Chapter 8

Beyond the visual: applying cinematic sound design to the online environment

Mark Ward and Linda Leung

Cinema is an alchemical interaction between its three constituent elements of sound, visuals and story. To be 'cinematic' is to combine these elements in ways that make the impact of the resulting experience larger than the sum of its parts. A movie might have lush production design and lighting, lyrical camerawork, sensual soundscape, or an engaging story, but in a 'cinematic' work these distinctions tend to recede to give way to an overarching feeling of salient wholeness. The 'cinematic' experience seems to defy dissection – to talk about one element is difficult without talking about all other elements – and it is this level of consolidation to which digital designers aspire when creating user experiences. Over time, cinema has refined the ability to design such holistic experiences, which seduce and immerse its audiences into the filmic world, engaging them emotionally and convincing them to suspend their disbelief.

However, while the mechanisms of engagement and immersion have been studied extensively in terms of story and visual image, the role of sound has enjoyed less attention in film theory. In spite of this theoretical blind spot, the film industry has nevertheless created an entire discipline of sound design. By comparison, the consideration of sound in particular types of digital interactive experiences is still in its infancy, perhaps equivalent to cinema's 'silent' era. Therefore, what can designers for computer screens learn from techniques of the silver screen in terms of sound design? How might sound contribute to the emotional life of digital interactive experiences as it does to a film?

In this chapter, we compare the development of sound design in cinema with that of digital interactive environments, ranging from computer systems to CD-ROMs to Web sites. Along the way, we examine the types of sonic problem-solving and conceptual tools used by filmmakers in the hope that digital designers need not reinvent, but rather benefit from a century's cinematic sound practices.

Loud and clear: making the historical case for sound

After a century of cinema how can it be that sound design continues to operate in an almost theory-free vacuum? The language and vocabulary of sound is not as sophisticated as for the moving image, and yet sound informs and anchors the moving image. Film is an *audio*-visual (AV) medium. In the term 'AV', audio precedes the visual: that is, not only is sound integral, it is prioritised. This reflects the developmental history of media, whereby the rise of radio and recorded music industries meant that audiences were already well practiced at mediated listening before their initiation into film.

Early cinema is commonly perceived as 'silent' (and thus a visual medium only) because no sound information was printed upon the filmstrip. In fact, a vast array of aural strategies were marshaled around the projected visual image, with many valiant but failed attempts to synchronise picture and sound (Abel and Altman 2001). For example, specific scores were composed and 'effects men' - even troupes of actors - were employed to provide a live accompaniment to the vision track. There was Thomas Edison and Eduard Muuybridge's joint venture that sought to fuse Muuybridge's moving-image system with Edison's gramophone. In pursuit of the holy grail of a single system of sound and image projection, competing systems appeared such as Edison's Kinetophone (which appeared first in 1895 and then in a re-designed form in 1913) and Gaumonts' Chronophone (1902). These proposed technology solutions to synchronizing sound with image were expensive, with theatrical venues reluctant to install costly sound replay systems, while a myriad of non-standardised equipment locked exhibitors into proprietary systems that reproduced sound of low quality. To place film sound in this historical and technological context goes a long way to explaining why cinema has been theorised as an intrinsically silent visual medium.

From this vantage point, similarities with the practice of sound design in digital interactive experiences become apparent. As in the beginnings of cinema, early online experiences of sound have been contained within the boundaries of the lowest technical denominator. Until relatively recently, the online experience was often disturbed by buffer under-runs and visual stutter, streaming video that displayed picture and sound out of sync, and poor sound quality due to bandwidth and compression issues. All of these instances are analogues of problems faced in cinema's early development. Likewise, digital experience designers could not rely on the

availability of appropriate technology for the distribution of sound. Therefore, the biases of cinema in privileging sight over hearing, and in being predisposed to visual dominance, have been repeated in digital interactive (and particularly online) media.

'The neglect of sound in mainstream interface design is striking. This can be partially accounted for by the lack of understanding about how sound is processed in the real world, and lack of inspiration about how to use sound innovatively.' (Macaulay et al 2000: 161)

Early innovations in the use of sound in digital interactive environments were made by Apple Computer, which deployed personal computers into the marketplace equipped with sound-cards, as well as plug-in or built-in microphones allowing user-generated material to be recorded and manipulated, or linked to other documents. Most importantly, however, Apple's Macintosh user interface was conceived of in visual and aural terms, embedding William Gaver's work in *auditory icons* (1986) into its philosophical core.

If interactive media can be seen as part of the historical trajectory of cinema, its industries can adapt the lessons, rather than reproduce the mistakes, from the past in terms of sound design. To echo the call-to-arms of *The Jazz Singer* (the first 'talking' film), "You ain't heard nothin' yet".

Sound and vision: how it works cognitively and experientially

Understanding how sound contributes to the emotional design of a film or digital interaction requires knowledge of the processes by which our senses gather data, translate this into perception, then make sense of those perceptions by integrating them into a coherent reality. The glue that binds this sensory data into perception, and then perception into narrative is emotion.

Emotion focuses on a particular object or person (Grodal 1997). They are specific and are informed by a reason: anger, hate, fright, fun, grief, panic. Emotions are normally functional; that is, they tend toward producing an action, such as 'fight or flight'. Emotions are dynamic, variable and very brief, lasting between 2 and 8 seconds. Beside bring short, they are intense, and unable to be altered whilst occurring. If the emotion system is sufficiently activated, the processed data will galvanise into a perception and percolate into our consciousness, thus

becoming available to our attention (Tan 1996). This means that designers and filmmakers have only a short amount of time to harness this emotional activation and engage their users and audience respectively.

In terms of evolution, dangerous and unpleasant experiences are more important to remember than pleasant ones, and so memories of our negative experiences are more strongly encoded, and therefore more easily recalled. A few brief seconds of intense emotion can set a mood lasting hours or even days. Filmmakers have become adept at instigating negative emotions and setting black moods both intentionally (as in horror movies where uncomfortable feelings are deliberately triggered) and unintentionally (through poorly made films). Similarly, digital experience designers have been guilty of unconscious, negative emotional design, as in when users become frustrated by a DVD interface that is difficult to navigate or a Web site that is so disappointing that they refuse to return. As Donald Norman (2004: 119) asserts in his book, *Emotional Design*, sound can also harm an experience: noise, especially, is a source of negative affect. Indeed, in addition to negative emotional responses, noise also has the capacity to cause actual physical pain.

LeDoux (1998) has described the pre-conscious processing of sound that occurs before it arrives in our awareness. Sound firstly reaches the 'ancient' part of the brain, the amygdala. This is a quick-and-dirty processor that allows reaction to a sound before it is 'heard'. It is a direct conduit to the emotions, capable of triggering the 'fight-or-flight' response. The sound also proceeds in parallel to the second processing centre, the relatively 'modern' auditory cortex. It takes twice as long to get to the cortex as it does to get to the amygdala. The auditory cortex 'listens' to what has been heard, employing cognitive scripts and ideas about what to do with the sound data. It is here the amygdala's response is either accepted or rejected. The effect of this double processing is that we emotionally *feel* a sound before we *know* we've heard it.

Paradoxically, sound is perceived in visual terms. In the mid-1990s the MIT Perception Labs conducted a number of experiments to determine the effects of sound quality on the perception of picture quality. The results of the experiments revealed that those watching standard

resolution TV with CD-quality sound identified this combination as superior in visual quality to a system employing HD quality images with lo-fi sound. [citation needed here]

Other examples of this cross-modal interaction can be seen in the work of Dutch cognitive psychologist Béatrice de Gelder. De Gelder conducted a multitude of experiments examining 'audio-visual fusion', a psychological mechanism whereby the modalities of hearing and vision fuse into one perception. Because of the principle psychologists call 'visual dominance', we perceive this fusion as essentially visual. In 2000, de Gelder published findings of an experiment where two groups of people were shown two discs of light, one moving left-to-right, and the other moving right-to-left so they seemed to pass through each other. Those who saw the discs mute saw the discs pass through each other. The group who saw the same visual action but with an auditory cue at the point where the discs overlap perceived the two discs to bounce off each other and travel back the way they came. [citation required here]

Another of de Gelder's experiments set up a series of light pulses accompanied by synchronised auditory cues. The subjects were to indicate when they became aware of the appearance of the visual flash. Whilst the visual sequence ran, the auditory bleeps were sped up or slowed down. The effect on perception was the sequence of visual flashes appeared to speed up and slow down in sympathy with the auditory cues. Both experiments demonstrate that sound controls and changes visual perception or what we think we see. [citation required here]

If sound design is a form of emotional or sensorial design, how can we describe its functions? While Chion (1994) and Holman (1997) have described a range of functions of sound on screen, we propose four that are particularly relevant to both film and interactive media:

i)   Emotional truth
ii)   Point-of-View (Point-of-Audition)
iii)   Storytelling
iv)   Physical experience

Arguably, the primary function of sound in both film and interactive media is the structuring and communicating of an ***emotional truth***. Emotion is the prism through which all other information is filtered and so should be the guiding principle by which a work is created – that is, it is the organising principle at the core of the work. This function harnesses the pre-attentional ability of sound to create mood, orienting the audience as to how visual information is to be received or interpreted. It is the key that the audience uses to make sense or meaning, and a tool for the user to check the veracity of the message they are receiving. It happens before rational logic, is pre-conscious and visceral and therefore has the capacity to affect us in a primal way.

Sound sets emotion, which in turn, sets mood. Mood can be described as a general state of mind without reference to a particular object. Mood is not directly connected with actions, and is more passive and basic than emotion. Mood is a more stable and longer-lasting process than emotion, allowing us to orient ourselves consistently to our environment without being on an 'emotional rollercoaster'. Mood smoothes things over, connecting our moment-to-moment existence so that the environment does not appear to be constantly changing around us. Mood is the afterglow of emotion, and it orients us toward experiencing the next similar emotion. As a result, mood sustains the likelihood of experiencing emotion, and the experience of emotion sustains mood (Frijda in Tan 1996: page?). Mood can be likened to the 'energy' of a film, or the 'feel' of a digital experience.

Sound also functions to establish ***point-of-view***, *or in this case* ***point-of –audition***. In film, point-of-audition has the role of simulating a character's feelings or thoughts by the choice of detail that is presented. Sound steers the audience through the available visual information, promoting what is important, and in which order. In a corporate Web site, this can be represented as the 'voice' of an organization, or it can create an acoustic environment or sense of place reflecting the 'buzz' of a company. Other digital experiences might allow sonic customization, whereby a user can self-select a mood, or express their taste or point-of-view, such as choosing either a male or female voice to provide spoken navigation instructions on an in-car GPS system.

The third function of sound is to structure ***storytelling***. This is the case whether they are rich, complex stories as told in live-action film, or the simple, key messages of a Web animation.

Sound 'completes' the picture to legitimise it as a version of the real world; that is, it relates the screen-world to the offline world. It also acts as a form of grammar connecting or separating visual blocks to create new units of meaning. Sound has the role of a binding or unifying element, similar to connective tissue. This is also true in a cross-media sense, where a television advertising jingle, when used also in a Web site, can enable immediate cross-media recall and brand recognition.

The fourth and final function of sound is *physical experience*. It contributes to the overall multi-sensory experiences that are expected in cinema and, we argue, in interactive media. These are audio-visual media in which sound affects vision, and vision affects sound. In addition, digital media generally requires tactile interaction, and so encompasses touch as well.

'Hearing is a way of touching at a distance' (Schafer cited in Cranny-Francis 2005: 59).

This is where sound can have a synaesthetic role in interactive media, where sound is felt as much as it is heard, as can be seen in video games where explosive sounds and images are accompanied by vibrations in the console. That the senses of hearing and touch are so closely aligned provides a compelling argument for sound design to be a critical component of interactive (that is, tactile) media development.

An anatomy of a sound design

Over the course of the last century, the design of sound in cinema has evolved into a specialised craft organized around four major aural categories:

- music
- ambient sound, or atmospheres
- sound effects
- dialogue (Murch, sound designer for *Apocolypse Now*, cited in Cranny-Francis 2005: 76)

Each of the above is deemed a discipline in its own right with its own artisans, spawning such roles as: production sound recordist, and boomswinger; ADR recordist; foley artist, and

recordist; editors for dialogue, ADR, foley, effects and atmospheres; sound designer; composer; and sound mixer.

The role of **music** in creating a sense of space and place is already well known. Both Cubitt (1998: 97) and Bull (2000) liken music to architecture in terms of how it is used to structure daily life through the construction of acoustic space. In cinema, this additionally serves as a storytelling function, not only in helping to express the world within the film, but also by articulating the emotional status of characters. According to Norman (2004: 115), the affective states produced through music are cross-cultural.

'The proliferation of music speaks to the essential role it plays in our emotional lives. Music acts as a subtle subconscious enhancer of our emotional state throughout the day. This is why it is ever-present…' (Norman 2004: 119)

Thus, music operates at both macro and micro levels of film. It can also be sub-categorised according to *source* (music from within the story-world), or *score* (music from outside the story-world). Likewise, this can be applied to the construction of online spaces, such as virtual worlds or corporate Web sites. A corporate Web site can be regarded as a world unto itself with its own characters or protagonists. Music is often part of the branding of large corporations in their broadcast media marketing (such as in advertising jingles), but is oftentimes overlooked in their online presence. Music has the capacity to maintain cross-media consistency in corporate branding as well as to quickly establish the qualities of a company's culture to the user. It is now also standard practice to highlight individuals and their roles within an organization on their Web site, and music can function as a signifier of the emotional life or 'vibe' of a company as well as the personalities of the people that work there.

This potential for music to affect experience lies in its impact on all 3 levels of brain processing:
1. Visceral in the initial pleasure of the rhythm, tune and sounds
2. Behavioural in the enjoyment of playing or listening
3. Reflective in the activity of analysing melody and lyrics (Norman 2004: 115).

Just as film scores can be enjoyed in their own right, digital interactive experiences can be similarly layered to allow music to take precedence over the visual image. Generally, a user can turn the sound or off on a Web site: perhaps it should also be possible to fade or 'mute' the visual. Nonetheless, as is evident from film, when music is combined with the visual image in ways that also affect the brain viscerally, behaviorally and reflectively, deep and memorable experiences are evoked. When the brain is influenced on all these levels, it suspends its disbelief, accepting the differences between its own world and the new one created for it. The user is persuaded to stop, look and listen.

The world within a film is not only maintained through music, but also **sound effects** and **ambient sound**. These sonic elements may complement or contradict the objects or actions within the frame or, indeed, the entire filmic world. In one sense, these elements are designed to mimic or reflect the auditory landscape *in real life* (IRL), and thus to sustain the fiction of the world on the screen, aiding audiences to suspend their disbelief. On the other hand, sound effects may are also designed to comment upon and critique the film-world.

Sound effects for film are often described as 'hard effects' as they are sounds literally cut hard in sync with the picture. But the definition of sound effects stretches to also include *foley* and *atmospheres*. Foley is usually recorded in a sound studio to picture playback and usually refers to body movement, such as clothing, footsteps, and other subtle motions. Foley is also a useful technique to record 'unnatural' sounds. It adds layers of 'reality' to actions or objects. Atmospheres, or ambient sounds, sonically represent the environment of the scene, such as distant traffic or wind through trees and telephone lines. Usually such ambient sounds contain information which construct and frame the emotional tone of the scene. All such sound can be 'real' or 'unreal' but most are not a literal translation of what one would hear in the visual location and situation. There is always some form of selection and filtration.

In terms of digital interactive environments, sound effects and ambient sound can be similarly applied whether it is a wholly virtual experience or representative of an offline product or experience. As mentioned previously, just as music can help depict a company culture, sound effects and ambient sound can portray the 'buzz' of the company. Web sites for nightclubs,

restaurants, cafes and bars can also benefit from this kind of emotional tagging, and while a Web site may not allow the user to taste the food or smell the coffee, it may provide some sonic and visual indication of the ambience, how busy and patronized the establishment is, or what a patron might expect. Sound design may even be able to evoke taste and smell synaesthetically. Maintaining consistency and meeting expectations is crucial to the user continuing to suspend their disbelief and sacrifice their time (whether willingly or impulsively) to partake in the world you have constructed for them.

It may be that the minimal expectation of sound on the part of users is the reason behind the general neglect of sound by digital experience designers. However, transcending expectations can also be the source of innovation. For example, Norman (2004: 53) describes a project that introduces sound to photography as a way of enhancing the experience of taking and looking at photos. HP labs in Bristol have been developing 'audiophotographic' photos that combine an audio track which captures:

- Ambient sounds recorded the moment the image is taken, assisting in recall of the event
- Music evoking memories of the event
- Spoken word, allowing interpretation or narration of the event.

Of all the components of sound that can be designed, **dialogue** or the spoken word is probably the least expected by online users. While it is often the major sound consideration in films, what relevance does it have to the online experience? Film dialogue is the product of a script that has undergone numerous iterations. The script editor must decide how to refine and condense the dialogue to within the movie length limit of approximately 120 minutes by dispensing with anything unnecessary to progressing the story efficiently. In the early days of multimedia design, following a film production model, there was reference to scripts to detail the content of CD-ROMs, with scriptwriters contracted to develop these (England and Finney 1999: 120-123). Interestingly, the legacy of this film production model meant that there was also more attention to sound design in the development of multimedia CD-ROMs and this has been inherited and extended by the games industry. But as digital media production became more Web-centric, the

methodologies evolved and scripts became content inventories. In Web development, what were traditionally constituent elements of a script (story, visuals and sound) fall under the umbrella of 'content' and largely refers to text. Thus, the content of a Web site can be equated with dialogue in film in that it has to be carefully crafted to tell a particular story or convey specific messages. In film, dialogue is literally between characters. In the online domain, dialogue could be conceived of as the conversation an organization wants to have with their customers, and the same questions asked by film script editors apply: what is the story (or message) that needs to be told? Is this content necessary to the story or message? How long is the user expected to engage with the content? Is the content engaging enough to hold the user for the expected amount of time? This sort of dialogue between company and customer does not necessarily have to involve sound. However, one example includes interactive phone systems which are already commonly used. While there is less evidence of this kind of oral / aural real-time interaction with synthetic agents online, perhaps it will become more widespread with the introduction of voice-command interface technology such as Fonix Speech *VoiceIn* (www.fonixspeech.com). By defining the parameters of verbal interaction, one could be said to be 'writing the dialogue script', providing flexibility in response to user demands and allowing a connection to the audience on a level where emotional performance becomes an active ingredient.

On a more literal level, dialogue as spoken word also has possibilities that have not been thoroughly explored in the online environment. Arguments for improving the accessibility of Web sites have included providing an audio alternative to screen-based text for people with visual impairment (see W3C Web Content Accessibility Guidelines). As many visually-impaired Internet users have screen readers which can then be output to speech synthesizers, this means that Web content should be designed not only for how it *reads* but also for how it *sounds*.

Summary
After a century of trial-and-error in film sound production processes, the interactive media industry can now cherry-pick the best of what the last 100 years of cinematic sound design practice has to offer and apply these techniques to the digital environment. Although sound design mechanisms in film have become nuanced and specialized, their capacity to compel the viewer/listener through aural seduction presents potentially powerful lessons for digital

designers, in spite of the constraints of experiencing sound on a computer screen compared with the cinematic silver screen. Sound effectively draws users into its world as 'we have eyelids but not earlids' (Welsch cited in Cranny-Francis 2005: 71): that is, we can block out the visual merely by closing our eyes, but it is much more difficult to escape the aural.

As we have detailed above, sound design is a form of emotion design. That is, it uses the design of sonic elements to elicit mood and emotions *from* its audience, or it represents those moods and emotions *to* its audience. The four functions of cinematic sound design that can be leveraged for digital experience design are:

i) **Emotional truth**

- Pre-conscious processing of sound creates mood, and mood is the filter through which all visual information passes
- Sound provides the key as to how the visuals are to be received and interpreted by the audience
- Sound emotionally encodes the content for meaning

ii) **Point-of-Audition / Point-of-View**

- Sound steers an audience's attention through visual information, promoting what is important and in which order, simulating perception and emotional labeling
- Sound reflects a point-of-view

iii) **Storytelling**

- Sound connects visual elements to create cinematic phrases, sentences, paragraphs, and ascribes meaning to each
- Sound 'completes' the visual image by creating a verisimilitude, thus aiding the audience to suspend their disbelief

iv) **Physical experience**

- As a result of cross-modal interaction, sound modifies the perception of weight, texture, and speed of on-screen objects, lending the visual elements legitimacy as a truthful version of

a concrete world (which has important application in 3D virtual reality environments)

- Hearing is a form of touchas it has a synaesthetic quality

- Sound can maintain the energy level of a sequence even if the visual elements seem 'slow' or 'empty'

- An immersive soundfield provides the basis for the visceral experience of spectacle, and the sublime experience of wonder.

References and Recommended reading

Abel,R. and Altman,R. (2001*) The Sounds of Early Cinema*. Bloomington, Indiana: Indiana University Press.

Bull,M. (2000) *Sounding Out the City: Personal Stereos and the Management of Everyday Life*. New York: Berg.

Chion,M. (1994) *Audio-Vision: Sound on Screen*. New York: Columbia University Press.

Cranny-Francis,A. (2005) 'Chapter 4 – Sound' in *Multimedia*. London: Sage Publications.

Cubitt,S. (1998) *Digital Aesthetics*. London: Sage.

England,E. and Finney,A. (1999) *Managing Multimedia: Project Management for Interactive Media*. Harlow: Addison-Wesley.

Fonix Speech *VoiceIn* [Online] Available:
http://www.fonixspeech.com/index.php

Gaver,W. (1986) 'Auditory Icons: Using Sound in Computer Interfaces', *Human-Computer Interaction: Journal of Theoretical, Empirical, and Methodological Issues of User Science and of System Design*, 2(2), pp167-177.

Grodal,T. (1997) *Moving pictures: a new theory of film genres, feelings, and cognition***.** New York: Oxford University Press.

Holman,T. (1997) *Sound for Film and Television*. Boston, Mass.: Focal Press.

LeDoux,J. (1998) *The Emotional Brain: the Mysterious Underpinnings of Emotional Life.* New York: Touchstone.

Macaulay,C.; Benyon,D. and Crerar,A. (2000) 'Voices in the Forest: Sounds, Soundscapes and Interface Design', *Towards a Framework for Design and Evaluation of Navigation in Electronic Spaces*, [Online] Available:
http://www.sics.se/humle/projects/persona/web/littsurvey/ch10.pdf

Norman,D. (2004) *Emotional Design*. New York: Basic Books.

Tan,E. (1996) *Emotion and the Structure of Narrative Film: Film as an Emotion Machine*. Mahwah, NJ: Erlbaum.