Semiosis in the Film Soundtrack:
Aural Perspective and Social Distance in *The Queen* Film Trailer

**Betty Noad and Len Unsworth** | University of New England, NSW

The emergence of digital technologies has changed the design of texts and our literate practices so that we now interpret and construct texts which have written, visual, audio and spatial dimensions for making meaning, that is, multimodal texts. Contemporary digital texts such as television advertisements, film trailers, video and television programs increasingly privilege sound features (speech, music, sound effects) to ‘design’ meanings and to ‘position’ listeners towards the interests of the composers of multimodal texts. Indeed digital texts that persuade, such as television advertisements and film trailers, particularly feature sound to build a convincing message about a product, for consumers. Sound now takes a significant place alongside language and visual images in the digital texts of our multimodal landscape (Baldry & Thibault 2006, van Leeuwen 1999), and will be a crucial part of texts that students must learn to critically understand and use.

This paper discusses the educational imperative for schools to begin developing students’ knowledge about sound as an integral communicative mode in contemporary digital texts. In acknowledging the paucity of research that frames and supports teaching about sound, this paper also argues the need for research that builds a rigorous basis for theorising the modal resources of sound, and informs pedagogical practice. In this educational research context, one such research perspective in relation to sound is examined in this paper. van Leeuwen’s theoretical modelling of sound as a social semiotic, as presented in his book *Speech, Music, Sound* (1999) is outlined, and it informs the exploratory research design reported in the final section of this paper. van Leeuwen’s assertion that the conceptual and technical description of semiotic resources of sound facilitates the interpretation of meaning is investigated. Using *The Queen* film trailer as a model, van Leeuwen’s methodology is applied to analyse the semiotic resource of loudness, and interpret the meaning of loudness. The educational implications of this investigation are discussed.

The educational impetus for teaching and researching in the area of sound

Current definitions of multimodal literacy (Jewitt & Kress 2003, Kress & van Leeuwen 2001) accordingly require explicit teaching that develops students’ knowledge about the meaning-making resources of multiple modes and the skills to interpret and compose multimodal texts in digital formats, enabling them to participate in the design of future texts. Syllabus imperatives in Australian states increasingly require students to learn multimodal literacy in this way, and to develop social, cultural and critical perspectives about the contexts which impact on the construction of multimodal texts. The New South Wales *7–10 English Syllabus* (Board of Studies NSW, 2001 p. 7–13) and the Queensland *T1–10...*
English Syllabus (Queensland Studies Authority, 2005 p. 7–12) demonstrate a significant response to the call for teaching multimodal literacy, requiring students to learn about the forms, features and structures of multimedia texts, and to learn to compose complex multimodal texts. The Victorian Essential Learning Standards (VCAA & Authority 2006 online) and the South Australian Curriculum Standards and Accountability Framework (South Australian Department of Education, 2001 online) require primary and secondary school students to learn to interpret, critically respond to and create digital multimodal texts.

Along with social imperatives for students to control multimodal texts comes the educational imperative that teachers access professional learning for effective teaching of multimodal literacy. To plan for systematic and explicit teaching about texts and technologies of today, teachers need comprehensive practical frameworks and metalanguages to plan pedagogical practices for teaching multiliteracies (Unsworth 2001). Currently teachers have access to comprehensive theories, frameworks and metalanguages to support teaching about linguistic features of texts (Halliday 1978, 1985). Kress and van Leeuwen (1996) have provided frameworks and a grammar in relation to the design of visual images. While syllabuses have made scant reference to the tools provided by Kress and van Leeuwen, teachers are gradually taking up their framework and metalanguage for explicit teaching about visual images, enabling them to become teachers of visual literacy (Noad 2004).

Despite syllabus demands, teachers do not have access to frameworks or metalanguages for teaching explicitly about how meanings are made in multimodal texts by the mode of sound. A review of current research literature highlights the dilemma: there is a paucity of research providing a cohesive theoretical framework for describing and explaining the textual resources of sound, or the role of sound in multimodal texts, for educational purposes. Indeed, any metalanguage for talking and teaching about the resources of sound is presented as tentative (McDonald 2003, van Leeuwen 1999) and tools to analyse the role of sound for making meaning together with other modes are only recently being developed (Baldry & Thibault 2006, Jewitt & Kress 2003, van Leeuwen 1999). To support the teaching of multimodal literacy, future research needs to focus on developing a credible basis that comprehensively theorises the mode of sound, and identifies how the resources of sound make meanings. In the section that follows, aspects of the theoretical modelling of sound as a social semiotic (van Leeuwen 1999) are presented, as an existing research perspective that potentially contributes to educational inquiry.

van Leeuwen’s theoretical modelling of sound

Multimodal texts make meanings by integrating resources from multiple modes of communication. Each mode, such as language, or image, or sound, has ‘a regularised organised set of resources for meaning-making’. Following Halliday’s theorising of language as a social semiotic (1978) and his study of systemic functional linguistics (1985), other researchers have modelled the semiotic resources of image (Kress & van Leeuwen 1996, O’Toole 1994), and gesture (Martinec 1998). In Speech, Music, Sound (1999) van Leeuwen theorises the mode of sound also as a social semiotic. While van Leeuwen acknowledges the possibility that the resources of sound have not yet reached the level of abstraction and specialisation which the resources of language and image have assumed in response to cultural use, he nevertheless suggests that the semiotic resources of sound are sufficiently developed to allow description of the way they systematically function to make meaning. In modelling sound this way, van Leeuwen is positioning sound in the developing discourse of multimodal semiotics, or systemic functional semiotics (Macken-Horarik 2006).
Sound is integral to multimodal and dynamic texts such as film. Some researchers have focused on ‘intersemiosis’ that is how the resources of multiple modes co-pattern and co-contextualise to produce meanings in film (Baldry & Thibault 2006, Martinec 2000, Ravelli 2000), for example how sound integrates with image and movement to produce text-specific meanings. In contrast, van Leeuwen (1999) focuses on providing a descriptive account of the range of ‘intrasemiotic’ resources available to the mode of sound that can produce meanings on a multimodal film or radio soundtrack.

In line with research in film, van Leeuwen conceptualises the intrasemiotic resources of sound as those included in speech, music and other sounds, where ‘other sounds’ refer to sound effects not produced by humans or musical instruments (Bordwell & Thompson 1993, van Leeuwen 1999). van Leeuwen identifies six major domains of sound which he asserts are common to speech, music and sound effects: ‘sound perspective, sound time and rhythm, the interaction of ‘voices’ (for example by taking turns or speaking, singing, playing or sounding together in different ways), melody, voice quality and timbre, and modality’ (van Leeuwen 1999, p. 9). According to van Leeuwen, sound perspective, sound time and rhythm, melody and voice quality and timbre derive their semiotic affordances from their ‘materiality’ (1999, p. 125), that is from the concrete material resources (human, musical and non-musical instruments) used to produce sounds. van Leeuwen argues that the semiotic resource of loudness, for example, realises perspective in sound; the semiotic resource of rhythm realises timing patterns in sound; the semiotic resource of pitch realises melody, and sound qualities are realised by all these semiotic resources as well as resources such as breathiness, or tension in sounds.

As well as providing a descriptive account of the semiotic resources of sound, van Leeuwen also investigates how these resources can be used to ‘say and do things with sound’ and how to interpret their meanings (1999, p. 9). Van Leeuwen theorises sounds as the actions and interactions of people, places and things in our environment, and that sounds are not things, nor can they represent things. Sound can present the hum of a car engine for example, but it cannot be the engine itself; sound can present the singing of a national anthem, but it cannot be the singers themselves. Following Martinec (1996), van Leeuwen distinguishes between the ‘presentation’ and ‘representation’ of sounds according to whether the source of the sound is identifiable or not (1999, p. 35). Sounds present action and interaction if the source is visibly presented, whereas sound represents action and interaction if the source is invisible but can be identified. The sound of a door slamming can be presented with an image of a door slamming shut, and the same sound may later represent a door slam despite the source being invisible.

As well as presenting action and interaction, van Leeuwen theorises that sound, like all other semiotic modes, can ‘create relations’ between the listener and the sound itself (1999, p. 12). According to van Leeuwen, relations of ‘formality’ are created by the system of social distance, which is realised by the semiotic resource of loudness (van Leeuwen 1999, p. 24). Musical timing can reflect the ‘affective relationship’ that people have with clocktime or the time of the metronome, whether they embrace or struggle with it for example (van Leeuwen 1999, p. 50). Interacting sounds (of speech and music) reflect the degree of interactivity and involvement that people adopt in an interaction, and how that involvement is affected by all kinds of ‘unequal power’ relations (van Leeuwen 1999, p. 71). Melody, realised by the semiotic resource of pitch, has the capacity to invoke emotions and feelings, thus creating affective relations with the listener (van Leeuwen 1999, p. 97).

Following a systemic functional approach (Halliday 1978, Martin 1992), van Leeuwen develops a system network for each of the semiotic resources of sound. Each system
network describes choices offered by a semiotic resource, that have semiotic value, or potential for meaning-making. To explain how each of the semiotic resources of sound realise meanings, van Leeuwen extends his theoretical modelling to include an interpretive framework which engages the concepts of ‘provenance’ and ‘experiential meaning potential’ (1999, p. 46) and the concept of social context. According to van Leeuwen, the potential meanings that can be ascribed to sounds derive from two sources: the provenance of a sound refers to ‘associations or connotations’ invoked about the people and place where the sound comes from (1999, p. 210) and the experiential meaning potential of a sound refers to ‘our experience of what we physically have to do’ to produce a particular sound (1999, p. 205). The experience and associations that we have with a sound generate meaning potential, but the actual meaning is then formed or clarified with reference to the context in which the sound is situated.

In the following section, van Leeuwen’s ‘System network of aural perspective and social distance’ (1999, p. 30) is presented for conceptual and technical description of loudness, as it informs the research design that analyses loudness in a phase of The Queen film trailer.

**Loudness as a semiotic resource of sound**

van Leeuwen theorises that the semiotic resource of loudness realises aural perspective and social distance in sound. According to van Leeuwen, perspective in sound ‘creates relations’ between the subject they represent and the receiver they address, that is, the listener, in two ways and both are related to distance (van Leeuwen 1999, p. 30). First is the relation of aural ‘perspective’, where loudness hierarchises simultaneous sounds and places them at different distances from the listener, so that the listener relates to sounds according to their importance, or relevance (van Leeuwen 1999, p. 212). Second is ‘social distance’ which applies to single sounds, where different sound qualities (such as volume, breathiness and pitch) create relations of differing degrees of formality between what is represented by the sound, and the listener (van Leeuwen 1999, p. 14).

van Leeuwen engages the concept of ‘soundscape’ to demonstrate how the semiotic resources of sound generally characterise the sounds that we hear around us, or the sounds that we hear on a film or radio soundtrack (van Leeuwen 1999, p. 15). Baldry & Thibault (2006) and van Leeuwen (1999) acknowledge that a number of sounds can occur in a soundscape, from different sources and different locations, and that these different sounds

---

**Figure 1. A system network of aural perspective and social distance. (van Leeuwen, 1999, *Speech, Music, Sound*, Figure 2.2)**

---
have different degrees of relevance to the listener. This paper examines how the systems of aural perspective and social distance, for example, position sounds in a soundscape into different relations of significance with listeners.

**Aural perspective**

van Leeuwen theorises that when sounds present or represent actions simultaneously, the semiotic resource of loudness realises the perspectival positioning of sounds, or ‘aural perspective’ (van Leeuwen 1999, p. 22). The system of aural perspective hierarchises simultaneous sounds and places them into groups at different distances from the listener so that the sounds assume varying degrees of importance, or significance, for the listener. Loudness realises aural perspective whether from ‘the levels of the sounds themselves, from the relative distance of the people or objects that produce them, or from the way a soundtrack is mixed’ (van Leeuwen 1999, p. 23). As well as being described as a ‘relative’ concept, loudness is also described as a ‘subjective’ concept.

Loudness is relative to the ‘position of the listener’ (van Leeuwen 1999, p. 17), and under normal (non-amplified) circumstances loudness is related to perceived physical distance, so that sounds will be heard in a range from loud to soft, and be perceived as close or remote. However the position of the listener can be manipulated by sound reproduction and replay technologies, such as the amplifier or the sound mixing board, so that a relation can be ‘created’ for the listener (van Leeuwen 1999, p. 17). In this situation a soft sound can be amplified so that it positions the listener to treat the loud sound as important.

Loudness, which is often understood as volume, depends ‘both on intensity and frequency’ (Kane & Sternheim 1978, p. 414). Intensity refers to volume, or the amplitude of sound waves measured in decibels (dB), so that large displacements produce high measures of decibels (loud volume) and small displacements produce low decibel measures (soft volume). Frequency refers to the number of sound waves passing a point per second, which closely relates to the pitch of a sound (Kane & Sternheim 1978, p. 414) and is measured in hertz (Hz). The pitch of a sound is characterised as the highness or lowness of a sound. The loudness of a sound therefore includes notions of both volume and pitch, which do not always vary together. Given, as well, that the perception of volume and pitch in human hearing varies with individuals, description of loudness as ‘subjective’ appears to be reasonable. The perception of loudness here can only refer to what most humans would normally hear as loud or soft, high or low sounds.

The concepts and tools that van Leeuwen selects for analysis of aural perspective in sound are more concerned to describe volume rather than pitch. In other words, when van Leeuwen refers to ‘loudness’, he is referring mostly to volume, regardless of how it is produced (van Leeuwen 1999, p. 23). To analyse sound in terms of volume and therefore aural perspective, the concepts of sound dubbing technicians in radio and television such as Murch (1985) and Schafer (1977) are considered by van Leeuwen, as each divides the soundtrack into three ‘zones’ – close, middle and far away (1999, p. 15). It is Schafer’s concepts that van Leeuwen selects and adapts as a tool for analysing aural perspective in sound. Accordingly, van Leeuwen indicates that simultaneous sounds can be divided into three groups, and be positioned as Figure, Ground and Field, which he defines as:

**Figure**: if a sound or group of sounds is positioned as figure, it is thereby treated as the most important sound, the sound which the listener must identify with and/or react to and/or act upon. **Ground**: if a sound or group of sound is positioned as Ground, it is thereby treated as still part of the listener’s social world, but only in a minor and less involved way. We are to treat it as we treat the familiar
faces we see every day … as a context we take for granted and only notice when it is not there any longer.

Field: If a sound or group of sounds is positioned as Field, it is thereby treated as existing not in the listener's social, but in his or her physical world. We are to treat it as we would treat the people that crowd the streets through which we walk, or the trees that populate the forest past which we drive.

(van Leeuwen 1999, p. 23)

The above concepts refer to sounds in a soundscape that can be perspectively positioned. However it is also the case that there is an opposite to perspective, and that is ‘immersion’, or ‘wrap-around’ sound (van Leeuwen 1999, p. 28). It is pointed out that immersive sounds are often characterised by ‘low frequency’ sounds, that is sounds that have a low pitch or tone, such as foghorns, which seem to come from everywhere at once and the listener feels fully integrated and immersed in the environment (van Leeuwen 1999, p. 28). A modern dance club is another example, where ‘low frequency sounds seek blend and diffusion rather than clarity and focus’, and the listener is immersed ‘in the centre of the sound … flooded by it’ (Schafer 1977, p. 118 in van Leeuwen 1999, p. 29).

Van Leeuwen provides a number of examples which demonstrate how his concepts about loudness and distance, and his terms, can be used as tools to identify the aural perspective of sounds in a soundscape, and thus how they relate to the listener in terms of importance, whether the sounds are static (not changing) or dynamic (changing over time).

In a soundscape where the relative loudness of sounds are ‘static’ (van Leeuwen 1999, p. 30), that is the sounds remain at the same level of loudness for a time, the sounds are described as being in Figure, Ground and Field position. For a city office worker for example, the sound of muffled outside traffic noise may be positioned as Field, the telephone rings and tapping of office computers may be positioned as Ground, and the talk of colleagues at a meeting may be positioned as Figure. Van Leeuwen notes that sound recording and replay technologies can manipulate aural perspective. For example in a rock concert situation, recording technologies can position the cheering and singing of thousands of fans as Field, the band accompaniment as Ground, and the apparently soft singing voice of the solo artist can be amplified to Figure.

In most soundscapes however, van Leeuwen theorises that sounds are ‘dynamic’ (van Leeuwen 1999, p. 18), that is, the loudness of a sound can change over time. Sometimes loudness can move us towards or away from a certain position, thus changing ‘our relation to what we hear’ (van Leeuwen 1999, p. 18). Van Leeuwen gives the example of a television documentary showing a Christ statue on top of a hill to the accompaniment of choral church music, then the camera tilting down to the busy expressway at the foot of the hill, and the roar of the expressway faded in to drown the church music, and ‘distanced the audience from the religious sentiments evoked earlier’ (van Leeuwen 1999, p. 18). For the most part, van Leeuwen refers to the ‘dynamics of sound’ in relation to the capacity of loudness, or volume, to identify sounds on a range from loud to soft, whether they are static or changing. In his ‘system network of aural perspective and social distance’ , the term ‘dynamic’ is used as a descriptor to indicate increased or diminished proximity from the listener, as in the examples that follow.

In Hollywood film, for example, the ‘typical aural schema’ is usually to present and represent music and action as background and screen dialogue as foreground, a schema realised by technological manipulation of volume (van Leeuwen 1999, p. 19). Here, audiences are positioned to relate to the most prominent sound, the screen dialogue, and to regard it as most significant. However van Leeuwen provides an example of how a different aural schema can change our perspectival position and thus the relation that we have with
sounds on a soundtrack. He cites Jane Campion’s film *The Piano* (van Leeuwen 1999, p. 19) where the music is Figure, representing throughout the film the inner emotions of loss and longing of Ada (the main character), and the sounds of gushing rain, screaming and the violent actions of Ada’s husband recede into Field. In this film audiences are positioned to adopt an interested attitude towards Ada and her emotions, and disregard her cruel husband. This perspectival and attitudinal hierarchising of sound at any one point in time is realised by dynamics (loudness), but there is also the possibility of dynamics changing over time, as described in the ‘Christ statue’ example above.

Whether sounds are static or dynamic, the perspectival positioning of sounds impacts on the way that listeners will relate to the sounds, and in this case relations of significance are created, causing listeners to understand that some sounds (or actions) are important and other sounds (actions) can be ignored. From a linguistic point of view, the system of aural perspective can be considered as realising Textual meanings, as aural perspective realises relations of salience or ‘relevance’ (Eggins 1994, p. 273), and identifies Theme. In sound, the most salient, or significant sound (in terms of loudness) is placed in the Theme position, positioned as the action that is most relevant to the listener.

**Social distance**

Sounds are produced to carry a certain distance, largely by using the semiotic resources of ‘loudness’ to do so (van Leeuwen 1999, p. 23). van Leeuwen theorises that this combined loudness and distance suggests ‘a set of possible social relations’ between sounds and listeners which may vary from ‘close’ (intimate, personal and informal distance) to more ‘distant’ (formal and public distance) (van Leeuwen 1999, p. 212). In other words, social distance, which refers to single sounds, creates relations of differing ‘degrees of formality’ between the sound that is presented or represented, and the listener (van Leeuwen 1999, p. 15). The single sound may be perspectivally positioned as Figure, Ground or Field, and can be considered as having an ‘absolute’ rather than ‘relative’ level of loudness (van Leeuwen 1999, p. 25).

To analyse and describe various relations of social distance, van Leeuwen engages concepts identified by Hall (1964, in van Leeuwen 1999, p. 24) which link physical distance to the sound of the voice and thus to social distance. For example, Hall suggests that at very close range (3 to 6 in.) ‘the voice will be a soft whisper, and the message ‘top secret’, for the ears of one (very special) person only’ (1964, 1966 pp. 184–5 in van Leeuwen 1999, p. 24). van Leeuwen uses Hall’s research to inform the construction of his own indicators of social distance, but conceptualises the ‘sound of the voice’ differently. He suggests for example that to indicate ‘personal distance’ the sound of the voice needs to include, other than low volume, qualities such as a ‘relaxed’ voice, such as when you open your throat and relax your voice the sound becomes relaxed and mellow, and ‘low pitch’ that is, a low tone of voice resulting from a relaxed voice (van Leeuwen 1999, p. 27). To indicate ‘formal distance’, van Leeuwen also suggests a ‘higher, tenser’ voice is needed, a higher tone of voice (pitch) that results from tension in the throat muscles (van Leeuwen 1999, p. 27). To indicate social distance, van Leeuwen argues that ‘the sound of the voice’ needs to include qualities such as tension and pitch as well as volume, referring to this sound as ‘voice quality’ (1999, p. 25).

Van Leeuwen’s indicators of the ‘system of social distance’ (van Leeuwen 1999, p. 27) are:

- Intimate distance: the relation between the sound and the listener is one of real or imaginary intimacy – what is presented or represented by the sound is
regarded as one would regard someone with whom one is intimate. In speech, intimate distance is realised by whispering of maximally soft voices.

- Personal distance: the relation between the sound and the listener is a real or imaginary personal relation – what is presented or represented by the sound is regarded as one would regard a friend with whom one can discuss highly personal matters. In speech it is realised by a soft, relaxed voice at low pitch and volume.

- Informal distance: the relation between the sound and the listener is a real or imaginary informal relation – what is presented or represented by the sound is regarded as one would regard someone with whom one has a businesslike but nevertheless informal encounter. In speech it is realised by a full voice at somewhat higher pitch and volume.

- Formal distance: the relation between the sound and the listener is a real or imaginary formal relation – what is presented or represented by the sound is regarded as one would regard people to whom one speaks in a formal or public context. In speech it is realised by an overloud, higher and tenser ‘projected’ voice.

- Public distance: the relation between the sound and the listener ‘stretches the limits’ and is regarded as one would regard someone who can only just be reached when one shouts at the top of one’s voice – hence it is realised by maximally loud sound.

van Leeuwen’s concepts of social distance are the tools that he engages for distinguishing between ‘choices’ afforded by semiotic resources of sound, to make meaning about social relations. The ‘system network of social distance’ developed by van Leeuwen distinguishes between these choices, and uses descriptors that refer to social distances in terms of the social relations they invoke, such as ‘intimate’, ‘informal’, ‘public’ (van Leeuwen 1999, p. 30).

The following examples demonstrate how social distance may be realised in ‘normal’ circumstances, where there is no technological recording or replay of sounds. A loud, projected voice which is addressing a crowd, such as a school principal addressing a school assembly of sixty students, may signal that the speaker is placed within ‘formal’ distance from the listener, suggesting that the actual social relationship constructed between speaker and listener is ‘formal’ (van Leeuwen 1999, p. 27). According to van Leeuwen these social relations (realised by the loudness of the sound and the actual distance) can also ‘extend to places and things’ (van Leeuwen 1999, p. 28), and he gives the example where softly lapping water can sound close enough to touch, creating relations of intimacy.

However, van Leeuwen suggests that the technologies of amplification and recording have broken the links between social distance and actual proximity: it has ‘uncoupled the two, and allowed them to become independent semiotic variables’ (van Leeuwen 1999, p. 25). Under these circumstances, social distance is indicated by ‘voice quality’ (van Leeuwen 1999, p. 25). In a film soundtrack for example, a soft breathy whisper that is amplified can stand out clearly against loud drums or brass sections, and is no longer for the ears of one very special person, but is audible to many. Here the sound is recognised as a soft breathy whisper by the voice quality of ‘breathiness’, that is the sound produced when one breathes and speaks at the same time; and the ‘relaxed’ quality of the voice (van Leeuwen 1999, p. 133), so that an intimate social distance is maintained. Van Leeuwen claims that the same ‘scales of social distance could be constructed for musical and non-musical sounds’ such as engines, and points to the examples of the put-put boat and the
screaming racing car or jet (van Leeuwen 1999, p. 25). The implication here is that the put-
put engine would be identified by one set of sound qualities and that the screaming jet
engine would be identified by a different set of sound qualities. According to van Leeuwen’s
suggestion, the sound of the screaming jet engine can be technologically manipulated to a
‘comfortable’ level yet still be recognised as a screaming jet engine by its sound qualities,
thus maintaining a public distance with the listener. In music, van Leeuwen asserts that the
 technological manipulation of the singing voice, which is often achieved by ‘close miking’,
can similarly engineer social distance. He provides the example of the ‘crooning style of
Bing Crosby’ which was designed to create an ‘intimate, personal relationship with fans’
even if the suggested relationship was an imaginary, and not a real one (van Leeuwen 1999,
p. 25). Van Leeuwen stresses that amplification allows singers to ‘address us as though they
have a personal relationship with us’, even though they may be physically remote (van
Leeuwen 1999, p. 25).

Analysis of aural perspective and social distance in Phase 2 of The Queen

This final section presents an analysis of aural perspective and social distance in a section of
The Queen film trailer (2006), demonstrating an application of van Leeuwen’s analytic
concepts and procedures to investigate sound perspective in film. The film trailer was
selected as a model for analysis, as it represents the kind of multimodal film text that is
accessible from the Internet, is interesting to students, and provides opportunity for critical
study. The Queen film trailer was released in 2006 by Mirimax Film Corporation, prior to
the release of the feature film in 2007. The feature film is about how the Royal Family, and
the Queen in particular, responded to the death of Diana, Princess of Wales. In the
following example of analysis (Table 1), Phase 2 of The Queen film trailer, the Queen
announces that no member of the Royal Family would make a public statement about the
death of Diana, as Diana was no longer considered a member of the Royal Family following
her divorce from Charles. Charles questions this reasoning, and Blair, the Prime Minister, is
astounded at the Queen’s decision for silence when the British public clearly wish the
Queen to speak to them.

In a departure from van Leeuwen’s scripting for notation, sound effects, speech and
music are annotated separately in this analysis, to enhance the visual display of sounds that
occur simultaneously. As well, Baldry and Thibault’s notation of a downward-pointing
arrow ↓ is engaged to indicate a ‘continuing’ sound (2006 p. 215). Italics have been used
to annotate music in terms of perspective and social distance, to visually differentiate them
from speech annotations.

In Phase 2 of The Queen film trailer, there are two ‘voices’ that alternate on the
soundtrack: the musical ‘voice’ and the spoken ‘voice’ of each character. One ‘voice’ is
made significant to the listener as the loudness is increased and aural information is
delivered, then faded down to accommodate the next prominent ‘voice’. Instrumental
music (a group of strings) ‘illustrates’ the verbally told story, and at the beginning of Phase
2 the music, as Figure, presents the anguish of the mourners and the actions of the Queen
as dramatic. Music then continues to be perspectivally postioned as Ground, while the
Queen, Charles and Blair express their views and develop the verbal theme of the conflict
created by Diana’s death. The speech of the Queen, Charles or Blair is always loud, and
perspectivally positioned as Figure, so that the listener treats each speech as important.
Cymbal clashes, shimmers and drumbeats occur in Ground position in between the speech
of each character, acting as a kind of ‘bridge’ between speakers (Goffman 1974, p. 147 in
van Leeuwen 1999, p. 32), and linking the action of one speaker with another. This analysis
reveals the aural schema of film trailers such as The Queen, with the speech of main
characters assuming Figure position and foregrounded as important, while the music for the
most part is positioned as Ground, except for occasions where the musical ‘comment’ is
positioned to be prominent and therefore relevant to the listener.

With the speech and music in alternating perspectival positions, the speech and music
also assume imaginary intimate social relations with the listener. Speech is always either
closely miked or amplified in some way so that the sound of the voice (in this case it is
mostly loudness) enhances a sense of closeness with the listener, and intimate, or close social
relations are assumed. In this text, voice qualities other than loudness are not a necessary
consideration for annotations of social distance which refer to speech. When amplified
instrumental music is introduced as Figure into the film trailer at the beginning of Phase 2,
the music is a single voice setting the ‘emotional temperature’ (van Leeuwen 1999, p. 111)
that listeners are ‘positioned’ to treat as significant, and to assume a close relation with.
When the instrumental music is reduced in volume to Ground, allowing speakers into

<table>
<thead>
<tr>
<th>Phase</th>
<th>Sound effects</th>
<th>Speech</th>
<th>Music</th>
<th>Perspect</th>
<th>Social Dist</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHASE 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>42-59</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>42</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>43</td>
<td></td>
<td>(Queen) 43 No member of the Royal Family will speak publicly about this</td>
<td></td>
<td>Figure</td>
<td>close</td>
</tr>
<tr>
<td>46</td>
<td></td>
<td></td>
<td>Instrumental music</td>
<td>Figure</td>
<td>close</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ff fast high pitch</td>
<td>Ground</td>
<td>med</td>
</tr>
<tr>
<td>Phase 2 b</td>
<td></td>
<td></td>
<td>1</td>
<td>Figure</td>
<td>close</td>
</tr>
<tr>
<td>47</td>
<td></td>
<td>47 Diana’s no longer a member of the Royal Family</td>
<td></td>
<td>Figure</td>
<td>close</td>
</tr>
<tr>
<td>49</td>
<td></td>
<td>(Charles)</td>
<td></td>
<td>Figure</td>
<td>close</td>
</tr>
<tr>
<td>50</td>
<td></td>
<td>50 What are you talking about?</td>
<td></td>
<td>Figure</td>
<td>close</td>
</tr>
<tr>
<td>51</td>
<td></td>
<td>(Queen) 51 Charles this a private matter</td>
<td></td>
<td>Figure</td>
<td>close</td>
</tr>
<tr>
<td>52</td>
<td></td>
<td></td>
<td></td>
<td>Ground</td>
<td>med</td>
</tr>
<tr>
<td>Phase 2 c</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>53</td>
<td></td>
<td>(Queen) 53 We do things in this country # quietly # and with dignity</td>
<td></td>
<td>Figure</td>
<td>close</td>
</tr>
<tr>
<td>54</td>
<td></td>
<td></td>
<td>Instrumental music</td>
<td>Ground</td>
<td>med</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ff fast high pitch</td>
<td>Figure</td>
<td>close</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Instrumental music</td>
<td>to Ground to fade</td>
<td>close</td>
</tr>
<tr>
<td>Phase 2 d</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>57</td>
<td></td>
<td>(Blair) 57 Will someone please save these people from themselves?</td>
<td></td>
<td>Figure</td>
<td>close</td>
</tr>
<tr>
<td>58</td>
<td></td>
<td></td>
<td></td>
<td>Ground</td>
<td>med</td>
</tr>
</tbody>
</table>
Figure position, listeners assume an informal relation with the music but are aware of its continuing commentary. Similarly, the sounds of cymbals, shimmers and drumbeats negotiate an informal social distance with listeners that nevertheless ensures that listeners orient to dramatic points made by the speakers.

The only time that this aural pattern is interrupted is at the end of Phase 2c, where the music stops completely, and silence is heard. Van Leeuwen does not elaborate on this aspect of sound. Other researchers have regarded silence as the absence of sound that is expected (Baldry & Thibault 2006, Pun 2006) suggesting that silence indeed has a role in negotiating meaning. In Phase 2d, silence is noted as the absence of music, which perhaps points to the surprise felt by audiences as the Queen resists the call for her to join the public mourning. However, the speech of Blair occurs as loud, and in isolation, and perhaps the silence is to accentuate his exasperation. The impact of silence as an aspect of the soundtrack is yet to be clarified in research, especially in relation to the social distances it may engineer.

Conclusion

Analysis of a section of the soundtrack of The Queen film trailer demonstrates that it is possible to identify how sound can construct aural perspective and relations of social distance between a listener and a sound, using the analytical concepts, tools and procedures modelled by van Leeuwen in Speech, Music, Sound (1999). Speech and music were configured as alternating turns in terms of loudness, so that the listener was positioned to treat each as important, and thus relevant to the film trailer story. Selective amplification of speech and music engineered relations of close social distance with the listener, so that the listener would assume an imaginary intimate relation with what was being ‘said’ with speech and music, and in the case of speech, who was saying it. By identifying aural perspective and social distance between sounds and the listener, van Leeuwen’s methodologies appear to successfully reveal part of the purposeful design of meanings made with sound, in multimodal film texts.

The theoretical modelling of sound as a social semiotic has pedagogical implications. Teachers can frame their thinking and teaching about sound in a way that articulates with the current theorising of language and image as multimodal semiotics. Students can learn to conceptualise and describe particular semiotic resources of sound such as loudness, for example, and use these descriptions to discuss the possible meanings that loudness realises in film texts. It remains, however, for further research to validate the application of van Leeuwen’s particular theoretical perspective in classroom contexts. It remains also for research to further explore the theoretical modelling of sound so that a rigorous basis is established that frames the teaching of multimodal literacy.

References


McDonald, E. (2003) *Sound as embodied behaviour: towards a model of music as a social semiotic system*.


