-generated images had to be degraded; their perfection had to be diluted to match the imperfection of film’s graininess”. In other words, the ‘too real’ nature of the digital presents a hyper-clean digital simulacrum that needs to be ‘noised up’ in order to be integrated into or accepted by our sense of the real. Manovich here is talking about visual degradation of the images, but he does not directly acknowledge the role of the soundtrack in anchoring these ‘too real’ images. The soundtrack can add the requisite degree of noise, dimensionality, presence, depth, rhythm, and flow to the hyper-clean images.

For The Phantom Menace, when the characters are shot up or blown apart, the sounds heard by the audience come from various old vacuum cleaners, jacks, drills and heavy pieces of iron. When the audience looks at the characters touching themselves or each other, the sounds that are heard come from Foley artists using pineapples or coconuts or even cantaloupes to substitute for skin surfaces; these sounds are, again, subsequently matched up with the appropriate image track during post-production. For the C-3PO character, many of the movement sounds are made by a large stainless steel salad bowl with wires, while the tread of the character’s footsteps is obtained by recording the Foley artist walking on various surfaces in the Foley studio whilst wearing

---

old-fashioned, 18-inch-long kids’ skis. The humming tone emitted by the lightsaber is a holdover from the original movie, where it was made by combining the sounds from the motor of an old 35mm movie projector and a sputtering TV picture tube. To simulate a swinging lightsaber in a duel, sound designer Ben Burtt played the original sound over a speaker, whipped a microphone past the speaker, and rerecorded the resulting whish. “You get a big Doppler shift in the sound, as if it's a sword swinging through air”, Burtt remarks.

The aesthetic of sound, of course, has been significantly shaped by the shift to digital, which has had a significant impact on the evolution of hyperreal sound. Since 1989, the most revolutionary aspects of sound production have been in the mechanics of how technicians record, edit, and mix the final soundtrack – namely, digital technology. As David Yewdall affirms, “This has changed the sound process more profoundly than nearly any other innovation in the past fifty years”. This has also impacted the soundful image. Digital sound plays a significant role in reinforcing the allure of spectacular images. Some advantages of digital storage of sound signals are extremely wide dynamic range and freedom from noise, very wide and flat frequency response, and very low distortion. Digital sound enables better definition: in technical terms, improved acuity and precision in rendering of detail. Michel Chion has described the notion of fidelity as “a purely commercial one”, which “corresponds to nothing precise or verifiable”; and in contemporary terms, “definition is (mistakenly) taken as proof of fidelity, where it’s not

64 Yewdall, Practical Art of Motion Picture Sound, 1999, p. 23
being confused with fidelity itself”. In other words, high definition sound has now come to be equated with fidelity. Definition is a function of the width of the frequency band (which allows us to hear frequencies all the way from extreme low to extreme high) as well as its dynamic range (amplitude of contrasts, from the weakest levels to the strongest). Digital sound has progressed in definition particularly through gains in high frequencies; “high frequencies reveal a new multitude of details and information, contributing to an effect of greater presence and realism”. Digital audio has been heralded as the last word in quality, the ultimate state-of-the-art in sound reproduction – beyond which there can be no improvement. This, of course, is not correct. “While digital systems do offer much reduced noise and distortion compared to analogue systems, they still do have noise and distortion. The types of distortion are quite different from those of analogue systems, and their objective measurement and subjective evaluation are difficult”. As Michel Chion has written, “The most highly perfected digital recordings are certainly quantitatively richer in detail than those of yesteryear, but they are no less coloured by the technical process – perhaps even more so”. Digital technology certainly does not offer a transparent, perfect recording of an original sound event – however, in many ways digital sound provides an effect of greater presence and realism.

It is clear that digital audio systems sound different from even the best analogue systems, and many people are convinced that they sound better, offering a clearer, more powerful sound. However, listening tests have been conducted where digital and high-quality analogue master tape recordings of the same performance are played side by side.
to a panel of listeners, and in many of these tests the analogue recordings have been described as being a little more ‘musical’ and pleasant, even though they are admittedly a little more noisy. After all, the point of classical sound recording technique is clarity and purity of sound in the reproduction, a clarity that is interrupted by noise and sonic distortion in the real world. (I want to suggest that ‘noise’ also provides a sense of body, of ‘presence’, different to the effect of transparent presence offered by digital sound. Perhaps now, in the age of mechanical reproduction, we should speak in nostalgic tones of the grain of the machine, the “hear-strip” of noise that accompanies sound. Digital can be disdained as being too ‘clean’, too pristine, lacking the sense of space, noise, and context.) With the fetishization of audio equipment, recorded sound has acquired a noise-free standardisation. The increasing perfection of the signal-to-noise ratio and the clarity of digital sound means that noise is banished (or silenced) in favour of a ‘clean’ sound, where (as I have said) sounds are capable of being increasingly selected, mixed, and highlighted for our listening attention. The machining of sound and hearing, facilitated by digital technologies and aesthetics, has meant that machines now supply sounds that we pay attention to, silencing other sounds as being mere noise; in addition, certain images are now particularly soundful, relegating other subdued images to the background.

With the adoption of digital sound technologies and aesthetics, various elements of the soundtrack – ambient sound-beds, sound effects, re-sounded voices, music, Foley noises, etc. – were to play important roles in blending and animating the new digital

---

70 This is Adorno’s phrase.
71 It has been pointed out to me that Aden Evens makes a similar point in his article ‘Sound Ideas’, in *A Shock to Thought*, edited by Brian Massumi (2002, pp. 171-187).
visual images and sequences. When Dolby technology first started to make popular inroads in 1975, and was developed into various digital sound systems, the benefits could be easily summarised – more definition, and noise reduction. Sound could now be mixed loud, with ‘perfect’ presence and detail. Digital advances and new achievements in speaker design and amplifier clarity allowed sound to be brought into the theatre at ever-increasing levels. Images now have the capacity to be louder, and more resounding than ever before – and the tendency for images to be more soundful, as images are fused with hyperreal sounds, has become the normal state of affairs. Digital soundfulness has a close relationship with silence and noise. “When Dolby first appeared”, Michel Chion observed in a 1998 lecture, it was treated “as a new acoustic space that had to be filled up to the greatest possible extent to make the best use of it”. In other words, the aesthetic of ‘fullness’ and ‘sound density’ was pursued, where films (such as the original Star Wars) attempted to take complete advantage of the rich aural palette that was offered. However, it soon would “become clear that the extra space created… was space to be emptied rather than filled”, and “it was quickly understood that what had been created was a new sound space to empty, a new silence. This new silence, which reigns around isolated words or sounds, imparts a particular new intensity to certain scenes”.72 This new silence, obtained by removing or suppressing (or silencing) unwanted background noise, imparts an added gravitas to the sounds that are presented, which are granted a new emphasis, along with the visuals. In his book The Voice in Cinema, Chion notes that before the developments of modern sound technologies and practices, “every sound and every silence in a film was embedded in a continuous background tone that provided a sort of

sonic continuum. Since Dolby increases dynamic contrast, it makes silence deeper, and from these silences the voice emerges differently.\(^73\) The voice, too, has been transformed by the transformations in noise and silence provoked by machined sound and by the machined fusion of sound and image.

In line with the adoption of hyperreal sound as the general template for a new model of sound, sound recording techniques have continued to emphasise close miking, with emphasis placed on intelligibility, clarity, and foregroundedness of the sound – ‘proper’ spatial perspective is largely disregarded. As a result, contemporary instances of recorded sound have tended to be spatio-temporally unspecific, legible beyond the original context, a permanent record of an ephemeral event. This is where it can be said that the notion of the writing of sound maintains a lasting influence – where a sound event has become comparable to a species of quotation, a transcription removed from its original context. The establishment of this recording style emphasises the recognisability and function of sounds outside of a specific context. Like the quotation, however, such sounds threaten to float free of the specific contexts within which they initially occurred and functioned. These disembodied sounds, which then possess the ghostly trace of an absent presence, infused with supplementarity and the echoes of their ‘original’ contexts, are then ‘reconnected’, recontextualised, or re-embodied (and received by an auditor within a particular situation, as a new sound event).

Today, the voice – close-miked, machined, and highlighted for our attention – is more soundful, more intimate, more forcefully persuasive, and more compelling; it possesses more presence. The voice has become clearly hyperreal, even when it is not

explicitly dubbed. The re-sounded voice is wrapped up in a cocoon of silence; typically, no other sounds or noises should disturb the sanctity of the voice (mix other sounds or noises into the background, or make sure you foreground the voice), as the re-sounded voice commands silence in its wake. As Chion points out:

Dolby helps to give a direct, close, and palpably physical presence to the voice, entirely changing the way we perceive it. More generally, it focuses finer attention on vocal texture, subtle variations of timbre, vibration of vocal cords, resonances. Multitrack sound helps to situate the voice in a more precise relation to other sounds that may be spread out in various directions in space.  

Via the machined fusion of sound and image, strong links have been created between the voice and the body, the voice and face. The recorded/replayed voice is no longer an isolated sound, no longer quite disembodied; the voice has been amplified and rendered more soundful, speaking from loudspeakers and also from screens, resounding with a clear and strong timbre, coming from images, re-embodied by referring to the image to which the re-sounded voice ‘naturally’ emanates. Voices and sounds are now re-embodied by images (are made by images), even as images now are given voices, produce sounds and noises, and are infused with music.

Curiously, though, sound effects maintain an enduring connection to analogue sources, much more so than visuals. Cubitt points out that mixing a film soundtrack typically involves both digital and analogue recording, as well as the manipulation, layering, and mastering of a large score of tracks and elements. However, “the connection with the analogue world, and the ability to hear with astute and creative ears the attack, mass, and decay of sounds as well as their potential credibility and emotional impact, are

---

74 Chion, 1999, p. 166.
highly prized skills in the world of film sound. Relatively few sounds, even musical sounds, are digitally originated in Hollywood". Tom Kenny talks about his infamous source of the gloop sound effect for T2’s T-1000 morphing transformations: “What’s amazing to me is… Industrial Light & Magic using millions of dollars of high-tech digital equipment and computers to come up with the visuals, and meanwhile I’m inverting a dog food can”. Analogue sources conveniently provide the organic noise and presence of base sounds, which can then typically then be later digitally manipulated in order to lend a distinctive soundfulness to anchor, animate, and give life and rhythm to moving images.

This resounding soundfulness, of course, can go too far. To paraphrase Baudrillard, the accusation can be laid that the contemporary film soundtrack becomes prostituted to the alluring audiovisual spectacle enabled by digital sound-image assemblages. As the rerecording mixer Richard Portman complains, “The pictures these days are overbuilt, there’s too much sound – they have sounds when there’s no need for it”. Yewdall cautions, “With so many technological goodies in postproduction and such powerful sound theatre sound systems, some movies are just too loud for too long. They lose their impact because the audience becomes numbed”. With sound employed for the mere sake of soundfulness, the shock and resonance of sounds infusing moving images has generally been lost. This can be related to one of Baudrillard’s few (typically

---

75 Cubitt, 2004, p. 266.
76 Kenny, 1991, p. 64.
77 “The circumstances in which computer-generated images are produced for the Hollywood cinema”, Michele Pierson says, “usually involve some sort of tendering process by whereby special effects companies bid for contracts in an extremely competitive market”. This has contributed to “cultural and technological pressures that have led to the production of imagery that increasingly looks the same”. (Special Effects, 2002, p. 157.) Unfortunately, soundtracks, sound effects, and film scores are also increasingly being assembled along production-line values.
78 In LoBrutto, Sound-On-Film, 1994, p. 44.
provocative) comments about sound, where he has claimed that stereo represents only stereo itself, as a hyperreal simulation of the real. Stereo does not represent sound; it represents only the stereo reproduction. There is a limit point – a ‘threshold’ – beyond which the increasing perfection of stereo, instead of bringing music closer, actually drives it further away.\(^{80}\) With the implosion of medium and message, Baudrillard says, “the medium itself is no longer identifiable as such”.\(^{81}\) Instead of listening to the sound, we are instead more intent on listening to and judging the quality of its stereo reproduction. Music disappears in its technical perfection – now, we come to listen to the sound rather than the song.\(^{82}\) In Kellner’s analysis, Baudrillard “suggests that in a situation of technological perfection, entities will cease being what they were previously”.\(^{83}\)

At this stage, I want to turn to the work of Paul Virilio in order to productively extend this line of discussion, which can be conceived in terms of the relationship between sound, silence, and the act of silencing. Virilio is perhaps best known as a thinker of speed, describing his theoretical project as being concerned with dromology, or the science of speed. For Virilio, speed is not itself a phenomenon, but rather a relation between phenomena.\(^{84}\) “In effect”, he states, “speed does not solely permit us to move more easily, above all it permits us to see, to hear, to perceive and thus to conceive more intensively the present world”.\(^{85}\) Virilio argues that the accelerated speeds of transmission and communication enabled by modern technologies led to a loss of immediate presence and a diminution of lived embodied experience. For Virilio, the ability to traverse the

---

\(^{85}\) *The Open Sky*, 1997, p. 12.
world at high speed, either via the accelerated rapidity of transports or in the instantaneity of telecommunication, is viewed in terms of a negation of the space, volume, or extension of the world. Virilio argues that dromology has a negative impact on the shaping of perception and of social or political space. He argues: “With acceleration there is no more here and there, only the mental confusion of near and far, present and future, real and unreal – a mix of history, stories, and the hallucinatory utopia of communication technologies”.86

In a short essay titled ‘Silence on Trial’, Virilio speaks of the “sonorization of the image” with regards to the death of art.87 My intention here is to expand and elaborate on Virilio’s formulation of the sonorization of the image – images are no longer silent, as they have been colonised by sounds. In an age where recorded sound (amplified, manipulated, and played back) enables us to hear everything – there is no such thing as silence, for even silence is a constructed hyperreal artefact. Sound film has shaped a new type of silence (by interrupting certain sounds, by suppressing noises) as well as made silence impossible (now, even images and things speak, and often in voices not their own). With the emergence of the talkies and sound film, “the idea of accepting the absence of words or phrases, of some kind of dialogue, became unthinkable”.88 Hearing and vision had to be synchronised; images have to be heard. Virilio makes his conclusion perfectly clear: “What can you say about the ‘talkies’ and about the sound-track that puts the finishing touches on the effect of mastery of the image track, except that they are a lot

86 The Art of the Motor, 1995, p35.
87 Armitage explains that “for Virilio, present-day sound art obliterates the character of visual art while concurrently advancing the communication practices of the global advertising industry, which have assaulted the art world to such a degree that it is at present the central dogma of the multimedia academy”. (Armitage, ‘Art and Fear: An Introduction’, in Virilio, Art and Fear, 2003, p. 15.)
more harmful than people realise?" What does hyperreal sound silence, as it supports the now-familiar impact of an audiovisual spectacular? To borrow Baudrillard’s formulation, hyperreal sound silences and murders the real, in a perfect crime. Silence is a voice, says Virilio, which “was thwarted, and definitively... by the arrival of the talkie – not by the arrival of the cinema, but by the talking cinema”. Images speak, and no longer are silent; indeed, images are no longer allowed to be silent.

Silence has become an unacceptable disturbance in the proper operation of soundful images. Virilio has claimed, “Today it is the spoken word which is logically withering away before the instantaneity of the real-time image”. (One might assume that he does not agree with Walter Ong’s hopes regarding secondary orality.) “Silence no longer has a voice. It lost its voice half a century ago” – Virilio here is referring to the advent of the talkies, of sound film, and of television – silence is no longer accepted, as silence itself has been silenced; “what is now regarded as obscene is not so much the image as the sound – or rather, the lack of sound”. Virilio is concerned with “a censorship of silence in an age awash with the obscenity of noise”. Virilio’s argument about silence is founded on this idea that media presentation comes to substitute for immediate reality – images speak, and this mediated soundfulness acts to silence the voice of the person in the street, to overwhelm the quiet of immediate contemplation. Sound film has become incredibly noisy (louder, more aurally spectacular, in the quest for dimensionality, presence, and attention). Virilio writes, “Once you have the talkies up and running, you can get walls, any old animated image whatever to talk. The dead too,
though, and *all who remain silent*. And not just people or beings, either, but things to boot!"⁹⁴ (Siegfried Kracauer has cited Cohen-Séat: “And I? says the leaf which is falling. – And we? say the orange peel, the gust of wind… Film, whether intentionally or not, is their mouthpiece”.⁹⁵) Sound film has provided reinforced sounds and alluring soundscapes. Sound film has provided soundful images which act as templates for audiovisual perception. Sound film is a *lure* for our attention. Sound film has supplied a sound-image repertoire with which to stock our memories. Sound film can be interpreted as a symptom of the clamour of our contemporary times. What is the wider impact of soundful images? How have sounds – facilitated by this invention called sound film – come to infuse images?

In our age of secondary orality/aurality, the ‘silence’ that has historically been a concomitant of visuality can be said to be in the process of being reshaped into something more ‘noisy’. More and more of us find that we want (or need) aural distraction when we are reading, exercising, relaxing, or even trying to go to sleep. Ambient music accompanies us while we make our rounds at the supermarket; many of us publicly despise it, but many of us also miss it when it is not there. The traditional sanctuaries of ‘silence’ – libraries, museums, and art galleries – have been colonised by archive tapes, listening equipment, voice-based exhibits, and artworks which can literally speak for themselves. At the movie theatre, more and more of us accustomed to the constant electronic buzz of modern life are unwilling, or unable, to be quiet. Silence has been transformed by technology, by noise, by different organisations of sound, by cultural perception, by amplification, by mediated sound, by the clamour of modern life, by the

⁹⁴ Ibid, p. 75.
⁹⁵ Kracauer’s *Theory of Film*, 1997, p. 45.
machined fusion of sound and image. Paradoxically, people have also learned to deny their attention to what they are hearing even while listening to it. ‘Pure’, ‘natural’ sound has become an ideal, a discourse that can be appealed to in order to conjure up associations of immediacy, nature, direct representation, etc. This ideal operates in conjunction with and in opposition to the construct of pristine, hyperreal, high-definition digital sound – hyperreal sound invokes this ideal by operating as its binary opposite.

Virilio writes of the “digital technology that has now sunk its teeth into the whole array of artistic disciplines, from the taking of photographs to the capturing of sounds”. 96 “Digital sound is aesthetic”, says Eisenberg. “Seamless, zipless, it does not need our attention. It lets us drift off”. 97 With the proliferation and spread of machined (recorded/manipulated/transmitted/replayed) sound, we are now unavoidably immersed in a contemporary acoustic environment of mass-mediated sound; with the sonorization of the image, mass-mediated sounds and voices are even clamouring to speak for images, to comment on images. As I have previously noted, Walter Murch believes that the re-association of sound and image “should stretch the relationship of sound to image wherever possible. It should strive to create a purposeful and fruitful tension between what is on the screen and what is kindled in the mind of the audience”. However, with present-day cinema, the tendency and danger is to “suffocate its subjects by its very ability to represent them” 98 – what this means is that because film appears to be ‘complete’, there is no room left for ambiguity, for imagination; the “excess of cinema” 99 provides hyperreal images and sounds that are passively absorbed by the audience.

---

97 Eisenberg, 2005, p. 212.
Sounds have come to extend, reinforce, and stabilise images in certain ways; the reassuring perfection and interpenetration of images and sounds have become established aspects of our everyday reality. Murch has commented, “The problem is that if you record the actual sound that goes with that space, it has nothing to do with the emotion of being there”.100 ‘Real’ sound becomes ignored and denigrated as noise if it does not ‘perform’ or signify (re-enact) a ‘significant’ event. Hyperreal sound, in its clarity and persuasiveness, has become the accepted norm. As Yewdall explains:

I can play you a recording of a rifle butt impacting the chin of a man’s skull. It is an absolutely authentic recording, yet it sounds flat and lifeless. On the other hand, I can edit three carefully chosen sound cues together in parallel and play them as one, and I promise that you will cringe and double over in empathetic pain.101

Virilio suggests that the special effects of the “new popular cinema” possess an “instantaneity and interactivity” which implies that the spectator is “no longer being provided with any kind of verisimilitude, but fed exclusively on the exhibition of accidents”.102 Of course, here he is speaking of visual effects, but his comment can also be applied to how images have become reinforced by sound. As connoisseurs of sounds as well as collectors of images, what we consume is less the ‘thing itself’ than an alluring and spectacular sound-image event that in a way becomes ‘real’ for us.

The power of images has been reinforced by mediated sound. Soundful images have acted to silence the masses, “who have lost their tongues”103 – for images speak, and only images – as the “art of speaking has bowed out before the ‘talking’ cinema and the

---

101 Yewdall, Practical Art of Motion Picture Sound, 1999, p. 6.
oratorical power of political tribune has been defeated by media culture. From now on, what speaks is the image – any image, from billboard images to images at home on the box.\textsuperscript{104} Hyperreal sound acts to silence through processes of selectivity and transformation and by providing a compelling realer-than-real spectacle that mediates the way we perceive other sounds and noises. It also acts to idealise and transform and present what was apparently silenced. Hyperreal sound modifies our conception of the past, just as it transfigures our present and future by establishing new relationships, memories, narratives, and dialogues. This ‘problem’, of course, is in addition to the more widely known discussions of how images have accumulated and proliferated, circulating through a profusion of screens. Virilio muses darkly, “Machine for seeing, machine for hearing… Media presentation dominates everywhere you turn”.\textsuperscript{105} It can be said that we have undergone a bombardment of images which have now superseded words in terms of their influence. This revolution, of course, has proceeded apace, as images are now soundful, resonating with sounds, noises, and voices. Virilio has cynically described it as “a ventriloquist’s art”.\textsuperscript{106}

This description can be productively related back to Chion, who tells the story of an anchorman who is commentating on a live television transmission of an air show. Forced to improvise, at one stage the anchor remarks, ‘Here are three small airplanes’ – and indeed, the image that is seen is of three tiny airplanes against a blue sky. Chion’s point is that the anchor’s commentary is not redundant. While he appears to be stating the obvious in terms of what we see on the screen, he is also determining how we see what is

\begin{itemize}
\item \textsuperscript{104} Ibid, p. 86.
\item \textsuperscript{105} Virilio, ‘Silence on Trial’, in \textit{Art and Fear}, 2003, p. 77, 82.
\item \textsuperscript{106} Virilio, \textit{The Accident of Art}, 2005, p. 38.
\end{itemize}
on the screen. If the anchor had said, ‘The weather is magnificent today, then “that’s what we would have seen in the image, where there are in fact no clouds”. In other words, the redundancy of the commentary “is illusory, since in each case these statements would have guided and structured our vision so that we would have seen them ‘naturally’ in the image”. The words that we hear provide added value to the images, rigorously framing what we see. Chion concludes: “if the film or TV image seems to ‘speak’ for itself, it is actually a ventriloquist’s speech”. The image of the three small airplanes in a blue sky therefore can be taken to be “a puppet animated by the anchorman’s voice”.107 The image, of course, does not offer a simple depiction of reality; the image does not really speak for itself. Images, rather, offer a particular mediated perspective upon the world,108 a perspective that is nevertheless understood through the discourse of the realistic depiction of reality.109 As Žižek observes, “the ideological struggle is fought out not only at the level of arguments but also at the level of images”.110 The point to emphasise here is that images are animated, shaped, and framed by voices, music, and sounds – which operate to anchor the flow of images by subliminally (or even blatantly) informing the audience just how the visual narrative ought to be interpreted and received.

108 By the late twentieth century, a range of contemporary critics had effectively problematized and refuted the realist paradigm around the image. Umberto Eco, for instance, rejected Barthes’ notion of a ‘pure image’ in favour of a semiotic analysis, firmly declaring that “everything which in images appears to us still as analogical, continuous, non-concrete, motivated, natural, and therefore ‘irrational’, is simply something which, in our present state of knowledge and operational capacities, we have not yet succeeded in reducing to the discrete, the digital, the purely differential”. (Eco, ‘Critique of the Image’, in Burgin, Thinking Photography, 1982, p. 34; Eco then proceeds to specify no fewer than ten categories of codes that can be applied to the photographic message, which should no longer lay claim to being a simple reproduction of ‘the real’.)
109 For the moving image, elements of montage, methods of framing, shot selection, lighting, etc., work to shape particular processes of making meaning, influencing specific ways of seeing, and providing a specific reading or interpretation of events.
These images are infused with voices, sounds, music – today, images cannot be silent, but are required to speak out in resounding fashion.

Virilio’s position can be read as an account of how sounds are now operating to hold images open in specific ways, as sounds now finalise images by fixing or freezing a particular sound-image articulation. However, this reification should not be understood as complete. While sound has typically come to be utilised to add a stable, standardised, predictable dimensionality and presence to the image, sound still possesses the potential to productively extend the image. As Siegfried Kracauer wrote several decades ago: “Supposing shrill screams or the blasts of an explosion are synchronised with images of their source and/or its environment: much as they will leave their imprint on the spectator’s mind… they may prompt him to scrutinise the [images] in a mood which increases his susceptibility to their multiple meanings”.\footnote{Kracauer, *Theory of Film*, 1997, p. 127-128; originally published in 1960.} For Kracauer, the true potential of the cinema, which has been “so greatly increased by technical innovations and scientific discoveries”, lies in its capability “to open up new, hitherto unsuspected dimensions of reality”.\footnote{Ibid, p. 8; Kracauer has been often unfairly criticized as a naïve realist. However, as Miriam Hansen points out, “Kracauer’s concern is not with authenticity or verisimilitude but rather with film’s ability to discover and articulate materiality”. (Hansen, ‘Introduction’ to Kracauer’s *Theory of Film*, 1997, p. xvii.)} Films “alienate our environment in exposing it”.\footnote{Ibid, p. 55.} Film reveals and defamiliarises the flow of life. “Intimate faces, streets we walk day by day, the house we live in – all these things are part of us like our skin, and because we know them by heart we do not know them with the eye”.\footnote{Ibid, p. 55.} Kracauer observed that: “The motion picture camera has a way of disintegrating familiar objects and bringing to the fore – often just in
moving about – previously invisible interrelationships”¹¹⁵ But I argue that sound also possesses this capability, engaging in a sophisticated and evolving articulation with the image, working to spark associations, revealing “previously invisible interrelationships”, prompting “memories of the senses” and “cataracts of indistinct fantasies and inchoate thoughts”.¹¹⁶ Sound retains the potential to bring images to life.

In summary, machined sounds and images are increasingly being fused together, and are creating events that operate as models of audiovisual (and perceptual) reality. And with the acceleration of media technologies, we can observe the continuing force of “this power of the image that eclipses reality”¹¹⁷ – even as images now increasingly become soundful, and resonant in new ways. Hyperreal sound, infused with transparent presence, filled with ‘perfect’ detail and clarity, can act to allow images to ‘speak’ resoundingly, or it can act to silence in a number of ways. It can silence by relegating other sounds and voices, which are judged to be unimportant and/or extraneous, pushing these other unimportant, unwanted sounds to the background. Audiences come to expect certain relationships between sound and image; certain ‘natural’ sounds are forgotten (or silenced), and new machined relationships are naturalised in a process of constructed realism. As new sound-image juxtapositions and couplings which were once strange and defamiliarising become familiar, frequently used soundful images will become naturalised, ‘unproblematic’ invocations of reality, thereby losing their power of estrangement, and providing hyperreal models of reality. Re-embodied sounds and re-sounded images are acknowledged as ‘real’ constructions, portraying ‘natural’ sound-

¹¹⁵ Ibid, p. 54.
¹¹⁶ These phrases are borrowed from Kracauer, Theory of Film, 1997, p. 165-166.
¹¹⁷ Solnit, 2003, p. 156.
image events. These soundful images are, of course, media representations which provide
us with signs and symbols that allow us to build up shared understandings and common
codes for interpreting the world. Representations begin with language – the primary
medium – but they extend into all other sorts of communication practices, including
audiovisual forms, which would clearly seem to be proliferating, becoming increasingly
influential in our contemporary age. Contemporary audiences have grown very familiar
with the modes and conventions of sound-image representations, perceptions, and events.
Here again we must confront the antinomy present within soundful images: it might
never be a ‘seamless match’, and there might always be an originary absence at the
juncture of sound-image events, but these hyperreal events, that construct and refer to
‘original’ events that never happened – these soundful images are beginning to insinuate
and reinforce a convincing illusion of presence, propagating what can be called ‘sincere
fictions’ that we, the audience, have come to willingly partake in.
Chapter 8: An audiovisual imaginary

This chapter argues that as mass-mediated sounds and images and soundful images persuasively affect the way we see, hear, and make sense of events around us, the stockpiling of sounds and images in our memory has resulted in an audiovisual imaginary. It is clear that as images can now speak to us in voices not necessarily their own, our audiovisual age’s moving-image culture is not just about how we have come to see in different ways. Theorists have commonly spoken of the imperialism of the image – now, however, it might be more precise to speak of an imperialism of sound and image. We live in a dense audiovisual environment, where images and sounds overlay one another, interacting, blending, and bumping up against the other, a world filled with resounding images, re-embodied sounds, soundful images, which have affected our engagement with the world through the new language of images and sounds. In examining the notion of an audiovisual imaginary (where soundful images are impacting upon the way we see, hear, and make sense of the world), this chapter focuses on the work of Gilles Deleuze. However, in order to appreciate Deleuze’s contribution, it is worth returning again briefly to Baudrillard and Virilio, as Deleuze offers a productive counterbalance to their perspectives. While Deleuze’s cinema books have typically been read in terms of their insight on visual images, it is important to recognise that for Deleuze the modern image is audiovisual, founded in many respects on the power of a disjunction of sound and sight. I wish to suggest that Deleuze offers another means of conceptualising the resonance between sounds and images.

In order to think about how mass-mediated sounds and images and soundful images are impacting on the way we see, hear, and make sense of events around us, a
fundamental insight is provided by the philosopher Charles Sanders Peirce. He advocated that we should think of our minds as signs mixed with mortal life, as he insisted that all thought must necessarily be in signs – Peirce observed, “only by external facts can thought be known at all. The only thought, then, which can possibly be cognised, is thought in signs. But thought which cannot be cognised does not exist. All thought, therefore, must necessarily be in signs”.¹ Peirce comments that: “In fact, therefore, men and words reciprocally educate each other; each increase in a man’s information involves, and is involved by, a corresponding increase of a word’s information”.² Signs allow the formation and articulation of concepts and ideas which enable us to experience, think about, and understand the world in particular ways. Signs have changed from Peirce’s time, and now are no longer predominantly in the form of the written or spoken word. Rather, signs now increasingly take the form of machined sounds and images: in this age, people and sound-images “reciprocally educate” each other, as sound-images shape how we experience, think about, and comprehend the world, and as we infuse sound-images with social and cultural meanings. Peirce’s theory of signs is of particular relevance to an age when intelligence can be stored in the media, where the media produce and circulate signs for our reception, which then in turn shape the way we view our reality.

This conceptual point by Peirce allows for a fuller appreciation of Paul Virilio’s argument that the cinematic image has come to shape a transformation of perception and a virtualization of experience. (Deleuze also draws heavily on Peirce’s semiotic model in

his cinema books.) For Virilio, the advent of cinema has marked the beginning of a generalised transformation of collective ways of seeing and representing. Virilio talks about the “appearance of the motor” – where motorised film projection emerges as a technology of viewing, and also disseminates an increasingly influential mode of perception and experience. Cinema and its audiovisual descendents, such as television, Virilio argues, have saturated our collective social and cultural experience, and these media have had a significant impact on our general habits of apprehending and conceiving reality. Virilio asks, “How can we resist this deluge of visual and audiovisual sequences… that endlessly bombard our imagination?”

Fredric Jameson has written that we have become accustomed to “the transformation of reality into images”. Images speak to us, and soundful images have inevitably affected the way we see, hear, and make sense of events around us. It is no exaggeration to say that “we are all part of a moving-image culture and we live cinematic and electronic lives”. Indeed, cinematic representations and cultures have had an immense impact on the ways we interact with and perceive the world. Our frequent encounters with cinematic and televisual discourses affect us in the most profound, socially pervasive, and yet personal way; cinematic media have contributed to shaping a significant evolution of our modes of temporal and spatial consciousness, as well as transfiguring our embodied sense of subjectivity and existential ‘presence’. We are utterly absorbed, even if (or perhaps especially if) soundful images pass by without being directly noticed.

---

As I have contended, our audiovisual age’s ‘moving-image culture’ is not just about how we have come to see in different ways. Rather, we live in a dense audiovisual environment, a world filled with resounding images, re-embodied sounds, soundful images, which have affected our “cognitive mapping”7 of the world through the new language of images and sounds. The question of the soundful image's fidelity to the real avoids the real point; soundful images now provide the template against which to comprehend the reality of lived experience. “In sum”, Virilio argues, “as they have mediated our engagement with the world, with others, and with ourselves, cinematic and electronic technologies have transformed us so that we currently see, sense, and make sense of ourselves as quite other as than we were before them”.8 Our audiovisual culture has led to a stockpiling of sounds and images in our memory; I am proposing that it is productive to describe this in terms of an audiovisual imaginary.

Our memories have been colonised by images that are no longer silent, sounds of images, and soundful images. In his essay, ‘The Work of Art in the Age of Mechanical Reproduction’, Walter Benjamin cites Paul Valéry: “Just as water, gas, and electricity are brought into our houses from far off to satisfy our needs in response to a minimal effort, so we shall be supplied with visual or auditory images, which will appear and disappear at a simple movement of the hand”.9 Jameson has declared that: “No society, indeed, has ever been saturated with signs and messages like this one”.10 Benjamin also cites Duhamel: “I can no longer think what I want to think. My thoughts have been replaced

---

7 I have borrowed this term from Fredric Jameson; however, I will be utilizing this phrase with a different emphasis.
8 Sobchack, ‘The Scene of the Screen’, p. 69.
by moving images". We perceive things only by retaining the memory of past perceptions and anticipating and connecting future perceptions. The past and the present (and even the future) have become nothing but a vast collection of (soundful) images, a multitudinous photographic (and audiovisual) simulacrum. Virilio has commented that our memories have been “turned into junkshops – great dumps of images of all kinds and origins, used and shop-soiled symbols, piled up any old how”. After all, Benjamin observes, “history decays into images”.

As André Bazin explains, “There is nothing to prevent us from identifying ourselves with the moving world before us, which becomes the world”. We are compelled to watch, for “[t]he image is there, in front of me, for me”: according to Roland Barthes, the image (including the sound) is a perfect lure. As Barthes says, “the image captivates me, captures me”. The spectator, as Jean-Louis Baudry also claimed, is “chained, captured, or captivated” by the spectacle projected on the screen in front of him or her in the darkened room. The cinema screen enabled audiences to take a journey through different spaces without leaving their seats; as film historian Anne Friedberg puts it, it created “a mobilised virtual gaze”. Virilio says that there has been a “generalised rise of pictures and sounds in the static vehicles of the audio-visual… From now on everything will happen without us moving, without us even having to set out”. Jameson emphasises: “movies are a physical experience, and are remembered as such,

---

16 Baudry, ‘The Ideological Effects of the Basic Cinematographic Apparatus’, p. 44.
17 Friedberg, 1993, p. 2.
stored up in bodily synapses that evade the thinking mind”. ¹⁹ He continues by saying that: “in the seam between the day to day; the filmic images of the night before stain the morning and saturate it with half-conscious reminiscence… film is an addiction that leaves its traces in the body itself”. ²⁰ Soundful images have come to leave indelible traces upon us. What needs to be pointed out is that soundful images have managed to escape from the confines of the darkened room; their influence has spread. These soundful images captivate us, populating our memories and forming an audiovisual imaginary through which we comprehend the world around us.

The modern image has, in many ways, been described as a spectacle. In the influential 1967 book *The Society of the Spectacle* (1967), Guy Debord argued that we are now presented with “an immense accumulation of spectacles”. ²¹ Michel de Certeau has echoed Debord’s rhetoric about our society of the spectacle, in arguing that “from TV to newspapers, from advertising to all sorts of mercantile epiphanies, our society is characterised by a cancerous growth of vision, measuring everything by its ability to show or be shown and transmuting communication into a visual journey”. ²² Nevertheless, this ‘visual journey’ is not silent. To reiterate, the spectacle is not just visual; the spectacle is audiovisual. As Virilio notes, we now live in an era “when the audio-visible of the mass media reigns”. ²³ After all, an audiovisual spectacular has become commonplace. Media theorists such as Paul Virilio have come to conceive of the mass

²⁰ Ibid, p. 2.
²¹ Guy Debord, *The Society of the Spectacle*, 1994, thesis 1; the book was originally published in French in November 1967 by the Paris publishers Buchet-Chastel. Debord elaborates: “The spectacle is not a collection of images; rather, it is a social relationship between people that is mediated by images”. (1994, thesis 4.)
media apparatus itself in terms of warfare, where we are bombarded with images and sounds, or we are supplied with images and sounds through a “logistics of perception”\textsuperscript{24}.

Michel Foucault disagrees with the notion of the society of the spectacle. “Our society is not one of spectacle”, he insists, “but of surveillance”.\textsuperscript{25} However, I want to suggest that surveillance should not be understood as being mutually exclusive with the spectacle. Rather, surveillance and the spectacle are connected (becoming increasingly interrelated) as the proliferation of media spectacles are subject to our constant (even if inattentive or blasé) surveillance. In our contemporary age, where acutely sensitive microphones have multiplied, it must be said that surveillance is, of course, as much aural as visual.\textsuperscript{26} Electronic and cinematic media have effectively enabled us to watch and be watched, hear and be heard. Virilio explains that we now have “a world that is constantly ‘tele-present’ twenty-four hours a day, seven days a week”.\textsuperscript{27} The world has become a series of soundful images, under constant surveillance, subject to our voracious gaze, made into spectacle; “our society has begun to offer us the world”\textsuperscript{28} as collectable and consumable images and sounds.

\textsuperscript{24} Paul Virilio’s term: “The logistics of perception… simply denotes the supplying of cinematic images and information on film to the front-line”. (John Armitage, in \textit{Paul Virilio: From Modernism to Hypermodernism and Beyond}, edited by Armitage, 2000, p. 9.) This logistics of perception has come to involve soundful images, not just visual images.

\textsuperscript{25} Foucault, \textit{Discipline and Punish}, 1979, p. 217.

\textsuperscript{26} In his book \textit{Discipline and Punish}, Foucault examined Jeremy Bentham’s 1791 treatise on a model prison (called the Panopticon), where Bentham proposed a circular arrangement of cells visible to a jailor in a tower in its centre, who was himself hidden from the sight of the prisoners by a system of shutters. The Panopticon, of course, would be studied by Foucault as a model and metaphor for surveillance. In an interview with Foucault, Michelle Perot highlights Foucault’s focus on visuality and the gaze by pointing out that Bentham also suggested the use of tin tubes to link the chief inspector with each prisoner through sound as well as sight (‘The Eye of Power’, in \textit{Power/Knowledge: Selected Interviews and Other Writings, 1972-1977}, edited by Colin Gordon, 1980, p. 154).

\textsuperscript{27} Paul Virilio, \textit{The Information Bomb}, 2000, p. 13.

The normalisation of the audiovisual spectacle is why Baudrillard insists that we have moved beyond the society of the spectacle.\textsuperscript{29} For Baudrillard, the spectacle is no longer spectacular, but rather commonplace and diffused throughout reality, “and one can no longer even say that the medium is altered by it”.\textsuperscript{30} The spectacle is no longer extraordinary, and hence has become an everyday aspect of how we view and engage with reality. What happens when soundful images leave the darkened cinematic halls and enter into the streets, our homes, and our everyday lives? Reality is ‘furnished’ with images and sounds and concepts, increasingly influencing the ways we see and hear by shaping perceptual habits and expectations, prompting particular patterns, links, connections, and emphases in our construction of meaning. Soundful images have become absorbed into our memories and unconscious, shaping the way we perceive the world around us, and even impacting upon the rhythms of everyday life. These soundful images are, of course, being supplied to us via a mass media apparatus that is virtually omnipresent in contemporary society. As part of this media age, we are constantly being inundated in a deluge of pseudo-concrete images, fleshted out and extended by sounds; in addition, we are also bathed in an ocean of machined sounds, which connect and link to images. The machined fusion of sound and image has led to ways of seeing and hearing that have gradually become second nature.

In an age of soundful images, a new model of perception has been gradually installed. Gilles Deleuze has shown that we always already work with an image of thought or some abstract notion of life.\textsuperscript{31} Sounds also carry images – ideas, concepts,

\textsuperscript{29} “We are no longer in the society of the spectacle”, he claims (Baudrillard, ‘The Precession of Simulacra’, p. 30).
\textsuperscript{30} Ibid, p. 30.
mental representations and associations. The sonorization of the image implies that sound ‘fleshes out’ the image (i.e., grants images depth, presence, temporality, spatiality, and dimensionality), and the image embodies sound (i.e., claims sound as its natural property, frames itself as the source of the sound). Theorists have commonly spoken of the imperialism of the image – now, however, it might be more precise to speak of an imperialism of sound and image. Schafer points out: “When sound power is sufficient to create a large acoustic profile, we may speak of it, too, as imperialistic”.32 After all, the soundscape of the world has changed: “Modern man is beginning to inhabit a world with an acoustic environment radically different from any he has hitherto known”.33 Increasingly, sound is making meaning in conjunction (connection, synthesis, and/or counterpoint) with the image. Perhaps we can say that machined sound now operates to amplify the power of the image.

Soundful images have become noisier, more clamorous, and louder. Walter Benjamin, who was remarkably sensitive to the special nature of the moving image, observed that “film corresponds to profound changes in the apperceptive apparatus – changes that are experienced on an individual scale by the man in the street in big-city traffic, on a historical scale by every present-day citizen”.34 For Benjamin, film not only ushered in a new form of artistic reproduction, but also a distinctive mode of aesthetic reception. The notion of film reception as ‘shock’ was of central importance for Benjamin, who argued that the experiences of violent impact, tactility, and shock are the hallmarks of film. However, despite Benjamin’s optimistic perspective with regards to

---

33 Ibid, p. 3.
the revolutionary potential of the mass media, it is evident that the shock of the visual spectacle has largely been absorbed, made natural, commonplace. A distinctive feature of the twentieth century and the entrance into our twenty-first century has been seeing and hearing the world in terms of representations. Audiovisual representations and simulations have become more than mere entertainment and escapism. They reflect – and often construct – the beliefs and myths we have about the world we live in, shaping the way we perceive and interact with the world around us. Audiovisual mass media influence behaviour; the lines between fantasy and reality are often blurred and there is an ongoing interplay between the two. Frequent engagements with cinematic events have accustomed many of us to the shock and impact of spectacular sounds and images, which have been diffused into and have become familiar components of our mass-mediated environments. The so-called pristine clarity of digital sound can be said to reinforce a hyperreal soundfulness that acts to silence ‘natural’ noises and sounds, as images that are not soundful are no longer quite as compelling. Our perception has largely become standardised in terms of a relatively stable serialisation of sounds and images, which have come to comprise our audiovisual imaginary, through which we experience and frame our world.

Baudrillard speaks of “the disenchanted proliferation of screens and the profusion of images”, where the image has come to be more persuasive and compelling than the original object. In “the perfection of the reproduction, in the virtual rendition of the real”, we come to “witness the extermination of the real by its double”.

35 Unfortunately there is insufficient room here to productively examine Slavoj Žižek’s productive readings of film, fantasy, the symbolic, the imaginary, and the Lacanian real.
that illusion has disappeared in the course of the evolution of the moving image from silent movies to sound, to colour, to high-tech special effects. “Cinema today knows neither allusion nor illusion”, he says: “it links everything on a hyper-technical, hyper-efficient, hyper-visible level. No blanks, no ellipses, no silence, just like television, with which cinema has become increasingly assimilated… We are moving ever closer to high definition, in other words to the useless perfection of images”. Baudrillard declares: “The closer we reach absolute definition, the realist perfection of images, the more its power of illusion is lost”.37 In a 2002 newspaper article, Garry Thompson makes an incisive comment that appears to perfectly illustrate Baudrillard’s point. Thompson writes that modern movie-goers have inevitably become jaded as they are now “conditioned to expect the incredible and to accept the unbelievable”, knowing that the magic of contemporary cinema makes virtually anything possible. “Movies, having finally acquired the unlimited power of illusion, have discovered they have lost one of its greatest powers, the power to amaze”.38 We have come to judge and perceive the world through the reality of images, which has become commonplace. “It seems the reality effect”, Paul Virilio complains, “replaces immediate reality”.39 Boorstin says: “We have become so accustomed to our illusions that we mistake them for reality. We demand them. And we demand that there be always more of them, bigger and better and more

vivid. They are the world of our making: the world of the image”, where we have come to believe “in the superior vividness of a… representation to a drab original”.41

Whilst writing about surrealism, Walter Benjamin described a time when: “sound and image, image and sound interpenetrated with automatic precision and such felicity that no chink was left for the penny in the slot called ‘meaning’”.42 Benjamin’s statement can be read as uncannily prescient as a forecast of the capacity of soundful images to lapse into what Baudrillard has termed the “excess of the cinema”, where in our audiovisual age, “images, spectacles, simulations proliferate and terrorise, fascinate, and mesmerise”.44 Soundful images, resulting from the perfect fit and ‘automatic precision’ of sounds and images, have become realer-than-real – as their increasingly influential reality effect shapes, closes off, and frames how we perceive the meaning of our ‘immediate’ reality. Assorted pessimistic commentators have lamented that spectacle has come to replace substance. Imagery has become more extravagant, more decorative. We have become super-saturated with images. Music, sound effects, voices, machine sounds and noises, all have become increasingly tumultuous in clamouring for our attention. The image has been reinforced by sound.

For instance, how do we create the ‘satisfying tidy impact and “closure” of a Hollywood explosion’? The sound designer David Yewdall offers a series of careful instructions:

To achieve a unique explosion sound, try taking several recordings of the same explosion, recorded on different tape recorders using different microphone

41 Ibid, p. 204.
42 Benjamin, One-Way Street and Other Writings, 1985, p. 226.
configurations, and then blend them together. Often, one set of microphones delivers a more ‘crackly’ character to the explosion, with little low-end. Another stereo pair has a rich bottom-end boom, with little high-end. A third stereo pair may have more mid-range, with a near overtone tail to it. Putting all three stereo pairs together during the sound editorial process delivers a combination that none of the stereo pairs alone could capture.45

Perhaps we can say that sound-image relations have largely become standardised in presenting a parade of soundful images, where ‘no chink has been left for meaning’ outside the carefully crafted packaging and interpenetration of image and sound.

In an article written a few days after 9/11, Slavoj Žižek refers to The Matrix, and provides a troubling illustration of our hyperreal audiovisual imaginary, which has come to be furnished by soundful images:

[W]hen the hero (played by Keanu Reeves) awakens into ‘real reality’, he sees a desolate landscape littered with burned ruins – what remained of Chicago after a global war. The resistance leader Morpheus utters the ironic greeting: ‘Welcome to the desert of the real’. Was it not something of a similar order that took place in New York on September 11? Its citizens were introduced to the ‘desert of the real’ – to us, corrupted by Hollywood, the landscape and the shots we saw of the collapsing towers could not but remind us of the most breathtaking scenes in the catastrophe big productions.46

The terrorist attacks on New York’s World Trade Centre were followed and accompanied by the broadcast of images and sounds sparking emotional reactions among a global audience. People across the world watched and heard the images and sounds, which were largely consumed as a ‘live’ audiovisual event and spectacle from a safe distance. Typical

45 Yewdall, Practical Art of Motion Picture Sound, 1999, p. 115.
Hollywood action blockbusters have, in an ironic reversal, come to provide a model, a framework, from which to comprehend the traumatic event of September 11 – itself an occurrence which has been defined by the images and sounds that have been broadcast and circulated throughout the world. After all, as Susan Sontag remarks, “notions of reality and spectacle precisely reinforce and infuse each other”, as the spectacle becomes “the universal category through whose forms the world is seen”\textsuperscript{47} – and, of course, heard.

Baudrillard implies that film and reality are fusing: “life is cinema”, he says,\textsuperscript{48} as today “the virtual camera is in our heads”,\textsuperscript{49} investing real-life from the inside – for we have “swallowed our microphones and headsets”\textsuperscript{50}.

Perhaps one of the features that helped make the terrifying and unbelievable televiral images of a plane flying into the side of the World Trade Centre seem more real for many viewers was the poor, ‘unrealistic’ sound quality. As David Toop pointed out, “Unlike the satisfying tidy impact and ‘closure’ of a Hollywood explosion, the sounds heard on television rolling news as the towers collapsed were fragmented, seemingly boundless, chaotic, resistant to understanding, intensely painful”.\textsuperscript{51} Frequent engagements with cinematic events have accustomed many of us to the shock and impact of spectacular sounds and images, which have been diffused into our mass-mediated environments, absorbed and become commonplace. Were the images and sounds of 9/11 simply unreal, or unconvincing (like a bad movie, and as compared with a good movie, which presumably would be realer-than-real, and convincing in its unreality)? Were these images and sounds not quite as seamless, not sufficiently spectacular, not containing

\textsuperscript{51} 2004, p. 80.
enough transparent ‘presence’ as measured against a typical cinematic presentation? Were the images and sounds too painfully ‘raw’, presented in terms of flawed, imperfect shots and sounds, as opposed to the unreal perfection and hyperreality of the typical Hollywood blockbuster?

Mediated sounds – now louder, more powerful, more pristine, more detailed, fuller and more spectacular – contribute to the perfection and presence of the virtual, realer-than-real image. Mediated sound now provides an essential dimensionality and presence to the virtual, ‘perfect’ image – sound is now packaged with the image. After all: “Great sound will make the picture appear to look better and be more visually exciting”.52 Images (reinforced by mediated sounds) have proliferated and been disseminated – and have come to install a new model of ‘real’. Sounds have come to extend, reinforce, and stabilise images in certain ways; the reassuring perfection and interpenetration of images and sounds have become established aspects of our everyday reality. Images have become incredibly noisy (louder, more aurally spectacular, in the quest for dimensionality, presence, and attention). Soundful images have provided reinforced sounds and alluring soundscapes. Soundful images act as templates for audiovisual perception, interacting with image to provide a lure for our attention. Our audiovisual culture has been supplied with a sound-image repertoire, an audiovisual imaginary, with which to stock our memories.

For Virilio, a “virtual theatricalization of the real world” has taken place, as “all our perceptions, all our impressions have been distilled for us”.53 Virilio’s pessimistic proposition that media presentation acts as a substitution for immediate reality can be

---

52 Picture editor Howard Smith, in Yewdall, *Practical Art of Motion Picture Sound*, 1999, p. 86.
53 *The Art of the Motor*, 1995, pp. 147, 23.
compared to Baudrillard’s reading of simulation, where signs have come to produce reality, where immediate reality is centred on the production, transmission, and proliferation of signs, and where the production of objects as the motor of social life has given way to the production and proliferation of mass-mediated signs. Unlike Baudrillard, Virilio believes reality does not disappear, but is rather displaced by another mode of reality, where media presentation has come to replace (or substitute for) an immediate reality, such as photography substitutes itself for real life, or film substitutes the static representation of the real with moving images, or virtual reality substitutes itself for ‘real life’.54 For Baudrillard, “reality disappears into hyperreality, for Virilio new technologies provide a substitute reality, a virtual reality which becomes more powerful and seductive than ordinary reality”.55 Despite their disagreements, Baudrillard and Virilio provide similar dystopian readings of the powerful technological, mediatized forces reshaping social experiences. Their interpretations can be compared to Marshall McLuhan’s more benign notion of media massage, which offers a more positive and utopian reading of the powerful technological, mediatized forces reshaping social experiences, as media technologies are interpreted as productive extensions of man.

Baudrillard, Virilio, and even Debord have in common the claim that human beings have been losing control of their objects and the object world, particularly in the form of mass media representations and signs, which have been taking on a life of their own (extending the classic theory of reification), thereby threatening the subject with loss of power and agency. But I think it is important to carve out a conceptual space for the

way we, as subjects, engage with these media representations and signs; after all, we should not forget the importance of theorising just how we produce, circulate, and receive media representations and signs. The image has been the subject of a variety of analyses and debates from different theoretical approaches, as a site of meaning-making that continues to grow in influence. It is time to acknowledge that images are now soundful, and that this aspect also requires investigation. Soundful images should be understood as a new language, founded on old models, where old critiques about the nature of ‘immediate’ reality renew themselves by being recast in new forms. After all, in a prior historical era, the complaint formerly was that the printed word had become more real than things in the world; now, the lamentation is that (soundful) images have become more real than reality.

The cinema books by Gilles Deleuze can be read as a counterbalance to the pessimism found in Baudrillard and Virilio. As David Rodowick points out, Deleuze makes the implicit assumption that contemporary culture has become predominantly audiovisual. For Deleuze, “the semiotic history of film is coincident with a century-long transformation wherein we have come to represent and understand ourselves socially through spatial and temporal articulations founded in cinema”, even if these articulations are now diversifying into other popular forms such as television, video games, etc.56 This observation can be related back to Benjamin’s point that cinema has not only ushered in a

---

new form of artistic reproduction, but also a distinctive mode of aesthetic reception: “film corresponds to profound changes in the apperceptive apparatus – changes that are experienced on an individual scale by the man in the street in big-city traffic, on a historical scale by every present-day citizen”.57 What Deleuze describes as “spatial and temporal articulations founded in cinema” and Benjamin calls “profound changes in the apperceptive apparatus” can be productively related to what I have been calling an audiovisual imaginary, as soundful images, founded with the coming of the talkies, help to shape our notions of space, time, and subjectivity. In this sense, the emergence of an audiovisual culture needs to be recognised as being “concomitant with the history of cinema and the history of film theory”.58

Deleuze’s cinema books have been largely received as complex commentaries on images. However, in these cinema books, Deleuze offers a conceptual toolbox that can be productively appropriated and used in terms of thinking about how cinema – in providing us with a sound-image repertoire – concretely produces a corresponding audiovisual imaginary. In my reading of Deleuze, I will move past an image-centric interpretation of Deleuze’s ideas, in order to bring to the surface and engage with a range of ideas about soundful images, where images need to be understood as not being merely visual, but also soundful. For Deleuze, the image components of the cinema comprise a moving “signaletic material which includes all kinds of modulation features, sensory (visual and sound), kinetic, intensive, affective, rhythmic, tonal, and even verbal (oral and written)”. But, says Deleuze, “even with its verbal elements, this is neither a language system nor a language. It is a plastic mass, an a-signifying and a-syntaxic material, a material not

formed linguistically”.59 Deleuze draws from Charles Sanders Peirce’s semiotic theory, which understands signification as an ongoing process, in order to deduce a theory of signs from material the cinema has itself historically produced. After all, as Deleuze argues, “Cinema itself is a new practice of images and signs”.60 It is evident that Deleuze’s theory of cinema “is quite different from traditional film theory, which always sees images as representations of other (transcendental) worlds”.61 Deleuze (drawing from the work of Henri Bergson) puts forward a conception of images that are not representations of something else but exist in themselves. This has been an insight that I have adopted – images need to be understood as being soundful, and these soundful images are not simply representations of an original event, but rather exist in themselves.

According to the representation model, images – as representations – are copies of an original reality. Philosophies of representation are based on the idea of a model and a copy (the original and the image, the essence and its reflection). There is also a clear distinction between the one who is looking (the subject) and that which is being looked at (the object). According to Deleuze, we should consider the image not as a representation but as an expression of mental relations.62 In this sense, he draws on Bergson, who remarks that “the object exists in itself and, on the other hand, the object is, in itself, pictorial, as we perceive it: it is an image, but an image which exists in itself”.63 Deleuze takes Bergson’s argument that “all matter is Image, and that the universe is defined as the

59 Cinema 1, 1986, p. 29.
60 Cinema 2, 1989, p. 280.
61 Pisters, 2003, p. 3.
62 “An atom is an image, which extends to the point to which its actions and reactions extend. My body is an image, hence a set of actions and reactions. My eyes, my brain, are images, parts of my body. How could my brain contain images since it is one image amongst others? External images act on me, transmit movement to me, and I return movement: how could images be in my consciousness since I am myself image, that is, movement?” (Cinema 1, 1986, p. 58.)
63 1911, p. xii
whole aggregate of images acting and reacting to one another on all their surfaces and in all of their parts”. What this perspective does is to dispense “with the usual way of distinguishing between subject and object. There is no object distinguishable from its image, which is to say, the set of actions and reactions that it incurs or to which it submits. This web of actions and reactions is ultimately universal. All bodies, regardless of size or distance, are potentially interconnected”. The basic philosophical problem for Bergson and Deleuze, then, “is not one of subject and object or inside and outside, but rather, how these two systems of images interact, how they are woven together in a perceptual and/or epistemological event”.64 As Pisters explains, “The image is not seen as a representation, an umbilical cord, but as a thought-provoking encounter”.65

Deleuze proposes that particularly with the emergence of the world of images, we need to comprehend the brain as a screen; “the brain is a screen” in the sense that it “is a form of relation, of interchange, of mutual synthesis between the brain and the universe”66. However, this conceptual means of overcoming the traditional object-subject binary should not simply be understood in terms of visual images. In addressing the issue of soundful images, I have read this complex Deleuzian idea in terms of an audiovisual imaginary – where soundful images can be understood as an expression of mental relations, as ways of perceiving, thinking about, and experiencing the world – we perceive, think about, and experience what we call reality through an audiovisual imaginary. Through the cinema, then, “we have come to live in a universe that is metacinematic. In it, all kinds of virtual (past and future) images are ‘stored’ and actual

66 Flaxman, ‘Introduction’ to The Brain is the Screen, edited by Flaxman, 2000, p. 16.
(present) images are constantly generated, and both types mutually influence one another. In this sense, we come to understand our past, present and future through a new ‘camera consciousness’ that has entered our perception”.67 Deleuze’s statement here can be revised slightly in order to fully bring out the audiovisual implications inherent in Deleuze’s argument. In our metacinematic universe, all kinds of virtual (past and future) soundful images are ‘stored’ and actual (present) soundful images are constantly generated, and both types mutually influence each other. In this sense, we come to understand our past, present and future through a new audiovisual imaginary that has entered our perception.

Rodowick notes that the basic question “informing Deleuze’s cinema books is this: how does a sustained meditation on film and film theory illuminate the relation between image and thought?”68 The answer lies with Deleuze’s central assertion that with its aesthetic practices, “cinema concretely produces a corresponding image of thought, a visual and acoustic rendering of thought in relation to time and movement”69 – the development of cinema should be related to a distinct shift in significatory practice, in terms of the development of new ways of imagining and imaging the world. The problem that we are faced with, however, “is that the world is increasingly composed of clichés that condition a whole network of conceptual reflexes. The world has become a bad film, Deleuze says, that we inhabit and that inhabits us as a ‘habitus’, a mode of regularity and ‘control’ (Burroughs)”.70 This warning returns us to the pessimism of Baudrillard and Virilio. Deleuze “depicts image practices as social and technological automata where

---
68 Rodowick, 1997, p. 5.
70 Flaxman, ‘Introduction’ to The Brain is the Screen, edited by Flaxman, 2000, p. 11.
each era thinks itself by producing its particular image of thought”. 71 He complains that contemporary cinema is dying “from its quantitative mediocrity”, where “there is no cerebral stimulation or birth of thought”; he cites Artaud in denouncing the “imbecile world of images”. 72

The problem, as diagnosed with the theoretical tools supplied by Baudrillard and Virilio, is the standardisation and reification of images and sounds, where our thoughts have been replaced by soundful images. For Deleuze, we have come to operate in accordance with fixed concepts that, like clichés, imprison our image of thought. Deleuze talks about the “civilisation of the image”, 73 pointing to how the cinema restricts its own potentialities, where in the process of making images civilised, images are reduced to clichés and rote concepts. This, of course, is not so far away from the bleak perspective espoused by Baudrillard and Virilio, in terms of how their arguments can be related to the pervasive forms of our contemporary sound-image repertoire, which have shaped an audiovisual imaginary that, in Baudrillard’s and Virilio’s terms, have come to act as simulation, or as substitute for an immediate reality. As Flaxman writes, the frenzy of images that characterises our audiovisual culture has “reduced us all to the feeling that we are unwilling repositories of and accomplices in a plan to populate the world with mindless images. Today, technology and capitalism conspire as never before to proliferate such clichés; accelerating, they overrun our homes, workplaces, supermarket checkout lines, and airports, hunting us down in our most private moments”. 74

71 Rodowick, 1997, pp. 6, 7.
74 Flaxman, ‘Introduction’ to The Brain is the Screen, edited by Flaxman, 2000, p. 9.
this dire circumstance, Deleuze is interested in “restoring hope in a possibility of thinking in cinema through cinema”.75

How does he propose to do this? Deleuze proposes that his cinema books offer a “logic” of signs: together, *The Movement-Image* and *The Time-Image* constitute a “taxonomy, an attempt at the classification of images and signs”.76 Deleuze's contribution is to examine the changing narrative trends that develop with classical to modern cinema – his work can be productively applied to postmodern cinema and digital culture. He presents “a new analysis of modern cinema whereby the latter is shown as a transitional stage between pre-war cinematic art and the aesthetic possibilities offered by a subsequently evolving digital audiovisual form of communication”.77 Deleuze’s film-historical analysis is particularly interested in this changing relationship between time and image. In his cinema books, he argues that change in narrative trends “necessarily entails the mutation of images; conversely, as images undergo mutation, the piecing together of time sequences is consequently altered”.78 In supplying this taxonomy of signs in cinema, Deleuze theorises the transition from classical (‘movement-image’) to modern cinema (‘time-image’). Again, the terminology of this taxonomy lends itself to an image-centric interpretation; however, the movement-image and the time-image can be (and should be) productively read in terms of an evolving transformation in sound-image relations.

It is interesting to see that Deleuze, in dividing the history of cinema into two major eras, marks the transition point at the late 1940s and early 1950s, with Italian neo-

75 *Cinema 2*, 1989, p. 165.
76 *Cinema 1*, 1986, p. xiv.
78 Ibid.
realism and the maturation of Hitchcock, instead of the more obvious shift from silent film to sound film. Kovács remarks: “The fact that the image cannot be divorced from time lies at the heart of Deleuze’s understanding of the cinema”. This is an opportune moment to remember that it was the synchronous soundtrack which has made cinema an art of time. For in order to maintain sync, the playing of a soundtrack requires the stabilisation of the projection speed for the image, and as a result: “Filmic time was no longer a flexible value, more or less transposable depending on the rhythm of projection”. After all, sound is fundamentally temporal, as sound creates a linear narrative through its audible beginning and ending (attack, duration and decay). Because of its transience and ephemeral nature, sound anchors us within the stream of time. An image, particularly a still image that does not give reference to the passage of time, can be coupled with sounds that imbue that scene with temporality. The moving image, by itself, is still abstracted, seemingly detached from the flow of life and time, and the infusion of sound creates a sense of duration. With this in mind, Deleuze’s preoccupation with the relation of time and image would seem to necessitate a particular attention on the role of sound. However, it is only with the emergence of the time-image that there is a new emphasis on sound: the soundtrack is no longer strictly contained by the image track.

Deleuze describes classical cinema as being characterised by the movement-image, which “is based on cause and effect, on organic linkages and teleological development, and on protagonists ploughing purposefully through the narrative space”. The movement-image, like everyday perception, connects a flow of different images and

79 Kovács provides an answer to this; see his ‘The Film History of Thought’, in Flaxman [Ed.], 2000, pp. 159-160.
82 Stam, 2000, p. 260.
sounds into ordered wholes. What defines movement-images is the degree to which these images and sounds and the links between them are rational; in classical Hollywood cinema, for instance, situations typically provoke actions that generate new situations, providing the constituents of the basic trajectory of narrative. The regularities and continuities of narrative are based upon “the regularities and continuities of a commonsense space-time”\(^83\) and Deleuze ascribes this regulatory ‘regime’ of the movement-image to a “sensory-motor schema”\(^84\), which “determines the spatial and temporal coordinates of the image by composing a line of action through rational intervals and organic construction”\(^85\). As Flaxman notes: “The sensory-motor schema is the mechanism of our relation to the world of images, the result of which is narrative”, where “the images procured are recognisable, capable of being linked to other images along a methodical, and ultimately normative, chain”\(^86\). The sensory-motor schema – shaped by and responding to a stable, normative sound-image repertoire (not merely a stable, normative flow of visual images) – gives rise to a particular way of imagining, imaging, and ordering the world. For Deleuze, it is the regime of the movement-image that has come to be responsible for the civilisation of the image, for what I have described as the reification and standardisation of the soundful image.

At the end of his first book on cinema, Deleuze argues that the movement-image, and especially the action-image (a category of the movement-image), has come to the end of its possibilities. Deleuze situates this crisis around the Second World War, which creates a break between the movement-image and a new form of image that he terms the


\(^{84}\) *Cinema 2*, 1989, p. 127.

\(^{85}\) Rodowick, 1997, p. 74.

\(^{86}\) Flaxman, ‘Introduction’ to *The Brain is the Screen*, edited by Flaxman, 2000, p. 5.
time-image. This does not mean that there are no more movement-images, but rather that a new type of image – the time-image – comes into existence. As Rodowick explains, “A new montage form emerges in the cinema of the time-image. Montage is based here on irrational intervals that ‘de-link’ images, as well as the relation between images and sounds, which are no longer limited by an image of the whole”. 87 The time-image associated with modern cinema is less concerned with linear cause-effect logic than the movement-image. Images, and images and sounds, “are no longer linked by rational cuts, but are relinked on to irrational cuts”. 88 The time-image is operated by “irrational cuts”, and whereas with the sensory-motor schema of the movement-image and everyday experience, images are synthesised or connected into ordered wholes, with the time-image the image is no longer perceived as an image of this or that. “Instead of one image after the other, there is one image plus another; and each shot is deframed in relation to the framing of the following shot”, Deleuze explains, and with this new conception of montage, “the cinematographic image becomes a direct presentation of time, according to non-commensurable relations and irrational cuts”. 89 In this sense, with the time-image, cinema takes us away from actualised objects and wholes to the very flow of images, where instead of connecting or synthesising images into meaningful progressions, the “image becomes a space for reading: seeing and hearing as decipherment rather than following an action; a legible image or lectosign to be read, rather than an action-image to be absorbed or reacted to”. 90 Sound, of course, needs to be recognised as an integral component of the time-image; Deleuze remarks that “the sound as well as the visual

87 Rodowick, 1997, p. 80.
90 Rodowick, 1997, p. 75.
elements of the image enter into internal relations which means that the whole image has
to be ‘read’, no less than seen, readable as well as visible”.91 The time-image thus
presents what Deleuze calls opsigs and sonsigs, or pure optical and acoustic images.

The time-image, then, which can impose an incongruous sound over a discordant
visual image, and can then cut from image to image and voice to voice, undoes the single
horizon of time from which ‘we’ view the world. In the time-image we do not see time as
a logical connection or progression but as diverging series and flows, as interval,
disruption or difference; cinema presents the way things do not hang together through
images in states of variation without organising observers or subjects. The question then,
as Deleuze puts it, “is no longer that of the association or the attraction of images. What
counts, on the contrary, is the interstice between two images: a spacing that causes that
image to be wrested from the void and fall back into it”.92 With the advent of the time-
image, the gap (or fissure) between images becomes primary. Deleuze writes:

Film ceases to be ‘images in a chain… an uninterrupted chain of images each one the
slave of the next’, and whose slave we are. It is the method of BETWEEN, ‘between two
images’, which does away with all cinema of the One. It is the method of AND, ‘this and
then that’, which does away with all the cinema of Being = is. Between two actions,
between two affections, between two perceptions, between two visual images, between
two sound images, between the sound and the visual: make the indiscernible, that is the
frontier, visible. The whole undergoes a mutation, because it has ceased to be the One-
Being, in order to become the constitutive ‘and’ of things, the constitutive between-two
of images.93

---

91 Cinema 2, 1989, p. 22.
92 Cinema 2, 1989, p. 179.
Deleuze argues that cinema possesses “a ‘dissociative’ force which would introduce a ‘figure of nothingness’, a ‘hole in appearances’”. For Deleuze, cinema – through the emergence of the time-image – has the power of releasing us from our tendency to organise images into some shared external world. With “the classic cinema, the image and thought are in mutual accord, the sensori-motor schema common to film and viewing allowing a ready passage between the screen world and the world of the spectator”. In modern cinema, however, the time-image opens up the possibility for us “to see, [to hear,] to feel, to sense, and finally to think differently”. Deleuze therefore values the cinema of time-images for the capacity to provoke thought to its proper activity, to articulate an unthought in thought that resists or even challenges those forces that freeze or curtail thought in the form of universal consensus or ‘common sense’. In this sense, Deleuze can be read to propose that the cinema of time-images possesses the capacity to give rise to a rich audiovisual imaginary, enabling a new way of thinking the world, a means of challenging, subverting, and reshaping restrictive traditional images of thought.

For Deleuze, modern cinema reveals that cinema is always already “made up of aberrant movements and false continuity shots. The direct time-image is the phantom which has always haunted the cinema, but it took modern cinema to give a body to this phantom. This image is virtual, in opposition to the actuality of the movement-image”. In the most recent French edition of Dialogues (in an added text called ‘L’actuel et le virtuel’), Deleuze suggests that every ‘actual’ image is surrounded by a mist of virtual

---

96 Flaxman, ‘Introduction’ to The Brain is the Screen, edited by Flaxman, 2000, p. 3.
97 Cinema 2, 1989, p. 41.
images.\textsuperscript{98} In contrast to the actuality and rational linkages and narratives of the movement-image, the time-image is founded on multiple virtual possibilities for de-linking and re-linking of images and sounds. The time-image is described as crystalline because it is multifaceted, possessing two poles: actual and virtual. “What we see in the crystal”, Deleuze writes, “is no longer the empirical progression of time as a succession of presents, nor its indirect representation as interval or as whole; it is its direct presentation, its constitutive dividing in two into a present which is passing and a past which is preserved, the strict contemporaneity of the present with the past that it will be, of the past with the present that is has been”.\textsuperscript{99} The time-image produces soundful images which unveil the force of time as change.

In the modern cinema, Deleuze remarks, with the emergence of the time-image, “visual images are re-enchained through the interstices of irrational cuts, and sound, too, is involved in such cuts”.\textsuperscript{100} Modern cinema is defined by a “dissociation” between sound and image, which “strengthens each of them”, and “which has to invent their new relationship”.\textsuperscript{101} Deleuze quotes Chion, who says about synchronisation: “Through it the image says to the sound: stop floating everywhere and come and live in me: the body opens to welcome the voice”.\textsuperscript{102} However, sound (and not only the voice) gains prominence and even independence at some points. Deleuze and Guattari remark that “sound invades us, impels us, drags us, transpires us”.\textsuperscript{103} Sound ‘tunes’, ‘de-tunes’, or ‘re-tunes’ (territorializes, deterritorializes, and reterritorializes) the listener, pulling us

\textsuperscript{98} Cited in Pisters, 2003, p. 3.
\textsuperscript{99} \textit{Cinema 2}, 1989, p. 274.
\textsuperscript{100} Bogue, \textit{Deleuze and Cinema}, 2003, p. 191
\textsuperscript{101} \textit{Cinema 2}, 1989, p. 247.
\textsuperscript{102} \textit{Cinema 2}, 1989, p. 332.
\textsuperscript{103} Deleuze and Guattari, \textit{A Thousand Plateaus}, 1987, pp. 347-348.
into different contexts, different worlds, different situations, evoking presence, identity, location, memory, conjuring up visions of various moments in time and place, compelling us to perceive or feel or act in particular ways. The modern image is audiovisual, founded in many respects on the power of a disjunction of sound and sight. Deleuze comments that “all the sound elements, including music, including silence, form a continuum as something which belongs to the visual image”. However, this sonic continuum can also possess the “specific autonomous power” of a “foreign body”. Sound can operate to make something visible in the image that was previously buried, opaque, occluded, or even formerly not present. Sound is capable of echoing and reinforcing the images, but also of engaging in an interaction with the images that is “independent of any common structure”. For Deleuze, particularly with the time-image, sounds should be understood as “a new dimension of the visual image, a new component”.

In other words, for Deleuze, image and soundtrack are relatively autonomous; while referring to each other, they nonetheless resist being reconciled into an organic whole. While this, at first glance, may appear to contradict what I have been arguing in terms of the relationship between sound and image, I wish to suggest that this can be read as another way of conceptualising what I have termed the resonance between sounds and images. Deleuze states that the modern audiovisual image is constituted by “a disjunction, a dissociation of the visual and the sound, each autonomous”, where “at the same time an incommensurable or ‘irrational’ relation… connects them to each other,

---

105 Cinema 2, 1989, p. 239.
without forming a whole, without offering the least whole”. This heautonomous relation he traces to “the collapse of the sensory-motor schema, which separates the visual image and the sound image, but puts them all the more into a non-totalizable relation”.\(^\text{107}\) As Rodowick explains, Deleuze adapts the notion of ‘heautonomous’ from the introduction to Kant’s Third Critique; by this term, Deleuze means that image and sound are distinct and incommensurable yet complementary.\(^\text{108}\) This ‘incommensurable complementarity’ between sound and image refers to how there is a complementarity between sound and image based on their strategic disassociation. In most narrative cinema, sound is a component of the image, where sound space is an extension of visual space. However, with the collapse of the sensory-motor schema, and the emergence of the time-image, sound space becomes autonomous, no longer simply a component of visual space, therefore transforming the out-of-field. Sound is freed from its dependence on images, no longer woven into the linking of actions and reactions. As images and images and sounds are linked by irrational cuts, sound and image become incommensurate, where each follows its own compositional logic. As Rodowick points out, this then “affects the visual images whose own seriality as discontinuous any-space-whatevers is enhanced”.\(^\text{109}\)

For Deleuze, the modern emergence of time-images, based on irrational cuts and de-linkages of images, is associated with the emergence of any-spaces-whatever. Emancipated from the logical coordinates of action and the rational linkage of images, the unity of space, once cohered by cause-and-effect movements, splinters into many disparate fragments. Any-spaces-whatever – “irrational, disconnected, aberrant,

\(^{107}\) *Cinema 2*, 1989, p. 256.
\(^{108}\) See Rodowick, 1997, p. 145.
\(^{109}\) Rodowick, 1997, p. 145.
schizophrenic spaces – no longer obey laws of traditional, commonsensical causality”.  

For Deleuze, an any-space-whatever is “a space that does not yet appear as a real setting or is abstracted from the spatial and temporal determinations of a real setting”. It is a space which has not yet been situated. As Deleuze explains, while they are figures of indetermination, an any-space-whatever is not an abstract universal, but rather “is a perfectly singular space, which has merely lost its homogeneity, that is, the principle of its metric relations or the connection of its own parts, so that the linkages can be made in an infinite number of ways. It is a space of virtual conjunction, grasped as pure locus of the possible”. An any-space-whatever can be an emptied space, or a space whose parts are not yet linked in a given trajectory of movement. An any-space-whatever, then, “is a spatial fragment or fragments whose identity and meaning remain part of an indeterminate multiplicity that nonetheless expresses a quality or power”. Rodowick comments: “The peculiarly abstract quality of any-space-whatevers returns us not to objects or things themselves, but to specific visual impressions”.  

Because of the heautonomous relationship between images and sounds, “modern visual images and sonic images must be ‘read’, in the sense that these images, disengaged and disconnected from their standard contexts, must be reconnected, re-enchained in ways that cannot be anticipated ahead of the appearance of the given images”. Sounds, acting as a foreign body, can inflect the image with meaning by making particular

110 Flaxman, ‘Introduction’ to The Brain is the Screen, edited by Flaxman, 2000, p. 5.
111 Rodowick, 1997, p. 63. Deleuze borrows the term from the French anthropologist Marc Augé (whom Deleuze incorrectly calls Pascal Augé in Cinema 1). Augé uses the term to help understand the effects of modern urban planning on the human psyche and interpersonal relations.
113 Rodowick, 1997, p. 64.
114 Rodowick, 1997, p. 66.
115 Bogue, Deleuze and Cinema, 2003, p. 188.
connections, resulting in an any-space-whatever by destabilising the image through inflecting it with potentially contradictory meanings and connections. Under the influence of sound, the image can cease to signify in a predictably visual manner, becoming interactive with the re-embodied sound. Or, sound can stabilise the dizzying rush of images, anchoring the any-space-whatever by allocating a space/time context, or by drawing spectator attention in particular ways. The practice of re-sounding images enables the possibility for sound to alter or shape the image’s narrative legibility, creating continuities or discontinuities not mandated by the image alone. The comments made by film director Kurosawa can be revisited in terms of the heautonomous (or resonant) relationship of sound and image. “Ever since the silent film gave way to the talkie”, Kurosawa has said, “sound has interfered with the image – and at the same time this flood of sound has become largely meaningless”. Kurosawa reiterates that cinematic sound should never function as mere accompaniment: “Real sound does not merely add to the image, it multiplies it”.116

The issue, then, is to fulfil the potential of sound-image assemblages, by allowing a rich resonance to ring out from the relationship between image and sound. Deleuze’s concept of the time-image can be used as a way to discuss the historical development and potential of soundful images. Deleuze values and privileges the time-image’s capacity for articulating change, becoming, differentiation, and the unthought, championing the time-image against the movement-image; he cites Paul Virilio’s ideas in commenting that the technology of the movement-image has been intimately and historically linked to the

logistics of warfare, propaganda, and ordinary fascism. To phrase this in terms of our discussion, with the emergence of the time-image, soundful images are allowed to resonate properly, to speak, to articulate and present new perceptions and thoughts that challenge the clichés of the traditional image of thought. However, with the movement-image, images are civilised, soundful images are silenced through meaningless floods of sound, and sound-image assemblages are packaged and imposed upon us by the mass media through a logistics of perception. “Mass mediatization”, in this way, functions through the “imposition of models”. It is not far-fetched to remark that Baudrillard’s pessimistic account of simulacrum should be analogised to what Deleuze views as the fundamentally totalitarian nature of the movement-image. The simulacrum, Baudrillard says, serves to act as a powerful form of “social control”. Simulation contains a model of ‘reality’ and meaning present “there in advance, inscribed in the code”, producing our experience, expectations, conception of the real, and behaviour. The “diffraction” and dissemination of its models and their unilateral imposition therefore plays a “regulative role”.

The problem with Deleuze’s model is that it is a linear reading, where cinema progresses from movement-images to time-images. As Rodowick notes, there is “the sense that, for Deleuze, the cinema of the movement-image has been fully realised while that of the time-image is emergent”. However, “there are few ‘pure’ examples of films where direct images of time predominate. Mixed or hybrid examples are more

---

120 Ibid, p. 59.
121 Ibid, P. 70.
common”.122 Indeed, despite Deleuze’s hopeful championing of the potential of the time-image, the clichés of the movement-image can be said to be far more preponderant, particularly when examining the typical contemporary Hollywood blockbuster. For instance, Sean Cubitt observes that recent films “like Star Wars, Episode Two: Attack of the Clones move from set piece to set piece along only the slenderest thread of narrative”; each of these set pieces “acts to trigger a rush of emotion: putting a clock on the action, staging the spin-off computer game as sport, escaping from ghoulish caverns. The scream, the laugh, the tear, the white knuckles, the racing pulse: the stimuli are clichés because the emotions they elicit and that audiences seek are clichés”.123 Not only do movement-images still predominate, they are desired and desirable.

Deleuze admits that films based on the movement-image continue to be made: “the greatest commercial successes always take that route, but the soul of the cinema no longer does. The soul of the cinema begins by undoing the system of actions, perceptions and affections on which the cinema had fed up to that point”.124 Deleuze, then, upholds the potential of the time-image, whose time has not, and perhaps will not, ever fully arrive. Cubitt argues that Deleuze’s time-image, which Deleuze renders the essential element of cinema, is merely another ‘effect’ or ‘technique.’ For Cubitt, no technique is essentially avant-garde, because any technique can be appropriated as a mere ‘effect’ and thus commodified. Cinema’s failure for Cubitt is its susceptibility to become a mere commodity, and the challenge for cinema in the 21st century, he says, “is the struggle for not yet finite, not yet infinite, ecological, human, and technological community”.125

122 Rodowick, 1997, p. 89.
124 Deleuze, Cinema 1, 1986, pp. 206-207.
(Despite his critique of Deleuze, Cubitt also maintains a rather hopeful if not utopian vision of cinema’s critical capability.)

The point to be drawn, however, from juxtaposing the accounts of Baudrillard, Virilio, and Deleuze, is that there is an ongoing struggle against the reification of thought and the channelling of flows of information in our contemporary information society, where soundful images have become a primary commodity and communicative tool for exchange. The audiovisibility of contemporary culture is defined by its particular stratifications and organisations of the space of visibility with that of utterability. “The visible and the utterable, in fact the whole audiovisual regime, are caught up in relations of power”.126 As such, the language of soundful images is constantly being negotiated and renegotiated, and our audiovisual imaginaries are constantly being shaped by various forms of sound-image assemblages. Soundful images can become reified and commodified, offering clichéd concepts and narratives – or they can open up new ways of thinking, new conceptual linkages and modes of thought, new opportunities for reflection and creativity. It remains a fascinating open-ended question as to the possible future (or futures) of the soundful image, this pervasive and persuasive mass mediated form that has become widely accepted, circulated, and exchanged, that has come to populate our world and our ways of communicating, that has come to powerfully impact upon the ways we perceive and engage with the world and those in it.

126 Rodowick, 1997, p. 197.
Conclusion

The production, circulation, and reception of soundful images in the modern world takes place on a scale that is increasingly global, increasingly constituting a simultaneous and unified imaginary across continents. During the twentieth century, and on into the twenty-first, the flow of information and communication on a global scale has become a regular and pervasive feature of social life. Arjun Appadurai argues “that electronic media... transform the field of mass mediation because they offer new resources and new disciplines for the construction of imagined selves and imagined worlds... these media... compel the transformation of everyday resource” by providing “resources for experiments with self-making in all sorts of societies, for all sorts of persons”.1 As more and more aspects of life are coordinated through electronic information flows, the space of mediation is becoming increasingly important; after all, “all realities are communicated through symbols”.2

Media and communications technologies have changed the shape of our world, as well as the form of our language. Marshall McLuhan has complained that: “For the past 3500 years of the Western world, the effects of media – whether it’s speech, writing, printing, photography, radio or television – have been systematically overlooked by social observers”.3 This, however, is no longer the case, as media studies have steadily garnered an increasing degree of attention. Nevertheless, in this era of mass-mediated sounds and images, the image has remained a primary focus for media theorists. Sound, on the other hand, has been comparatively neglected; especially the impact of sound on

---

1 Appadurai, Modernity at Large, 1996, p. 3.
the image, despite the important, influential, and evolving articulation and interaction that is taking place between images and sounds. In this dissertation, I have not sought to defend the field of sound at the expense of the image. Rather, I have suggested that we attend to the evolving articulation between sound and image, which signals an important new area of theorisation that I have addressed here through the general example of film sound. It is my hope that perhaps the ‘civil war’ between sound and image will transition into a more sensitive appreciation and discussion of the evolving relationship between sound and image.

In this dissertation, I have offered a perspective on the emergence, evolution, and naturalization of what is clearly the most influential contemporary mass media form, the soundful image. This dissertation has been based upon the central notion of re-sounding images, a semi-neologism which encapsulates the argument that the grafting of ‘unnatural’ sound (sound not of the real source or quite simply mediated sound) onto an image creates a ‘soundful’ image in which the relation between image and sound is not natural, nor indexical, not representative, not reproductive, but rather resonant, which is to say that there is a natural affinity to the pairing that creates a new aesthetic – or media – (hyper)reality.

The dissertation has put forward the argument that sound now operates through a close relationship with image; and that these sound-image articulations (or soundful images) have become naturalised, powerful, and extremely familiar aspects of our mediatized reality. Sound is fluid and malleable: sound can be edited, re-rearranged and put in different synchronous relationships with the image. With the increased possibilities of re-recording, numerous different tracks of sound can potentially be separately
manipulated and then recombined. The ‘magic’ of sync sound is now largely taken for granted, as images and sounds interact and are fused together within the perception of the audience, presenting a coherent mass-mediated reality. Sound comfortably moulds itself around the image; this articulation lends both image and sound added degrees of meaning.

The articulation of sound and image can be traced to an initial historical machined separation (where sounds and images are recorded separately), followed by a post-Edisonian machined fusion (as sounds and images can be recombined to form soundful images, which are more than simply the sum of their parts). The new language of sounds and images and soundful images has allowed voices, sounds, and music to be re-embodied through the image, bringing moving pictures to life, as sound contributes presence and dimensionality to images, installing a believable sense of performed authenticity. The cinema has facilitated and disseminated the development of a new model of sound in conjunction and articulation with the evolving flow of images; hyperreal sounds have become ubiquitous, increasingly fusing with the image, and operating as key components of soundful images. A new type of meaningfulness is being imparted to sounds, which are increasingly being associated with images (as sound becomes the sound of an image). This opens up an important new perspective on the relationship between sounds and images – sounds, voices, and music are increasingly implying images, or creating expectations or anticipations with regard to images. Sounds, voices, and music are more and more referring to images, being associated with images, converging with images, invoking images as their source – in a new model of cinematic
sound, where music, voices, and sounds are increasingly becoming visual and visualised. Mass-mediated sounds and images have populated our memories, as soundful images have unavoidably come to affect the way we see, hear, and make sense of events around us.

Having offered a particular history or take on the emergence, evolution, and naturalization of the soundful image, questions remain to be asked, opening up avenues for prospective further research. In what direction is the articulation of sound and image moving today? What are the possible futures of the impact of soundful images in an evolving age of audiovisuality? The cinema has offered and continues to offer an exemplary range of soundful images, but it is arguably being overtaken by new cultural forms. As the ways in which we perceive and engage with the world around us continue to be shaped by the persuasive force of mediatized images and sounds, the evolving articulation between sounds and images must be recognised as playing a crucial role in our understandings of the pervasive influence of contemporary media forms. As this dissertation has argued, the relation between sound and image is being re-balanced in new and important ways, and as we strive to come to terms with the widespread effects of our media environments, it is time to pay sustained attention to the widespread implications and histories of re-sounded and resounding images. After all, the language of soundful images, already prominent, can be clearly seen to be spreading even further, shifting into different shapes and forms (seen and heard on digital television, multi-media games, real-time streaming media players, etc.) – which cannot be ignored.

---

4 “Media has turned the world into a theatre”, as David Toop says, and “film’s hyperreal theatre of sound animates everyday noise outside the cinema”. (Toop, Ocean of Sound, 1995, p. 260, 256.)
References and Bibliography


Altman, Rick (Editor), 1992, *Sound Theory Sound Practice*, Routledge


Arnheim, Rudolf, 1957, *Film as Art*, University of California Press

Bálázs, Bela, 1972, *Theory of the Film: Character and Growth of a New Art*, translated by Edith Bone, Arno Press


Baudrillard, Jean, 1983, *Simulations*, translated by Paul Patton and Paul Foss, Semiotext(e)


Baudry, Jean-Louis, ‘The Ideological Effects of the Basic Cinematographic Apparatus’, *Film Quarterly* 28.2, Winter 1974-75


Bayard, Caroline and Knight, Graham, ‘Vivisecting the Nineties: An Interview with Jean Baudrillard’, Part 1, *CTHEORY 18*, nos. 1-2, August 1995


Benjamin, Walter, 1985, *One-Way Street and Other Writings*, translated by Edmund Jephcott and Kingsley Shorter, Verso


Bogdanovich, Peter, 1993, *Orson Welles and Peter Bogdanovich*, Harper Collins


Brophy, Philip (Editor), 1999, *Cinesonic: The World of Sound in Film*, Southwood Press

Brown, Royal S., 1994, *Overtones and Undertones: Reading Film Music*, University of California Press

Buhler, James; Flinn, Caryl; and Neumeyer, David (Editors), 2000, *Music and Cinema*, Wesleyan University Press


Burdick, Alan, ‘Now Hear This: Listening Back on a Century of Sound’, *Harper’s Magazine* 303, no. 1804, July 2001


Buskin, Richard, ‘Making Contact’, interview with Randy Thom and Tom Johnson; article found online at [http://www.filmsound.org/randythom/contact.htm](http://www.filmsound.org/randythom/contact.htm), accessed on July 28, 2005

Buskin, Richard, 'Soundings from The Phantom Menace', *Studio Sound*, July 1999


Cage, John, 1999, Silence: Lectures and Writings, Marion Boyars Publishers, originally published in 1968 by Calder and Boyars


Coffman, Joe W., ‘Art and Science in Sound Film Production’, JSMPE 14, February 1930, p178-179


Cook, Pamela, 1996, *Fashioning the Nation: Costume and Identity in British Cinema*, British Film Institute


Deleuze, Gilles, 2000, ‘The Brain is the Screen’, in *The Brain is the Screen: Deleuze and the Philosophy of Cinema*, edited by Gregory Flaxman, translated by Marie Therese Guiris, University of Minnesota Press, p365-373


Collin Chua
Re-Sounding Images


Donnelly, K. J. (Editor), 2001, *Film Music: Critical Approaches*, The Continuum International Publishing Group, Inc

Donnelly, K. J., 2005, *The Spectre of Sound: Music in Film and Television*, BFI Publishing


Eisenstein, Sergei, 1957, *Film Form and Film Sense*, Cleveland: Merieian


Flaxman, Gregory (Editor), 2000, *The Brain is the Screen: Deleuze and the Philosophy of Cinema*, Minneapolis: University of Minnesota Press


Foucault, Michel, 1979, *Discipline and Punish: The Birth of the Prison*, translated by Alan Sheridan, Peregrine Books


Kalinak, Kathryn, 1992, *Settling the Score: Music and the Classical Hollywood Film*, University of Wisconsin Press


Leyda, Jay (Editor), 1988, Eisenstein on Disney, translated by Alan Upchurch, London: Methuen

Leyda, Jay, 1972, Kino: A History of the Russian and Soviet Film, Allen and Unwin

LoBrutto, Vincent, 1994, Sound-On-Film, Praeger Publications Text


Macksey, Richard, and Donato, Eugenio (Editors), 1972, The Structuralist Controversy: The Languages of Criticism and the Science of Man, Baltimore


Massumi, Brian, ‘Realer than real: The Simulacrum According to Deleuze and Guattari’, Copyright No.1, 1987


McLuhan, Eric and Zingrone, Frank (Editors), 1995, Essential McLuhan, Routledge


Michelson, Annette (Editor), 1984, *Kino-Eye: The Writings of Dziga Vertov*, University of California Press


Musser, Charles, 1994, *Thomas A. Edison and his Kinetographic Motion Pictures*, Friends of Edison National


Romney, Jonathan, and Wootton, Adrian (Editors), 1995, *Celluloid Junkies: Popular Music and the Movies since the 50s*, BFI Publishing


Stam, Robert, 2000, *Film Theory: An Introduction*, Blackwell Publishers


Toop, David, 1995, *Ocean of Sound – Aether Talk, Ambient Sounds and Imaginary Worlds*, Serpent’s Tail


Weis, Elisabeth and Belton, John (Editors), 1985, *Film Sound: Theory and Practice*, Columbia University Press


Williams, Alan, ‘Is Sound Recording Like a Language?’, *Yale French Studies* 60, 1980


Zurbrugg, Nicholas (Editor), 1997, *Jean Baudrillard, Art and Artefact*, Institute of Modern Art