Soundscape Mapping
Comparing Listening Experiences

Iain Peter McGregor

Submitted in partial fulfilment of the requirements of
Edinburgh Napier University for the degree of Doctor of
Philosophy

December 2010

Abstract

The perceived auditory environment is an increasingly important part of people's everyday interactive experiences. While sound design is an established discipline in media such as video games and cinema, this is not the case in Human-Computer Interaction (HCI). HCI designers are rarely trained in sound design, and may not make the most effective use of sound in the design of interactions. Even when sound is at the centre of a design it is rarely evaluated to compare the experiences of designers and listeners. This dissertation reports work conducted to develop a way of comparing sound designers' intentions for a sound design with the experiences of listeners.

Literature on methods of measuring, classifying and visualising sound was reviewed, as well as approaches to sound design in different forms of media and computing. A published method for representing auditory environments was selected for preliminary studies. The four studies addressed to the difficulties of describing auditory environments and how they might be visualised. Two surveys were conducted in order to identify attributes of sound that would be meaningful to 75 audio professionals and 40 listeners. A way of classifying and visualising sound events and their distribution in physical environments was developed and evaluated.

The soundscape mapping tool (SMT) was trialled with sound designs from a range of fields within media and computing. The experiences of both the designer and listeners were captured for each of the designs using the SMT. This work demonstrated that the SMT was suitable for capturing the intentions of 10 sound designers and the experiences of 100 listeners. The trial also provided information about how the SMT could be developed further. The dissertation contributes evidence that auditory environments can be abstracted and visualised in a manner that allows designers to represent their designs, and listeners to record their experiences.

Acknowledgements

I would very much like to thank my supervisory team for their excellent support throughout this project. My Director of Studies Professor David Benyon for his calm insightful guidance and Dr Phil Turner for his immeasurable insight and patience. As well as Dr Alison Crerar and Professor William Buchanan for their timely advice and support.

Within the School of Computing I am extremely grateful for the opportunity to collaborate with such excellent researchers as Dr Susan Turner, for some of the data gathering stage (concurrent verbalisations), and Dr Grégory Leplâtre for the first sound design illustration. I would also like to thank all of my colleagues, as well as the numerous participants for being so generous with their time.

I am very appreciative for all of Ian McLeod's grammatical corrections, and his willingness to read and re-read iterations.

I would like to finish by thanking my wonderful wife, Libby McGregor, who makes everything possible, as well as my fabulous little family: Jade, Holly, Coco, Emmeline and Bella, who make everything worthwhile.

Associated publications

McGregor, I., Crerar, A., Benyon, D., & Leplâtre, G. (2006a). Workplace Soundscape Mapping: A Trial of Macaulay and Crerar's Method. ICAD 2006. London: Department of Computer Science, Queen Mary, University of London.

McGregor, I., Crerar, A., Benyon, D., & Leplâtre, G. (2006b). Mapping Workplace Soundscapes: Reifying Office Auditory Environments World Forum for Acoustic Ecology. Hirosaki, Aomori, Japan: Soundscape Association of Japan.

McGregor, I., Crerar, A., Benyon, D., & Leplâtre, G. (2007). Establishing Key Dimensions for Reifying Soundfields and Soundscapes from Auditory Professionals ICAD 2007.

McGregor, I., Crerar, A., Benyon, D., & Leplâtre, G. (2008). Visualising the Soundfield and Soundscape: Extending Macaulay and Crerar's 1998 Method Proceedings of the 14th International Conference on Auditory Display. Paris: IRCAM.

McGregor, I., Leplâtre, G., Crerar, A., & Benyon, D. (2006). Sound and Soundscape Classification: Establishing Key Auditory Dimensions and their Relative Importance ICAD 2006 London: Department of Computer Science, Queen Mary, University of London.

McGregor, I., Leplâtre, G., Turner, P., & Flint, T. (2010). Soundscape Mapping: a tool for evaluating sounds and auditory environments The proceedings of the 16th International Conference on Auditory Display (ICAD-2010) Washington D.C., USA: International Community for Auditory Display (ICAD).

Table of contents

Index of figures

Index of tables

1 Introduction

Sound is one of the easiest ways to augment any environment, and has been used extensively as a method of communicating information, whether intentionally or not (Delage, 1998). The skills of a sound designer are often required to create sounds that neither distract nor induce fatigue (James, 1998). At present this expertise is not easily conveyed, being based upon years of practical experience and knowledge transferred through apprentice style training, rather than readily communicated in published texts (Yewdall, 1999, Hannan, 2003). Brewster (2008) raised the issue of the vast majority of information within computing still being conveyed visually, despite successful research into the use of non-speech sounds going back to the early 1990's. Robare and Forlizzi (2009) highlighted a lack of design theory regarding guidelines for sound design within computing, despite the number of products which replay sound having increased dramatically since 2000. Robare and Forlizzi go on to argue that the way in which interaction designers have used sound has not improved. They proposed that the situation was unlikely to become better until sound design was part of the concept stage of software design (ibid.).

Sound designers have an extensive understanding of the ways in which sound can be used to manipulate listeners' experiences, which for example within the film industry has been formalised by authors such as Altman (1992) and Chion (1994). Presently sound designers only have a limited number of techniques to understand what pre-exists in the variety of auditory environments that their sounds inhabit. There are also a restricted number of methods with which sound designers can formally measure their design's impact. Designers have tried to mitigate the effects of pre-existing sounds and different types of environment on sound designs rather than evaluate these effects and adjust sound designs accordingly. There has been a reliance on either acoustic isolation for auditoria, volume control for private environments or headphones for public environments (Bull, 2000). Techniques for evaluating sound include simple noise pollution measurements (American National Standards Institute, 1994), and elicitations of semantic interpretations from listeners

(Ballas & Howard, 1987), or what Bohme (2000) refers to as 'object-orientated' descriptions and Metz (1980) terms the 'sound of what'. Listening tests are commonplace in the field of product design where experienced listeners (those who have previous experience of listening tests) are preferred, the results of which have been published since 1956 (Bech, 1992, Soderholm, 1998, Engelen, 1998). However, listener testing is limited to products such as audio reproduction equipment and vacuum cleaners, and has not migrated into mainstream media, and only partially into computing (Bech and Zacharov, 2006).

Barrass and Frauenberger (2009) state that in the field of auditory display design, designers need to consider the end-users and the context of use, as auditory displays might be used in a wide variety of environments. An individual inhabits a unique soundscape, based on their previous experiences and interests, and as such will provide unique responses, (Dubois, Guastavino and Raimbault, 2006). Maps created by multiple inhabitants could provide an insight into the typical versus the individual experience. The designer's perspective could be compared to that of individuals, or a typical response for a specific environment. This could allow an anthropocentric approach to the design of auditory systems suitable for shared auditory environments.

The term soundscape is analogous with landscape in that it represents an individual's unique experience of inhabiting an auditory environment (Schafer, 1977). Traditional methods for measuring auditory environments revolve around descriptions of the quantifiable loudness, pitch and timbre as well as sound events' duration and spatiality (Altman, 1992). Attempts have been made to communicate the experience of inhabiting soundscapes, most notably through maps, the first instance being by Granö in 1929. There is little evidence of adoption of these methods by professional audio practitioners, who concentrate on a sound's physical manifestation rather than its perception by a unique listener, which, Augoyard (1998) points out, is a laboratory abstraction.

Interest in the concept of the inhabited soundscape, and how this can be used within the traditional field of acoustics has gradually increased. The Positive Soundscapes Project was funded by the Engineering and Physical Research Council (EPSRC) and began in 2006 (EPSRC, 2006). This multidisciplinary approach incorporated both scientific and artistic practices and aimed to re-evaluate environmental sound from the listener's perspective. It further sought to extend the paradigm of noise control, as well as engender positive sound design (Davies *et al.*, 2009). The European Cooperation in the field of Scientific and Technical Research (COST) set up an action plan, TD0804: Soundscape of European Cities and Landscapes, in 2008 to create 'soundscape assessment and indicators' as well as 'tools to support designers and decision makers in planning and reshaping urban/rural spaces' (COST, 2008). Schiewe and Kornfield (2009) stated that this work was not sufficiently ambitious and should include 'the geography of sounds', as the field is currently 'highly neglected' adding that this will challenge traditional cartography. There is also work being conducted on an international standard for the Perceptual assessment of soundscape quality (ISO, 2010).

Visualising listeners' experiences in the form of a map allows the 'amplification of cognition', that is to say a large amount of information can be stored in a smaller, more easily accessible form, as well making patterns easier to observe (Card, Mackinlay & Shneiderman, 1999). Mapping allows the transfer of temporal data into a single image, as well as making it relatively easy to compare maps with different content (Borgmann, 1999).

Soundscape mapping could also be used to test how an additional sound-generating object would affect the pre-existing environment. For example, confirming whether a sound event would physically or conceptually mask other sound events, such as a series of verbal alert messages making it difficult to monitor a telephone conversation that a listener is trying to attend to. Likewise, the technique could be applied to test augmented environments, for example, helping to understand how wearing a single earpiece affects the interpretation of the pre-existing environment, as is commonly used by the British Police Service. Furthermore a soundscape mapping tool (SMT) could be used to test how complex auditory interfaces affect traditional working practices and environments, such as an auditory display within a commercial vehicle.

Soundscape mapping could also be used when developing auditory interfaces. It could inform the designer about what an interface has to compete with sonically, and what it might successfully replace. Delage points out that if listeners have any chance of interpreting the meaning of new sounds then they have 'to be in the range of what they already know' (1998, p.72). A virtual soundscape map could test a spatial environment, to establish whether the spatial cues are being perceived as expected, as well as to check whether all of the auditory elements of an audio display are appropriate and clearly heard, under different hardware and operating conditions.

1.1 Research aim and questions

The aim of this research is to develop and trial a method that can allow sound designers to represent their designs and listeners to communicate what they are attending to.

In order to achieve this aim the following three research questions have to be addressed:

1. What attributes are important to both sound designers and listeners when describing sound?
2. How can a soundscape be classified and visualised so that it is meaningful to designers?
3. How can soundscape mapping be used by designers to compare their intentions for a sound design with the experiences of listeners?

The first question addresses the identification of attributes of sounds. In order to be able to compare listening experiences it is important to develop a classification that is meaningful to both designers and listeners. It is necessary to establish which attributes allow designers to represent their sound designs. It is important to identify attributes that are understandable to listeners so that their experiences can be captured. If both designers and listeners use the same attributes to communicate what they are listening to then different experiences of a sound design can be compared.

The second question is concerned with the formalisation of attributes into a classification. Appropriate scales have to be established so that individual listening experiences of soundscapes can be visualised in a meaningful form for sound designers. Different methods of visualising sound events need to be surveyed in order to identify which forms of display are most appropriate for each attribute.

The third question addresses how soundscape mapping could be used to compare listening experiences. In order to establish how soundscape mapping could be used by designers it is important to survey the ways that sound is designed for different forms of media. A range of sound designs have to be mapped so that procedural problems associated with the capture of designers' and listeners' responses can be identified and resolved. Methods of combining responses have to be investigated so that groups as well as individuals' listening experiences can be represented.

If all three of the research questions are successfully addressed then it could be possible for designers to use soundscape mapping as a method for evaluating their sound designs as well as the auditory environment that their design augments. The experience of listening to a design or an auditory environment as a layperson could be introduced into the design process for traditional media and computing. Soundscape maps could show if listening experiences are similar, as well as how experiences differ. Soundscape maps could also highlight the impact of any auditory augmentation upon the pre-existing auditory environment.

1.2 Dissertation structure

The dissertation can be broken down into four distinct sections: introduction, literature review, studies and conclusions. Chapter 1 introduces the work, Chapters 2 and 3 incorporate the literature review, and Chapters 4 and 5 describe preliminary studies and requirements gathering. The soundscape mapping tool is illustrated in Chapter 6 and trialled in Chapter 7. The final chapter (8) draws conclusions and introduces further work.

Chapter 2 defines key terms such as auditory environment, soundfield and soundscape, and provides an overview of current methods for representing soundfields and soundscapes. This literature review provides context and a foundation for the later studies, illustrating published research into describing the experience of listening. Classifications of soundscapes in terms of environment, functions, attributes and descriptions are considered, and methods of quantitative and qualitative visualisation are explored.

In Chapter 3 current methods of sound design for traditional media and computing are discussed. In order to create a tool that could be used by sound designers it was important to survey current practice. The review provides an insight into the approaches to sound design adopted within different industries. The review also highlights attributes of sound that designers manipulate as part of their everyday practice.

The preliminary studies reported in Chapter 4 establish what some of the procedural difficulties are when trying to capture the experience of inhabiting a soundscape. A study using a previously untested method for visualising soundscapes was conducted in order to establish its suitability and propose improvements or further developments. A series of interviews with listeners, where they used their own vocabulary to describe their memories of sound, were coded in order to identify attributes that were used when describing sounds. These attributes were used to extend the prototype method using published methods of soundfield measurement and soundscape classification, as identified in the literature review. These studies provided the starting point for the development of the SMT. The studies highlighted some of the problems associated with capturing and visualising soundscapes, and suggested possible solutions.

In order to establish a common set of attributes for classifying sound events it was necessary to survey both designers and listeners, the results of which are presented in Chapter 5. Current practice and terminology used for designing and evaluating sound was sought through a questionnaire that was completed by auditory professionals. The next stage was to ascertain what attributes and terms were used by laypersons when they described what they were listening to. A series of concurrent verbalisations were provided by everyday inhabitants' experiences of a computing centre, under different reproduction conditions. All of the verbalisations were recorded, transcribed, and subsequently coded to generate a list of auditory attributes, which could then be compared with the results from the audio professionals.

Data from the survey and concurrent verbalisations were combined with the literature review in order to create a first iteration of the soundscape mapping tool (SMT). The SMT was illustrated using the design and evaluation of an in-car auditory display in Chapter 6. The results were used to refine the tool prior to testing with designers from the fields of traditional media and computing.

In order to test the suitability of the tool from the designers' perspective, as well as to provide examples of the SMT's use, professional sound designers and listeners took part in an evaluation in Chapter 7. A number of modifications were proposed, specifically a reduction in the number of attributes and the simplification of the visualisation, although designers did agree that tool allowed them to compare a design with the experience of listeners.

Chapter 8 includes a summary of the thesis, conclusions, further work and future applications. This chapter also details the strengths and limitations of the research, in addition to the thesis' contributions to the field of sound design as a whole. Further work was identified such as listener tests with each of the remaining attributes individually, to establish the most appropriate scale as well as to refine the descriptions in order to improve the internal validity of responses. A proposed method of visualisation of the differences between designers' and listeners' responses also needs to be validated. Three future applications of the research are identified: auditory display design for commercial vehicles, sound simulation evaluation in virtual environments and auditory display design for hospital environments. The soundscape mapping tool allowed the comparison of listening experiences for sound designs, but was found to be complicated and time consuming. Simplifications have been identified along with commercial applications.

2 Classification and visualisation of soundscapes

The basis of the research conducted for this dissertation falls into three broad areas: establishing attributes for describing sound, classifying and visualising soundscapes, and comparing designers' intentions for a sound design with the experiences of listeners. This chapter describes how sound can be measured in terms of soundfields and perceived as soundscapes. Published methods of classification and visualisation of both sound and soundscapes are explored. Definitions for auditory environments, soundfields and soundscapes are provided, and attributes for describing sound are identified.

Listening and hearing are not the same (Handel, 1989). Sterne (2003) states that Matthieu-François-Régis Buisson first highlighted the differences between passive hearing and active listening in 1802. Szendy (2008) argues that we must hear but we can choose to listen. Madell and Flexer (2008) define hearing as the acoustic mechanism of sound being transmitted to the brain, and listening as the process of focusing and attending to what can be heard. Listening is an active process where conscious choices are made about what to attend to (Barker & Watson, 2000). Attending to a sound event or series of sound events necessarily distracts a listener, affecting what else they can attend to (Ashcraft, 2006). Hearing can be considered as the physiological process of sound detection with listening as the subjective interpretation of the meaning of, or the reaction to, sound events (Blesser & Salter, 2007). Listening is a consciously controlled act (Truax, 2001) and sound designers select and manipulate sounds in order to inform listeners about what they should be attending to (Beauchamp, 2005). It could be argued that listening is an active process and that hearing is a passive process.

2.1 The relationship between soundfields and soundscapes

The four characteristics of sound are generally described in terms of loudness, pitch, timbre, and duration, (Rossing, Moore & Wheeler, 2002). All four characteristics present problems with regards to physical measurement. Any communication of measurement requires qualification to ensure accuracy, as there are often different methods for measuring what appear to be identical attributes of sound (Everest, 2002). One of the most commonly confused measures is the decibel (dB), which is a measure of power ratio. Decibels can be used to measure electrical power (dB W), acoustical power (dB PWL), electrical voltage (dB V), acoustical pressure (dB SPL) and even hearing level (dB HL) (P. Brown, 2002a, Haughton, 2002). P. Brown (2002b) warns that sound level measurements require additional information to prevent the measure from becoming meaningless through ambiguity.

Human perception of loudness is subjective. As the sound pressure level (SPL) of a sound event increases, the relative percentages of the perceived frequencies varies, predominately in the perception of frequencies from 20 – 200 Hz, which are also called bass. As the SPL rises, especially above 85 dBu (decibels unterminated), then a more uniform experience is derived than when listening at 35 dBu (Plack & Carlyon, 1995). This difference in perception led to the development of equal loudness levels or contours, measured in phons (a measure of perceived equal loudness) (see Figure 2-1). Equal loudness contours helped to identify the wide variation in perception of frequency content relative to SPL (Hartmann, 1997).

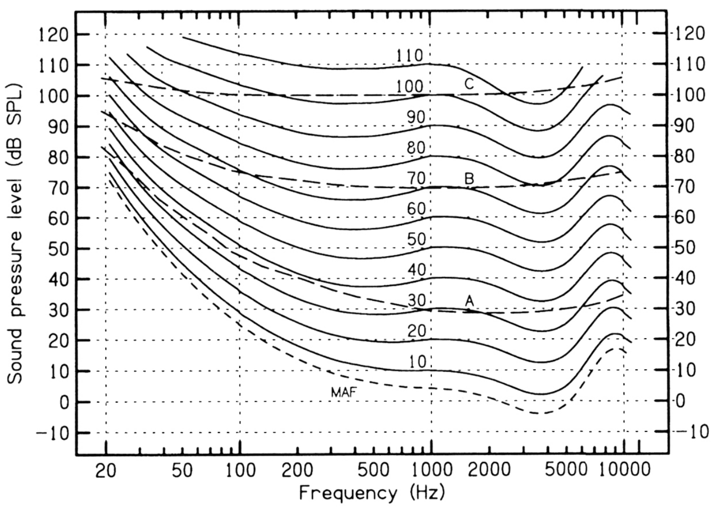


Figure 2-1: Equal-loudness contours and A, B and C weighting curves (Moore, 1997, p.56)

There are three widely accepted weightings for SPL all based on the logarithmic decibel scale (dB), A, B and C (see Figure 2-1). The different weightings correspond to the differences of human hearing at different SPLs. The most widely used weighting is A, which represents a quieter level of listening experience (20 – 55 dB) (Everest & Pohlmann, 2009). The same A weighting is commonly used no matter the SPL of the source being measured and can produce some inaccurate results, especially at high levels (White & Louie, 2005).

Frequency is measured in Hertz (Hz) and represents the number of times per second a medium fluctuates above and below the ambient condition. The ambient condition is the natural vibration of the medium (P. Brown, 2002a). Pitch is the subjective experience of whether a sound is high or low (Rossing, 2007). A medium that vibrates slowly would have a low pitch, and a fast vibration would sound high. Pitch cannot be directly equated with frequency, as it can vary slightly with intensity (Berg & Stork, 2005). Within the field of psychoacoustics pitch is measured in mels, but it is not a one-to-one relationship, as the scale is 's-shaped' rather than linear with only 1 kHz at 40 phons having exact correspondence with 1000 mels. This means that mels are 'often at odds with musical scales' (Gelfand, 1998). Whilst measurement in mels is scientifically accurate, a mismatch with listeners' experiences has meant that Hz is more commonly used when referring to pitch (Houtsma, 1995).

Timbre has been defined as a way of differentiating between sound sources if they have the identical loudness and pitch (ANSI, 1994). The timbre of a sound is the way in which fundamental frequencies and harmonics are combined according to their relative amplitudes and timings (Handel, 1989). Timbre is dependent on the way in which a sound is created and is rarely capable of being reproduced accurately. This is due to its dependency on additional factors that alter the physical wave, such as room acoustics and air absorption (Everest, 2002). Timbre is related to frequency spectrum, and spectrum analysers are used in the study of instrument timbre. However, spectrograms only display frequency versus time and intensity for a typical condition and cannot truly represent the timbre of a sound source (Rossing, Moore & Wheeler, 2002).

The perception of, and the actual duration of a sound event can be very similar, but are not identical (ibid.). A sound can drop below an audible level, but the object might still be vibrating and producing an inaudible but measureable sound. Temporal measurements are normally made using seconds or fractions of seconds, although they can also be calculated in samples, which are based on the sampling frequency of the device being used (Berg & Stork, 2005).

An auditory environment refers to all of the measurable vibrations within a volume that are vibrating between 20 – 20000 Hertz (Hz), and 0 - 150 decibels sound pressure level (dB SPL), that are potentially audible to the human ear (Haselgrave, 1995). A soundfield or sound field can be defined as the auditory environment surrounding a particular sound source. A soundfield represents the quantifiable characteristics of a sound source or event (Ohlson, 1976). Soundfields are normally considered in terms of sound pressure level (SPL), duration, location and frequency range (Ballou, 2002). Sound reveals information to individuals inside a 360-degree environment, unlike light, which presents information in the anterior visual field of 180 degrees (Ong, 1967). A listener can hear a soundfield passively, but a soundfield does not represent the active experience of listening (Truax, 2001).

A soundscape can be defined as the surrounding auditory environment that a listener inhabits (Rodaway 1994, Schafer, 1977, Porteous & Mastin, 1985). The soundscape surrounds the listener and is an anthropocentric experience (Ohlson, 1976). In contrast, a soundfield is the measureable area surrounding an audible object and represents what might be heard (Rodaway, 1994). Acoustic Ecology refers to the study of the way in which listeners relate to their soundscapes, and how this affects its character (Westerkamp, 2000). The Handbook for Acoustic Ecology sets out to provide definitions of terms used in the study of sound from multiple disciplines, in order to describe 'every aspect of the acoustic environment' (Truax, 1978).

The concept of soundscape is not new. Granö was the first to differentiate between the study of ‘sound’ and ‘noise’ in 1929. He mapped auditory phenomena with reference to the ‘field of hearing’ rather than ‘things that exist’. Granö did not use the term soundscape, instead the concept of proximity was applied, which represented the area immediately surrounding an inhabitant (Granö, 1997). The concept of soundscape was revisited in 1969 when Southworth tried to establish how people perceived the sounds of Boston, and how this affected the way they saw the city (Southworth, 1969). Schafer (1977) and Truax (2001), as part of the World Soundscape project, attempted to formalise the concept using descriptions derived from existing terms such as soundmarks, rather than landmarks. Schafer (1993) argued that all soundscapes should be designed or regulated to display what he terms high-fidelity (distinct easily interpreted sounds), rather than low-fidelity (indistinct, difficult to interpret sounds).

Soundscape research has increased in popularity recently with UK and European projects such as The Positive Soundscapes Project (EPSRC, 2006) and Soundscape of European Cities and Landscapes (COST, 2008), as well as the current development of an international standard for the Perceptual assessment of soundscape quality (ISO, 2010). It is now understood that focusing on purely noise measurement and reduction does not equate with improving listeners acoustic comfort. The projects mentioned above are considering the soundscape as a whole rather than single sound sources which require to be silenced, such as traffic or industrial noise. It has been proposed that the type of sounds that make up a soundscape and the characteristics of each sound play a significant role when it comes to improving people’s quality of life (COST, 2008).

2.2 Classification of soundscapes

A number of methods have been developed, within a variety of fields that could be used to classify sounds or soundscapes. Porteous and Mastin (1985) argued that the classification of soundscapes allows the subjective cataloguing of sound events in order to provide descriptive content as well as meaning and value. They state however that there are still problems with classifying sounds due to the broad range of individual perceptions.

R. Murray Schafer (1977), within the World Soundscape Project, developed a clear terminology for describing sounds within a soundscape (see Table 2-1). Schafer's intention was to study the acoustic environment so that soundscapes could be improved through understanding and design. Sound events that were positive could be retained, whilst those that are considered detrimental to society would be eliminated. Schafer's approach ensured that soundscapes could be hi-fi (High-Fidelity), with its associated clarity, rather than lo-fi (Low Fidelity), where man-made sounds typically dominate and sound events require amplification in order to be heard.

Term	**Description**	**Example**
Keynote	sounds that are fundamental to an environment	traffic on a road
Signals	sounds that are actively listened to	announcement over a PA
Soundmark	sound unique to the environment	Palace of Westminster's Big Ben
Archetypal	historical, often 'mysterious' sounds	creak of ancient wood
Centripetal	a gathering sound	church bell
Centrifugal	a scattering sound	fire alarm
Hi-Fi (High Fidelity)	sounds clearly heard against background	countryside
Lo-Fi (Low Fidelity)	difficult to hear individual sounds unless amplified	city centre

Table 2-1: Schafer's 1977 terminology

Sounds were classified according to their impact upon the listener, whether it was key to an environment (keynote) or if it was attended to (signal). If listeners thought that a sound was distinctive to a particular environment the sound was termed a soundmark. If a sound had some historical association then it was regarded as archetypal. Consideration was also given to the sound's effect whether it had the effect of gathering people (centripetal) or dispersing them (centrifugal). Without doubting the relevance and importance of classifying soundscapes, Rodaway (1994) raised concerns about Schafer's reliance on visual metaphors.

Schafer (1977) also developed a system suitable for field notes, as he realised that Pierre Schaeffer's (1966) sound object classification only worked for single musical objects. Schafer enhanced Schaeffer's classification through additional information about a sound's setting, estimated distance, intensity, distinctiveness, ambiance, occurrence, and environmental factors such as reverberation or displacement. In addition Schafer designed a complex two-dimensional visualisation of sound that denoted attack, body and decay horizontally, and duration, frequency fluctuations and dynamics vertically.

Schafer went on to generate a catalogue recording information about the evolution of the soundscape from ear-witness accounts contained within literature. An example of which is a reference made by Cicero in around 70 BCE to his dislike of the sound of saws being used. From this catalogue Schafer was able to track the gradual change from natural sounds to those associated with technology (Schafer, 1977). Schafer's framework concentrated on the allusions to sound and on the listener's perception of a source's function and meaning (see Table 2-2). Sonnenschein (2001), a film sound designer, adapted Schafer's work in order to propose a form of the classification suitable for the film industry. There is still no accepted standard for categories when it comes to naming sounds for a sound effects library in the film industry. This lack of standardization has become less of a problem with the adoption of searchable metadata within audio files, but it is still an issue when a category-based convention is used, as the categories normally vary from company to company (Viers, 2008).

	I. Natural	**II. Human**	**III. Society**	**IV. Mechanical**	**V. Quiet & Silence**	**VI. Sounds as Indicators**
A	Creation	Voice	Rural Soundscapes	Machines		Bells & Gongs
B	Apocalypse	the Body	Town Soundscapes	Industrial & Factory Equipment		Horns & Whistles
C	Water	Clothing	City Soundscapes	Transportation Machines		Time
D	Air		Maritime Soundscapes	Warfare Machines		Telephones
E	Earth		Domestic Soundscapes	Trains & Trolleys		Other Warning Systems
F	Fire		Trades, Professions & Livelihoods	Internal Combustion Engine		Other Signals of Pleasure
G	Birds		Factories & Offices	Aircraft		Indicators of Future Occurences
H	Animals		Entertainment	Construction & Demolition Equipment		
I	Insects		Music	Mechanical Tools		
J	Fish & Sea Creatures		Ceremonies & Festivals	Ventilators & Air Conditioners		
K	Seasons		Parks & Gardens	Instruments of War & Destruction		
L			Religious Festivals	Farm Machinery		

Table 2-2: Schafer's 1977 classification according to referential aspects

Gabrielsson and Sjögren (1979) argued that perceived sound quality should be able to be described through separate perceptual dimensions. The authors were concerned with a lack of knowledge about the relationship between the physical parameters of sound reproduction systems and listeners' perceptions of sound quality. Gabrielsson and Sjögren argued that traditional measures of sound quality (frequency response, signal to noise ratio, distortion, etc.) were insufficient to explain the listening experience, even to sound engineers. Gabrielsson and Sjögren used adjective ratings, similarity ratings and free verbal descriptions of listening experiences to derive a list of attributes that could be used to describe the perceived sound quality. They identified attributes that were associated with clarity (clearness/distinctness), dynamics (loudness), spectra (brightness-darkness, sharpness/hardness-softness, fullness-thinness), space (feeling of space, nearness) and noise (disturbing sounds). These attributes were found by Gabrielsson and Sjögren to be ideal as a starting point for the aesthetic evaluation of sound events, as well as for sound reproduction systems (see Table 2-3).

Dimensions	**Adjectives/Expressions**
Clearness/distinctness	clear, distinct, clean/pure, rich in details diffuse, muddy/confused, blurred, noisy, rough, harsh, rumbling, dull, faint
Sharpness/hardness-softness	sharp, hard, shrill, screaming, pointed, clashing, loud soft, mild, calm/quiet, dull, subdued
Brightness-darkness	bright dark, rumbling, dull, emphasised bass
Fullness-thinness	full thin
Feeling of space	feeling of space, feeling of room, airy, wide, open closed/shut up, narrow, dry
Nearness	more or less near more or less distant
Disturbing sounds	Noisy/hissing, crackling-crunching, whistling-whizzing, wheezing -
Loudness	loud, sharp, hard, painful -

Table 2-3: Gabrielsson and Sjögren's 1979 perceived quality of sound reproducing systems

Gaver (1993) advocated an ecological approach to classifying sounds according to their audible source attributes in order to aid the design of auditory icons. He argued that sound events are generated by some combination of interacting materials: solids, gasses or liquids, and these interactions can be classified in terms of their rate of occurrence. Vibrating objects can be considered in terms of impacts, scraping and others, with the former representing short collisions and the latter more sustained contact. Aerodynamic sounds include explosions and continuous events, again separating impulsive from persistent. Liquid sounds consist of dripping and splashing, this time separating impulsive from intermittent.

Gaver's classification can be extended into a more thorough map in order to illustrate the qualitative nature of the sound events heard through everyday listening (see Figure 2-2). Sound events can be placed in three overlapping areas related to the type of interacting material. The level of complexity of a sound event, which is broken up into basic level (fundamental), patterned, compound and hybrid sources, affects the positioning in the map. The more complex sound events are located towards the centre, the less complicated are positioned towards the edge of the map.

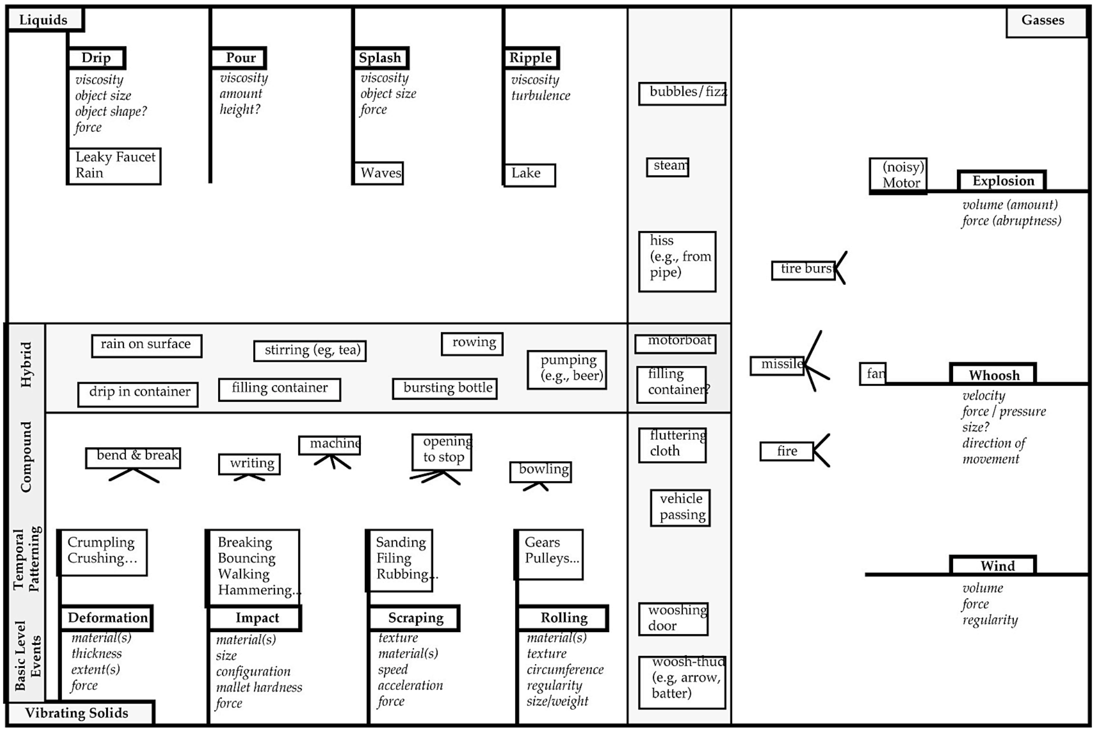


Figure 2-2: A 'map' of everyday sounds. (Gaver, 1993, p.14)

Gaver (1993) acknowledged that this classification was not complete, citing the voice, electricity and fire as possible simple sonic event additional candidates. Gaver also stated that any definitive classification of a source was questionable due the qualitative nature of listening. However, this alignment of the physical actions with everyday language does give a form of eliciting psychoacoustical responses, with a high degree of detail when patterned, compound and hybrid events are included.

Houix et al. (2006) extended Gaver's work in order to provide a framework for sound synthesis algorithms so that sound events could be simulated from known physical interactions between objects, to develop a taxonomy of the sound. Houix et al. started with solids, liquids and gasses, but this time split the solids into models of fracture, impact and friction; liquids into bubble and flow; and gasses into turbulence and explosion. This allowed the description of basic events and textures such as crumpling, hitting and rolling for the impact model, from which processes such as walking, crushing, bouncing, dropping, breaking, and sliding could be derived.

Amphoux (1997) was interested in the sonic memory, perception and interpretation of shared city spaces and developed the *EMP* model in order to capture the contextual sonic identity. Amphoux's model was derived from the work of the geographer Berque (1990) who was researching the connection between society and environment. In Amphoux's EMP model *E* stands for Environmental listening or the spatial, semantic and physical context. *M* denotes Milieu listening or the social context, and *P* is Paysagère or landscape listening (Abdulkarim & Abu-Obied, 2001). Amphoux argued that it is essential to consider all three forms of listening using a repertoire of qualitative criteria (see Table 2-4). Each first level criterion (E, M and P) has a second level of three criteria (9), every second level criteria has a three further criteria (27), with an additional fourth layer that contains a total of 51 criteria. The fourth level of criteria is incomplete and is thought by Hellström (2003) to be extendable with an unlimited number of criteria.

This complex hierarchical model provides a way of classifying the sonic identity of a city, and is used as part of a sonic identity chart. The sonic identity chart is a table that is used by a researcher to summarise recordings, questionnaires and interviews with listeners about their relationship with a specific location. The repertoire of sonic criteria is used during the fourth and last stage of the sonic identity chart, this is known as the objectification of qualitative criteria, and forms the most substantial part. Criteria are chosen, as necessary, from any of the categories within the repertoire of sonic criteria in order to represent the environment as a whole. The objectification can also be annotated with natural phrases for example, 'monotonous traffic sound heard at a distance'. The intention is to use a shared set of criteria that are both subjective and universal in order to systematically classify what listeners experience in a specified environment (Hellström, 2003).

(E) Environmental listening: Acoustic criteria of quality	Spatio-temporal criteria	Scale	Volume Opening Sonic relief
		Orientation	Laterality Frontality Verticality
		Atemporality	Rhythmicity Off-time Eventuality
	Semantic-cultural criteria	Publicity	Anonymity Shared knowledge Public-private relationship
		Collective memory	Sound anchoring Time donor Personification
		Naturality	Naturalism Intentionality Narrativity
	Criteria linked to sonic material	Reverberation	Reverberant space Megaphone space Tautology space
		Sonic signature	Sound emblem Sonic stereotype Sonic postcard
		Sonic metabolism	Compositional clarity Distinctness Complexity
(M) Milieu listening: Sonic criteria of qualification	Criteria of evaluation	Artificialisation	Functionalism Machinisation De-realisation
		Banalization	Standardisation Erasing Indifferentiation
		Stigmatisation	Folklorisation De-humanisation Abjection
	Criteria of idealisation	Privatisation	
		Metropolisation	Patrimonisation Mondanisation Humanisation
		Naturalisation	
	Criteria of imagination	Visualisation	Miniaturisation Picturalisation Visual pregnancy
		Aesthetisation	Theatricalisation Projection Musicalisation
		Fabrication	
(P) Landscape listening: Sonic criteria of qualitativeness	Criteria of representativeness	Typicity	Emblematic legibility Legibility of a sonic stereotype Typicity of a sonic postcard
		Rarity	
		Authenticity	
	Criteria of expressiveness	Internalisation	Evacuation Latency Incarnation
		Belonging	
		Immersion	
	Criteria of reflexiveness	Schizophony	
		Symphony	
		Ediphony	

Table 2-4: Amphoux's 1997 Repertoire of sonic criteria derived from Hellström, 2003 and Abdulkarim & Abu-Obied, 2001.

Prior to his work on Amphoux's Repertoire of Sonic Criteria, Hellström (1998) developed a method of exploring auditory environments that he called the Sound Profile of Place. Hellström thought that creating an inventory of a site and classifying the sounds within it would provide a thorough understanding of a sonic environment that could be used to influence the design of future environments. Hellström's classification was based on the work of phenomenological architectural theorists Werne (1987) and Norberg-Schulz (1976). In terms of space, Hellström applied Lynch's (1960) elements of form to sonic space (complex structures/sound groups) and his own classification for sonic structure or character (simple structures/individual sounds) (see Figure 2-3).

Sources of sound within the overall soundscape were first identified and then classified as either stationary (s.s.) or mobile (m.s.). Hellström then applied a set of seven categories to each of the sounds identified within the overall soundscape. Each of the categories worked in terms of opposition: enclosure vs. extension, man-made vs. natural, present vs. past, local vs. general, figure vs. ground, order vs. chaos, and static vs. dynamic. This meant that a source of sound such as a fountain might be classified as a stationary source of sound that has a character that is static and homogeneous with hard articulation and strong intensity. The fountain may be considered to have space defining boundaries as well as being a local landmark. The fountain could possibly have large variations over night, day and seasons in addition to being regarded as both ground and order.

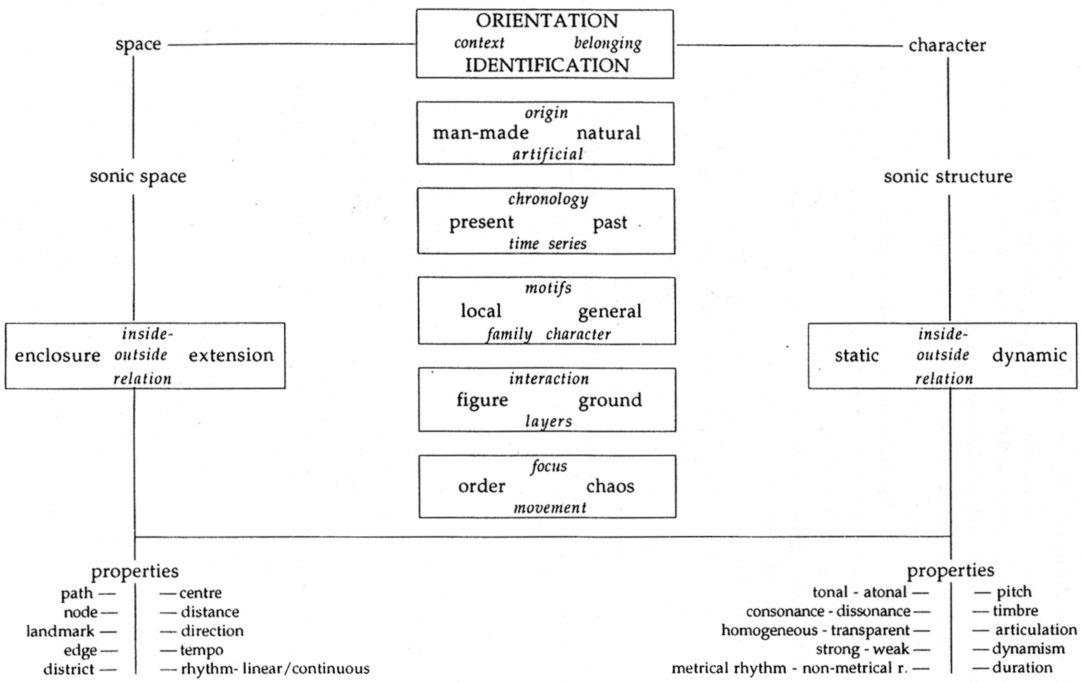


Figure 2-3: Sound Profile of Place (Hellström, 1998, p.32)

It is not necessary for all of the categories to be applied to each sound source, only the categories representative of the sound are applied. After classification a general summary was included where factors such as the differences in sound sources according to the time of day or season are identified. Hellström hoped that this form of mediation would contribute to the future design of sonic environments, through the work of acoustic designers, though there is little evidence of the method having been adopted or developed further (ibid.).

Delage (1998) debated the role of sound and ergonomy relating to the auditory feedback provided by manual tools, and developed a list for the interactive function of sounds (see Table 2-5). He found that listeners derive meanings from sounds, which can be classified into nine categories. Delage argued that this form of auditory feedback provides real-time information about actions based on only what is heard. Delage proposed that sound could inform the listener about how well, or what action a machine is performing or prompt a user into action. Although this list is confined to physical objects, it is potentially applicable to all sounds that communicate information to a listener, whether intentionally or not, such as those commonly associated with auditory displays. For example, the appointment reminder sound in email software could be classified as being of assistance to users.

Function	Description
Warning	Be careful
Assisting	Don't forget
Incitement	I am ready you can use me
Monitoring	In hospitals. In industry
Reassurance	You did OK
Forgiving	Try it again you'll succeed in the end
Guiding	Pedestrians at a crossroad
Protecting	Your car or house
Relaxing	So that you perform better

Table 2-5: Delage's 1998 interactive functions

Another early soundscape classification was by Macaulay and Crerar (1998) who proposed a system that could be used to preview the auditory environment prior to the development of an auditory display, as well as provide a shared language for comparative studies. The model provided interactive systems designers with a framework for classifying the constituents of soundscapes according to sound type, information category and acoustical information.

Sound type was subdivided down into music, speech, abstract and everyday. The information categories, which were guided by a theoretician in audio-visual relationships, Chion (1994): visible, hidden, imagined, patterns, passing of time, emotions and position in Euclidean space, allowed an insight into the information content provided to the soundscape inhabitant. Finally the model included acoustical information, which was sourced from Ferrington (1994), who cited Schwartz (1973) as the inspiration: foreground, contextual or background. Foreground represented sounds that listeners actively engage with, contextual sounds underpin the foreground, and background sounds are the ambient sounds often not attended to (see Table 2-6).

Sound Type	**Example**
Music	Any type of identifiable music, radio/stereo
Abstract	Unusual sounds not normally experienced, video recorder chewing a tape
Speech	Conversation
Everyday	Identifiable recognised sounds.
Information Category	**Example**
Visible entities and events	The phone ringing
Hidden entities and events	The photocopier round the corner being used
Imagined entities and events	Something big is happening on the political desk (it has gone quiet).
Patterns of events/entities	Someone is batch copying a large document
The passing of time	It's nearly deadline time (because the shift change is happening)
Emotions	The sports desk sub-editor is unhappy (tapping)
Position in Euclidean/acoustic space of entities/events and of the listener	The editor is at the foreign desk behind me (can hear his voice)
Acoustical Information	**Example**
Foreground	Computer beep to attract your attention.
Contextual	Door opening (Help you orient to the nature of your environment.)
Background	Whine of disk drive providing reassurance or information about the state of the world.

Table 2-6: Macaulay and Crerar's 1998 workplace soundscape mapping classification

G. W. Coleman, Macaulay and Newell (2008) extended the original 1998 Macaulay and Crerar method to create a sonic mapping tool suitable for participatory design. They developed a three-tiered approach where individual participants (both designers and end-users) map the sounds in a given physical environment in order to inform the design of audio elements for homogenous domains. Sounds are identified and classified into type: musical, everyday, speech or unknown, and then according to their information category. The primary consideration was whether sounds are visible or hidden, and then if the sounds were signals, actions or emotions. Signals and actions were sourced from product sound classifications by Jordan (2000), with Jordan's navigating being changed to actions, and emotions being retained from Macaulay and Crerar (1998) (see Table 2-7). The acoustical information categories of foreground, contextual and background were retained without any changes.

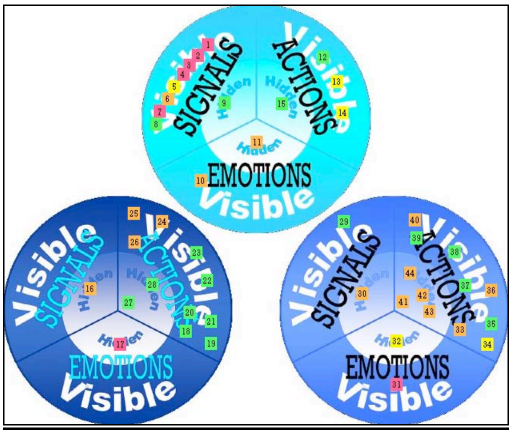
Information Category	**Definition**	**Examples**
Signal	Sounds which inform the listener with *specific* information about a particular event	Car alarm, telephone ringing, zebra crossing
Action	Sounds which are heard but do not provide immediate or apparent information to the listener	Traffic, air conditioner, trees rustling in the wind
Emotion	Sounds which are perceived to express some form of emotion, or which arouse particular emotions in the listener	People laughing, fingers tapping on a desk to indicate annoyance/stress, bird song

Table 2-7: Definitions of Information Categories (G. W. Coleman, Macaulay & Newell, 2008)

The resultant sound event classifications are represented visually on a sonic map in a manner similar to a radial tree chart. Three circles depict the different acoustical categories, each circle is subdivided according to information category, and individually numbered coloured squares represent the type of sound, which are cross-referenced to a description. An example of a sonic map representing a group's experience of an exterior soundscape is shown in Figure 2-4.

It is possible to calculate the classification of each sound event based on its colour and position within the map. Therefore the mobile phone scrolling sound event (31) was classified as musical visible signal and foreground, and is visualised as a pink square located in the outer circle in the top left sector of the light blue circle. In contrast the laughter (32) has been classified as speech hidden emotions and background, and is a yellow square located in the inner circle of the bottom sector of the far right circle. From the visualisation it can be established what the group was attending to (circle/acoustical category) whether they could see the object (rings/visibility), what information they gained from the sound (sectors/information) and how they interpreted it (colour/type). There is no spatial information, which is normally associated with cartography and maps specifically.

The second stage of the sonic mapping tool relates to aesthetics, in that they are either pleasing (sound romances) and or displeasing (sound phobias). Participants choose five sounds, which they consider significant to the audio domain and provide reasons for their choice. The purpose of this technique is to guide designs towards aesthetically pleasing results, as well as highlight positive and negative associations.



ID	Description	ID	Description
1	Mobile phone scrolling	23	Woman in high heels
2	Phone off	24	High pitched bus engine
3	Text getting received	25	Rattling delivery lorry
4	Text done	26	Paper rustling in wind
5	Shouting – not sure where from	27	Van/car revving
6	Strange horn sound	28	Public buses
7	Music from mobile phone	29	Lorry beeping as reversing
8	Car horn, close	30	Traffic lights
9	High pitched car horn	31	Music from someone's headphones
10	Biplane	32	Laughter
11	Jet plane	33	Coughing
12	Lawnmower	34	General chatter
13	Guy singing	35	Spark of a lighter
14	Sneezing	36	Bike chain rattle
15	A rattling truck	37	Carrier bag rustling
16	Car alarm	38	Taxi tyres hitting curb
17	Birds	39	Parked car starting
18	Dry leaves blowing along ground	40	Car door slamming shut
19	Motorbike driving past	41	Sounds of loading/unloading
20	Wind in trees	42	Tapping a stick
21	Traffic on main road	43	Keys jingling
22	Land Rover/jeep	44	Loud bang/clang

Figure 2-4: Group one's Sonic Map (from the outside of the art college building) pink = musical sounds, green = everyday, yellow = speech, orange = unknown (G. W. Coleman, Macaulay & Newell, 2008, p. 7)

The final stage of the tool is an earwitness account, which is a term coined by Schafer (1977) to describe written accounts of historical soundscapes. Participants relate their listening experiences in writing, and these descriptions are appended with the associated sonic map. These descriptions are typically around 250 words in length and range from narrative versions of the sounds chosen for the sound romances and phobias to constructing short stories with plot character and even humour. The authors acknowledged the shortcomings of this method, highlighting the difficulties of visualizing a soundscape. They found that the maps provide an insight into the auditory environment, rather than accurately represent the perceived, shared, soundscape (G. W. Coleman, Macaulay and Newell, 2008).

Mason (2002) studied the perceived spatial attributes in recorded sound in order to develop objective methods of subjective attributes. A verbal elicitation exercise was used where participants were asked to use single words or short phrases to describe the spatial differences between sounds. It was found that the terms could be classified into four categories: Dimensions, Position, Envelopment and Reverberation-related.

Dimensions included wider, bigger and deeper, whilst position related to further behind and outside head. Envelopment incorporated surrounding and more enveloping, with boomier, large space and wetter all being associated with reverberation. Upon refinement of the method using a repertory grid technique Mason (2002) found that responses could be grouped into Width/diffuseness/envelopment, positional, timbral/frequency based and others. He proposed that these descriptors could be used to provide subjective data about perceived spatial attributes in order to complement traditional objective measurement techniques. Whilst this study was based upon recorded sound, as the study was built upon techniques for describing concert hall acoustics, all of the attributes could be applied to the spatial aspects of sound events occurring in the physical environment.

Alexanderson and Tollmar (2006) proposed a phenomenological approach in order to study how sounds were utilised within working environments, so that interactive soundscapes could be created. Rather than use an existing classification system, themes were identified from participatory soundscape interpretation (PSI) sessions, where operators commented on recorded sounds with accompanying video from their workplace environment. It was found that operators used sound in the workplace in three distinct ways for identification, notification, and social awareness. The researchers found that three information categories could be translated into which sound events were either distinguishable and/or meaningful, as well as for identifying suitable areas for auditory augmentation.

Design workshops were used to discuss concepts with operators about possible ways of gathering data about their workplace auditory environments. The proposed approaches were named SonicProbe, SonicRep, and ScapeNav (see Figure 2-5). The SonicProbe was a handheld device to be used by participants that combined a stereo recorder with a still image camera for capturing soundmarks. A soundmark was any device that made a sound that they considered important.

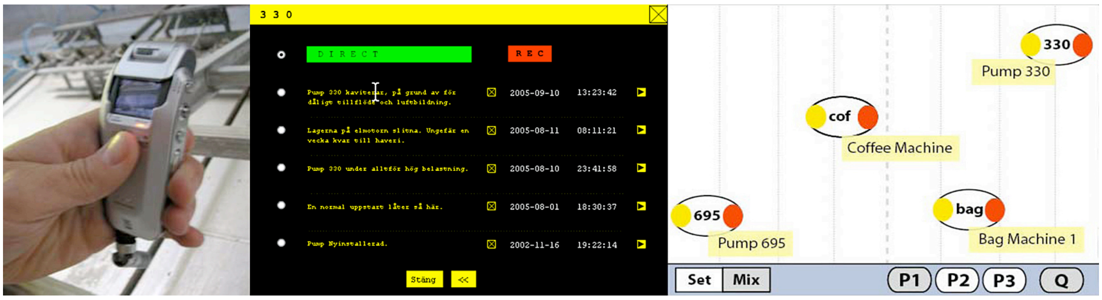


Figure 2-5: SonicProbe, SonicRep and ScapeNav prototypes (Alexanderson & Tollmar, 2006, pp. 256–7)

The SonicRep was a way of bringing all of the recordings and images together within a single software album, so that information be reviewed easily, as well as extended. Extensions included adding recordings taken of the same source under different operating conditions, as well as providing additional textual information in order to explain the context of the sound event. The ScapeNav was an interactive portable device, where virtual audio hot spots were generated from the audio recordings, so that participants could interact with them at relevant physical locations, as a form of auditory augmentation. This four stage approach provided an insight into the way sound events were experienced, as well as allowed operators to compose their own virtual soundscapes (ibid.).

Summary

Despite there being a variety of technqiues for classiying sound and soundscapes there has been a lack of adoption in the broader design community. Different methods have been developed and then sometimes developed futher by a single researcher or group, but often the methods lie dormant. There is some consistency across methods such as breaking down complete soundscapes into individual identifiable sound events and then classifying them. Only Gabrielsson and Sjögren (1979) and Mason (2002) differ from this approach as they were concerned with reproduced sound as a whole, but both of their methods can be used for classifying single sound events.

Recurring attributes can also be identified such as spatial, dynamics, temporal, spectral, aesthetics, clarity, material and interaction. In terms of spatial attributes Schafer (1977) was concerned with a sound's estimated distance and its environmental factors such as reverberation or displacement. Gabrielsson and Sjogren (1979) identified the feeling of space and nearness, while Amphoux (1997) referred to orientation and reverberation. Helstrom (1998) specified enclosure vs. extension as well as centre, distance and direction and Mason (2002) highlighted width/diffuseness/envelopment. Dynamics attributes were highlighted by Schafer (1977) as intensity, Gabrielsson and Sjögren (1979) identified loudness, Amphoux (1997) was concerned with scale and Hellström (1998) specified strong – weak. Temporal attributes included duration (Schafer, 1977) atemporality (Amphoux, 1997) and metrical rhythm – non-metrical rhythm (Hellström, 1998). Spectral attributes related to both frequency (Schafer, 1977) and timbre in the case of brightness – darkness and fullness-thinness (Gabrielsson and Sjögren, 1979). A full sound has a broader range of spectra, while a thin sound has a much narrower range. Hellström (1998) was concerned with both pitch (tonal – atonal) and timbre (consonance – dissonance) and Mason (2002) referred to timbral/frequency based dimensions.

Aesthetics were considered by Gabrielsson and Sjögren (1979) who specified disturbing sounds and Amphoux (1997) identified a criteria called aesthetisation. The disturbing sounds related to audio artefacts such as hissing, whilst aethetisation referred to the judgement made by the listener along aesthetic lines. Clarity was specified in terms of hi-fi or lo-fi environments by Schafer (1977), and as clearness/distinctness by Gabrielsson and Sjögren (1979). Amphoux (1997) referred to sonic metabolism (compositional clarity, distinctness and complexity), and Hellström (1998) thought of clarity in terms of a sound being homogeneous or transparent. Material and interaction are closely linked, with Houix et al. (2006) developing Gaver’s 1993 work. It is necessary to establish the material in order to specify the interaction, although the material could be inferred from the interaction as dropping could not be applied to a gas, and a liquid might be described as pouring or dripping.

Additional factors have also been considered by researchers, such as a sound’s meaning, its type, and level of engagement. The meaning attached to a sound event by a listener was present in 7 of the 11 methods of classification. Schafer (1977) delineated sounds as soundmarks or archetypal, Amphoux (1997) was concerned with the collective memory, whilst Hellström (1998) referred to context and belonging. Delage’s 1998 interactive functions specifically addressed the auditory feedback provided by a sound. G. W. Coleman, Macaulay and Newman (2008) extended Macaulay and Crerar’s 1998 method separating out a sound’s visibility from signals, actions and emotions. Alexanderson and Tollmar (2006) classified sounds according to their use for identification, notification or social awareness.

The type of sound related to whether it was either man-made or natural (Hellström, 1998) or to its artificialisation (Amphoux, 1997). G. W. Coleman, Macaulay and Newell (2008) changed abtsract to unknown, within Macaulay and Crerar's 1998 method, but retained music and everyday. The final factor was the level of engagement, Schafer's (1977) signals were sound events that listeners actively attended to. Hellström (1998) split this into the familiar gestalt figure and ground. Macaulay and Crerar (1998) took it further by splitting it up into foreground, contextual and background which were retained by G. W. Coleman, Macaulay and Newell (2008).

Schafer's 1977 classification has been the most commonly applied, mostly by academics within the field of acoustic ecology. Amphoux's 1997 method is the most extensive form of classification, and is again confined to the acoustic ecology community, neither methods have seen significant uptake outside the field. Almost all of the authors, except for Schafer, accept that their approaches are incomplete. All of the methods are targetted at trained listeners who either report their own responses or interpret other listeners' experiences. This lack of an accepted method, as well as the identification of the merits of the field of soundscape studies has been highlighted by the European COST 2008 action plan, who aim to provide the scientific underpinning in order to take the field forward.

2.3 Visualisation of soundscapes

Visualisation can be used to create long-term representations of the short-term phenomenon of sound. Visualising sound is nothing new. Stone Age carvings found in the chamber at Newgrange in the Boyne Valley in Ireland show concentric rings and waveforms. The patterns correspond with the numbers of nodes and anti-nodes of the natural (or primary) standing wave (110 Hz) measured inside the structure (see Figure 2-6). Devereux (2001) argues that the carvings represent the waveforms rendered visible when the midwinter (solstice) sunrise sent beams of light into the chamber revealing motion in particles floating in the air. It is possible to physically visualise standing waves in a similar manner to the chamber at Newgrange by using an ultrasonic Galton Whistle placed within a parabolic reflector. The force is strong enough to create standing waves capable of levitating cork chips (Everest, 2001).

Figure 2-6: A range of rock-art inside the chamber at Newgrange – concentric rings, lozenge patterns, zigzags (Devereux (2001), p. 64h)

The physical properties of sound have a long tradition of visualisation starting with Thomas Young's *phonoautograph* in 1807, which utilised a stylus to notate the vibrations of a tuning fork in a wax drum, creating a waveform. The waveform plotted amplitude against time, which is the de-facto form of representation for editing sound in modern digital audio workstations (DAWs) (see Figure 2-7).

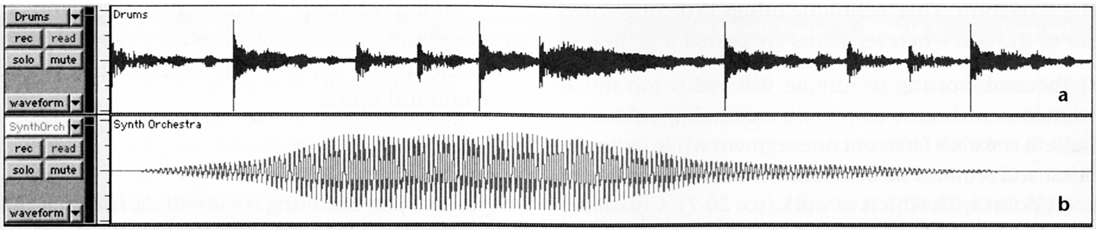


Figure 2-7: Waveforms showing percussive (a) and sustained sounds (b). (Alten, 2011, p. 425)

A waveform display of amplitude versus time can provide information about duration, frequency, fluctuations and dynamics (Roads, 1996). Duration can be calculated by making note of the start and end points of the wave according to the horizontal timescale. The frequency of a simple wave can be established by counting the number of complete cycles within 1 second. Fluctuations can be seen as the variation in the horizontal scale and dynamics can be calculated according to the horizontal scale that is typically displayed in decibels (dBs).

Another established form of visualisation is that of the spectrogram, which is sometimes referred to as a sonogram. A spectrogram plots frequency in hertz against time, usually in seconds (see Figure 2-8). Intensity is represented either through colour or greyscale (Kruth & Stobart, 1999). This method is routinely used in bioacoustics to help identify masked sound sources, as it is relatively easy to remove the background noise from a spectrogram. Spectrograms are rarely used within sound design, due to the intense amount of processing required in order to display it in real time, and the complexity of the results. More commonly, sound pressure level is plotted against frequency; this can be combined with time to form a perspective plot or waterfall diagram (see Figure 2-8) (Roads, 1996).

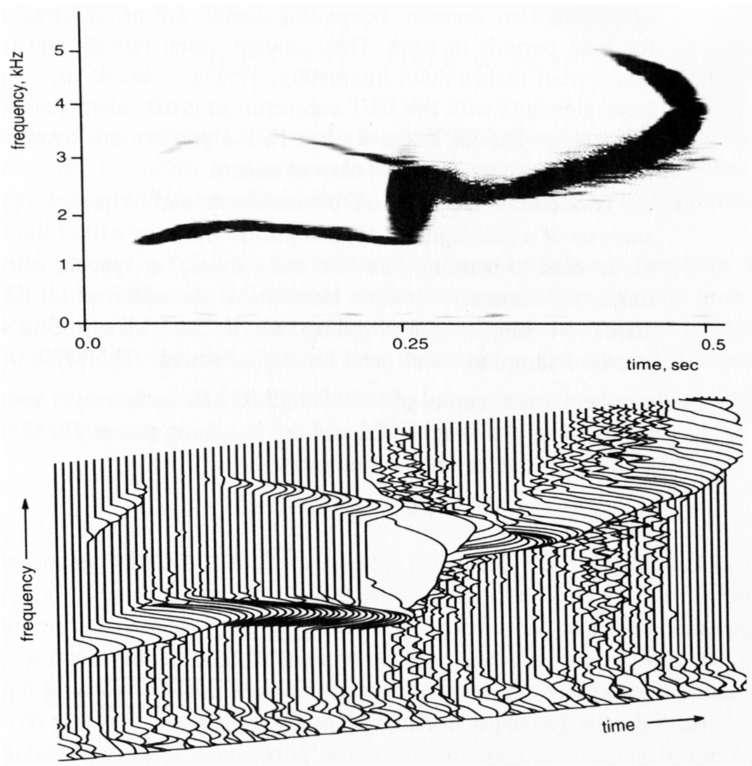


Figure 2-8: Standard spectrogram (above) and waterfall version of spectrogram (below) showing the call of a male northern cardinal (Steiglitz, 1996, p. 212)

Other forms of sound visualisation include directivity pattern where frequency response is plotted against degrees on a circle. Directivity pattern is used both for displaying the response characteristics of loudspeakers and microphones, and can also be displayed in 3D as a balloon, illustrating either amplitude or phase. The directivity pattern provides information about the optimum positioning for both a microphone and a loudspeaker. The technique also highlights any potential discrepancies such as dead zones, which are an essential factor when it comes to matching for stereo pairs (Rumsey, 2001).

Within the field of building acoustics superposition diagrams illustrate the direction and frequency of travelling waves in reverberant fields, using parallel lines and arrows (Pierce, 1981). Superposition diagrams are especially useful for highlighting interference patterns in room acoustics, potentially illustrating where constructive and destructive interference occurs. Tristimulus diagrams are used to plot the relative loudness of three different parts of the audio spectrum in order to display timbre over time (Howard & Angus, 2006). Tristimulus diagrams are commonly used to display the timbre of different musical instruments, as it is possible to differentiate fundamental frequencies from mid and high frequency partials. Both superposition and tristimulus diagrams are highly specialised and rarely used in other audio fields.

Schafer (1977) produced a three-dimensional representation where a simple sound could be plotted against the hearing abilities of an individual and still retain information about timbre, dynamics and time. This method is a precursor to some of the complex forms of visualisations found in contemporary acoustical simulators. In order to illustrate average sound pressure levels in dB A across a geographic area isobel maps were routinely used on the World Soundscape Project (see Figure 2-9). SPL readings were regularly taken throughout an environment and where they were the same they were joined using contour lines (Schafer, 1977). Figures were taken from either actual readings or extrapolated, as is more common in noise prediction software, where very often the results can be coloured and animated (Navcon, 2007). The World Soundscape Project also went on to plot individual sound events against time, in both hours and months (Truax, 2001).

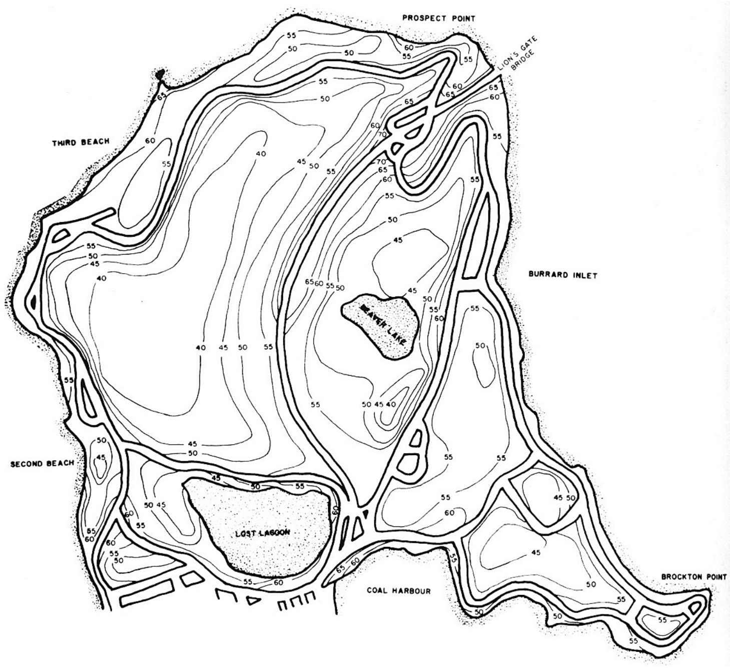


Figure 2-9: Isobel map of Stanley Park in Vancouver (Schafer, 1977, p. 264)

Three attributes of sound have established methods of quantitative visualisation (dynamics, spectra and spatial) and the fourth (temporal) can be inferred as both dynamics and spectra are commonly plotted against time. Attributes are never plotted singly; visualisations are always a combination of two or more attributes. Dynamics are predominantly visualised through waveform displays, but isobel maps or noise maps, where dynamics are plotted against spatial attributes are becoming increasingly popular for displaying sound pressure levels for cities (Klæboe, Engelien, & Steinnes, 2006). Dynamics are also included as part of spectrograms which are the most common form of displaying spectra. Spectrograms have become popular with sound designers as an interface for audio restoration (Lukin & Todd, 2006). Directivity patterns display spectra against spatial attributes, superposition diagrams relate spatial against dynamics and tristimulus display spectra and time.

Information about the qualitative aspects of a sound event or soundscape are more commonly found in textual form (Rice, 2001). Sound designers' maps usually resemble lists where sounds such as neon buzz and gunshot are written next to time on their corresponding audio tracks. These sound maps are often more accurately referred to as cue sheets, as they provide no information about position, or relative levels (Sonnenschein, 2001). To compensate for this lack of spatial and level information many individual designers develop their own forms, few of which are decipherable by others, highlighting the fact that there is no single classification approach within the film industry, nor in any of the other forms of traditional media.

When the qualitative aspects of sound events or soundscapes are visualised it is predominantly in map form. Granö, a Finnish geographer, created one of the earliest examples of a soundscape map in 1929 (see Figure 2-10). Granö identified sources of noise (people, birds and cows) and mapped the 'field of hearing' for the island of Valosaari. Identifying where each sound source was located, and then calculating the isopleth or contour line at 25 metres from the source established the field of hearing. Hatching was used to display the area where each source would be clearly audible. Areas that were greater than 25 metres from the source where the sound would be relatively quiet were left unhatched. The map illustrates the dynamics attributes of the island, showing where multiple sound sources might be heard. Areas where more than one sound source might be heard are visually shown as using overlapping hatching. Rodaway (1994) considered that Granö differentiated between the foreground immediate soundscape that was related to the visual experience and the background distant soundscape that was purely auditory, as the visual cues were hidden. Rodaway went on to argue that this figure-ground framework related to Schafer's (1977) classification of sounds as signals (figure) or keynotes (ground).

Figure 2-10: Auditory phenomena of the Valosaari medium. Sounds and noises: 1 = Produced by people always in summer; 2 = produced by people sometimes in summer; 3 = produced by people frequently at all times of the year (boating route, ice road); 4 = bird song in spring and summer; 5 = clanging of cow bells (less often mooing of cattle or bleating of sheep) in summer. The hatched area for each auditory phenomenon terminates at the 25-m phenomenal curve. (Granö, 1997, p. 127)

Ballas and Howard (1987), while comparing the similarities with the perception of speech and environmental sound, utilised a state transition diagram, more commonly associated with syntactic structure of grammar or capturing machine behaviour, in order to create organized sound sequences. This method allowed for all of the dramatic variations, which could occur from a single sound source, to be notated in a single diagram, and subsequently helped to prove their theory of environmental sounds being processed in a manner similar to speech, that is basing any interpretation on expectation and context.

Southworth (1969), a landscape architect and city planner, proposed that the visual experience of a city was interrelated with the auditory experience. Southworth's map of a part of Boston details when the soundscape is considered distracting, uninformative, or even dull (see Figure 2-11). The caption details nine conditions all with unique markings that were overlaid onto a simplified map of the Boston soundscape. Graphics were used to represent the relationship between the visual and sonic identity. Strong visual and sonic identities were represented through a densely hatched area with a black border. Dull visual and sonic sequences were borderless and sparsely hatched. Other dimensions included a lack of temporal continuity, which was illustrated by a series of dots, with distracting and uninformative sounds shown as a series of closely overlapping circles, and areas with sonic involvement having parallel waves. The final two dimensions of whether sounds were difficult to differentiate, or were well-related to the city, were shown by a series of joined curves. The curves for the difficult to differentiate areas terminated in dots, the curves for the sounds well related to the city ended in arrows. Southworth's map is the first published example representing the shared experience of a soundscape.

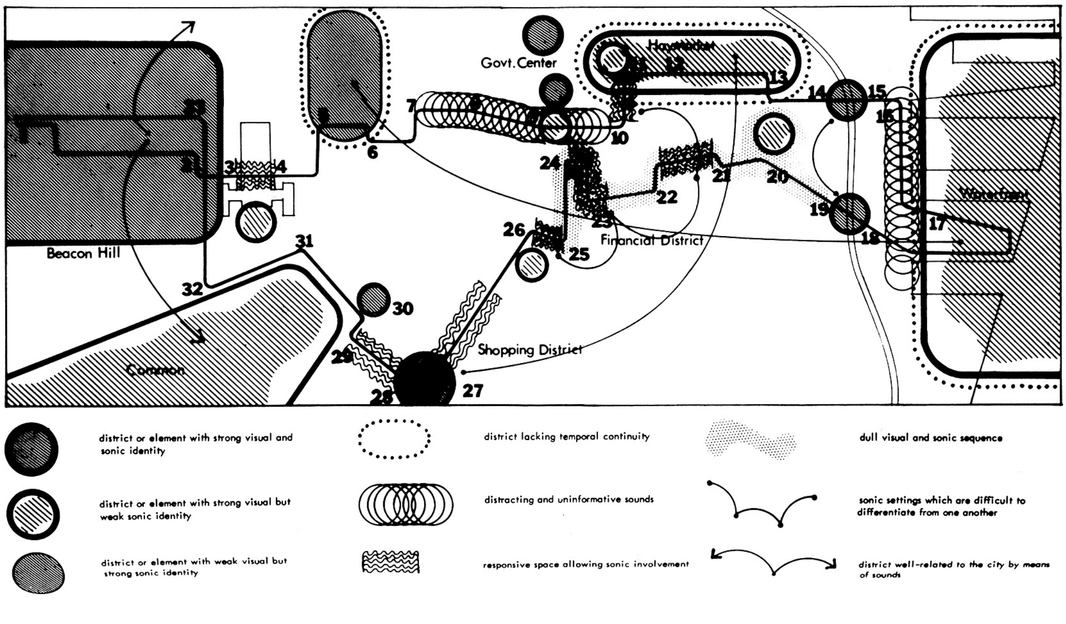


Figure 2-11: Evaluation of part of the Boston soundscape (Southworth, 1969, p. 66)

Whilst studying the soundscape of Kanda in Japan, Torigoe (2002) created a map of the incoming sound sources which were external to the immediate environment under study, but yet still clearly audible within its boundaries (see Figure 2-12). This hand-drawn sketch contained an outline of the area, with pictorial and textual descriptions of the sound sources linked to the map via arrows to denote their relative positions, and is similar in form to maps used by the World Soundscape Project.

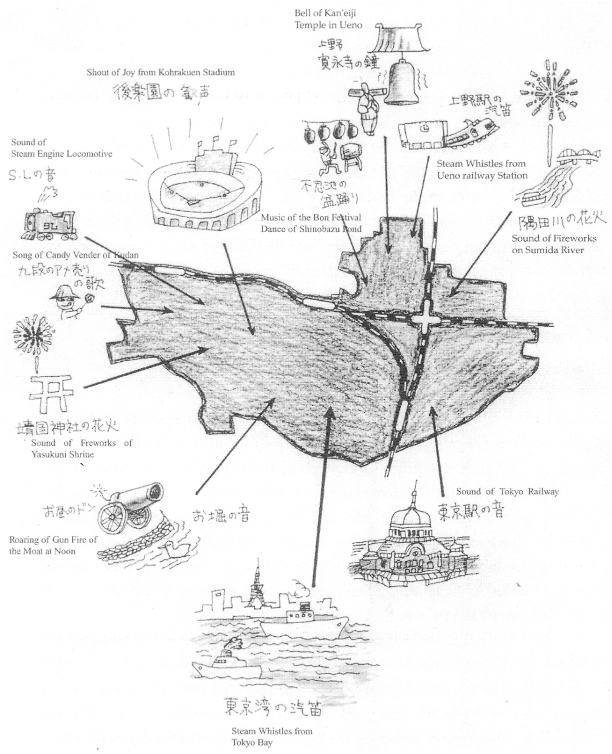


Figure 2-12: Incoming sounds (Torigoe, 2002, p. 46)

Hedfors (2003) extended the theme of representing multiple attributes of sound in map form. A hand drawn map was created during the development of a prototype computer tool to represent the audio qualities of landscape architecture. Icons were overlaid, hatched and coloured onto a traditional map of a formal garden in Uppsala, Sweden, with the colours being based on the work of Mahnke (1996). Skilled listeners who were either experts in landscape architecture/planning or music or acoustics made the classifications (see Figure 2-13). The different colours within the map illustrate the perceived qualities of sounds. Symbols are used for static sound sources. Information about duration and rhythm are included in the legend, which splits up the sounds in to categories according to whether the source is stationary, mobile, background or a wall. Blue is used to denote soft or distant sounds, green for moderate, grey for fatiguing sounds, yellow for positive, gold for caring, and red for powerful. Colours such as orange, emerald and mauve are used when sounds fall between two of the primary categories. Orange has the properties of both red and yellow so is powerful and positive. Within the map it can be seen where the soft sound of the fountain is audible as well as the tiring fan and even the moderate to soft reflections of sound off the walls of some of the buildings surrounding the park.

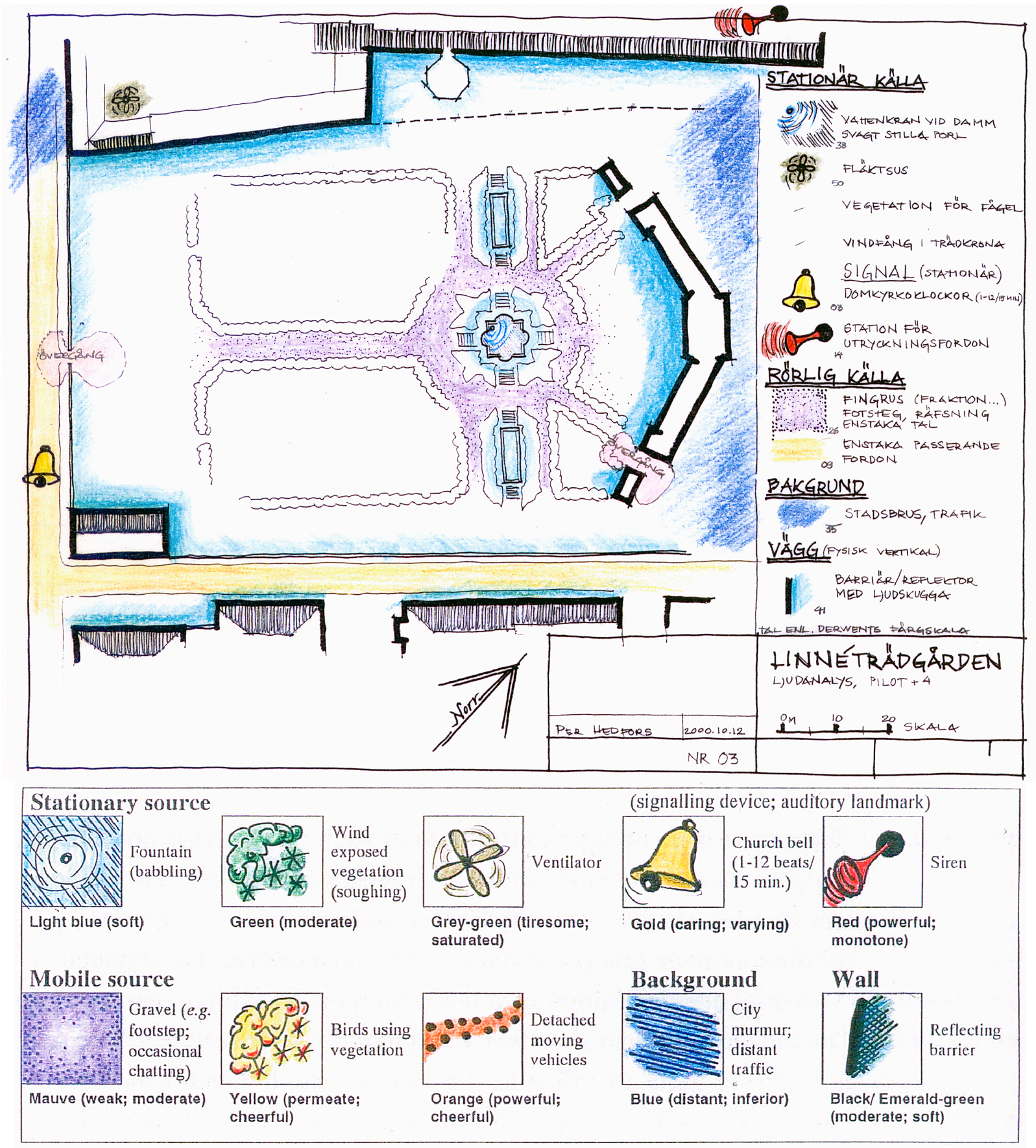


Figure 2-13: Layout of Linnegarden with the nearest houses and preliminary legends. The colours refer to sonic atmospheres freely, according to Mahnke (1996) (Hedfors, 2003, p. 99, and Hedfors & Berg, 2002, p. 106)

Jeon, Kook and Jang (2006) also used a map in order to illustrate their design or a Spontaneous Acoustic Field Reproduction System (SAFRS), which created a variety of sounds according to different climatic conditions. Jeon, Kook and Jang visualized the movement of individuals along with the audible range of the sound sources and their intended effect (rest, interest), in addition to traffic noise. This approach shows a link between acoustical simulations and soundscape mapping as it considers auditory elements already present within the environment, as well as the impact of sounds used to augment the pre-existing auditory environment.

Visualisations have been included successfully within acoustical simulations since the 1960s. Krokstad, Strøm and Sørsdal (1968) were the first to graphically represent the distribution of reflected sound in concert halls. Modelling methodologies for acoustical simulations include physical, computational, empirical and hybrid models (Cheenne, 2002). Physical scale model acoustical simulations were first adopted in 1933, with computational models introduced in 1965, in 1986 the Bose Modeler became the first commercially viable computer–aided design (CAD) solution (Bose, 2010). The most commonly used software today is the Windows based Enhanced Acoustic Simulator for Engineers or EASE by Renkus-Heinz (2007). Information about a building's structure (in the form of AutoCAD 2D and 3D files), as well as its acoustics, can be combined with a variety of manufacturers' loudspeakers, to visualise and auralise the way in which specific frequencies behave according to the position of the loudspeakers within the building. The results can be displayed various manners such as ray tracing where the direction and intensity of a sound is shown as either a straight line, cones or pyramids complete with reflections overlaying the wire view of the building. The listener's perspective can be displayed using a hedgehog or isometric approach, where the same information is shown without the building, and from any position.

The use of acoustical simulations within sound reproduction design is gradually replacing subjective experience, despite the high associated costs. Similar software is utilised regularly for building acoustics, but there is no analogue for sound design. There is a long tradition of predictive formulae in the field of acoustics. In terms of sound design, there have been few studies beyond that of a sound event's physical properties. Without any formulae as to how individuals hear a soundscape based on their interests and experiences it is almost impossible to model.

Giaccardi, Eden and Fischer (2006) used a Geographical Information System (GIS) to display either an individual's or a combination of individuals' soundscapes. The GIS formed part of their work on a virtual museum called *The Silence of the Lands* where listeners could share experiences of soundscapes. The authors used the GIS linked to audio recordings and GPS data created by participants to apply principles of Envisionment and Discovery Colloratory (EDC). EDC is a framework designed by Arias *et al.* (2000) to allow multiple participants to create and share contextual information and artefacts related to a design problem. Small groups used this tangible social interface as a form of collaborative design to create ideal virtual soundscapes of an open air space (see Figure 2-14). The authors' main concern was the preservation of 'natural quiet' through cultural negotiation. Colour was used to show areas with different levels of agreement about levels of noise. Unfortunately the authors found that the system, which was an early prototype, did not sufficiently engage the local community, which they suggested was due to insufficient time for participants to become familiar with the system (Giaccardi *et al.* 2006).

Stratoudakis and Papdimitriou (2007) also used a GIS in order to reconstruct a soundscape, using a combination of audio recordings, still images, soundfield measurements (dynamics, spectral, spatial and temporal) and soundscape classification (source description, meaning and origin) made over a year. Each identified sound event was described according to its source and then classified as either background or foreground, as well as being either human, biological or geophysical. All of the data was combined so that participants could navigate an interactive sonic map of an area around a lake in Greece. Where sampled data was not available for a specific area or time the results were extrapolated from the nearest points, allowing a continuous scale to cover the entire area under study. The system enabled both thematic and composite mapping, so that attributes could be viewed either separately, or in specified combinations. Within thematic mapping red was used for human, green for biological and blue for geophysical sounds. Users could navigate spatially as well as along a timeline, so that they could see the appropriate visualisation and hear the associated foreground and background sounds. The authors believe that system not only provides a realistic listening experience, but could also be used to experiment with different sound sources and easily compare different soundscapes. The number of measurements (21) outnumbers the classifications (3), which means that the tool is currently weighted towards representing a soundfield rather than a soundscape. But as it is a database system, listeners' responses could be added to extend the number of attributes.

Figure 2-14: Prototype of the Tangible Social Interface (EDC) (Giaccardi *et al.* 2006, p. 12).

Valle, Lombardo and Schirosa (2009) also adopted a technique of soundscape generation, this time using graphs of sound objects that could be mapped. Their four-stage GeoGraphy model began with the classification, analysis and recording of a physical soundscape, through to production of a database, then analysis and finally generation of a simulated soundscape. The authors classified sound objects as atmospheres, events or sound subjects. Atmospheres represented sound objects that could not be broken down into identifiable sources, based on the work of Bohme (2004). Events were used to describe clearly identifiable sources that are easily isolated. Sound subjects represented a complex source that contains a sequence of related events. They also applied three continuous parameters based on the work of Wishart (1994): dynamics, reverberation and brightness. Dynamics were based on the musical scale of pianissimo possible *ppp* to forte possible *fff*. Reverberation was quantified in terms of dry versus wet, and brightness in terms of bright versus dull.

Recordings were combined using omnidirectional techniques that represented the soundscape as a whole, with highly directional recordings of sound objects in order to create a virtual soundscape. Listeners could navigate through using a graphical user interface (GUI) in the form of a sound map (see Figure 2-15). Individual sound objects are described as graphs, which were mapped according to their coordinates within the physical environment, thus forming a map of graphs. As listeners proceeded through the map, the audibility radius changed and affected what could be heard, simulating the physical soundscape. Valle, Lombardo and Schirosa (2009) argue that unlike other soundscape generation systems such as Listen (Warusfel & Eckel, 2004) and Tapestrea (A. Misra, Cook & Wang, 2006) their system can reproduce a listener's experience of a soundscape, as the soundscape generation is based upon listeners' classifications of previous experiences.

Schiewe and Kornfield (2009) surveyed the field of what they referred to as audio cartography and found that the visualisation of sound had been for the most part disregarded, and that the work that had been conducted was limited in its scope. They argued that acoustic geography should incorporate both subjective and measured dimensions of the acoustic environment in spatial terms. Schiewe and Kornfeld argued that descriptions and measurements should be combined from soundscape research (Schafer, 1977), acoustics (Heckl & Muller, 1994) and psychoacoustics (Zwicker & Fastl, 1999). From the soundscape research the authors were interested in hi-fi and lo-fi, sound marks and sound events. Within acoustics the following were identified as suitable measures: sound pressure level, velocity, intensity, source wave propagation and frequency. From the field of psychoacoustics a number of variables were chosen: loudness, pitch, sharpness, rhythm, annoyance, melodiousness, roughness and fluctuation (Schiewe and Kornfeld, 2009).

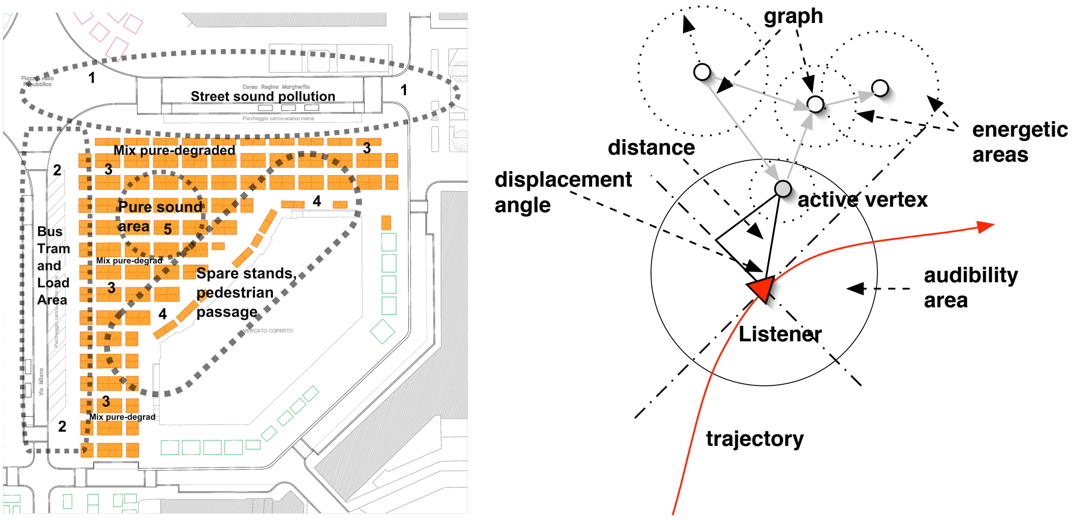


Figure 2-15: Map indicating sound zones, and listener in the map of graphs. (Valle, Lombardo and Schirosa, 2009, pp. 5-6).

Schiewe and Kornfeld suggested that even the European Cooperation in the field of Scientific and Technical Research (COST) action was insufficient (COST, 2008). Their proposal was to further develop visual descriptive annotation systems in conjunction with acoustic annotation systems to help overcome the incompatibility of the senses (hearing and sight). The annotation systems would allow the full potential of cartographic techniques to be used for visualising sound. The cartographer Bertin's (1983) visual variables formed the starting point for the creation of a style guide that would be meaningful and easy to interpret to both experts and the public without requiring any training (see Figure 2-16). Schiewe and Kornfield (2009) proposed that the results of this work could then be used as a form of urban sound cartography that they believed would be the first system to combine geographical and soundscape data visually.

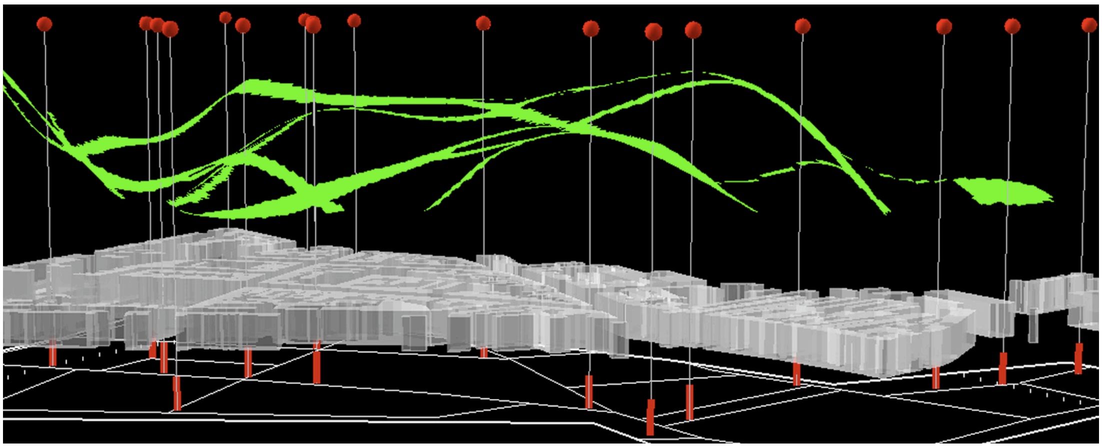


Figure 2-16: 4D map animation shows the spatiotemporal propagation and intensity of diverse urban sound sources, e.g. sounds of traffic and sounds of human activity etc. (Kornfeld, 2007).

Summary

Visualisations of soundscapes appear to fall into two distinct groups, techniques that only include listeners' experiences (Southworth, 1969, Ballas and Howard, 1987, Torigoe, 2002, Hedfors, 2003, Giaccardi, Eden & Fischer, 2006) and methods that combine both soundfield measurements and listeners' experiences (Schafer, 1977, Stratoudakis and Papdimitriou, 2007, Valle, Lombardo and Schirosa, 2009, Schiewe and Kornfield, 2009). The methods that combine both measurements and descriptions only use limited data about the listener's experiences when compared to the total number of measurements. Both approaches rely predominantly on the identification and meaning of sound sources along with the following attributes: spatial, dynamics, temporal and spectral. It is uncertain how all of the visualisations have been created, as most authors reported preliminary work without including fully annotated examples.

2.4 Conclusions

This chapter provided an overview of auditory environments, their measurement in terms of soundfields, and their classification as soundscapes; it finished by illustrating methods of representation. Definitions were provided for auditory environments, soundfields, soundscapes and mapping. While the measurement of soundfields in terms of SPL and frequency has been ubiquitous, spatial and temporal dimensions have not been so widely considered, except in terms of audio reproduction in stereo and surround sound systems. The visualisations associated with soundfields are mostly concerned with SPL either in waveforms, or noise maps. In terms of soundscapes there has been little take-up of methods of classification or visualisation, despite its comparatively long history. There are many attributes that have been considered, but no all-encompassing method has been created in order to represent a soundscape.

3 Designing sound for different media

In this chapter sound design approaches are discussed according to media type: theatre, radio, film, television, video games and auditory displays. Sounds are designed for a wide variety of purposes. Within traditional media, sound design is not concerned with representing real sound events, as sometimes this is impractical to achieve. Sound is often used as a sleight-of-hand (Chion, 1994), making the audience believe that something has happened. For example, broken bones may be represented by carrots and celery sticks being snapped, or fingers pressing down on cornstarch simulating wolves walking through snow. Within the different types of entertainment media, there is a natural progression as each new form of media borrows from the previous incarnation. For example, video game sound designers have adopted many of the techniques associated with film sound (Newman, 2009). Video game sound designers have added interactivity so that some of the sound events are directly controlled by gamers' actions, whilst other sounds remain passively experienced within noninteractive sequences (K. Collins, 2008). Within computing, auditory display design has split between user interface feedback and multivariate data presentation (Alexanderson & Tollmar, 2006). That is to say sounds that are designed to either inform a listener about the actions taking place within a user interface, or to communicate data relationships in order to aid interpretation.

3.1 Theatre

The use of sound design in theatre extends back to the earliest documented performances of the *Passion* in Egypt around 2000 BCE, where actor-warriors staged mock battles. Real weapons were used and all of the associated sounds were generated through accurate combat recreations, to such a degree that fatalities occurred (Gillam, 2006). The sound design in this case was a by-product of using authentic props, but the desire for auditory augmentation reaches through the history of theatre in order to extend the theatrical space. One reference to machinery being used to create sound effects was the canon that burnt down the original Globe theatre in 1613. The canon had been used regularly during plays to suggest battles, or in this case the entrance of Henry VIII (Thomson, 1992).

In contemporary theatre, sound design concentrates on three areas: dialogue, sound effects and music. The first area being to ensure the intelligibility of actors, since the introduction of cinema and especially television theatre audiences have become less patient with poorly projected dialogue. All sound design starts with ensuring that the cast are clearly audible, which usually involves amplification to prevent articulation loss (Burris-Meyer, Mallory & Goodfriend, 1979; Walne, 1990). Amplifying an actor's voice is not new, ancient Greek masks were sometimes designed to amplify the voice (Flickinger, 1926). The modern equivalent of the Greek mask is a miniature wireless microphone connected to an amplification system. Sound designers are expected to amplify the minutest murmur in order to maintain the audience's attention (Finelli, 2002).

The second area of theatre sound design is sound effects, which are usually designed to be either realistic or stylistic (Kaye & LeBrecht, 2000). Sound effects are then mixed with the dialogue and music, in order to support the performance. Dialogue audibility is always expected, even when replicating an environment where dialogue would be inaudible, such as on the deck of a ship during a storm (Leonard, 2001).

This optimisation of clarity and volume of the source through amplification can quickly cause additional problems, especially in a theatrical venue with poor acoustics (P. Coleman, 2004). The sound designer in theatre has to consider intelligibility above all else. In some cases this means that everything has to be reproduced at very high levels, as is common in musical theatre (Wollman, 2006). In other situations, such as dramatic theatre, only the sound effects and pre-recorded music are amplified, and even then only gently, in order not to overwhelm the dialogue (Leonard, 2001). Emphasis is on the perceived realism of all of the audible elements within the natural acoustics of the auditoria, but with scope for the adjustment of dialogue levels according to the effect of the audience presence and theatre size (Marshall, 2009).

Realistic sounds have a function, they can imply unseen characters, or more commonly provide auditory cues for actors to interact with, such as a telephone ringing. Realistic cues can be effective about communicating the era, time and location of a dramatic production. Realistic design can also be selective and representational. Representational effects are just as realistic but scarce, being only used when absolutely necessary. For example, the brief sound of a cuckoo will place the action outdoors during the springtime (Kaye & LeBrecht, 2000).

Stylistic sounds have a psychological purpose and relate to exaggeration, distortion or conceptualization, either in the abstract or the absurd. The abstract represents the designer's portrayal of their reality. Sounds being altered but not beyond recognition. A car braking will still be understandable, but emphasis might be made on the screeching brakes, which become more akin to a human scream. Absurd sound effects can disrupt the suspension of disbelief, reminding the audience that they are in the artificial environment of the theatre. The designer has to ensure that there is still a relationship to the original sound source otherwise the effect might sound comedic rather than surreal (ibid.).

Brock (1950) identifies four distinct categories of sound effects within theatre productions: framing, underscoring, transitional sounds and specific cues, which have been present since Elizabethan theatre. Framing marks the start and end of each act, referring to what is about to or has taken place and is not linked to the acting, but it can be very effective in preparing an audience. For example the sounds of a Victorian street populated by horses, carriages, beggars, children playing, and a busker, suggests that we are outside a city-centre house, possibly about to enter. Underscoring acts as the emotional indicator for the action and goes unheard by the actors, with the audience often subconsciously hearing it in the background. Underscoring is useful when introducing characters, or commenting upon emotional states. The low rumble of thunder can suggest that a character is in a dark mood. Transitional sounds act as the bridge between one scene and the next, usually indicating the passage of time or the change of location, and are commonly used to set the scene or mask set movement. The segue is a subset of transitional which represents the transition within a scene. Clocks chiming or bells tolling are popular for transitional sounds (ibid.).

Specific cues represent sounds that are part of the actors' world, and emphasis is placed on spatial accuracy, so that sounds emanate from the correct source, for example the telephone sounds like it comes from the handset on the stage. All of the sounds are informative in nature and can be split into spot effects and ambience. Spot effects are identifiable sounds such as the sound of a tap running while an actor is simulating washing dishes. Ambience is used to describe the mix of sounds that set the audio backdrop (ibid.). Subtle vehicle traffic reminds us we are in a modern city, and the sound of a highly muffled television suggests that we are in a densely populated area such as a block of flats. The sound of a slightly muffled one-way conversation suggests that there is an additional person within the building who might enter at any point. Ambience is also an excellent approach for setting the scene even before the curtain rises (Waaser, 1976). By varying the pitch and volume of the off stage sound effects relative to those in the foreground, the perceived size of the space can be varied (Napier, 1962). A quiet telephone ring, supposedly emanating from a distant room, can make a house appear larger, and this approach can compensate for a small stage or a simple set.

All of these types of sounds can be mixed together with music and voiceovers to either focus or distract the audience's attention, as well as to provide a sense of realism (Bracewell, 1993). Sound design is prevalent throughout all types of theatrical productions, the only variation being the extent of naturalism, which is whether the sound appears to emanate directly from the loudspeakers or from the actors and objects (Collison, 2008). Sonic scenic illusion has been considered to be impossible, with sound design being most effective when it is used to 'affect the emotional state of the audience' (Appia, 1918). It has also been proposed that it is not possible to 'communicate intellectual ideas' using sound design in theatre, as there are problems with source identification and realism (Thomas, 2000). However, these concerns have not prevented sound designers from being a regular part of a theatre production team (Leonard, 2001).

3.2 Radio

Radio is unique among the other forms of media in that it relies on sound in isolation to communicate information, atmosphere and emotion. Radio has often been referred to as theatre of the mind for the reason that sounds are designed to create images (Mott, 1993). Crook (1999) has defined the listener's imagination as the fifth dimension within radio drama, the first four being dialogue, music, sound effects and post-modernist inclusion of pre-existing recordings. Like theatre, radio from its mainstream inception in the 1920s through to the late 1950s was produced live, and had to borrow much from the theatrical world when using sound to involve the listener (Mott, 1993).

Sieveking (1934) states that radio sound effects can be realistic, symbolic or impressionistic. Signposts contained within dialogue, such as a reference to rain can be confirmed by the sound of heavy rain falling onto a tiled roof. This is similar to anchorage (Barthes, 1977) where captions provide links between images and their context (Crook, 1999). Sound effects can also be used as signifiers with the sound itself being the signpost. A train whistle might suggest to the listener that a train is nearby. When the sound of multiple footsteps in a reverberant environment is added, the listener is transported to a train station with the suggestion of travel and parting. Abstract rhythmic non-musical sound can be used to symbolise emotions, for example, an almost inaudible sound of thunder suggests a character's internal unrest. Impressionistic denotes dreamlike sounds that indicate the character's inner fantasy world, such as gentle wind suggesting that the character is dreaming. One of the most influential sound effects of all is the lack of sound, silences lasting as long as five seconds can be very effective (Connelly, 2005). However, dramatic pauses can be perceived as dead air implying a technical fault (Reese, Gross and Gross, 2006).

Spatial indicators are important in radio, where relative proximities and movement between both the actors and the listener can be conveyed either by varying proximity to the microphone or fading up or down audio tracks (Crisell, 1994). The angle at which the sound source is presented to the microphone is as important as its dynamics (Esslin, 1987). Directional microphones are the mainstay of radio production, where off axis sounds are considerably attenuated as well as coloured due to uneven frequency response (Eargle, 2001).

Sound design is not confined to drama productions; jingles are regularly used in all types of programmes as a form of punctuation to indicate a change of subject or transfer of location, or even just to break up a long interview. Jingles vary from simple 2 second stingers, through to complicated music, effects and dialogue constructs, usually termed sweepers. Jingles are used in order to reinforce either the station's or the programme's brand or identity, through repetitive exposure (Fleming, 2002, Keith 2007).

Sound effects are routinely used in news programmes in order to paint pictures or illustrate a story, with more complicated auditory constructs used to create drama (McLeish, 2005). If sound effects are over-used in news programmes, it can often lead to complaints on traditional stations (Starkey, 2004), but when applied to features sound effects are regarded as trend setting (Keith, 2002). A more practical use is to record actuality where the background atmosphere is recorded in order to use as a bed to hide dialogue edit points (Beaman, 2006).

However it is in radio drama productions that sound design is extensively applied, through the use of signposts. If this is done successfully then there is no need to describe the setting in the script, as listeners have extensive experience of polysonic environments (McWhinnie, 1959). Signposts only work if the sounds are readily recognisable by the listener (Reese, Gross & Gross, 2006). The amount of reverberation applied to a voice or sound provides information about whether the scene is set indoors or outdoors, as well as the size and type of room (McLeish, 2005). Certain sounds are associated with location or time of day, a cockcrow indicates daybreak in the countryside (de Fossard, 2005). Actors perform some of the mundane effects themselves during the recordings. For example handling a telephone handset, which aids naturalism, as all of the relevant cues are in correct relationship to the actors and microphones (Crook, 1999).

The simple act of fading down an atmospheric sound effect, leaving a small gap, and then fading up a new atmosphere indicates both a new time frame and location (Beaman, 2006). The length of the pause can indicate the amount of time that has passed, a shorter gap implies almost immediate continuity, whilst a longer gap can suggest months or even years. A pragmatic approach is often adopted in drama where sound effects are mixed unnaturally low to aid dialogue clarity as well as to help prevent audience fatigue, and subsequent disengagement (Starkey, 2004). Radio, more so than any other medium, generates remarkable interest by the listeners in the accuracy of sound effects. A simple mistake of a bird that is heard to sing in the wrong month will generate a number of letters of complaint (McLeish, 2005, Beaman, 2006).

3.3 Film

When creating soundscapes for films, dialogue almost always takes precedence, effects and music often vie for prominence within a mix. Purcell (2007) argues that the precedence of dialogue extends to picture as well - very few films will communicate key story information through images, dialogue being, almost always, the medium through which stories are told. Music often provides the emotional backdrop, while sound effects supply much of the context. All three combine to form a soundtrack that ‘is communicable and valid but unanalyzable’, designed to elicit emotions, invoke feelings and set moods (Doane, 1985). Cinema is a transitory experience. If audiences have to stop and think about the soundtrack then the illusion falls apart (Sergi, 2004).

Sound effects were part of films even before recorded soundtracks were available. Theatre organs in the 1920s had a built in toy counter that could recreate vehicle horns through to birdsong. Japanese theatres had Benshi, who narrated films as well as creating vocal effects (Rose, 2009). Wesley Miller, who at the time was the chief sound engineer for MGM, found that when reproduced sound started to become the norm in the late 1920s, audiences did not appreciate the need to have to remain silent in order to be able to hear the dialogue. It was not expected that a relaxing experience should require concentration. Miller (1931) correctly predicted that they would 'fall into the habit of learning how to enjoy our product', stating that the most natural effects are the result of 'intelligent and studied artifices'. The artificial nature of soundtracks is now the norm, with disparate, isolated audio elements being 'worldized' by re-recording mixers. Making sound effects seem as if they occurred naturally at the time of action, ensures they belong to the world that the film inhabits (Sturham, 1974).

Most films adhere to the principle of starve the eye and feed the ear. During the introduction to scenes or transitions between scenes 'atmos[phere]' plays an important part. One of the most useful applications for the designer is the ability to colour the sounds, i.e. affecting each sound's timbre so that the sound itself provides cues about the environment. A radio playing music with a lot of reverberation can suggest a large space - add the sound of gently lapping water and the listener thinks of an indoor swimming pool. Aesthetic cues are also put to use: the crackles and pops of a record player place the music within the scene, and therefore make it diegetic, with inherent narrative importance, but the intention of these additional sounds is not to be consciously heard but rather felt (Beauchamp, 2005).

Everything that can be seen on the screen that can potentially make a sound has to be able to be heard, without imposing upon the dialogue. Therefore the first thing any sound design team does is to list the sounds that are either essential or might enhance the film. This proves often to be of great advantage to filmmakers, as it can transform the artificial into the real: polystyrene rocks can have weight when rumbling towards the lead character, and painted backdrops can appear animated. These sounds are commonly synchronised with the action and provide all of the physical cues about a sound generating object such as its mass, velocity, composition and vibration. The film director Robert Bresson stated that he will replace an 'image with a sound whenever possible' as a 'sound always evokes an image; an image never evokes a sound'. A further example can be seen in *The Empire Strikes Back,* where the director Irving Kershner uses a pneumatic sound to suggest a door opening, in actuality this sleight of hand was a single shot of a closed door cut straight to an open door without any intervening movement (Chion, 1994). Chion refers to synchresis as being the bonding of a sound to a visible source. Synchresis allows Foley artists to represent exploding human heads with watermelons in horror films. Synchresis is a form of analogy, where isomorphic or iconic sounds can stand in for real-world sounds, as long as they match the audience's perception of what the source's timbre and dynamics should be (Ament, 2009).

Motifs are the most useful form of sense making sounds, making connections to story, plot and mood (Burwell, 2003). A clock ticking can suggest a heart beating regularly, implying that a character is trying to calm themselves before a confrontation. The groans of a nonexistent character as someone kneads bread could show that the actor is taking out their anger on the bread as a substitute for the unseen person. Ironic sounds are the mainstay of comedy, the miniature pistol that produces the sound of a canon, while the much larger gun creates merely a pop.

An experienced sound designer limits the number of key sounds that an audience has to interpret. This is achieved through judicious mixing, when after being introduced, a sound may be dropped in volume, or its frequency balance altered so that there is no spectral overlap. More often a sound is dropped, emulating habituation, as the sound has moved from providing information about object and event to effectively becoming the a background object. This form of mixing guides the audience through a series of key sounds so that the audience can make sense of the complex auditory environment they are inhabiting. This selection is essential, as irrelevant sounds have to be ignored in order to interpret what is either most useful or most interesting (Bordwell and Thompson, 1985).

Sound design in films has little to do with the presentation of reality. Sound design is a technique, which if used successfully, manipulates perceived reality, enhancing the narrative and contributing towards the suspension of disbelief and placing the viewer within the environment of the story. Designers are not concerned with recreating sounds, accuracy is confined to the genre of documentary. Sound is used to extend the screen, highlight truths and obscure lies, through the use of synchrony and conventions.

3.4 Television

Contemporary television shares much of the sound design approaches of cinema, and with each generation of technology, gradually edges closer to film's production values (Nelson, 2007). Early television sound owed more to radio, as early broadcasts were live rather than pre-recorded, preventing the use of post-production techniques that are ubiquitous in cinema (Mott, 1993). Television, like radio, has also to contend with the pre-existing auditory environment, as well as typically poor domestic room acoustics. An effective sound design can help viewers ignore the sounds going on around them, as well as transport the viewer into the programme's reality (Rose, 2009).

Alkin (1989) argues that sound has a greater role to play in television than in film due to the comparatively small screen size. Alkin states that sound provides 'most of the information and substance'. The greater real estate and higher resolution when an image is projected on a cinema screen can provide considerably more pictorial information, making it easier to interpret what is taking place. This means that television sound has to be truly equal with picture. Unlike film, where most of the sound is created during post production, television sound has a greater emphasis on capture during production, concentrating less on manipulation and more on representation (Holman, 2002).

Television relies on sets for interior work to a greater degree than film, as multiple camera setups are often not practical outside a studio environment. Rooms that are supposed to be in different locations, are normally constructed next to each other with adjoining walls. The sets, having no front or top, often sound the same, lacking the ambient acoustic cues that provide context and perspective. These missing audio elements are easily added during the recording. Reverberation units place the dialogue in the real environment. Sound effects, cued at appropriate points, recreate the buzz track that would naturally occur if the material had been shot on location. This approach would rarely be adopted in the film industry (Lyver, 1999).

Holman (2005) emphasises the importance of the four dimensions of a sound track: frequency range, dynamic range, space and time. Televisions have a number of limiting factors that affect the sound design; due to limited loudspeaker size there is a limited frequency response. Low listening levels can make hearing bass difficult. Even when compensated for, the increased bass content can overtax reproduction systems. Domestic interiors' acoustics routinely colour the sound, as well as being poorly isolated from the surrounding environment. Within a cinema the acoustic environment is more controlled. The dynamic range of television sound is also limited, with the additional problem of highly compressed advertisements that can make programme material sound too quiet by comparison. Internal speakers distort at lower levels than commercial external arrays, reducing the potential dynamic range. The dynamic range is also hampered due to the domestic nature of television reproduction in shared environments. Space is limited, stereo is the norm, but surround sound is comparatively rare, which means that all of the sound elements have to emanate from the same two sources, making spatial positioning of sound sources limited. Time is inhibited due to advert breaks and the episodic nature of the medium. However, all of these limitations can work to the broadcasters advantage. Viewers have considerably more control over the experience through volume control and time shifting. If something was inaudible then the section can either be replayed or the volume increased. There is also less of an expectation for spatial information, as the prevalent desire is to minimise the technology's footprint, which negates a wider spatial stage (Wyatt & Amyes, 2005).

Television's strength is its intimacy emphasizing clarity above all. Sounds are ordered sequentially moving from focus to focus, rather than being layered, with a greater emphasis upon the gaps (silence). Dialogue recording concentrates on providing density and intensity, with spatial cues being de-emphasised, and sound events commonly used to expand the real estate, increasing the apparent screen size. Consistent use is made of the way in which auditory perspective is controlled, a sound source appears closer or further away according to its frequency content. For example when a cardioid microphone is closer to the sound source there is both an increase in bass (proximity effect) as well as high mid frequencies (presence), and these mimic the experience of a listener being close to a sound source. The apparent distance or auditory depth of field can be affected by the percentage of reflections introduced into the signal path, the more reflections compared to the direct sound the further away the source appears (Wyatt & Amyes, 2005).

Aural masking is kept to a minimum, so ambiences or room tones are represented by minimal sounds: such as a low continuous fan-like rumble reminding us that we are on a space ship. Separation is treated simply, panning being used sparingly for specific effects and pre-recorded music kept in stereo, rather than summed to mono, in order to prevent phase cancellation problems upon reproduction. The illusion of continuity is important within television sound design. If the audio were to match the picture exactly, especially when jump cuts are used, then the audience would quickly become disorientated, so audio transitions are slow and smooth unless a jarring effect is desired (Zaza, 1991).

3.5 Video games

Audio is considered indispensible within video games, its active nature aiding immersion when used with accurately synchronized visual imagery (Laurel, 1991, Guerraz & Lemordant, 2008). Sound within a game can transform the listening experience from being passive to active, as sound provides cues about possible dangers to a gamer's avatar (K. Collins, 2007). Jorgensen (2008) argues that sound can also aid usability as well as affect a player's performance. Both cinema and audio interfaces have influenced the design of game audio, as together they balance a need for usability with a cohesive narrative structure (Jorgensen, 2006).

Sound design for video games takes a different approach to that of film or television, but shares a lot of the production techniques. Where game audio differs is in the method of reproduction. Games rely heavily on what is called 'mixing on the fly' or interactive mixing. Unlike a linear narrative, where sound events can be auditioned in order to ascertain effectiveness, within a game the element of randomness has been introduced through adaptive or interactive audio. Sound within games has to be designed for a wide number of variables. Sound is typically divided into three distinct categories: dialogue, music and effects, all of which are triggered individually according to the player interaction (Sanger, 2004).

Dialogue within games can be related to avatars within the game as well as gamers in real life (IRL). Communication can be conducted through interpersonal communication or voice recognition (Fohlmann, 2004). Technologies such as Dolby's Axon not only allow players to chat with each other online, but also spatially locate the voices according to the game play, using head related transfer functions (HRTFs) (Dolby, 2009). Dialogue is used as an auditory backdrop as well as method of communicating key information (Chandler, 2005). Using dialogue as a backdrop can create problems with intelligibility, due to a clash with the music and sound effects, and is commonly resolved by providing subtitles (Griffiths, 2009).

As with television, listeners can experience the sound through different reproduction systems, from mono through to 7.1 surround sound, even within the same game platform. In real time, the volume, panning and frequency of a sound can be controlled in stereo, whereas in a 3-D space (surround sound) sounds can appear to move freely around the listener along x, y and z axes (Turcan and Wasson, 2004). Audio spatial cues contribute to immersion within games in a manner similar to cinema. The audio relates to the perspective of the virtual camera, allowing players to easily navigate between first and third person perspectives, without becoming disorientated (Stockburger, 2003). The main difference between films and video games is spatial accuracy. Films are more dependent on scattered sounds, whilst games require precise spatial cues in order to maximise the player's in-game survival (Wilde, 2004). Breinbjerg (2005) believes that spatial cues can be either architectural, relational, or place. Architectural cues refer to the quantifiable elements that inform the gamer about the dimensions and surfaces of an environment. Relational cues address the relative positions of the avatar and any sound objects within the environment, whilst place cues are concerned with context, and the sounds which make a game site specific invoking connotations and images.

Sound effects can be a powerful tool within games, as they convey information concurrently about both the game play environments and objects, unlike music or even speech where it can be difficult to interpret more than a single stream at a time (Röber and Masuch, 2004). It is important to differentiate between sound events associated with a gamer's avatar and those of other characters (Friberg & Gardenfors, 2004). Sound objects can be thought of as having direct, indirect or environmental communication. Sound can be generated by a gamer's actions, or by a follow-on event triggered by the game, or completely independently in order to provide a sense of the world the character inhabits (Bernstein, 1997). Sound effects can also be thought of as signals that accurately portray sound events, or referents that symbolise actions (I. Ekman, 2005). Interpretation of sound can be further complicated in multi player games as one player hears a sound event but does not attend to it, whilst another player listens and responds (Grimshaw, 2008).

In order to ensure that repeating sounds, such as pistol reloading and firing, do not bore the listener too quickly, randomised elements are used for all of the signature sounds. A simple system of random except last sound file played is used (Lecky-Thompson, 2002). Sounds are constructed in the same manner as within the film industry, (a palette of raw sounds augmented with enhancements) however a greater number of alternatives are provided. The alternative sounds are essential in order to reduce the number of audio repeats (Brandon, 2005). Repetition is a contributing factor when it comes to users either switching off the audio, or the game entirely. This randomisation approach is also adopted for speech so that every time the player comes across the same character they do not hear the same 'barks'. A newer approach is to use granular synthesis, where individual sounds are segmented into grains, which are sometimes shorter than 100 milliseconds. Pools of grains can then be recombined in a wide variety of ways in order to change their temporal, spectral, dynamic and spatial properties. Granular synthesis is not confined to any one form of sound being suitable for dialogue, music and sound effects (Paul, 2008).

Foley, named after Jack Foley, is the technique for recording sound effects in synchrony with picture, and is used extensively during the sound design process to provide realism. Often sounds are exaggerated in order to improve player gratification. Sounds are designed to provide the player with auditory feedback about their own actions, as well as to either prepare them or distract them from an event about to occur. In particular designers focus on the volume, mass, speed and interacting materials of the sound generating source. Sounds have to play in an appropriate sequence in order to allow the gamer to focus their attention. Care has to be taken in order to prevent the player's auditory environment from becoming saturated with too many overlapping sound events. Sounds rarely remain unaltered, as each designer seeks to create maximum effect, this being achieved through temporal, spectral, and dynamic manipulation of mixed sources (Childs, 2007).

Whilst the techniques of sound design are almost identical within the film and games industries, the original sounds from a film are not always suitable for a game. For the game version they can be sonically too rich in order to work well in a runtime mix with at least 130 sounds pre-loaded at any one time. One solution is to pare down the quantity so that only necessary sounds are triggered. Concentrating on the player's point of view means that not every action within a game has to have a corresponding sound event (Boyd, 2003).

Within the games industry, soundscape is a commonly accepted term for ambience. Ambience denotes environmental sounds, which consist of two types of elements: continuous and periodic. Continuous sounds are normally audio loops with varying frequency and dynamics. Periodic elements are typically environment specific randomized one-shot sounds. Attention is paid to ensure that sounds remain at an indeterminate distance, so as to be perceived as background sounds. Ambient sounds are played continuously throughout the game in order to help keep the player immersed within the game play. Any silence could allow a gamer's attention to wander, making them aware of the physical world around them (Marks, 2001).

In terms of dynamics there has been an over reliance on audio compression in games. There are two main reasons for this: limited mixing capabilities in games, and a desire to be louder than the previous game. Over compression is a problem referred to as the loudness wars within the music industry (Katz, 2002). A common technique in the film industry is to drop the volume prior to any cues that are to appear loud. This drop in levels increases the apparent volume, as listener perception is based on comparison rather than absolute levels (Bridgett, 2008). The adoption of working reference levels, such as those set by THX (2010), in conjunction with subtractive mixing, will improve the clarity of the individual sound elements and significantly improve game-play (Bridgett, 2008).

There are a number of parameters pertinent to designing the spatial dimensions of sound events. A sound cone or spatial projection of the source can affect both volume and frequency according to the player's orientation to the sound source. Minimum and maximum distances can be set in relation to the listener. The minimum marks the closest point at which volume starts to drop, and maximum marks the furthest point where the sound becomes either inaudible or barely audible. The fall-off itself can also be adjusted to anywhere between linear to exponential. Sound propagation is also considered, just as graphics are seen from the position of the virtual camera, audio is experienced from a virtual microphone. Through the technique of acoustical modelling, direct path audio is augmented with echo and reverberation. Environmental geometry and material composition are calculated in real time in order to create early and late reflections, diffusion, occlusion or transmission along with their material related frequency colourations. The designer inputs details about the size, shape and absorption rate of individual rooms, as well as the type of transitions (threshold or position) that occur when moving between environments (Boer, 2003).

The most common way of integrating audio into a game is to create zones or triggers. A zone denotes a predefined area that can be any shape or size, where a sound event or looped sounds play while the character is within the area. A trigger is usually more precise, acting as start or stop for either loops or one shot events. Characters are usually split into two: player characters (PC) and non-player characters (NPC). Sound events are normally associated with specific animations, and are triggered at specified frames within the animation sequence. This applies to both types of characters as well as all of their actions, irrespective of whether it is a sword being removed from a sheath, or a simple footstep. Objects are dealt with according to their type. Sound events can either be attached to frames within an animation, or to the object itself as part of a scripted event when more complex interaction is required (Brandon, 2005). Within games, unlike other forms of media, sound effects have priority over music and dialogue. Sound provides valuable information to the gamer about what is happening in their immediate environment, and beyond what is immediately visible on the screen in front of them.

3.6 Auditory Displays

Auditory displays have been defined by Kramer (1994) as an interface between users and computer systems using sound, and are considered a natural extension of the way in which sound is used in the physical world. Auditory displays differ from auditory interfaces in that they are only one way. An interface allows audio to be used as input as well as an output, whereas a display is only an output (McGookin and Brewster, 2004). Auditory displays can be split into the user interface audio and audio used in visualisation. User interfaces include earcons, auditory icons, sound enhanced word processors (text to speech), and other applications, whilst sound in visualisation includes audification, sonification, and auralisation (Vickers, 1999).

It has been argued by Brewster (2008) that audio within computing is still underutilised, despite its potential wealth of information. Cohen (1994) highlighted the need to use sound professionals rather than computer scientists, in order to ensure an aesthetically pleasing blend of sounds has been emphasized. Concerns have been raised about users not being considered sufficiently in the field of auditory display design. Barrass and Frauenberger (2009) state that designers need to consider the context of use, as applications might be used in a wide variety of environments. Fagerlönn, J., & Liljedahl, M. (2009) warn that end users may not feel confident enough to provide informed feedback about sound designs.

3.6.1 Speech

Speech is used both as a form of input as well as output within auditory interfaces. Speech recognition can be geared towards individual words used as commands through to consecutive words where the results are converted to text. Text to speech is a speech synthesis system commonly used to aid universal access, and is built into most operating systems. Products such as Amazon's Kindle now have a read-to-me function, with limited controls, in addition to its ability to play audio books (Amazon, 2009). There are problems with the pronunciation of words and numbers according to their context, which a few systems overcome by using a pronunciation dictionary that checks the context (Riley *et al.*, 1999). There are prosodic considerations to be included such as intonation and stress, but there is always a balance to be achieved between what is considered aesthetically pleasing and realistic, as opposed to comprehensible (Eide, 2006).

Spearcons have been proposed as an effective method of augmenting visual interfaces. Spearcons are speech synthesized words or phrases speeded up between 1.4 – 2 times, without an associated pitch change (Walker, Nance and Lindsay, 2006). Spearcons are an alternative to text to speech (TTS) and are based upon the discovery that blind listeners can still understand speech when it was replayed at a rate of 1365 words per minute, which is 2.8 times faster than standard TTS (Asakawa *et al.*, 2003). Spearcons can be easier to learn than earcons, as each word or phrase has a distinct sound whilst still being directly related to the textual source (Palladino and Walker, 2007), however spearcons have yet to be adopted by the broader auditory display research community.

3.6.2 Audification

Data can be represented by sound by two related methods of display each with decreasing levels of directness: audification and sonification. Audification is applied to sound generated by data with only minimal manipulation. Audification usually takes the form of pitch shifting and signal processing, in order to bring the data within the human range of hearing. There is no new sound generated, what is heard is a rendition of the original waves. The concept was first introduced as a way of listening to seismic sounds in order to distinguish between earthquakes and nuclear tests (Speeth, 1961). It was later found that a greater than average number of trained listeners were able to interpret the difference (Frantti and Leverault, 1965), and that audification was an appropriate technique for testing errors in seismological measurements (Hayward, 1994). Audification is only considered suitable as a companion to visualisation, and that its strengths lie in areas in which visualisation is weakest (Dombois, 2001).

3.6.3 Sonification

Sonification refers to a technique for transforming data into an audible stream that is analogous with data visualisation (Kramer *et al.*, 1999). It can be argued that a sonification method must be objective, systematic, reproducible as well as suitable for use with different data (Hermann, 2008). Data can be split into auditory streams where each stream is linked to a specific audio variable such as pitch, volume, note duration, fundamental wave shape, attack (onset) envelope and overtone (harmonics) wave shape. This can make the data not only more informative, but potentially increase the amount of information able to be transmitted concurrently (Bly, 1992). Interactivity can also be incorporated, allowing users to specify which area of the data is to be sonified, with stereo panning and variations in reverberation providing additional information about the horizontal orientation and relative depth of icons (Smith, Bergeron & Grinstein,1990).

Sonification uses a number of synthesis techniques for representing data characteristics (Scaletti and Craig, 1991) (see Table 3-1). An alternative to the prevailing paradigm of mapping dimensions to parameters is model-based sonification (MBS). MBS differs in that there is no sound until the end user excites or triggers the model. Data is mapped to objects that generate sound according to the type of interaction. The argument is that in the physical world objects do not make any sound unless an action is performed. MBS provides information about the type of excitation, and that the technique of applying virtual physics to sonify data makes the process more intuitive for listeners (Hermann and Ritter, 1999).

Technique	Description
Shifter	Data is converted to acoustic wave and shifted to become audible
Mapper	Stream of data affects series of audio parameters
Analyzer	Data affects parameters of audible sounds
Combiner	Sums, differences or products of multiple streams
Comparator	Stereo panning of two streams to compare differences
Marker	Sounds triggered when predetermined criteria are encountered
Histogram	Frequencies applied to attributes and magnitude for relative intensity

Table 3-1: Sonification synthesis techniques

A number of concerns have been noted about sonification, sonified data should be accurately presented and perceived in a form that is appropriate for the set tasks. Sonification should require a level of training that is proportionate to the benefits, and the suitability of the sounds should be thoroughly considered in order to ensure use (Barrass and Kramer, 1999). In general terms, pitch changes are beneficial for variations in numerical values, temporal variations convey differences in quantities effectively, and changes in loudness are useful to convey key points or changes in a data stream. In addition, contrasting timbres help prevent individual streams of data being merged, temporal patterns provide a natural translation for time series data, and presenting data sequentially in short bursts imparts a worthwhile approach for the comparison of data. Anything over three concurrent data streams is difficult to comprehend, especially if there is insufficient difference between timbres. Care should also be taken when using discrete loudness levels for detailed data, as it can easily become distorted or inaudible (Flowers, 2005).

Sonification design patterns, inspired by Alexander (1979), have been speculatively proposed in order to create sonifications that support the necessary functionality (Barrass, 2003). If used extensively, this would lead to a Pattern Language that could be used to study sonification in general. However, there are many difficulties associated with evaluating different types of interactive sonification, due to the sheer number of alternative approaches. One of the main problems is establishing a system's learnability versus its usability (Hermann and Hunt, 2005). A number of simple prototype systems have been successfully created and tested (Holland and Morse, 2002, Ramakrishnan and Greenwood, 2009), although there has been limited adoption of sonification in general, along with other forms of non-speech audio (Nees & Walker, 2009).

3.6.4 Auralisation

Auralisation has been referred to as a synonym for sonification (Gaver, 1997). It is also a method of aurally rendering soundfields, an approach commonly used for simulating architectural acoustics or loudspeaker installations (Kleiner *et al.* 1993). Auralisation is intended to enhance (M.H. Brown and Hershberger, 1992) and eventually replace visualisation (Alty, 1995). It is often based upon polyphonic techniques used in western music since the Middle Ages. This reliance on musical principles means that individual parts are as easily identifiable as the whole (Blattner *et al.*, 1992).

Auralisation can be used as a method of aiding programming and debugging tasks (Francioni, Albright & Jackson, 1991, Digiano & Baecker, 1992). Musical auralisation, when used carefully, is effective in representing concurrent processes without them becoming confused by the listener (Alty, 1995). It has been found that it is easier to find bugs, with little impact upon speed, but there is a noticeable increase in the workload of the participants (Vickers, 1999). A knowledge of sound design is often required in order to utilise early systems effectively, and it has been suggested that in order for the field to progress successfully it has to move beyond interaction design principles into the wider field of musical aesthetics (Vickers, 2004). This corresponds with more general advice that sound used in auditory displays should be as aesthetically pleasing as sounds found in nature or classical music (Gaver, 1997).

3.6.5 Earcons

Earcons can be defined as nonverbal audio messages directly related to icons (Sumikawa, 1985, Blattner *et al.*, 1989). A pun on icon is used to indicate the difference in medium (Buxton, Baecker and Arnott, 1985). Short, discernable, musical phrases, or motives, allow numerous alarms to be understood concurrently (Patterson, 1982). There are two types of earcons, two fundamental configurations and four methods of creation. Representational earcons refer to sounds that symbolise existing sounds, and can be thought of as auditory icons, which are discussed later. Abstract is applied to earcons when motives are used. The modular nature of earcons allows grouping within families, so that motives representing related information sound alike, and those conveying disparate information are dissimilar.

One-element earcons can be either single-pitch, or single-motive. Single-pitch applies when only a single note used, but the earcons still have pitch, duration and dynamics. Single-motive earcons have the additional parameters of rhythm, timbre and register. One-element earcons normally represent basic operating system actions such as file opening and closing, or error messages. Compound earcons can be constructed by either combining, inheriting or transforming a number of audio pieces. Combining adds the elements sequentially, and inheriting varies the motives according to which of the maximum of five hierarchical levels the earcon belongs. Transformation applies to earcons that retain the same motive for each level, but have it slightly altered according to timbre, dynamics or register, whereas rhythm stays the same, and pitch should only be altered if necessary (Brewster, 1994).

Earcons have to be memorised by the listener in order to successfully map audio sequences to specific functions, and the level of difficulty varies with each method of creation (Blattner *et al.*, 1989). Representational earcons such as the recognisable sound of a piano 'catch phrase' are the simplest to learn (Brewster, 1994). The total number used within a system is necessarily limited due to the large number of unique potential mappings and the difficulty in applying them to abstract functions. Combined earcons provide an easier way of remembering sound events if the first level is already understood. Hierarchical earcons can be very difficult to learn, both because of the sheer number of possible combinations, but also due to the complex nature of the alterations. The process of learning the implied meaning can take place whilst the user is navigating through a sonified hierarchical interface, without the need for initial training (Leplâtre and Brewster, 2000).

The arbitrary nature of mapping earcons prevents users applying their own previous experiences, which means that each set must be learned anew, and guidelines have been created to aid their design (see Table 3-2). There is also a tendency for earcons to sound like musical phrases, which does not suit workplace environments, and can quickly become annoying through repetition. Earcons are often too long in order to optimise identification (Gaver, 1997). The reliance upon an end user's memory, which is inherent in the design of earcons, limits their potential. As hearing problems increase, memory becomes correspondingly erratic, and accurate recognition of earcons decreases (Rabbitt, 1990, G. W. Coleman, 2008).

Guidelines	**Authors**
Vary rhythm with notes no faster than quavers	Patterson, 1982
Use pitch with another parameter unless notes are at least 2 octaves apart	Brewster, Wright & Edwards, 1994
Changes from major to minor mode convey positive/negative emotions	Lemmens, 2005
Use musical instrument timbres with complex harmonics	Brewster, Wright & Edwards, 1994
Avoid instruments with similar timbres	Brewster, Wright & Edwards, 1994
Intensity must be between 10 - 20 dBs above threshold	Patterson, 1982
Minimum pause of 0.1s between earcons	Brewster, Wright & Edwards, 1994
Varying spatial location allows simultaneous playback of earcons	Brewster, 1994
Extreme variations make earcons more demanding	Edworthy *et al.*, 1989, Brewster, 1994
Dissimilarity of earcons enhances interpretation but can be displeasing	Leplatre, 2002

Table 3-2: Guidelines for developing earcons

3.6.6 Auditory icons

Auditory icons and earcons appear to occupy the two ends of a presentation continuum with auditory icons being representational and earcons being abstract, but this distinction is often blurred. When auditory icons are created without a direct real world sound they are necessarily abstract, and when earcons remind listeners of sounds experienced outwith an interface, they attach their own representational interpretation (Brewster, 2008). Auditory icons, such as the delete sound when emptying the recycle bin, are utilised by designers of auditory displays to reveal data about systems through everyday listening. They typically take the form of pre-recorded samples that are analogous to the action being performed (Gaver, 1986). Auditory icons have been described as just another form of earcon, specifically representational or 'caricatures of normally occurring sounds' (Blattner *et al.*, 1989).

The advantage of auditory icons over earcons, is their familiarity, they are easy to learn, and have been used successfully by the film industry in the form of Foley for decades (Begault, 1994). Auditory icons can also be effective when used in combination with earcons, again due to their familiarity (Brewster,1994, Fitch and Kramer, 1994). Auditory icons have been found to be more precisely correlated with meanings than earcons. Earcons are most likely to be misinterpreted by inexperienced users. This disparity between auditory icons and earcons decreases as participants are made aware of the hierarchical nature of the sound design, as well as with increasing familiarity (Lucas, 1994, Jones and Furner, 1989).

Auditory icons use three possible forms of representation: symbolic, metaphorical and nomic, and they typically fall between two of the three categories. This affects their ability to be learnt, nomic being the easiest and symbolic the hardest. Symbolic sounds are based on conventions (i.e. telephone bell) where there is no natural mapping to its physical properties. Metaphorical auditory icons rely on the comparability of the sound with the object or action. These can be either metonymic, where the sound represents the object, or structure mapped, where one attribute represents another, such as pitch variation indicating fluctuations in values. Nomic represents the norm, where the sound represents the source, although a single source might have many different sounds, as well as sounds being interpreted as being generated by a variety of potential sources (Gaver,1986). Genre sounds from dramatic productions have been proposed as effective auditory icons. These symbolic sounds are easily recognisable through previous repeated exposure (Cohen, 1993). Unfortunately the cultural specificity of genre sounds make them difficult for listeners to adopt their new meaning within an interface, as the sounds retain their original associations. Despite this problem they are still appropriate for interface design, as they can be both entertaining and effective (Gaver, 1997).

Auditory icons are optimal when short in duration with a wide range of frequencies. Their corresponding lengths, levels and timbre should be similar. Cues need to be evaluated for their recognisability using natural language, and those not easily named, tested for their learnability. Conceptual mappings require testing by varying the concept represented by the sounds, and finally usability tests should be performed with the resultant auditory icons within interfaces (Mynatt, 1994). The inclusion of sound increases as well as decreases reaction times. Graham (1999) argued that auditory icons improve performance, whilst earcons have the opposite effect. This echoes the assertion that the inclusion of sound in computer games affects player's performances, both in positive and negative terms (Edworthy, 1998).

The annoyance factor of frequent exposure to auditory icons has also been highlighted, and earcons have been suggested as a suitable alternative (Bussemakers and de Haan, 2000). In order to help prevent the overall auditory interface from being too prominent it has been suggested that ambient sounds need to be used as a foundation, with more intense cues to sitting on top (Cohen, 1994). Real world sounds map most accurately the intended functions, and abstract sounds are considered more pleasant and appropriate (Sikora *et al.*, 1995).

3.7 Conclusions

Sound designers routinely manipulate the attributes of sound as part of their everyday practice. The soundtrack in traditional media is commonly split into dialogue (speech), music and sound effects during production (Adams, 2009, Buhler, Neumeyer & Deemer, 2009, Butler, 2007). Sounds can be designed to make an audience aware of certain actions as well as to allow events to go unnoticed through masking, such as a hiding awkward scene cuts (Cancellaro, 2006, Dakic, 2007). Foley artists will use stand in objects made of similar materials with equivalent interactions to represent sound events that cannot be captured any other way. Examples include squeezing liver in a bag to represent ET walking and a cantaloupe melon being ripped open for flesh rips (Ament, 2009, Viers, 2008).

The length of a sound can be used to convey a character's emotions, a longer bell ring can suggest impatience (Kaye & Lebrecht, 2000). The length of a silence (or lack of sound) can also be useful to convey the passage of time and even a change of location (Beaman, 2006). Changing a sound's pitch can make objects sound larger or smaller, and can also alter the age or gender of a character (Beauchamp, 2005, M. Collins, 2003). The dynamics of sound are used to suggest the intensity of an action, but it is not always necessary to use high levels, a short period of silence or soft sound either side of a louder sound can increase the apparent volume (Kerins, 2010, Whittington, 2007, Yewdall, 1999). Spatial cues, such as panning, can provide an insight about what a character is attending to (Beck & Grajeda, 2008, Kerins, 2010).

Designers of auditory displays are concerned about sounds being considered informative rather than uninformative or noise (Buxton, 1989, Brewster, 2008). The aesthetics of a sonification are thought to affect its usefulness. If a design is too pleasing it becomes musical and listeners are distracted (Vickers & Hogg, 2006). However, if a sonification is displeasing it can become annoying (Henkelmann, 2007). Clarity is an important issue for video game sound design as it can allow a player to identify what is going on and react accordingly (Bridgett, 2008, I. Ekman & Lankoski, 2009). Emotions are an increasingly important for the design of auditory displays, positive sounds can convey that no action is required whilst negative alerts can indicate that urgent action is required (Larsson, 2010). It is thought that emotional sounds are responded to more quickly and attended to for longer, making them ideal for auditory displays (Schleicher, Sundaram & Seebode, 2010).

Certain forms of media place greater emphasis on manipulating specific attributes of sound events. Within computing, greater emphasis is placed upon making sound events informative, whereas within film more attention is paid to the dynamics of the sound events, whilst in video games spatial cues are sometimes the most important. However, every attribute of sound identified above is manipulated to some extent during the sound design process for the different media.

The two forms of sound for computing discussed within this chapter have had a distinct difference in end-user and developer uptake. Sound within video games is ubiquitous, despite being beholden to the visual image, and game sound design techniques are now being used for theme parks and installations (K. Collins, 2008). The uptake of audio within mainstream computing has been poor (Brewster, 2008). Little work has been conducted on audio displays within workplace or domestic contexts. Most of the work conducted has concentrated on non-speech audio, which equates to music and sound effects within traditional media. Developers of earcons are careful to state that their abstract sounds are not music, despite being musical, and that to make their interpretation more effective earcons should be dissimilar and therefore less musical as a whole (Leplâtre, 2002). This is in contrast to sonification where is has been advocated that the aesthetics of music should be taken into consideration (Gaver, 1997, Vickers, 2004).

Within traditional media Radio and Television have to accommodate for the pre-existing auditory environment, whilst Theatre and Film have less to contend with due to acoustically treated auditoria and over amplification (Banham, 1995, Eargle & Foreman, 2002, Nisbett, 2003, Rose, 2009). Within the home it is assumed that the audience will listen at quieter levels, where attention may be limited (Holman, 2000, Nisbett, 2003). Having discussed sound design approaches according to media type the next chapter describes some preliminary studies that investigated the procedural difficulties encountered when comparing listening experiences.

4 Preliminary studies

The previous chapter discussed different methods of designing sound for media. It was found that although sound design methods are well established in traditional media and video games, this was less evident within computing. This chapter describes preliminary studies that were conducted in order to establish what some of the procedural difficulties are when trying to compare listening experiences. The four studies addressed capturing, classifying and visualising listeners' soundscapes.

A prototype soundscape mapping tool was trialled in order to evaluate it as a basis for developing an extended version. The prototype's original paper did not specify a procedure to follow or include a form of visualisation, so procedures were trialled and a visualisation developed. In order to identify which attributes of sound should be included within a soundscape mapping tool a series of interviews were conducted with listeners. The interviews were used to identify attributes for describing sound suitable for extending the prototype tool. The extended prototype soundscape mapping tool was used to combine and compare listeners' experiences of an open plan office. The tool was extended to include measurements of sound events, as well as additional classifications. Surround sound recording was trialled as a method of allowing listeners to experience a similar soundfield and listeners' classifications were combined so that maps could be created to show both individual's and groups' experiences. The results from these preliminary studies identified procedures to follow in order to elicit and visualise listeners' experiences of soundscapes.

4.1 Prototype soundscape mapping tool

Macaulay and Crerar's 1998 method was selected for this preliminary study, as its aim was to describe a soundscape using maps for the benefit of auditory display designers. The authors suggested that both fieldworkers and designers could use the tool, but the practical procedure was not specified. The method was not put into practice or evaluated in the original paper. This study proposed a procedure to follow for capturing and visualising listeners' experiences based on Macaulay and Crerar's 1998 method.

4.1.1 Method

The prototype method incorporated three distinct phases: identification, classification and visualisation. Sound events within an auditory environment were identified by listeners, and were subsequently classified by listeners using Macaulay and Crerar's (1998) criteria (see Table 4-1). The results were then visualised by the researcher so that listeners' responses could be compared.

Sound Type	**Example**
Music	Any type of identifiable music, radio/stereo
Abstract	Unusual sounds not normally experienced, video recorder chewing a tape
Speech	Conversation
Everyday	Identifiable recognised sounds.
Information Category	**Example**
Visible entities and events	The phone ringing
Hidden entities and events	The photocopier round the corner being used
Imagined entities and events	Something big is happening on the political desk (it has gone quiet).
Patterns of events/entities	Someone is batch copying a large document
The passing of time	It's nearly deadline time (because the shift change is happening)
Emotions	The sports desk sub-editor is unhappy (tapping)
Position in Euclidean/acoustic space of entities/events and of the listener	The editor is at the foreign desk behind me (can hear his voice)
Acoustical Information	**Example**
Foreground	Computer beep to attract your attention.
Contextual	Door opening (Help you orient to the nature of your environment.)
Background	Whine of disk drive providing reassurance or information about the state of the world.

Table 4-1: Macaulay and Crerar's 1998 criteria

Identification and classification

Three different approaches for eliciting listener responses were adopted in order to suggest which was the most appropriate. In the first approach listeners were asked to list in written form all of the sound events that they experienced during a specified time period while remaining stationary. All of the sound events were then classified verbally after the specified time period with the researcher using Macaulay and Crerar's 1998 criteria. The researcher asked the listener to classify each sound event individually according to its sound type, information category and acoustical information (see Table 4-1). In the second approach the researcher listed all of the audible sound events while the participant listened. The participant then classified only the sound events that they were aware of in the same manner as the first approach. In the final approach the researcher listed and classified all of the sound events, emulating the procedure inferred from the original Macaulay and Crerar (1998) paper. For all three approaches the researcher generated the visualisations.

Visualisation

The visualisation was based on a two-dimensional representation created by Macaulay, Benyon and Crerar (1998), (see Figure 4-1). It was not immediately obvious from either the original diagram or the narrative how the representation could be used to map a soundscape. As this was a trial of an existing method, it was decided that representations of attributes from the original model would be retained where practical.

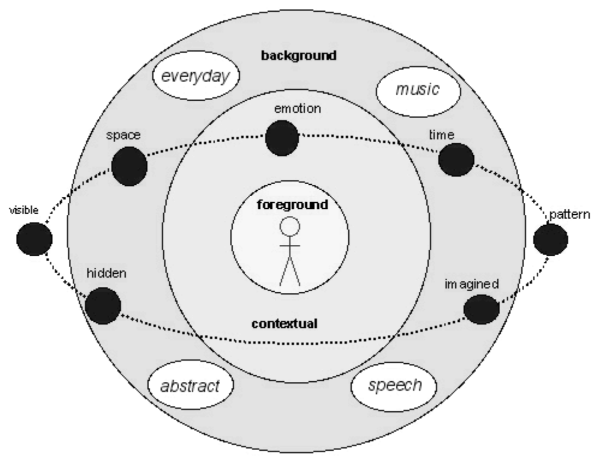


Figure 4-1: Two-dimensional soundscape map by Macaulay, Benyon and Crerar (1998)

In the prototype concept a series of concentric circles represented the acoustical information. The use of an ordinal grid emulated the experience of the listener being located in the centre of their soundscape, as suggested by the stick figure, with the foreground being closest. This ordinal grid was retained, but the contrast of the greyscale values used for the bands was increased to improve differentiation. The stick figure was removed, as it was unnecessary to notate the point of listening, which would remain constant in successive diagrams. Macaulay, Benyon and Crerar (1998) represented the information categories using labelled black circular shapes distributed along a dotted circular line. It was decided to change the information categories visualisation so that the circular shapes could be used to represent individual sound events. In order to retain the spatial positioning a nominal grid of seven labelled segments was incorporated to show the acoustical information.

Originally the sound type was represented using ellipses spaced evenly within the background ring. In the updated visualisation sound type was displayed by labelling each cross-referenced circular shape with a symbol (see Figure 4-2). Qualitative point symbols were adopted as they are commonly used to represent an object in cartography (R. P. Misra and Ramesh, 1989). *Music* was illustrated with quavers ♬ , *abstract* a sequential series of three numbers **123**, *speech* a sequential series of three letters **abc**, and *everyday* by an everyday item, which in this case was an apple .

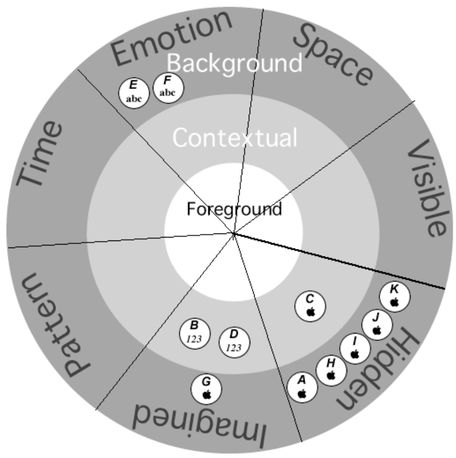


Figure 4-2: Updated soundscape map

In order to help prevent the image becoming cluttered with long descriptions, sound events were cross-referenced using capitalised letters. Sound events were placed onto the diagram according to participants' classifications. Colour was used to either identify different participants' responses or indicate the quantity of responses for each sound event.

Three of Bertin's (1983) six retinal variables were applied (shape, value/grey scale and hue/colour) along with his two locational variables (position). Only the cross-referenced descriptions, sound events and sound type required a key for the individual maps, as the acoustical information and information categories were labelled directly on the map. Further information about the colour was required when results were combined.

Participants and locations

Thirteen participants took part in the study in 12 different locations. Locations included university, commercial and domestic environments. All of the participants who experienced the university environments were postgraduate computing students. Two commercial environments were chosen, a photographic lab and a 50s style diner, in order to provide a broader range of auditory environments, as well as to test the method with listeners who might use auditory feedback within their work. The participants from the photographic lab and the diner were all full-time members of staff who had worked in their environments regularly for a number of years. The final participant who took part in the domestic environments was the author. None of the participants had known hearing problems.

All of the listeners were assured verbally that their responses would be anonymised so individuals could not be identified. Listeners were informed that they did not have to answer all of the questions and could stop at any time. Verbal permission was obtained to make use of the results for this dissertation and associated publication. Permission was not sought from the inhabitants of the rooms under study as no recordings were made, and no one was identified in the responses.

4.1.2 Results

In the first approach participants were asked to write down all of the sound events in their environment for at least 15 minutes. Each of the eight participants spontaneously closed their eyes and stopped what they were doing in order to consider their responses. Participants subsequently opened their eyes to confirm what they had listened to in each instance, before re-closing their eyes to continue. The number of identified sound events varied from 7 to 21 (see Table 4-2).

Participant	Location	Environment	Duration	# Events	Procedure
P01	Library	University	15 mins	8	1
P02	Computer Room 1	University	15 mins	8	1
P03	Computer Room 2	University	15 mins	7	1
P04	Computer Room 2	University	15 mins	11	1
P05	Staff Canteen	University	40 mins	16	1
P06	Staff Common Room	University	30 mins	21	1
P07	Computer Room 3	University	15 mins	17	1
P08	Computer Room 3	University	15 mins	8	1
P09	B/W Darkroom	Photographic Lab	55 mins	25	2
P10	Colour Printing	Photographic Lab	85 mins	35	2
P11	Reception	Photographic Lab	20 mins	20	2
P12	50s style Diner	Diner	180 mins	59	2
P13	Kitchen	Domestic	180 mins	53	3
P13	Study	Domestic	60 mins	46	3

Table 4-2: Summary of participants, locations and events

In the second approach participants were observed while working in either the photographic processing lab or the diner, durations ranged from 20 minutes to 3 hours. The number of identified sound events varied from 20 to 59 (see Table 4-2). Note was made of all the sound sources/events by the researcher, with classification taking place immediately after the listening period. If the participants were not aware of any of the sound events then the event was noted as ‘not aware’ and omitted from the map. Two of the participants were aware of all of the sound events, whilst the other two participants were not aware of a limited number (four and three respectively). Participants were asked if there were any sound events that they thought were missed, no additions were suggested.

The final approach was adopted in a domestic environment where the researcher acted as the listener. Within the kitchen 53 sound events were identified and 46 were identified in the study (see Table 4-2). Written record was made of each new sound event that was heard as normal work was conducted. This approach allowed a longer time period to be studied without an observer that might alter the subject’s working practices. A large number of sound events were identified but tasks such as chopping vegetables in the kitchen and typing in the study took longer to complete.

Classification

All of the 334 sound events listed by participants were classified using the Macaulay and Crerar (1998) criteria (see Table 4-1), and were then visualised. Responses fitted easily into the available categories, and all of the categories were used (see Table 4-3). Within the sound type, participants readily understood the criteria of *music* and *speech*[1]. Occasionally if the speaking was *background* it was classified as *everyday*. The terms *abstract* and *everyday* were not consistently applied; there were three main interpretations. The first interpretation was *abstract* representing a sound that the participant was familiar with but thought was unusual, or differed from the norm. *Everyday* was applied to a sound event that was closer to what the listener expected. The second interpretation was in terms of natural or artificial, *everyday* representing natural, and *abstract* representing artificial. The final interpretation was that of identifiable for *everyday* and unidentifiable for *abstract*.

	Library P01	Computer Room 1 P02	Computer Room 2 P03	Computer Room 2 P04	Staff Canteen P05	Staff Common Room P06	Computer Room 3 P07	Computer Room 3 P08	B/W Darkroom P09	Colour Printing P10	Reception P11	50's Style Diner P12	Kitchen P13	Study P13	Combined	Response	Rank
Sound Type																	
Music			14%			5%				3%	5%	2%		4%	2%	43%	4
Abstract	25%		29%	18%		24%	47%	50%	24%	22%	5%	39%	13%	23%	23%	86%	2
Speech	13%	13%	14%	18%	13%	5%	12%	13%	4%	3%	10%	7%	4%	6%	7%	100%	3
Everyday	63%	88%	43%	64%	88%	67%	41%	38%	72%	72%	80%	53%	83%	67%	67%	100%	1
Information category																	
Visible	38%		13%		38%	38%	47%	38%	44%	83%	50%	44%	36%	22%	40%	86%	1
Hidden	50%	25%	63%	55%	44%		24%	50%		14%	25%	27%	4%	4%	18%	86%	3
Imagined	13%		13%	27%	6%	52%							6%	4%	7%	50%	5
Patterns		38%					12%		48%		10%	22%	49%	57%	25%	50%	2
Time		13%													0%	7%	7
Emotions		25%	13%	18%	6%	10%	18%	13%	4%		15%	7%	4%	7%	7%	86%	4
Position					6%				4%	3%			2%	7%	2%	36%	6
Acoustical Information																	
Foreground	13%					5%	59%	38%	69%	60%	30%	36%	58%	39%	39%	71%	1
Contextual	50%	63%	29%	27%	88%	19%	29%	13%	24%	29%	35%	31%	34%	50%	36%	100%	2
Background	38%	38%	71%	73%	13%	76%	12%	50%	7%	11%	35%	34%	8%	11%	25%	100%	3

Table 4-3: Summary of classification frequency

Within the information category, listeners applied *visible* to sound sources that could either be seen, or easily identified, even if the source was not observable. *Hidden* was used by listeners for sound sources that were not visible, as well as when a

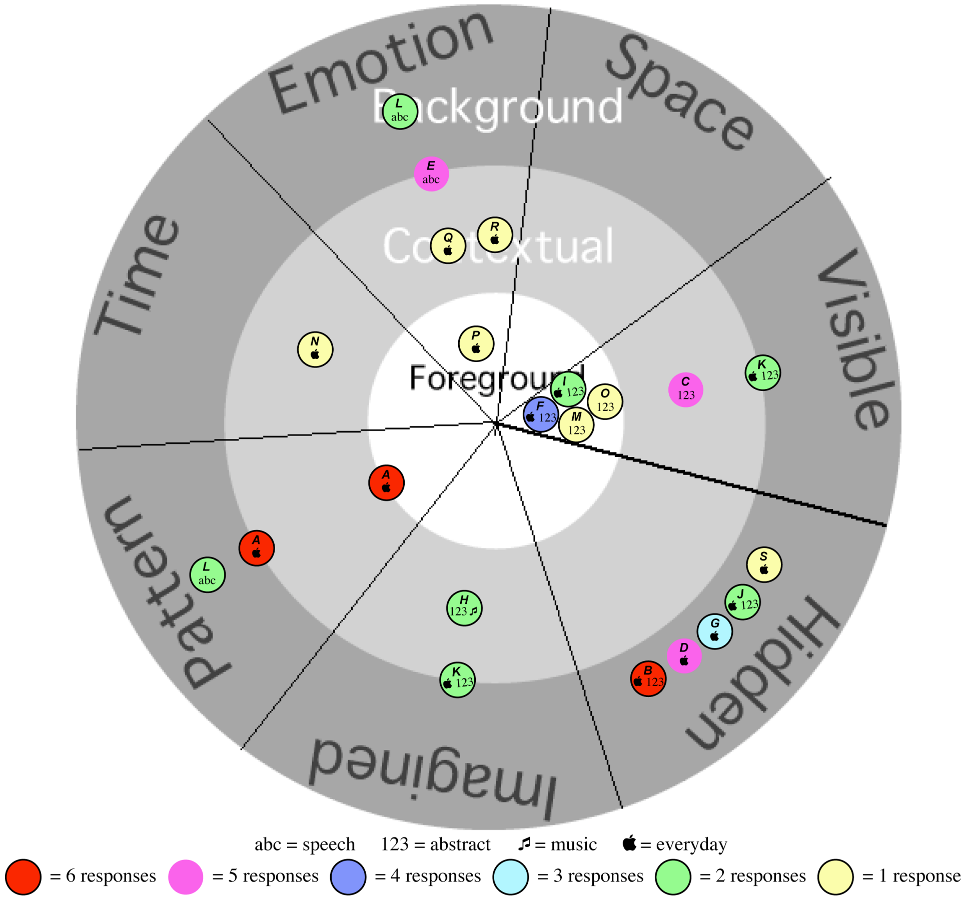
[1] Criteria have been italicised to aid identification.

sound source was observable but could not be easily identified. *Imagined* was applied when an estimate was being made by the listener about the source of the sound. *Pattern* denoted either a series of connected sounds over a short period, or an irregular long-term sound. The *passing of time* was applied only once, and referred to a sound event that acted as a reminder to the listener that it was time to go home. Listeners when referring to speaking as either a *contextual* or *foreground* sound event used the criteria *emotions*. *Emotions* also included actions that informed the listener about someone's mood, specifically impact sounds. *Position* was applied by listeners to moving objects, rather than indicating where a stationary object was located.

The application of acoustical information did not match the original aims of the paper. According to Macaulay and Crerar (1998), a *foreground* sound event provides little information about what is going on in the world around the listener, such as a beep. A *contextual* sound event informs listeners about what is going on contextually in their acoustic environment, and a *background* sound event provides reassurance about everything else that is occurring in the vicinity. The results from this preliminary study more closely represented Amphoux's 1997 levels of listening. F*oreground* sound events were actively monitored and interpreted (sonic symbols). *Contextual* sound events told the participants about the place they were inhabiting (sonic ambience). *Background* was applied to sound events that had been habituated.

Visualisation

Two different approaches were auditioned for combining listeners' responses. Within this preliminary study three of the environments were computer labs, which were experienced by five listeners. The three computer labs were combined to form a soundscape map of a typical computer lab (see Figure 4-3). Descriptions such as: 'chair shuffling', 'chairs moving', 'chair rolling', 'chair creaking' and 'chair squeak' were combined into 'chair movement'. The modal response was chosen for each attribute. If there were two results for the mode both were included, as in the sound type of the ‘chair movement’ which was both *everyday* and *abstract*. Record was made of the number of responses and the descriptions were sorted in descending order. Colour was added to indicate the number of responses per sound event.

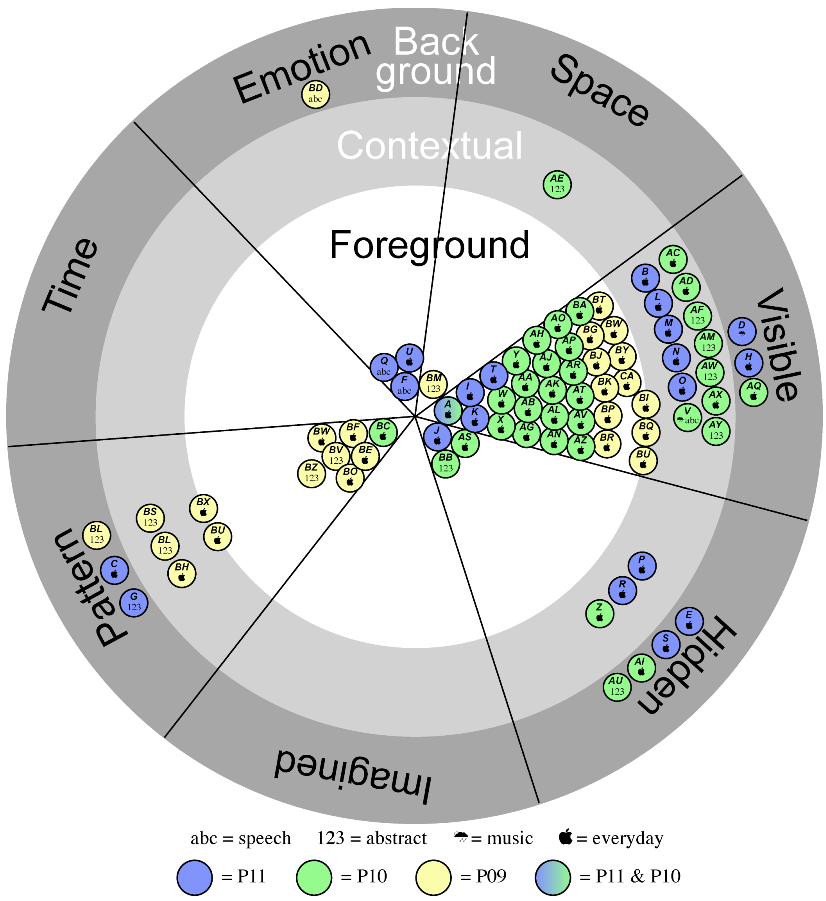


A	Keyboard Typing	**H**	Computer alert	**O**	Purse velcro
B	Chair movement	**I**	Mouse movement	**P**	Laughter
C	Computer Fan	**J**	Air conditioning	**Q**	Coughing
D	Door opening & closing	**K**	Paper movement	**R**	Sighing
E	People talking	**L**	Whispering	**S**	Foot movement
F	Mouse clicking	**M**	Computer hard drive		
G	Traffic	**N**	Bag being zipped up		

Figure 4-3: Visualisation of a typical response to a typical computer lab

With the photographic processing lab it was possible to create a map that represented the entire auditory environment, from the perspective of the listener most familiar with each area. Only a single sound event was shared, that of the telephone, which both P10 and P11 classified as *everyday*, *visible* and *foreground* (see Figure 4-4).

Colour was used in this instance to indicate which listener's perspective was being displayed. This way of combining the results does not compare listening experiences; it merely proposes a way of representing larger auditory environments.



A	Telephone ringing	AA	Paper door closing	BA	Slide Mount
B	Doors opening to lab/computer room	AB	RA4 paper processor Temperature beep	BB	Hairdryer (drying prints)
C	Coffee machine	AC	Closing paper insert lid	BC	Trays banging
D	Radio	AD	Paper bag rustling	BD	Radio 4
E	Customer's mobile phone	AE	Footsteps on floor	BE	Paper towel
F	Customer	AF	Trapped air in paper box	BF	Mixing chemicals while measuring temperature
G	Chair noises	AG	Light switch	BG	Ilford 2150 RC Processor fan
H	Door alert	AH	Bremson enlarger focusing switch	BH	Processor warming up
I	Telephone ringing	AI	Bremson fan	BI	Processor ready for next print
J	Telephone hands free dialling	AJ	Keypad buttons	BJ	Brochure for timings
K	Fax ringing	AK	Keypad confirmation of settings	BK	Tapping bottom of processing tank
L	Modem dialling	AL	Locking lens into position	BL	Splash of fluid
M	Keyboard tapping	AM	Racking enlarger up and down	BM	Throwing empty canisters into metal bucket
N	Till beeping	AN	Aperture selection	BN	Handling plastic/paper bags
O	Cash drawer	AO	On/off switch (Bremson)	BO	Air canister
P	Switch receipt	AP	Inserting film carrier	BP	Easel adjustments and opening/closing
Q	Conversation	AQ	Enlarger easel	BQ	Running water cleaning film
R	Traffic	AR	Air buster	BR	Enlarger on/off switches
S	Fan heater	AS	Staff knocking to warn approach	BS	Enlarger fan
T	Printer	AT	Revolving door	BT	Timer confirmation
U	Keys	AU	Water pressure gauge (omnipro)	BU	Timer countdown (seconds)
V	Stereo	AV	Print finished beep (omnipro)	BV	Click of light switch
W	Enlarger controller buttons (Buick)	AW	Paper handling	BW	Printing paper box opening/closing
X	Exposure transport x 4	AX	Door banging	BX	Water run off from tanks
Y	On/off switch (Buick)	AY	Nitrogen Generator	BY	Squeegee film
Z	Fan (Buick)	AZ	Trimming prints	BZ	Cleaning squeegee
				CA	Knocking excess water off reels

Figure 4-4: Visualisation of combined responses to the photographic processing lab

4.1.3 Discussion

Macaulay and Crerar's 1998 mapping tool proved easy to use for both the researcher and participants. The combination of categories covered every perceived sound event, and confirmed which attributes were consistently applied. Issues were identified with the application of the sound type. Listeners did not consistently apply *abstract* and *everyday*, the terms were used to indicate the familiarity, artificiality or identifiability of a sound event. It is proposed that *everyday* is changed to *other known* and *other unknown* replaces *abstract*, in order to indicate the identifiability of a sound event. It was also found that the acoustical information more accurately represented listeners' level of listening, rather than the complexity of a sound event, but this was applied consistently.

When considering how a soundscape could be classified and visualised, a number of omissions were discerned. The first omission was quantity. An auditory environment might have 20 inhabitants, but there was no measure of whether only one person was talking or everyone was talking. Information was also missing about how often these conversations took place, whether conversations were continuous or intermittent, or if conversations were concurrent or isolated sound events. The spatial location of a sound event was also omitted. In some cases sound events were equally spaced around the inhabitant, on other occasions sound events were clustered in a single location. The last two omissions were how loud and how high or low in terms of pitch sounds were.

It is only partially accurate to refer to the visualisations as maps. Whilst the acoustical information, and information categories are displayed using locational variables, neither relate to their spatial arrangement in the perceived soundscape. If a sound event had been classified as being both *visible* and *pattern*, two instances had to be included on the map, giving the impression that a greater number of sound events were present. This was not a problem when a sound event was considered to be *foreground* and *contextual* as the sound event was positioned on the line between the two rings. Unfortunately when a sound event was classified as *foreground*, *contextual* and *background* two instances had to be included.

Having briefly trialled Macaulay and Crerar's 1998 method a procedure for researchers to follow was identified, and a form of visualisation created. A modification to the method was proposed along with ways for combining listeners' responses.

4.2 Modified prototype soundscape mapping tool

In the previous study Macaulay and Crerar's 1998 prototype soundscape mapping tool was trialled and a modification proposed to the sound type. The preliminary study reported in this section expands the number of listeners, tests the modification, and extends the visualisation to include some spatial information about sound events. In order to investigate whether the modified prototype method was suitable for comparing listening experiences four types of office soundscape maps were generated from the 18 listeners' responses. Maps were created for each listener, and responses were combined to create maps that represented a single occupancy office, shared occupancy office and a typical office. This approach allowed a comparison of soundscapes according to the level of occupancy, as well as a typical experience.

4.2.1 Method

As identified in the previous study, Macaulay and Crerar's 1998 method can be split into identification, classification and visualisation. Participants were asked to verbally list each sound event that they could hear during a 15-minute period, excluding those made by the researcher. Participants were then provided with a copy of the modified classification (see Table 4-4) and asked to classify each sound event. In the previous trial the criteria *everyday* and *abstract* were not applied consistently by respondents, so the criteria were replaced, for this study, by the terms *other known* and *other unknown*. After classification all of the sound events were visualised by the researcher.

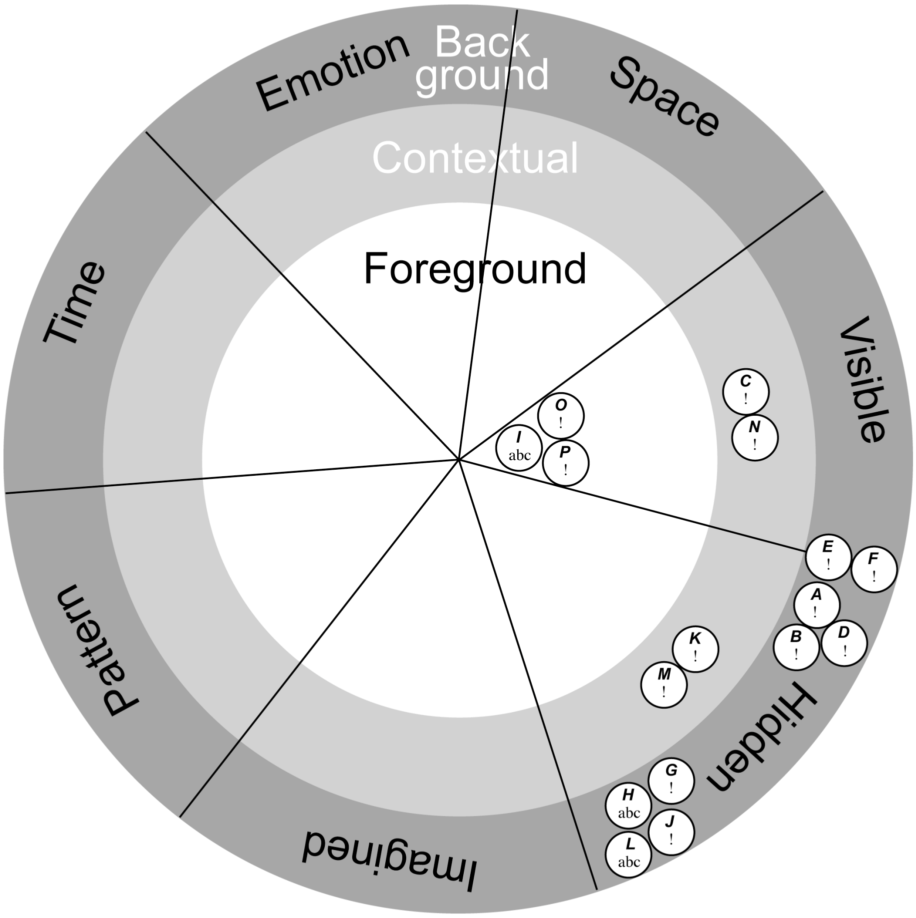
Sound Type	Example
Music	Any type of identifiable music, radio/stereo
Speech	Any form of speech
Other known	Identifiable sounds
Other unknown	Unidentified sounds
Information Category	**Example**
Visible entities and events	The phone ringing.
Hidden entities and events	The photocopier round the corner being used
Imagined entities and events	Something is happening in the children's room, it has gone very quiet.
Patterns of events/entities	Someone is batch copying a large document
The passing of time	I can hear children coming out of the school gate.
Emotions	The boss is unhappy I can hear his teeth grinding.
Position in Euclidean/acoustic space of entities/events and of the listener	The editor is at the foreign desk behind me (can hear his voice)
Level of Listening	**Example**
Foreground	Computer beep to attract your attention. (Sounds within the foreground of your experience of the soundscape).
Contextual	Door opening (Help you orient to the nature of your environment.)
Background	Whine of disk drive providing reassurance or information about the state of the world.

Table 4-4: Modified prototype soundscape mapping criteria

In the visualisation shown in Figure 4-5 all of the elements are identical to those from the previous study except for the sound type. Each concentric circle denotes the acoustical information, and the seven labelled segments refer to the information categories. The sound type is illustrated by labelling each circle with a symbol. *Music* is two notes ♬ , *other known* an exclamation mark **!**, *speech* a series of letters **abc** and *other unknown* by a question mark **?**. As in the previous study the sound events were cross-referenced with letters within each circle to prevent the image from becoming too cluttered.

Colour was not used for individual's maps; instead it was confined to maps with combined responses to aid differentiation. Individual hues represented the quantity of responses for each sound event. Different shapes were used to show whether the sound event was created by the participant (circle), or was an interior (square) or exterior sound event (polygon). In the previous study the spatial nature of a sound event was not specified, which meant that it was sometimes unclear whether a sound event was created by the listener, or was located within or outwith the listener's immediate auditory environment. Only a single shape (circle) was used for individual's maps (see Figure 4-5). As previously stated, this approach used three of

Bertin's (1983) retinal variables (shape, value/gray scale and hue/colour) in conjunction with his two locational variables (position).



abc = speech ? = other unknown ♬ = music ! = other known

A	Fans	**G**	Other People Typing	**M**	Fire alarm going off
B	Other computer noises	**H**	Speaking	**N**	Stapling
C	Curtains	**I**	Conversations	**O**	Buttons on/off (keys, light on)
D	Busses/Lorries	**J**	Walking in corridors	**P**	My Telephone
E	My Own Breathing	**K**	Doors		
F	Rustling of other people's paper	**L**	Speaking/Shouting outdoors		

Figure 4-5: Visualisation of shared office soundscape by P03

All 18 participants were Edinburgh Napier University employees, none of whom specialised in sound design. Participants were chosen for their varying degrees of room occupancy. Seven participants worked in single occupancy offices and 11 participants had offices where they shared with up to 11 colleagues. Offices were located on two different campuses and varied in size and configuration.

Verbal permission was sought from all of the inhabitants of the rooms under investigation. Participants were assured that their responses would be anonymised to prevent individuals from being identified. Participants were informed that they did not have to answer all of the questions and could stop at any time. Verbal permission was obtained to make use of the results for this dissertation and associated publication.

4.2.2 Results

A soundscape map was created for each of the 18 participants based on the visualisation detailed above. Responses were combined in order to create a single list of sound events that was used to derive three different sets of data. The first set was for the single occupancy responses, the second set specified multiple occupancy responses and the final set was for all responses in order to illustrate a typical office.

Table 4-5 shows a summary of the classifications applied during the mapping process. Within the sound type 85% of the responses were related to *other known*. Only eight of the participants referred to *speech*, which represented 11% of the overall events reported. There were 7 *other unknown* sound events out of 156 unique events. All of the *other unknown* sound events referred to exterior sound events where the sources were not visible. An estimate of the source was made in each case, such as the ‘suggestion of water outside, things passing through a puddle’. There were no instances of *music*.

With regards to the information category 62% of the events were *hidden* with only 23% being *visible*. This meant that 85% of the responses were classified according to their visibility. The *visible* responses predominantly applied to sound events that occurred inside the office, such as the ‘squeak of [the] seat as I lean back’. *Hidden* often pertained to sound events that occurred outside of the office, such as ‘somebody banged a door next door’. *Imagined* referred to unseen exterior auditory events, such as ‘vans loading and unloading outside’. *Patterns* included ‘speech’ as well as ‘activity on the pavement’. *Time* was represented through exterior sounds that reminded participants about the outside world, as in the case of 'low background noise of plants, trees and wind.' The sound of 'traffic outside the window' was used to subconsciously monitor the time of day, as the increased levels reminded one respondent that it was time to go home. *Emotions* were related to people such as 'xxxx clicking on keyboard'. *Position* was chosen when inhabitants were made aware of the spatial dimensions of their auditory environment, an example of which was a 'door closing in the distance'.

Participants	**Events**	**%**	**Rank**	**Respondents**	**%**	**Rank**
Single Occupants	60	38%	2	7	39%	2
Shared Occupants	96	62%	1	11	61%	1
Sound Type						
Music	0	0%		0	0%	
Speech	17	11%	2	8	44%	2
Other known	134	85%	1	18	100%	1
Other unknown	7	4%	3	4	22%	3
Information Category						
Visible	64	23%	2	16	89%	2
Hidden	175	62%	1	17	94%	1
Imagined	13	5%	3	8	44%	3
Patterns	8	3%	4	5	28%	4
Time	8	3%	4	5	28%	4
Emotions	9	3%	4	4	22%	6
Position	5	2%	7	2	11%	7
Acoustical Information						
Foreground	28	16%	3	12	67%	3
Contextual	35	20%	2	13	72%	2
Background	108	63%	1	18	100%	1
Multiple Classifications						
Double Classifications	34	22%		11	61%	
Triple Classifications	6	4%		3	17%	
Quadruple Classification	1	1%		1	6%	

Table 4-5: Summary of individual classifications using the modified prototype method

When classifying the acoustical information 63% of the sound events were *background*, 20% *contextual* and 16% *foreground*. Twelve of the 18 participants experienced *foreground* sound events, whereas all of the participants classified *background* information. *Foreground* sound events were typically the 'telephone' and 'conversations' that the listener took active part in. *Contextual* consisted mostly of 'people' and 'doors'. 'Computers' and 'traffic', along with the majority (63%) of the sound events, were classified as *background*.

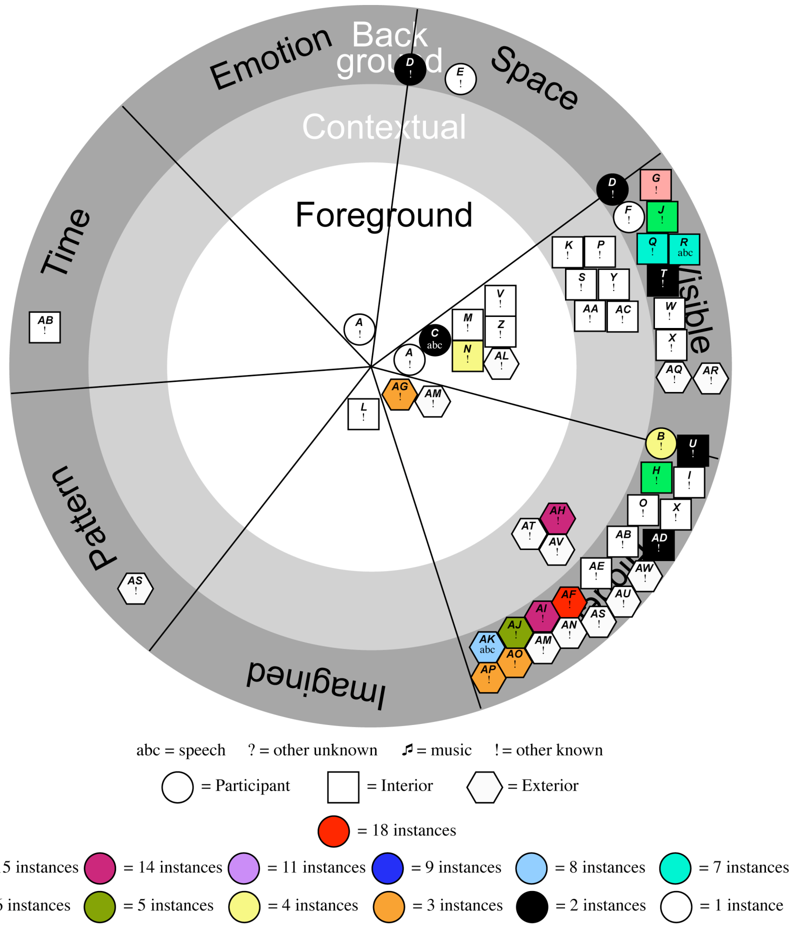
Twenty-seven percent of the listeners' responses were multiple classifications. Only two involved the sound type and were both *speech* and *other known,* where the sound of someone talking had been combined with the other sounds that person generated.

Of the 24 multiple classifications of information category all, except one, involved visibility (*hidden* or *visible*), in combination with another choice such as *emotion* or *passing of time*. The majority combinations of acoustical information double classifications were *foreground/background* or *contextual/background*.

When creating the soundscape maps the first stage was to establish whether the description of the sound event referred to a sound event generated by the participant, or the sound event was located within or outwith the office. The results were combined to take the number of sound events down from 156 to 49 (a reduction of 69%) (see Table 4-6). Record was kept of the number of participants who had contributed to each classification within each group, and the classification of each sound event was calculated individually for each group. The mode was calculated for each combined sound event, with equal values retaining both classifications. Figures 4-6 and 4-7 show the results from the three types of offices: combined, single occupancy and shared (multiple).

	Single Occupancy	Shared Occupancy	Combined	Single Occupancy	Shared Occupancy	Combined
Summary						
Participants	7	11	18	39%	61%	100%
Sound Events	28	37	49	57%	76%	100%
Double Classifications	7	5	8	25%	14%	16%
Location						
Participant	4	5	6	14%	14%	12%
Interior	10	22	25	36%	59%	51%
Exterior	14	10	18	50%	27%	37%
Sound Type						
Music	0	0	0	0%	0%	0%
Speech	2	3	3	7%	8%	6%
Other known	26	34	46	93%	92%	94%
Other unknown	0	0	0	0%	0%	0%
Information Category						
Visible	13	19	26	46%	51%	53%
Hidden	14	18	24	50%	49%	49%
Imagined	1	1	1	4%	3%	2%
Pattern	0	1	1	0%	3%	2%
Time	1	1	1	4%	3%	2%
Emotion	3	1	2	11%	3%	4%
Position	2	0	2	7%	0%	4%
Acoustical Information						
Foreground	8	7	11	29%	19%	22%
Contextual	3	9	9	11%	24%	18%
Background	19	22	30	68%	59%	61%

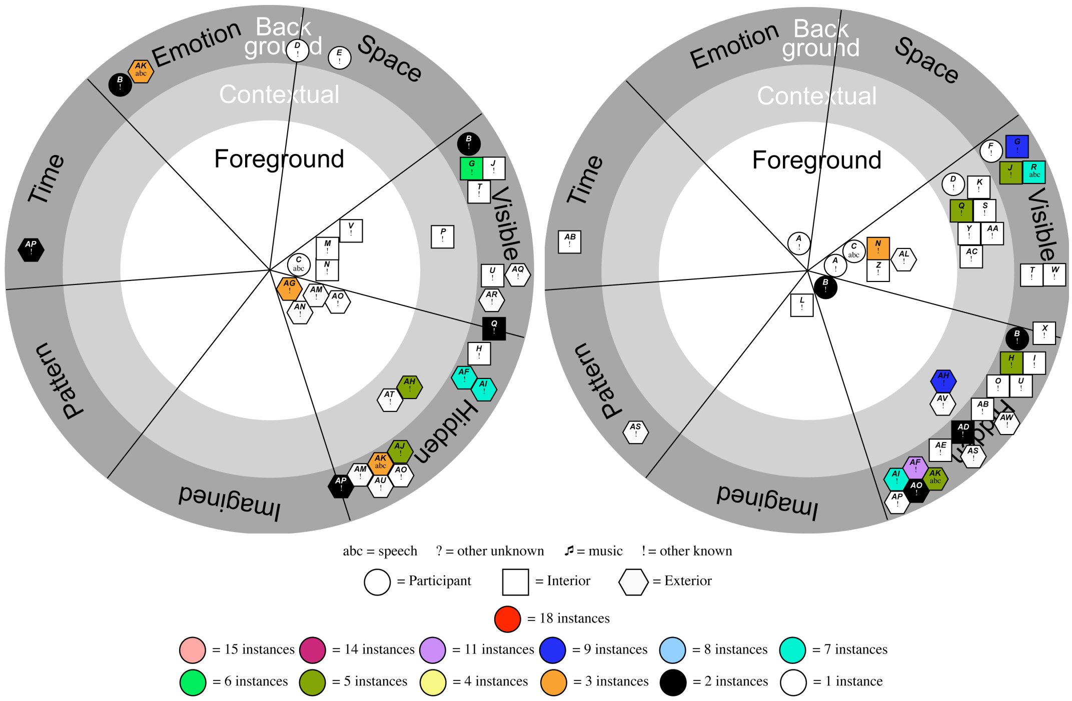
Table 4-6: Summary of sound classifications by type of environment



A	Participant Stomach rumbling	**R**	Inhabitants' Speech	**AH**	Doors opening and closing
B	Participant Breathing	**S**	Inhabitants' Sniffing	**AI**	Movement of People
C	Participant Talking	**T**	Clock Ticking	**AJ**	Indiscriminant Sounds
D	Participant's rustle of clothing	**U**	Air conditioning	**AK**	People Talking
E	Participant's Feet on carpet	**V**	Drawers opening	**AL**	Telephone Ringing
F	Participant's Tinnitus	**W**	Pen writing	**AM**	Photocopier
G	Computer fan	**X**	Paper Rustling	**AN**	Knocking on door
H	Computer Drives	**Y**	Stapling	**AO**	Birds
I	Computer Screen	**Z**	Lunch box opening	**AP**	Wind
J	Keyboard Typing	**AA**	Door opening & closing	**AQ**	Rain
K	Mouse Clicking	**AB**	Electrical Hum	**AR**	Window tapping
L	Computer Alert	**AC**	Curtains	**AS**	Drilling
M	Printer	**AD**	Radiator Water Flow	**AT**	Hand dryer
N	Telephone Ringing	**AE**	Fluorescent lights' hum	**AU**	Keyboard Typing
O	Telephone Click	**AF**	Traffic	**AV**	Fire alarm
P	Table Creaking	**AG**	Aeroplane Flying	**AW**	Central Heating System
Q	Chair movement				

Figure 4-6: Pictorial representation of typical office soundscape

The *single occupancy* offices included 57% of the total sound events, compared to 76% within the *shared* offices. The figures for the single occupancy office suggest a space where inhabitants have greater control over the auditory environment. In comparison, the results for the shared office recorded a greater number of sound events, possibly due to the increase in the number of inhabitants.



A	Participant Stomach rumbling	**R**	Inhabitants' Speech	**AH**	Doors opening and closing
B	Participant Breathing	**S**	Inhabitants' Sniffing	**AI**	Movement of People
C	Participant Talking	**T**	Clock Ticking	**AJ**	Indiscriminant Sounds
D	Participant's rustle of clothing	**U**	Air conditioning	**AK**	People Talking
E	Participant's Feet on carpet	**V**	Drawers opening	**AL**	Telephone Ringing
F	Participant's Tinnitus	**W**	Pen writing	**AM**	Photocopier
G	Computer fan	**X**	Paper Rustling	**AN**	Knocking on door
H	Computer Drives	**Y**	Stapling	**AO**	Birds
I	Computer Screen	**Z**	Lunch box opening	**AP**	Wind
J	Keyboard Typing	**AA**	Door opening & closing	**AQ**	Rain
K	Mouse Clicking	**AB**	Electrical Hum	**AR**	Window tapping
L	Computer Alert	**AC**	Curtains	**AS**	Drilling
M	Printer	**AD**	Radiator Water Flow	**AT**	Hand dryer
N	Telephone Ringing	**AE**	Fluorescent lights' hum	**AU**	Keyboard Typing
O	Telephone Click	**AF**	Traffic	**AV**	Fire alarm
P	Table Creaking	**AG**	Aeroplane Flying	**AW**	Central Heating System
Q	Chair movement				

Figure 4-7: Pictorial representation of single (left) and shared (right) occupancy office soundscapes

In the location category the participants identified a similar number of *participant* sound events, 14% in each case, suggesting that the level of occupancy might not have an impact on the participants' awareness of their own contribution to the soundscape. The level of occupancy did appear to have an effect on the relative percentage of *interior* and *exterior* sound events that listeners were aware of. There were a greater percentage of *interior* events in the shared offices (59%) compared to the single occupancy offices (36%), as well as a lesser amount of *exterior* sounds (27%), than reported by the single occupancy participants (50%).

The percentage of sound type was almost identical for each type of office. The information categories were also similar, with slightly fewer *emotions* and no *positions* in the shared office, and slightly greater *visible* events, 51% compared to 46%. The main differences can be seen in the acoustical information; the single occupancy offices had a greater percentage of *foreground* and *background* sound events in comparison with the shared offices, which had a greater percentage of *contextual* sounds.

4.2.3 Discussion

The modified prototype soundscape mapping tool again proved easy to use, with the combination of categories covering every perceived sound event. The increased number of listeners had no appreciable effect upon the ease of use. The tool showed the relative percentages of type, category and acoustical content. As with the previous trial, the acoustical information more closely represented Amphoux's 1997 levels of listening rather than the richness of the information being gathered.

The modification to the sound type improved the consistency of listeners' sound type classifications. In the previous study three different interpretations were applied to *abstract* and *everyday*. In contrast, *other known* and *other unknown* were used only when a sound event was neither *speech* nor *music*. *Other known* was chosen when the sound event was clearly identifiable such as a 'computer hard drive going tick, tick'. *Other unknown* was selected when listeners were unable to identify the sound event as in the case of 'unidentifiable low sounds out in [the] corridor'. It was proposed to retain this modification for the extended version of the soundscape mapping tool. The soundscape maps for the single and multiple occupancy offices allowed the comparison of the soundscapes for two different environments. However different listening experiences of the same auditory environment were not compared. A method of comparing listeners' experiences of the same environment will be tested in the extended prototype soundscape mapping tool.

Identifying whether the sound event was created by the listener, or was interior or exterior to the auditory environment, provided limited spatial information without cluttering the visualisation. However, additional information such as that provided by Cartesian coordinates is required if the visualisations are to be considered as cartographic maps. It was proposed to include Cartesian coordinates of sound events as part of the extended prototype soundscape mapping tool.

Omissions were again evident within the study, these included: dynamics, quantity, spectral and temporal attributes. In order to identify appropriate methods for extending the prototype method a series of interviews were conducted with listeners, as reported in the following section.

4.3 Listener interviews

This section reports a preliminary study that investigates what terms listeners use to describe sounds. Attributes of sound were derived from these listeners' responses in order to extend the prototype soundscape mapping tool. Schubert (1975) suggested that the most important aspect of listening is the identification of source and action. Gaver (1993) termed the identification of source and action as everyday listening, and the representation of a sound's attributes as musical listening. Blauert and Jekosch (1997) highlighted the problem of listeners routinely referring to fewer than four attributes when listening musically. Guastavino and Dubois (2006) found that musical listening was only possible when the sources and actions of sound events could not be easily identified. Handel (1989) proposed that three out of the four approaches which listeners use to identify sounds rely on previous experience. Rodaway (1994) agreed with Handel's thesis, arguing that in order to understand a sound or a soundscape a listener must compare it to their previous experiences.

Asking about memories of sounds provided a wider range of experiences, compared to eliciting descriptions about isolated sounds or a single auditory environment. The approach reported in this section was similar to a study conducted by Raimbault and

Dubois in 2005. Raimbault and Dubois used a questionnaire during one-to-one interviews to elicit verbal descriptions of soundscapes from town planners and city-users. Responses were coded, and it was found that only the planners used technical vocabulary, both groups provided generic expressions, with city-users providing more comparisons and descriptions of human activity.

4.3.1 Method

The purpose of this study was to generate a lexicon that could be classified to identify appropriate attributes to extend the prototype soundscape mapping tool. Face to face interviews were chosen instead of a survey to achieve a higher response rate, to reduce the number of 'no answers' and to try and prevent question confusion (Babbie, 1990). Interviewees given a chance to elaborate may include more insightful comments that might not have been captured in a survey (Lazar, Feng & Hochheiser, 2010). Twenty questions were created, 7 of which addressed auditory displays, 7 related to sound events, and 6 referred to auditory environments (see Table 4-7). Questions were informally trialled with lecturers and research students to establish the questions' suitability. Rather than establish a list of the different types of auditory displays, sound events or auditory environments participants had been exposed to, the intention was to discover what terms interviewees used when describing past experiences of sounds and auditory environments.

The same 18 participants who took part in the study reported in the previous section were interviewed individually in their normal office environment. The structured interviews took an average of 30 minutes each. Interviews were recorded, transcribed, and then coded using ATLAS.ti software (ATLAS.ti, 2008). Interviews started with questions about office equipment such as telephones, computers and any other auditory interfaces the interviewee might have experienced. The interview went on to query the impact of sounds that the attracted the interviewee's attention, relaxing, stressful and informative. Questioning finished by discussing the office's auditory environment in general and asking participants about what they would like to change or control.

Auditory Displays	
1	Have you ever used auditory interfaces?
2	Is the sound currently switched on, on your computer?
3	Have you altered the settings on your phone?
4	Do you enjoy using headphones?
5	If auditory interfaces were discrete and private, i.e. you had your own personal space and the noises created were completely inaudible to the rest of the environment and you didn't have to wear headphones of any format, would you use more?
14	Do you use auditory feedback in your daily life?
15	What would you like auditory feedback of?
Sound Events	
7	What type of sounds attract your attention?
8	What type of sounds do you find relaxing?
9	What type of sounds do you find stressful?
10	What type of sounds communicate information to you?
11	Can you describe a couple of sounds that are rich in communication, outside of speech and music?
12	How would you describe the sounds your computer makes?
20	What terms are you aware of for measuring or describing audio?
Auditory Environments	
6	Do you prefer to work in silence or with a musical or some other background noise?
13	How would you describe your work environment from an auditory point of view?
16	Could you give an example of an auditory environment, where you think the auditory elements are satisfactory?
17	Could you give an example of an auditory environment, where you think the auditory elements are unsatisfactory?
18	If you could control your auditory environment how would you alter it?
19	Could you describe an auditory experience from the past, excluding a concert?

Table 4-7: Questions posed to office inhabitants

Verbal permission was sought from all of the inhabitants of the rooms under study. Participants were assured that their responses would be anonymised to prevent individuals from being identified. Participants were informed that they did not have to answer all of the questions and could stop at any time. Verbal permission was obtained to make use of the results for this dissertation and associated publication.

Coding was conducted using the transcriptions from the interviews. Codes were assigned to quotations using an open approach where codes were suggested by the quotations, rather establishing a pre-defined set prior to coding. Once the first pass was completed and the codes were set, a second pass was made in order to ensure that each interview was referenced using the complete set of codes. A lexicon of terms used by participants to describe sounds was derived from the transcriptions using a spreadsheet. The lexemes were then classified, according to the codes derived from the open coding of the transcriptions, in order to establish a list of attributes suitable for extending the prototype soundscape mapping tool. Classifying the lexemes as well as coding the transcriptions helped ensure that descriptors for attributes used terms that were part of the listener's domain. Using the codes as a starting point for the identification of attributes for describing sound helped to reduce the total number of attributes, as well as providing confirmation for the original identification and interpretation of the codes.

4.3.2 Results

The results can be considered in three parts: codes, lexicon and attributes. Codes represent the categories assigned to quotations derived from the interview transcriptions and the codes have been underlined in the text below. The lexicon represents all of the lexemes or terms that were used by listeners to describe sounds and the lexemes are in bold. The attributes represent the attributes of sound that were derived by classifying the lexicon and the attributes have been italicized.

The 40 codes derived from the transcriptions can be divided into three groups (see Table 4-8): 100% of respondents, majority of respondents and minority of respondents. Within the 100% of respondents group source was predominant. Source represented any identifiable source that the interviewee referred to when describing a sound. Sounds discussed in response to the questions were not confined to those inside the office environment, participants often referred to external sources such as 'seagulls' and 'traffic'. Typical sources were described in generic terms such as a 'computer', 'telephone' and 'people'. Specific sources were only applied to named individuals, rather than objects, even when discussing the shared environment. Two references were made to the material of sound sources as being 'metallic' or 'wood'.

The code type was applied when referring to an abstract concept without identifying a specific source such as 'music', 'noise' or 'speech'. Action included all physical actions that generated a sound such as 'pouring', 'footsteps' and 'blowing'. Force was only mentioned five times, and might be considered as a subset of action.

Dynamics were detailed in terms of 'silent', 'quiet' or 'loud'. Alternative terms for dynamics included 'background' when referring to low levels of listening rather than spatial aspects, and 'noisy' when the sound was considered excessive without being directly related to pollution. Onomatopoeia included words such as 'clanking', 'click' and 'whine'. Informative referred to sounds which communicated a single state, or sequence of information as in 'signals', 'alarms' and 'cue'. Evocation was used to classify when a sound induced a 'mood' or 'emotion'.

Code	Respondents	%	Instances	%	Rank
Source	18	100%	509	25.68%	1
Type	18	100%	258	13.02%	2
Action	18	100%	205	10.34%	3
Dynamics	18	100%	198	9.99%	4
Onomatopoeia	18	100%	170	8.58%	5
Informative	18	100%	77	3.88%	6
Pollution	17	94%	66	3.33%	7
Spatial	17	94%	60	3.03%	8
Relaxing	16	89%	19	0.96%	9
Arresting	16	89%	18	0.91%	10
Temporal	15	83%	76	3.83%	11
Spectral	15	83%	64	3.23%	12
Natural	14	78%	37	1.87%	13
Environment	14	78%	19	0.96%	14
Stressful	14	78%	17	0.86%	15
Aesthetics	10	56%	31	1.56%	16
Emotions	10	56%	19	0.96%	17
Preference	9	50%	17	0.86%	18
Artificial	8	44%	15	0.76%	19
Content	7	39%	13	0.66%	20
Masking	7	39%	9	0.45%	21
Context	6	33%	20	1.01%	22
Quality	6	33%	8	0.40%	23
Quantity	5	28%	10	0.50%	24
Recipient	4	22%	7	0.35%	25
Clarity	4	22%	5	0.25%	26
Force	4	22%	5	0.25%	26
Pleasure	4	22%	5	0.25%	26
Room acoustics	4	22%	4	0.20%	29
Interest	3	17%	3	0.15%	30
Motivate	3	17%	3	0.15%	31
Evocation	2	11%	4	0.20%	31
Familiarity	2	11%	2	0.10%	33
Material	2	11%	2	0.10%	33
Mechanistic	2	11%	2	0.10%	33
Complexity	1	6%	1	0.05%	36
Dispersion	1	6%	1	0.05%	36
Effect	1	6%	1	0.05%	36
Gender	1	6%	1	0.05%	36
Privacy	1	6%	1	0.05%	36

Table 4-8: Codes resulting from transcribed terms

The majority of interviewees referred to pollution, both to the pollution created by others, as well as the impact that the interviewees had themselves on the auditory environment. References were made to personal responses to what participants considered being polluting sounds, from ‘irritating’ through to ‘annoying’ and finally ‘hate’. Spatial was always considered in relation to the interviewee’s position, such as ‘behind me’, ‘outside my office’ or the vague ‘out there’.

When relaxing sound events were described, terms included ‘relaxing’, ‘soothing’ and ‘peaceful’. This contrasted with stressful events that were only referred to with the single descriptor ‘stressful’. Motivate applied to stimulation, but was only applied to music. *Arresting* covered ‘urgency’ and ‘arousal’ as well as ‘arresting’. *Temporal* and *spectral,* like dynamic, were referred to in binary terms. Temporal descriptors included ‘consistent’ or ‘occasional’, with specific references to times of the day. *Spectral* attributes were referred to as ‘higher’ or ‘lower’ along with the generic, ‘tone’, ‘pitch’ or ‘frequency’.

Natural sounds were mentioned more often than artificial or mechanistic. Generally the natural sounds were regarded more favourably than the recorded or machine generated sounds. Aesthetics fell into positive or negative terms rather than passive, with slightly more instances of negative descriptions such as ‘offensive’, ‘piercing’ and ‘discordant’ compared to positive: ‘lovely’, ‘daintily’ and ‘pleasant’. Emotions were also expressed with polar responses, based around positive or negative emotions such as ‘happy’, ‘aggression’ or ‘distress’. Environment specified an identifiable location as the sound source rather than the more generic spatial. Environments included cities, buildings and rooms as well as outdoor locations such as ‘rivers’ and ‘gardens’. Room acoustics, whilst being rarely mentioned, referred to whether the room affected the sound positively or the room had poor ‘insulation’ that was related to pollution. Preference was indicated through simple terms such as ‘like’ or ‘dislike’, with the more specific pleasure related in terms of ‘pleasing’, or 'amusing'. Interest referred to whether the sound was 'boring' or had any relevance, without indicating pollution.

The codes contributed by the minority of the respondents are possibly more interesting to sound designers, as they represent responses generally more difficult to elicit from listeners. Content was applied to verbatim quotes of conversations this differed from *context*, in that the latter provided information about the context in which the listener interpreted the sound, rather than merely reporting it for example 'people talking if they were waiting to go into the lab'. Whereas recipient specifically related to whom the sound event was intended for.

Masking referred to sounds which were either generated by the participant in order 'to kill off other things' or sounds which listeners became 'attuned to' thereby masking themselves. Familiarity was expressed in terms of 'being used to it' and 'surprising'. Quality exclusively applied to the source producing the sound in terms of 'low', whereas clarity was related to the sounds themselves, again in negative terms being 'confused' or 'chaotic'. Quantity related to either one to three or 'lots' with no values in between.

The remaining codes only had single instances. Complexity, in this case 'simple', could be considered part of *aesthetics*. Dispersion was related in technical terms as 'unidirectional', and was applied to speech. Effect referred to a sound being 'used to speed up the heart rate'. Gender was applied to a 'girl screaming', but generally people were referred to in generic terms without specifying their gender. Finally, privacy could be related to recipient, in that the content was 'private' and not intended for the listener.

A lexicon was generated directly from the transcriptions by listing every word that was used to describe a sound. The number of instances for each participant was also noted and the list was ranked according to frequency. Singular and plural descriptions such as 'noise' and 'noises' were kept separate. The five most commonly used terms were ‘music’, which was the only word in the lexicon to be used by all of the participants, ‘people’, ‘noise’, ‘background’, and ‘telephone’. Only the first three terms were referred to by more than 89% of the respondents, all of the remaining 1238 words were used by less than 79% of the participants, and almost 57% of the total words were used only once, examples include ‘echoes’, ‘irritant’ and ‘sneeze’.

Using the codes derived from transcribed phrases in context, individual lexemes from the lexicon were classified in order to provide an insight into the attributes that participants used when describing sound. Table 4-9 shows that this method generated eight attributes with a 100% response, with a further eight above 50%, with only two below 50%. The 18 attributes derived from the classification of the lexicon can be grouped into *description*, *physical characteristics* and *perceptual characteristics*. Descriptions include *source*, *action*, *onomatopoeia* and *content*. Physical characteristics include *dynamics* (volume), *spectral* (pitch), *spatial* (location/movement), *temporal* (time), *quantity* (complexity) and *room acoustics*. Perceptual characteristics comprise: *effect* (impact), *emotions* (positive/negative), *interest* (relevance), *aesthetics* (pleasing/displeasing), *clarity* (distinct/indistinct) *pollution* (noise), *preference* and *privacy*.

Code	Sum	Frequency	Percent	Mean	Median	Mode	Rank
Source	949	18	100%	53	49.5	84	1
Content	554	18	100%	31	32.5	21	2
Action	331	18	100%	18	17	26	3
Dynamics	296	18	100%	16	16.5	8	4
Spatial	286	18	100%	16	13	6	5
Onomatopoeia	283	18	100%	16	14	14	6
Aesthetics	213	18	100%	12	11.5	1	7
Pollution	135	18	100%	8	6.5	4	8
Temporal	167	17	94%	10	8	4	9
Effect	88	17	94%	5	5	1	10
Spectral	135	15	83%	9	8	11	11
Emotions	51	14	78%	4	3	5	12
Quantity	46	13	72%	4	3	1	13
Clarity	19	12	67%	2	1	1	14
Preference	39	10	56%	4	3	1	15
Interest	37	10	56%	4	4	5	16
Privacy	5	3	17%	2	1	1	17
Room Acoustics	1	1	6%	1	1		18

Table 4-9: Attributes derived from lexicon classification

The results from the lexicon classification could then be used to extend the modified prototype soundscape mapping tool. In terms of descriptions the *source* and the *action* of a sound event are commonly present in the accounts provided by listeners when identifying what they experienced. The identification of sound events could be made more explicit by specifying the *source* and the *action*. *Content* is similar to sound type, especially in terms of *speech* and *music*, with all of the non-speech and non-music sound events being classified as either *other known* or *other unknown*. Formalising *onomatopoeia* into categories is difficult, a comparison of published lists, such as Peterson and Gross (1972), Cano *et al.* (2004), and Sundaram and Narayanan (2008) show a low level of positive correlation. Gaver's 1993 classification of interacting materials may provide an alternative approach to onomatopoeia through classifying a sound event's audible source attributes.

The physical attributes of sound events are partially represented in the modified prototype method. *Dynamics* could be inferred from the acoustical information with *foreground* sound events perhaps being perceived as louder than *background*, but the relationship is not explicit. Measuring the sound pressure level in decibels might be a more appropriate approach. *Spectral* attributes are not part of the prototype tool and could be measured in hertz. *Spatial* attributes are included in the modified tool, but as a single option within the information category. Carlile (2002) considers that direction and distance are necessary to convey spatial perception. The *spatial* attributes could be captured with Cartesian coordinates. The prototype method records whether a sound event has *temporal* significance, but it does not record when a sound event starts, finishes or its duration. Making note of the start and end times for sound events would allow both duration and the sequence of sound events to be captured. *Patterns* in the prototype tool relate to a somewhat limited extent *quantity*, implying more than one sound source within the grouping, but an accurate number is not captured. Listing each sound event separately may provide an accurate indication of *quantity*.

The perceptual characteristics of sound events are also addressed to a limited extent in the prototype soundscape mapping tool. *Emotions* are included but there is no indication about whether they are positive or negative. *Pollution* is not considered, nor is the *effect* of a sound event. A qualitative understanding of a sound event's information content could provide an insight about what *effect* a sound event had upon a listener, such as defined by Delage (1998). Delage's interactive functions could be extended to include noise, in order to convey *pollution*, as it can be argued that noise is unwanted sound. Noise might also indicate *aesthetics* in that the sound event is displeasing, whereas a pleasing sound event could be considered relaxing or inciting. Extending the interactive functions further to include neutral may also indicate *preference*. Desirable sound events could be specified according to their information content such as *assisting*, *warning* or *guiding*, undesirable sound events might be classified as either *neutral* or *noise.*

Clarity is obliquely related to both acoustical information and sound type, an unclear sound event might be *other unknown* and *background* whereas a clear sound event may be *other known* and *foreground*. *Interest* is more closely related to acoustical information; *foreground* sound events are those that the listener attends to, whereas *background* often represents habituated sound events. *Privacy* and *room acoustics* were only referred to occasionally by the interviewees, and are not represented in the prototype soundscape mapping tool.

4.3.3 Discussion

The interviews revealed how seldom descriptions of sounds go beyond object-orientated identifications, bearing out Ballas and Howard's (1987) experiences. An indication from this series of interviews is the reliance on the source when describing sound, as Metz (1985) states, when individuals are describing sounds they are picturing the sound's source. Of the 18 attributes of sound that were established in this study, 4 were currently present as part of the prototype tool (*source*, *content*, *action* and *interest*). Six attributes could be inferred from other attributes within the prototype tool, but were not explicit: *dynamics*, *spatial*, *temporal*, *emotions*, *quantity*, and *clarity*. Eight attributes were omitted from the prototype tool: *onomatopoeia*, *aesthetics*, *pollution*, *effect*, *spectral* and *preference*. *Privacy* and *room acoustics* were not part of the prototype method, but had insufficient responses to be considered. It was proposed to extend the prototype tool by measuring the quantitative attributes, as well as adding two forms of qualitative classification. The *temporal*, *spatial*, *dynamics* and *spectral* attributes could be measured. Gaver's 1993 interacting materials might provide a suitable alternative to onomatopoeia. Delage's 1998 interactive functions, with the addition of neutral and noise, could convey a listener's experience of *pollution*, *effect* and *preference.*

4.4 Extended prototype soundscape mapping tool

As a result of the interviews, it was possible to identify attributes for describing sound that were missing from the prototype soundscape mapping tool. This section details how the extended tool was used to model listeners' experiences of a shared office environment. The physical attributes of sound events (*dynamics*, *spectral*, *spatial* and *temporal*) were omitted from the prototype and can be measured. The perceptual attributes *aesthetics*, *effect*, *pollution* and *preference* were also omitted and might be conveyed by extending Delage's 1998 interacting materials to include neutral and noise. *Onomatopoeia* was also missing and may be communicated by Gaver's 1993 classification of interacting materials. All of the attributes that were already within the prototype tool were identified within the listeners' descriptions and were retained.

The extended prototype soundscape mapping tool involved four stages: capture, measurement, classification and visualisation. Surround sound recording apparatus was used to capture the auditory environment, and measurements of sound events were conducted using established techniques. Listeners classified the sound events using published but untried methods. Visualisations were created using cartographic principles derived from the work of Bertin (1983).

Listening is essentially, but not exclusively, a spatial experience, listeners utilise spatial cues in order to help identify discrete sound sources in complex auditory environments (Bregman, 1990). Mapping was chosen as the form of visualisation as there is an established tradition of mapping soundscapes commencing with Granö in 1929 through to Schiewe and Kornfeld (2009). Monmonier (1993) believes that cartography can be a visually effective method for gathering, collating and interpreting qualitative data.

4.4.1 Method

The soundscape mapping tool consists of four stages: capture, measurement, classification and visualisation as shown in Figure 4-8. Capture involved creating a floor plan and recording the auditory environment. The recording allowed intermittent and new inhabitants to experience the same auditory environment, as well as the notation and measurement of sound events. Classification elicited details about listeners' perceived soundscapes. Visualisation provided a method of interpreting the responses from individual listeners as well as groups of listeners.

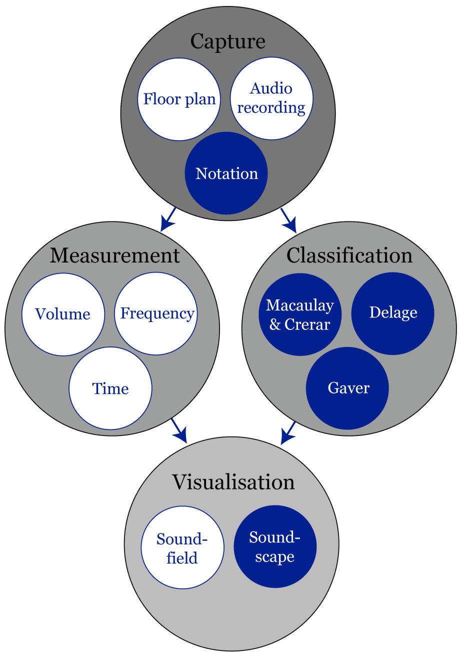


Figure 4-8: Extended prototype soundscape mapping tool

Capture

The capture stage involved the creation of a floor plan and a surround sound recording. Measurements were made of the room and all fixed objects, such as desks, filing cabinets, windows and doors, and were converted onto a floor plan with a scale of 100:1. A grid of cells, each representing 50cm by 50cm was overlaid onto the floor plan with additional cells added around the perimeter to allow the notation of sound events that originated from outside of the office. The grid was numbered in the same manner as an Ordnance Survey map starting from the bottom left with 0 0. Thus the room being modelled occupies 1 1 to 20 17 (see Figure 4-9).

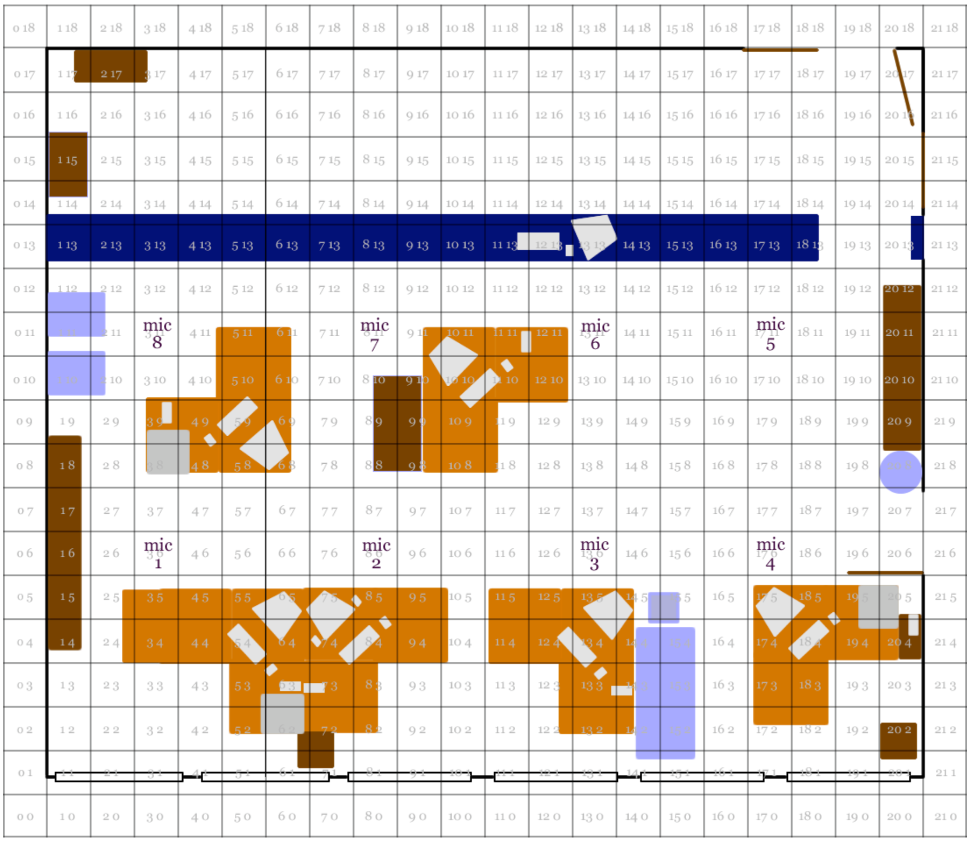


Figure 4-9: Floor plan of administrative office with fixed objects (bookcases, computers, desks, filing cabinets, water dispenser, worktop) and microphone positions (mic 1-8)

An eight-channel surround sound recording was made to create as accurate a reproduction as possible. Eight omnidirectional microphones were mounted in windshields on stands at 1.5m in height, and were positioned as shown in Figure 4-8 (mics 1-8). The windshields enclosed the microphones suspending them from the support enabling shock mounting that reduced the effect of unwanted vibrations. The recording was made in a 30-minute pass. A separate eight-channel pre-amplifier was used to minimize distortion and ensure consistency in both dynamics and frequency. Each channel was recorded uncompressed at 96kHz and 24 bits, providing a theoretical dynamic range of 144 dB, ensuring that the full audible range was covered. A sampling rate of 96 kHz was chosen so that the short time delays, with an accuracy of 15 microseconds, that listeners rely on in order to accurately locate sounds, could be reproduced (Al'tman *et al.*, 2004). Microphones were erected 4 hours prior to recording to allow inhabitants to become familiar with them, as well as to identify if physical obstructions were caused and to check recording levels. No announcement was made when the recording started or finished.

All departmental members of staff who might enter the room were contacted in advance, either personally or via e-mail to seek permission for the recording. Only three students were present during the 30 minute period, all of whom provided verbal permission to use the recording for research purposes immediately after they left the environment. The approach of obtaining permission after the recording was chosen as potentially over a 1000 students might choose to enter the room making it difficult to obtain universal permission. Notes about sound events were taken by the author during the recording to aid later identification. Individual conversations were not transcribed, and were only listened to by the author and the 18 participants.

Measurement

For the reproduction four sub-bass units supplemented eight compact monitors, which allowed for more accurate positioning of sound events than a 5.1 system. The sub-bass units compensated for the reduced frequency transmission range associated with compact monitors, as well as helped to alleviate problems with room modes during reproduction (Toole, 2006). Room modes can artificially add nodes (phase cancellation) and antinodes (phase reinforcement) to the reproduction that were not present during the recording (Jones, 2002).

All of the audible individual sound events were noted by listening back to the recording using the surround sound system. Note was made of Event, Source, Start Time, End Time and Location. Events represented the actions that generated the sound such as typing. Sources were identified according to the person or object generating the sound. Start and end times of sound events were noted in hours, minutes and seconds, and rounded up to the nearest second. The spatial location of sound events was noted as a grid reference, with any variation also noted.

After identifying all of the individual sound events recordings and sound pressure level (SPL) readings were taken of the original sound sources. The quantifiable attributes of a sound event are independent of inhabitants' perceptions, but each instance is still unique. Measurements are only representative of a single instance of a sound event, as the complex interaction of materials and other sound events will affect each occurrence. In order to isolate the sound event from the acoustic background recordings were made in mono, using the built in microphone on an SPL meter. Careful attention was paid to proximity to reduce the effect of colouration from the microphone being too close or too far away from the sound source.

In order to be able to calculate an approximate sound pressure level, A scale peak readings were taken for each sound event. The SPL meter was mounted on a tripod and the distance was altered until the peak was at least 6dB above the auditory background. A difference of 6dB meant that the sound event was double the SPL of the auditory background. Knowing the distance from the source allowed the SPL level at one metre to be calculated using the inverse square law. The formula:

$$dB\ SPL = Max\ SPL + (20 \times Log\ (distance1/distance2))$$

was used with the awareness that reverberation often amplifies a sound after a certain distance, which varies according to frequency and location (Foreman, 2002).

The frequency range of sound events was calculated by passing the recording of each sound event through a spectrogram within Metric Halo's *SpectraFoo* software (Metric Halo, 2008). The lowest and highest frequencies within the signal that was 6dB above the auditory background were noted. Measurements were collated into: *Event, Source, Time (Start & Stop), Location, SPL* and *Frequency Range* (see Table 4-10). Sound events were then grouped together into a candidate sound event list of *Event, Source, Time Period and Location*, suitable for questioning inhabitants.

Event	Source	Start Time	End Time	Location	dB A	Hz
Radio playing	Radio	00:00:00	00:30:00	13 3	40	100 Hz - 7 kHz
Traffic	Vehicles	00:00:00	00:30:00	0 0 - 21 0	66	20 Hz - 5 kHz
Typing	P05	00:00:00	00:00:05	19 3	54	800 Hz - 10 kHz
Typing	P03	00:00:00	00:00:10	11 9	56	800 Hz - 10 kHz
Mouse clicks	P01	00:00:04	00:00:06	12 3	40	3.5 - 16.2 kHz

Table 4-10: Example sound event measurements

Classification

Eighteen participants took part in the study: 6 regular inhabitants, 6 intermittent and 6 new inhabitants. The six regular inhabitants had experience of working in the office for at least 1 year, so were familiar with the environment. The six intermittent inhabitants were recruited from teaching staff within the department who responded to e-mail requests. All of the intermittent inhabitants had more than 1 year of experience of visiting the administrative office, typically daily. The six new inhabitants were also university employees, and were again recruited via e-mail requests. None of the new inhabitants had previously visited the office or met any of the inhabitants, so were unable to identify individuals within the recording. The new inhabitants simulated the effect of someone entering the auditory environment for the first time and trying to make sense of what they were listening to.

Verbal permission was sought from all of the participants. Participants were assured that their responses would be anonymised to prevent individuals from being identified. Participants were informed that they did not have to answer all of the questions and could stop at any time. Verbal permission was obtained to make use of the results for this dissertation and associated publication.

Regular inhabitants did not listen to the surround sound recording, as all of the interviews could be conducted in situ ensuring the correct auditory perspective. In order to provide intermittent and new inhabitants with a similar experience listeners were exposed to the surround sound recording of the office. After the intermittent and new inhabitants had listened individually to the recording they were asked to classify each sound event according to the criteria in Table 4-11.

Sound Type	**Example**
Music	Any type of identifiable music, radio/stereo
Speech	Conversation
Other known	Identifiable recognised sounds
Other unknown	Unidentifiable unrecognised sounds
Information Category	**Example**
Visible entities and events	The phone ringing
Hidden entities and events	The photocopier round the corner being used
Imagined entities and events	Something big is happening as it has gone quiet
Patterns of events/entities	Someone is batch copying a large document
The passing of time	It's nearly deadline time (because the shift change is happening)
Emotions	A person is unhappy (tapping or slamming)
Euclidean Position	Person is moving around you
Acoustical information	**Example**
Foreground	Computer beep to attract your attention.
Contextual	Door opening (Help you orient to the nature of your environment.)
Background	Whine of disk drive providing reassurance or information about the state of the world.
Interactive functions	**Example**
Warning	Be careful
Assisting	Don't forget
Incitement	I am ready you can use me
Monitoring	In hospitals, in industry
Reassurance	You did OK
Forgiving	Try it again you'll succeed in the end
Guiding	Pedestrians at a crossroad
Protecting	Your car or house
Relaxing	So that you perform better
Neutral	No relevant information
Noise	Unwanted
Interacting materials	**Example**
Impact	Door is slammed or object is dropped
Scraping	Pen writing, Paper Rustling
Other vibration	(specify)
Explosion	Car backfiring
Continuous aerodynamic	Wind, voice
Other aerodynamic sound	(specify)
Dripping	Water from tap
Splashing	Washing up
Other liquid sound	(specify)

Table 4-11: Sound event classification for extended prototype soundscape mapping tool

Three forms of classification were applied: a modified version of Macaulay/Crerar's sound types, information categories and acoustical information (1998), an extended version of Delage's (1998) interactive functions and Gaver's (1993) interacting materials. The interactive functions and interacting material classifications were added, as they provided additional information about what people were listening to that was missing from the prototype tool. Delage's interactive functions were extended through the addition of neutral and noise categories. *Neutral* denoted that the sound event had no information content, whereas *noise* indicated that the sound was unwanted, and might be considered a pollutant. Gaver's interacting materials was extended to include other aerodynamic, other vibration and other liquid in order to identify if any additional categories were required.

Visualisation

All of the attributes were displayed on a single map in the manner of a geographic information system (GIS). Maps were created for each listener. Individual's responses were combined for all of the regular inhabitants, as well as the intermittent, new and finally combined (all listeners), resulting in four maps. Four different groups of listeners were identified in order to test the tool's suitability for combining results and comparing listening experiences. When combining the results the modal response was employed. As in the previous studies reported in this chapter, when no single value could be established multiple values were included. All of the maps were generated as vector graphics within Adobe's *Illustrator* software (Adobe, 2009b).

As with the prototype procedure, Bertin's 1967 theory of cartographic communication was used to create the visualisation (Bertin, 1983). His eight visual variables were split into six retinal: shape, size, value, orientation, hue and texture, and two locational: x and y coordinates. Bertin proposed that the visual variables be applied to point, line and area symbols, but he did not include typography. Monmonier (1993) suggested that Bertin's variables were also suitable for type, which could also act as symbols, but he warned that hue and texture would interfere with legibility, and therefore should not be used. Monmonier stated that size would directly translate to point size, value to style, shape to typeface, and orientation to the angle of the text. An additional attribute was required to be displayed, than was possible using both Bertin's and Monmonier's approaches. A symbol border was used that included the retinal attributes of size, value, shape and hue.

Each sound event was given two capital letters cross-referenced to descriptions. Sound pressure level (SPL) was represented by size, the louder the sound the greater the diameter that was directly proportional to the SPL (see Figure 4-10). Servigne, Kang and Laurini (2000) suggested that varying the radius of circles would be appropriate to display different intensities of noise. D. Gibson (2005) also illustrated a sound's volume within a mix using size. D. Gibson argued that allowing sound events to visually overlap mimicked the effect of masking when a higher SPL sound occurred at the same time and spatial location as a lower SPL sound.

Frequency was represented through colour, with the frequency directly mapped to the visible spectrum. 20 Hz was blue and 20 kHz red, with all of the intermediate frequencies spaced in a logarithmic scale across the spectrum. D. Gibson (2005) cross-referenced different frequency ranges with colour, although he assigned low frequencies to red, and high to yellow. Matthews, Fong and Mankoff (2005) used coloured icons to illustrate pitch for visual displays of sound for individuals with hearing impairment, blue represented low, and red high pitch. High values, although normally for SPL, are commonly displayed as red, with low being blue. This can be found in fast fourier transformation (FFT) spectrograms, as well as sound level meters where the quietest levels are represented in blue, moving through green and yellow to red when a sound has exceeded the maximum level (Katz, 2007). Temporal attributes were not represented on the maps but could be represented on an animated version. Objects might appear and disappear according to their presence during the recording, scrolling along a timeline could be used to establish which sound events were present concurrently.

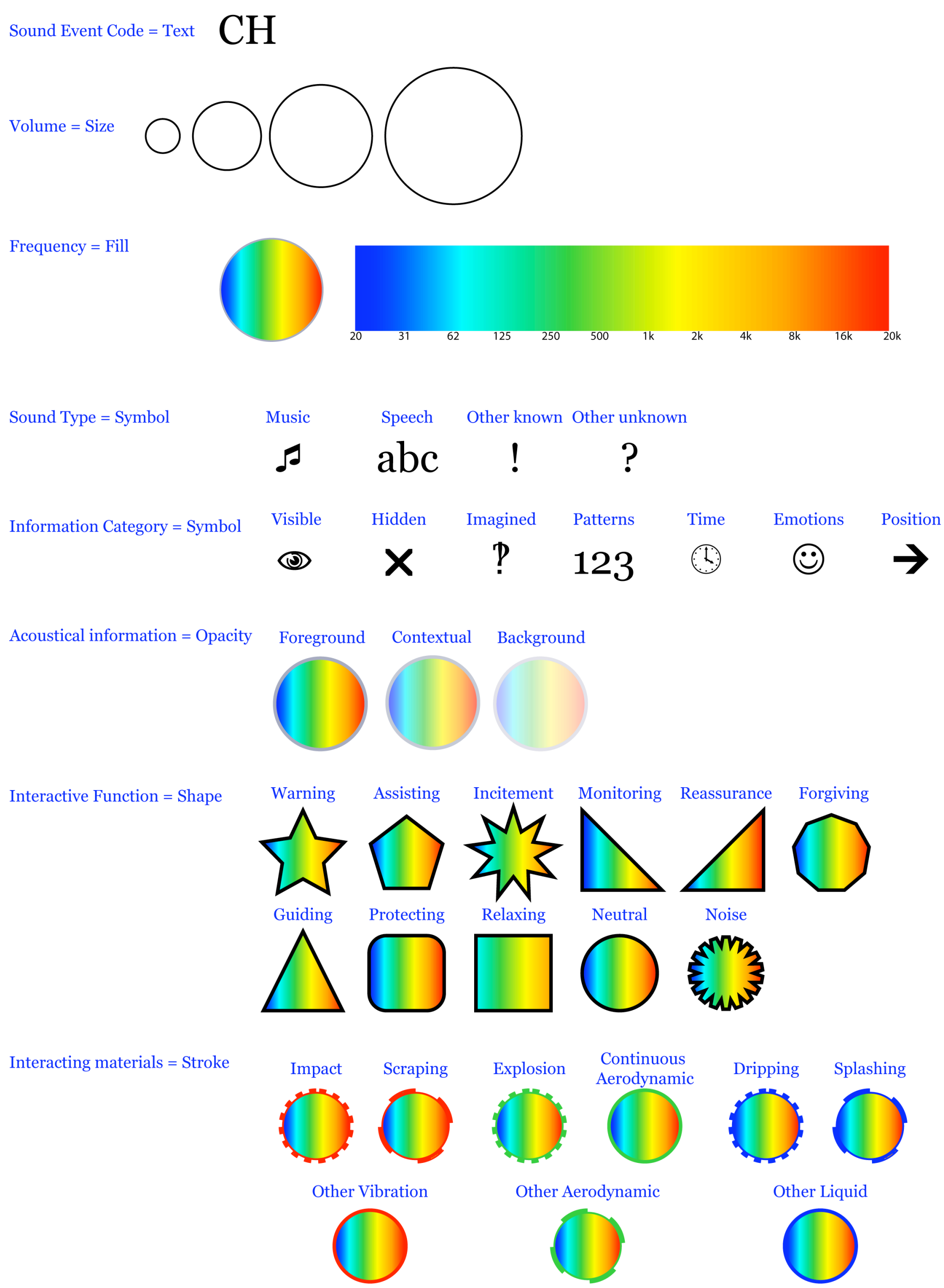


Figure 4-10: Visualisation key

Sound type was represented through symbols: *music* being two quavers ♬ ; *speech* as the letters 'abc'; *other known* was an exclamation mark '!' and *other unknown* a question mark '?'. Information categories were icons: *visible* was an eye 👁; *hidden* a cross ✕, *imagined* a interrobang, which is a question mark overlaid with an exclamation mark ‽. *Patterns* were represented by three consecutive numbers '123', with time being visualised by a clock face 🕒. A smiling face was used for *emotions* ☺, and an arrow for *position* ➔. Acoustical information was shown through the opacity of the fill colour, *foreground* had 100% opacity, *contextual* 66% and *background* 33% opacity. *Interactive function* took the form of different shapes. *Neutral* was a circle, *noise*, *incitement* and *warning* were stars with 18, 9 and 5 points respectively. *Relaxing* and *protecting* were both squares with the latter having rounded corners. *Monitoring* and *reassurance* were right angle triangles in opposite directions, and the remaining classifications of *guiding*, *assisting* and *forgiving* were an isosceles triangle, pentagon and nonagon respectively.

The colour of the borders was used to denote the *interacting materials*. These could be grouped into three: solid (red), liquid (blue) and gas (green) each with a subcategory of short (short lines), intermittent (longer lines) or continuous (continuous line). If a sound event moved, then the start and end points were both displayed and joined by a line that used the same coding as the interacting material.

When combined responses included two or more classifications multiple symbols were used for *sound type* and *information category*. Shapes were split into half and rejoined to illustrate each *interactive function* or show different levels of *acoustical information*. Multiple classifications of interacting materials were shown with multiple borders, so a solid green line denoting *continuous aerodynamic* might be surrounded by a series of short red lines signifying *impact*.

4.4.2 Results

The results can be split into capture, measurement, classification and visualisation. Capture involved the creation of a floor plan, recording the auditory environment and notating all of the audible sound events. Measurement involved establishing the sound pressure level, frequency range, start time, end time and grid reference for each of the audible sound events. Classification involved listeners classifying each sound event that they were aware of using Macaulay and Crerar's 1998, Delage's

1998 and Gaver's 1993 classifications. Visualisation involved the measured sound field and listeners' classified soundscapes being displayed on a single map.

Capture

The capture stage involved the creation of a floor plan, audio recording (as shown in Figure 4-9). The floor plan was useful during the classification process as it allowed listeners to recall different sound events according to its position within the auditory environment. The grid size was appropriate at this stage, but when it came to noting the positions of sound sources, degrees within the cells could have increased the spatial accuracy. Only a single inhabitant referred to the recording equipment whilst the recording was taking place, and there was one instance of mobile phone interference. The recording quality was sufficiently high to note all of the sound events without difficulty. The 30-minute recording contained 435 distinct sound events that were emitted by 139 unique sound sources. Sound events ranged from stationary sources that were continuously audible, such as a radio playing, through to intermittent moving sources, an example of which was an individual talking while walking, as well as single events such as water being dispensed.

The highest recorded SPL was a window being closed at 68 dB A, and the quietest was a person stapling at only 31 dB A. Speech fell between 36-64 dB A, depending upon context. All of the sound sources had relatively low levels, with only 15% of the sound events being above 60 dB A, and 45% being below 50 dB A. Almost the full audible frequency range was present. 20 Hz was generated by the traffic, and 19.6 kHz was present in the form of harmonics when some coins were dropped on a desk. Sound events varied in length from less than 1 second for stapling through to a continuous 30 minutes in the case of the radio playing.

Classification

The listeners' awareness of each sound event was established prior to classification. The regular inhabitants were aware of 57% of the sound events, the intermittent inhabitants indicated a 59% level of awareness, and the new inhabitants were aware of 61% of the sound events. Listeners were aware of sound events such as people speaking, footsteps and papers being shuffled. Listeners were unaware of subtle sound events such as mumbling, whispers or someone shuffling in their chair.

Within the sound type classification 71% of the sound events were classified as *other known*. *Other unknown* only accounted for 2% of the responses, and was not used at all by the regular inhabitants. *Speech* represented 25% of the sound events and *music* was chosen for 3% of the sound events. *Other known* sound events included water dispensing, printing and drawers being opened and closed. *Other unknown* was applied to sound events that the new inhabitants had not previously experienced but could recognise once a description was provided, such as the receipt printer.

When questioning the *intermittent* and *new* inhabitants, listeners were asked if they thought the specified sound event would have been *visible* or *hidden* if they had been in the physical office. *Visible* represented 63% of the total sound events across all three groups. The new inhabitants considered a greater percentage of the sound events to be hidden (36%) compared to the regular group (11%). Regular inhabitants cited an equal percentage of *emotions* and *position* (both 11%). *Emotions* and *position* represented only 4% and 3% of the sound events according to the regular inhabitants. The new inhabitants only regarded 1% of the sound events as being *emotions*, and did not use *position* to classify any of the sound events. *Imagined*, *patterns* and *time* had average instances of only 1%, 2% and 1% respectively. Speech was always classified as being *visible*, irrespective of group. Drawers opening and closing were considered *hidden* by the new inhabitants, but not by the regular and intermittent inhabitants. *Imagined* was used for vehicle movement and typing. *Patterns* were applied to footsteps and typing. *Time* was chosen by two of the intermittent inhabitants for classifying sighs. *Emotions* were selected for singing and sighs, but not for speech. *Position* was used by regular inhabitants for footsteps, as well as for actions preparatory to leaving, such as a person putting on their coat.

There was a greater difference between three groups for the classification of acoustical information. Regular inhabitants classified 49% of the sound events as *background* and 38% as *foreground*. The intermittent inhabitants assigned 51% of the sound events to the *contextual* category. The new inhabitants responses were more evenly spread across the three categories, with *background* being the most frequent at 44%. Speech was predominantly *foreground*, doors opening and closing were considered *contextual*, and both the traffic and the web radio were *background*.

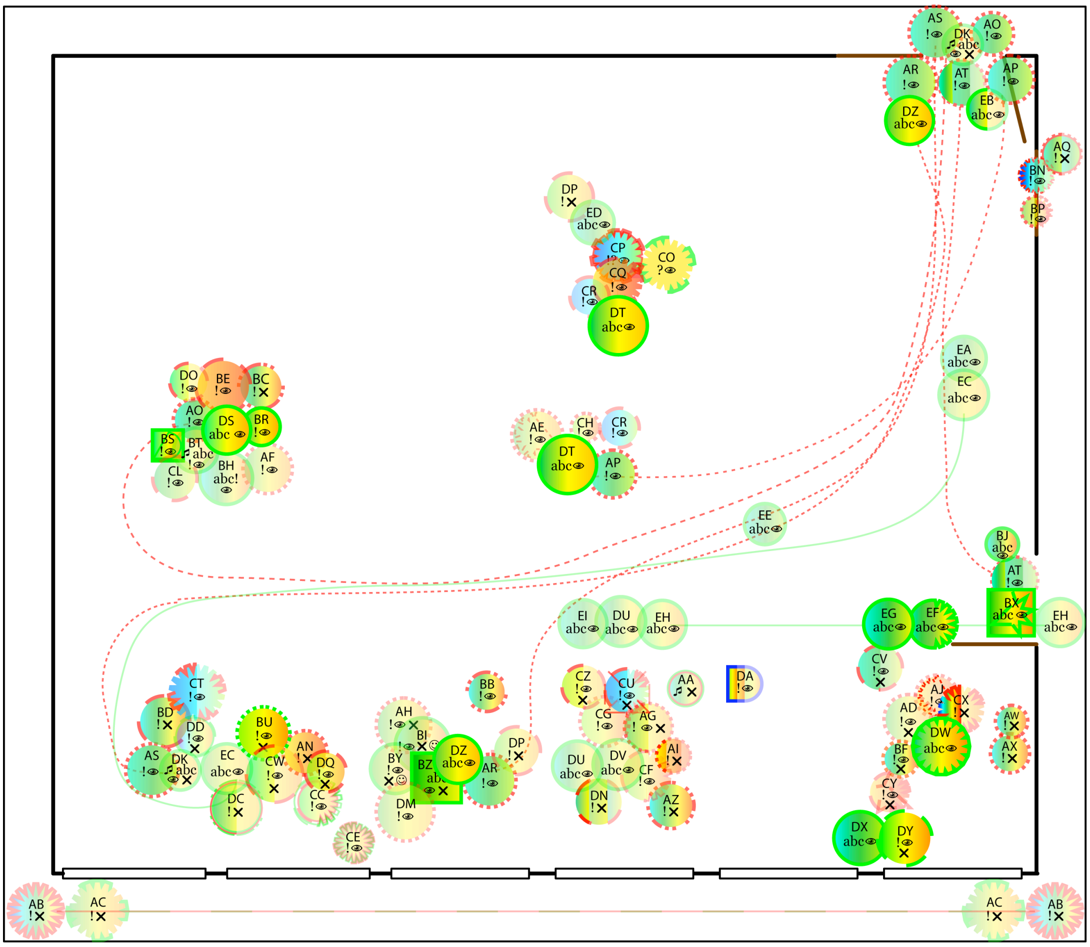
For the interactive functions *neutral* was the most commonly used response across all groups (54%) followed by *noise* (19%). *Neutral* was chosen almost uniformly for speech. *Noise* was attributed to traffic, the hand scanner and receipt printer. Regular inhabitants selected *warning* for 10% of the sound events; examples included a slam and door bangs. Footsteps were considered *assisting* by one of the intermittent inhabitants, and a regular inhabitant found the sound of their own typing and mouse clicking to be *assisting*. The same inhabitant classified the sighs, sneezes and exclamations made by the other inhabitants to be *incitement*. New inhabitants found 11% of the sound events to be *monitoring* such as sighs, sniffles and hiccups. Regular inhabitants regarded none of the sound events as either *monitoring* or *reassurance*. New inhabitants made the most use of *reassurance* with 7% of the sound events falling within this category, examples include a person coughing and clearing their throat. Only three sound events were classified as *forgiving*: a laugh, a sneeze and a sigh. *Guiding* was used for 2% of the total sound events, one regular inhabitant found the footsteps to be *guiding*, whilst a different regular inhabitant considered some speech to be *guiding*. There was only a single instance of *protecting*, which was applied by a regular inhabitant to some speech. Regular inhabitants considered 10% of the sound events to be *relaxing*, examples included the web radio and some singing.

The most common forms of classification for the interacting materials were *impact* (29%) and *continuous aerodynamic* (21%). *Impact* was chosen for sound events associated with computer interaction such as typing and mouse use. *Continuous aerodynamic* was applied to the traffic and speech. *Scraping* was selected for 15% of the sound events and included key rattles and sorting out sheets of paper. *Other vibration* accounted for 14% of the responses, and was applied to the mobile phone ringing. *Explosion* was selected for 3% of the responses, and was used to classify a sneeze. *Other aerodynamic* was applied to 15% of the total sound events and included both the traffic and the web radio. *Dripping*, *splashing* and *other liquid* were rarely used, as the only liquid within the office was the water dispenser.

Visualisation

A map was created for each participant, as well as each of the four combined groups: regular, intermittent, new and combined. Only the walls, windows and doors were retained from the original floor plan onto which the sound events were overlaid. If a listener or group of listeners were unaware of a sound event it was omitted. The measurements, acoustical information (opacity) and interactive function (shape) (see Figure 4-11) could be viewed on the macro scale. The symbols (sound type and information category), as well as the strokes (interacting materials) required to be viewed at a larger scale, especially when they represented multiple classifications (see Figure 4-11). Multiple classifications occurred when there was no single value for the mode within one of the categories. Multiple classifications occurred in all four groups and all of the categories. The new inhabitants had the largest percentage (57%) of sound event multiple classifications. The intermittent inhabitants had an instance of 37% multiple classifications, and the new group had 49%. The combined group had an instance of 30% sound event multiple classifications.

Some of the symbols were more successful than others: *music*, *speech*, *other known*, *other unknown*, *visible*, *time* and *emotions*, all had readily identifiable symbols. *Hidden*, *imagined*, *patterns* and *position* did not have established symbols, making them potentially difficult for the designer to remember.



Code	Event	Source	Code	Event	Source	Code	Event	Source	Code	Event	Source
AA	Web Radio	Computer	BD	Drawer close	P02	CG	Stapling	P01	DJ	Sucking sweet	P06
AB	Movement	Traffic	BE	Key rattles	P04	CH	Stapling	P03	DK	Singing	P02
AC	Movement	Vehicle	BF	Stamp	P05	CI	Stapling	Student 1	DL	Places object in bin	P05
AD	Typing	P05	BG	Slam	P02	CJ	Put on coat	P04	DM	Drop bag	P06
AE	Typing	P03	BH	Sigh	P04	CK	Shuffling	P01	DN	Zip of boot	P01
AF	Typing	P04	BI	Sigh	P06	CL	Shuffling in bag	P04	DO	Zipping open bag	P04
AG	Typing	P01	BJ	Sigh	Staff 1	CM	Shuffling in bag	P06	DP	Zip	P06
AH	Typing	P06	BK	Pen drop	Staff 2	CN	Moving chair	P04	DQ	Pushing telephone buttons	P02
AI	Mouse clicks	P01	BL	Pen drop	P01	CO	Beep	Hand scanner	DR	Dialling tones	Telephone
AJ	Mouse clicks	P05	BM	Pen drop	P02	CP	Receipt printing	Receipt printer	DS	Speaking	P04
AK	Mouse clicks	P03	BN	Door bang	C34b	CQ	Receipt tears	Receipt printer	DT	Speaking	P03
AL	Mouse clicks	P04	BO	Door bang	Corridor	CR	Leafing Paper	P03	DU	Speaking	Staff 2
AM	Mouse clicks	P06	BP	Door being opened	c34b	CS	Leafing Paper	Student 1	DV	Speaking	P01
AN	Mouse clicks	P02	BQ	Mumbling	P04	CT	Leafing Paper	P02	DW	Speaking	P05
AO	Footsteps	P04	BR	Hiccup	P04	CU	Leafing Paper	P01	DX	Speaking	Student 2
AP	Footsteps	P03	BS	Yawns	P04	CV	Straightening paper	Student 1	DY	mobile phone	Student 2
AQ	Footsteps	C34b	BT	Humming	P04	CW	Shuffling paper	P02	DZ	Speaking	P06
AR	Footsteps	P06	BU	Sneeze	P02	CX	Scrunching up paper	P05	EA	Speaking	Man
AS	Footsteps	P02	BV	Sniffles	P06	CY	Tearing paper	P05	EB	Speaking	P12
AT	Footsteps	Staff 1	BW	Whispers	P02	CZ	Rummaging sheets of paper	P01	EC	Speaking	P02
AU	Space bar	P05	BX	Exclamation	Staff 1	DA	water dispensing	Staff 3	ED	Speaking	Student 1
AV	Rummaging with boots	P01	BY	Exclamation	P06	DB	Coins	Desk	EE	Speaking	Staff 4
AW	Drawer open	P05	BZ	laugh	P06	DC	Clearing throat	P02	EF	Speaking	Staff 5
AX	Drawer close	P05	CA	Sellotape rip	P01	DD	Cough	P02	EG	Speaking	Student 3
AY	Drawer open	P01	CB	Sellotape rip	P03	DE	Cough	Corridor	EH	Speaking	Staff 1
AZ	Drawer close	P01	CC	Printing	Printer	DF	Standing up	P03	EI	Speaking	Staff 6
BA	Drawer open	P06	CD	Printing	Printer	DG	Window closing	P02			
BB	Drawer close	P06	CE	Printing	Printer	DH	Handling enquiry book	Staff 4			
BC	Drawer close	P04	CF	Sort out sheets	P01	DI	Opening sweet wrapper	P06			

Figure 4-11: Extended prototype soundscape map representing new inhabitants' responses

The maps were suitable for illustrating some of the differences in listening experiences through comparison. Within the magnified areas in Figures 4-12 and 4-

13 it is shown that the new inhabitants were aware of 11 sound events compared to the regular inhabitants who were only aware of 6 of the sound events. The new inhabitants classified the people speaking (DW and DX) as *foreground* whereas the regular inhabitants considered the speech to be *background*.

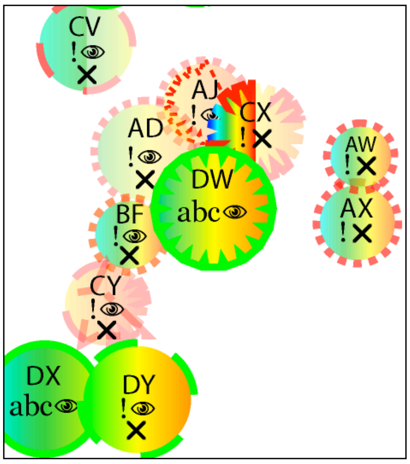


Figure 4-12: Magnified area of extended prototype soundscape map representing new inhabitants' responses

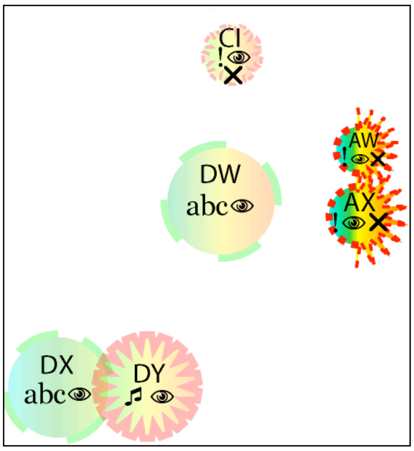


Figure 4-13: Magnified area of extended prototype soundscape map representing regular inhabitants' responses (identical coordinates to Figure 4-12)

4.4.3 Discussion

The soundscape mapping tool was found to be suitable for comparing listeners' experiences of the open plan administrative office. The use of the surround sound recording provided a consistent auditory environment for listeners to experience, as suggested by the similar percentages of sound event awareness across the different groups. The application of sound type was fairly consistent. Differences arose with vocalisations and mobile phone ring tones that reflect different experiences rather than an issue with the classification.

A visual bias was suggested within the information categories with 87% of the combined responses being classified as either *visible* or *hidden*. As identified in previous studies reported in this chapter the acoustical information more accurately represented Amphoux's 1997 levels of listening. The difference between the balance of *foreground*, *contextual* and *background* sound events was evident and might suggest the importance of specific sound events.

The interactive functions were extended for this study with the addition of *neutral* and *noise*. The original nine criteria were only occasionally selected, with *neutral* and *noise* representing 73% of the combined responses. Further studies are required in order to establish if the inclusion of *neutral* and *noise* is appropriate. The responses for the interacting materials suggest that the current nine types might be replaced by: solid, air and liquid each with impulsive, intermittent and continuous states. All of the existing criteria could be categorised according to the substance and pattern of vibration, an *impact* being an impulsive solid sound event, whereas *splashing* might be an intermittent liquid sound event, and *continuous aerodynamic* would become a continuous gas sound event.

Combining listeners' responses allowed the comparison of different levels of familiarity with the auditory environment. Multiple classifications arose and highlighted where responses were inconsistent. The measurements provided consistent values for all of the sound events that listeners were aware of. The choice of mapping for the visualisation allowed all of the attributes to be displayed within a single image. The temporal attributes were omitted and could be included using animation, but movement was still included. The study reported in this section has shown a method that might be used to compare listening experiences. Whilst the tool was trialled with listeners, no sound designers took part in the study.

4.5 Summary

The purpose of the preliminary studies was to establish what some of the procedural difficulties were, and to provide solutions for, comparing listening experiences. The difficulties can be broken down into capture, classification and visualisation. When capturing sound events surround sound recordings might allow listeners to experience a similar auditory environment. Asking listeners to list sound events in real time might prevent them from performing other tasks or may increase the amount of time taken to complete tasks. Preparing a list of candidate sound events rather than asking for individual descriptions could allow listeners to classify the same sound events making responses easier to compare. Recordings could be used to provide a candidate list of sound events for listeners to classify. Noting the source and action of each sound event may provide sufficient information for listeners to identify individual sound events.

In terms of classification listeners appear to be comfortable about classifying sound events from memory allowing responses to include reduced or musical listening. Limiting the scale of an attribute may improve the consistency of responses. Within the information for the shared office study 87% of the responses fell into two of the options (visible and hidden). In the same study 73% of the responses also fell into two of the options (neutral and noise) for the interactive functions. Calculating the mode might be an appropriate way of combining listeners' responses. The level of familiarity with an auditory environment could be an appropriate way of grouping listeners' responses, which might be in terms of regular, intermittent, new and combined (all). Specifying the material and interaction of a sound event may be a useful way of classifying onomatopoeic descriptions. This would also reduce the complexity of the interacting material attributes by having two attributes (material and interaction) with three options rather than one attribute with nine options.

For the visualisation it appears to be possible to visualise multiple sound events within a single map. Eight attributes may be included for each sound event as part of the map. When two sound events occupy the same coordinates it may be possible to distribute them evenly within a cell with minimal overlapping. Possible visual variables for visualising attributes of sound include: volume, fill, symbol, opacity, shape and stroke. Finally, potential methods have been identified of visually indicating when more than one option within an attribute has been chosen.

This chapter described a trial of a prototype method for classifying and visualising soundscapes in order to identify procedural problems and identify possible solutions. Attributes for describing sound that may be important to listeners were identified, and an example was provided of how an extended soundscape mapping tool could be used. No sound designers took part in these preliminary studies. However, all of the sound events were captured using published methods.

5 Designers and Listeners

In the previous chapter preliminary studies were discussed. A prototype soundscape mapping method was developed and trialled in order to investigate some of the problems associated with comparing listeners' experiences. However, as no designers took part in the preliminary studies, the method's suitability for comparing designers' intentions for a sound design with listeners' experiences has not been established. This chapter reports the results of two studies designed to establish attributes for describing sound that are important to both sound designers and listeners. A survey was conducted with audio professionals, and concurrent verbalisations were sought from listeners. In addition, the audio professionals were queried about methods that they used for classifying and visualising sound, as well as other factors that might inform the design of the soundscape mapping tool.

5.1 Pro audio questionnaire

The study reported in this section attempts to identify attributes of sound that might be important to sound designers. The study also tries to establish methods of classifying and visualising sound that are currently used by audio professionals. Seventy-five audio professionals completed a questionnaire created to elicit definitions of the words noise and soundscape, and to identify methods of measuring, classifying and visualising: sound, architectural acoustics and hearing abilities.

5.1.1 Method

A self-administered questionnaire was chosen as the method for gathering data. A survey allowed a greater sample size and diversity of respondents, without the travel or time zone restrictions associated with face-to-face or telephone interviews (Gosling *et al.*, 2004). Self-selection bias meant that the results reported here only represent the participants' responses rather than the wider community of audio professionals (Gray *et al.*, 2007). Direct e-mailing of a document was selected instead of a web survey as there are occasionally technical difficulties such as server crashes/freezes as well as privacy issues (Reynolds, Woods & Baker, 2007). An open-ended approach was chosen for all of the questions, except question 6, in order to allow freedom of response. Example responses were provided for 11 out of 20 questions. Ranking of answers was used in half (10) of the questions. Two of the questions used ranking to indicate which responses were most important. Eight questions included ranking to indicate frequency of use.

The first eight questions related to the respondents' backgrounds (see Table 5-1). Question 1 addressed respondents' job titles, followed by their responsibilities (question 2), and relevant qualifications (question 3). Question 4 asked about participants' own experiences of hearing loss. Brashears-Morlet, Santucci and Morlet (2006) found that 77.1% of musicians had some form of hearing loss, and that only half of those with loss were aware of it, which was considerably above the USA national average of 12.3% (Schoenborn & Heyman, 2008). Little work has been conducted on the level of hearing loss amongst other types of audio professionals.

Background	
1	What is your job title?
2	What are your main responsibilities?
3	What is your highest qualification relevant to sound?
4	Have you ever experienced any hearing loss, if yes, how would you describe it?
5	If textbooks or journals are relevant to your work, please list the ones you would recommend?
6	Please rank the importance of the following three factors within your work.
7	Please give definitions of noise and rank them according to relevance to your field?
8	Have you ever encountered the term soundscape? If yes, please define your understanding of it? If no. what does it suggest to you?
Quantitative Measurements	
9	What terms are you aware of for measuring audio?
10	What terms are you aware of for measuring room acoustics?
11	What terms are you aware of for measuring the hearing abilities of listeners?
Qualitative Measurements	
12	What terms are you aware of for describing audio?
13	What terms are you aware of for describing room acoustics?
14	What terms are you aware of for describing the hearing abilities of listeners?
Notation Classification and Visualisation	
15	What methods are you aware of for notating sounds?
16	What methods are you aware of for classifying sounds?
17	What graphical methods are you aware of for representing sounds?
Final Questions	
18	If there were a technique to represent am auditory environment, what parameters (characteristics) would you want or expect to be included?
19	What auditory environments would you be interested in capturing?
20	Would you like to elaborate further upon any of your answers, or make any points that might be interesting to follow-up?

Table 5-1: Questions posed to audio professionals

Question 5 requested information about relevant textbooks and journals in order to expand the literature review. Question 6 asked about the importance of three factors with regards to the respondent's own area of work. Pre-existing sounds, room acoustics and the hearing abilities of listeners are all contextual factors that might affect the experience of listening. Questions 7 and 8 were directed at finding definitions for the terms noise and soundscape.

Questions 9 - 11 related to measurements for audio, room acoustics and hearing abilities of listeners, followed by descriptions for the same fields (questions 12 – 14). These six questions identified relevant practices within the different fields, with ranking to show their frequency of usage. Information about formal and informal methods of notation, classification and visualisation was sought, so that the extent of the methods application and relevance to specific fields could be established. The questionnaire concluded with questions (15 – 20) about what parameters or characteristics respondents would expect to be included in a representation of an auditory environment, what auditory environments respondents would be interested in capturing, and a request for additional comments. These final questions helped identify what sound events and soundfields to capture for future studies, as well as addressing issues that might have been omitted from the questionnaire.

After trialling with participants from the School of Computing, the questionnaire was e-mailed as an unsolicited Microsoft Word document, to approximately 2000 auditory professionals, until 25 responses each had been obtained from individuals working in the three professions of interest: acoustics, computing and design (see Appendix A). E-mail addresses were gleaned from published papers, membership rolls, newsgroups, and web sites. The response rate was 3.75 % and care was taken not to e-mail a candidate more than once. Respondents included authors of textbooks in their fields, established practitioners within international companies, and cited researchers at universities with a track record in relevant fields. No attempt was made to match additional criteria, such as age, sex or country of domicile.

Participants were informed that the study was part of a research project about looking into the possibility of creating a computer-based system for noting, classifying and visualising auditory environments. Participants were also informed that a greater understanding of the auditory contexts in which designers currently operated was being sought, which would help to ensure that new sounds introduced, and sound designs, would be both audible and appropriate for listeners and the shared auditory environment. The purpose of the questionnaire was indentified as being a tool for finding out what methods and terminology audio professionals currently used when noting, classifying and representing sounds. Responses were e-mailed as text files, which were transferred to a spreadsheet. Results were compiled under four headings: acoustics, computing, design and combined.

5.1.2 Results

Seventy-five individuals completed the questionnaire. Respondents were placed into three equal sized groups according to their responses about their roles and responsibilities: acoustics, computing and design. The first group comprised practitioners in acoustics from a variety of fields, such as architectural acoustics and psychoacoustics. The second group, designers came from a variety of disciplines and their work was concerned with the creation of audio. The final group of computing practitioners were involved with either developing interfaces that incorporated audio or authoring software to manipulate audio.

Academics formed the largest part (54%) of both the acoustics and computing groups, whilst practitioners formed the majority (96%) within the design group. The job title most common amongst the combined respondents was that of a researcher (25%), followed by sound designer (24%) (see Table 5-2). Within the acoustics and computing groups 40% of the respondents were researchers; within the design group 72% of the respondents were sound designers. The computing group consisted of 44% practitioners (software developers/engineers, consultants or project managers), and 56% academics (researchers or professors).

Combied job titles	**n**	**%**
Researcher	19	25%
Sound Designer	18	24%
Professor	9	12%
Consultant	7	9%
Software Developer	6	8%
Sound Engineer	3	4%
Phonetician	2	3%
Software Engineer	2	3%
Sound Artist	2	3%
Engineering Physicist	1	1%
Flight Surgeon	1	1%
Noise Program Coordinator	1	1%
Physicist	1	1%
Pro Audio Manager	1	1%
Project Manager	1	1%
Psychologist (Engineering)	1	1%

Table 5-2: Job titles of the 75 respondents

When the responses were combined the three fields of employment most commonly represented were Music (13%), Software Development (13%) and Psychoacoustics (12%) (see Table 5-3). With the acousticians the largest subgroup was psycho-acousticians (36%), this was balanced by the combination of architectural and building acousticians with noise and vibration acousticians (40%), the rest of the group was made up equally of phoneticians, physicists and physiologists.

Combined fields	**n**	**%**
Music	10	13%
Software Development	10	13%
Psychoacoustics	9	12%
HCI	7	9%
Architectural & Building Acoustics	5	7%
Noise & Vibration Acoustics	5	7%
Theatre	5	7%
Game	4	5%
Film	3	4%
Multimedia	3	4%
Neuroinformatics	3	4%
Phonetics	2	3%
Physics	2	3%
Physiology	2	3%
Technology Development	2	3%
Television	2	3%
Radio	1	1%

Table 5-3: Combined employment fields of the 75 respondents

The categories of responsibility represented by the participants were development (60%), research (39%), administration (27%), education (23%), evaluation (23%)

and sales (4%) (see Table 5-4). Some of the respondents had more than one area of responsibility. Of the 39% of respondents who had research responsibilities, 29% were involved in sound design, 25% in management and 23% in teaching.

Main responsibilities	Acoustics		Computing		Design		Combined	
	n	%	n	%	n	%	n	%
Development	7	28%	15	60%	23	92%	45	60%
Research	15	60%	12	48%	2	8%	29	39%
Administation	8	32%	10	40%	2	8%	20	27%
Education	4	16%	10	40%	3	12%	17	23%
Evaluation	9	36%	2	8%			11	15%
Sales	1	4%	1	4%	1	4%	3	4%

Table 5-4: Main responsibilities of the 75 respondents

Overall, 61% of the respondents had been formally trained; the remaining 39% of the respondents attributed their expertise to industrial experience. Within the acoustics group 76% had formal training, with 44% of the respondents holding PhDs (see Table 5-5). When only the computing and design groups are considered then the split is more even, 56% of the computing participants had formal training and 44% had industrial experience. Fifty two percent of the designers had formal training and 48% had industrial experience.

Sound related qualification	Acoustics		Computing		Design		Combined	
	n	%	n	%	n	%	n	%
PhD	11	44%	8	32%	3	12%	21	61%
Masters	5	20%	3	12%	4	16%	12	16%
Bachelors	3	12%	3	12%	4	16%	10	13%
HND					2	8%	2	3%
Formal	19	76%	14	56%	13	52%	46	61%
Industrial experience	6	24%	11	44%	12	48%	29	39%

Table 5-5: Sound-related qualification of the 75 respondents

With respect to hearing loss 55% of the participants had no current hearing impairment. The remaining 45% of participants had hearing loss ranging from frequency loss through to two cases of severe deafness (see Table 5-6). The figure of 45% is higher than the 6.7% of the UK population between 16 and 60 having some form of hearing impairment (RNID, 2007). The incidence in this sample might be due to increased exposure to sound through work, and possibly a greater awareness or regularity of ear testing. Four participants reported more than one form of hearing impairment, which were either permanent (tinnitus and frequency loss) or temporary (threshold shift and excessive ear wax). Specific frequency losses were identified, which were all in the mid to high range (1kHz - 13.5 kHz), and in a single case was greater than 40 dB. Hearing loss was linked to a number of causes: senility, over-exposure (instruments and firing ranges) and physical damage due to accidents or health issues. The design group had an elevated incidence of frequency loss, but no cases of moderate or severe deafness. The computing group had the lowest level of hearing problems. The acoustics group contained four respondents who were either moderately or severely deaf.

Hearing loss	Acoustics		Computing		Design		Combined	
	n	%	n	%	n	%	n	%
None	10	40%	12	48%	8	32%	30	40%
Frequency loss	5	20%	7	28%	9	36%	21	28%
Temporary threshold shift	2	8%	3	12%	6	24%	11	15%
Tinnitus	4	16%	1	4%	2	8%	7	9%
Mild deafness	1	4%	2	8%	2	8%	5	7%
Moderate deafness	2	8%					2	3%
Severe deafness	2	8%					2	3%
Ear wax	1	4%	1	4%			2	3%

Table 5-6: Hearing losses of the 75 respondents

Respondents were asked to provide definitions of noise and rank them according to relevance to their field. A variety of definitions were provided, e.g. other speech, disorganised sounds, and pink noise, which were subsequently classified. Three categories of noise were shared across all three groups: preference (47%), artefacts (40%) and spectral (28%) (see Table 5-7). The most common definition was unwanted sound (44%), which was classified within preference, but there was no consensus as to a single definition beyond unwanted sound, which corresponds with Hellström's findings (2003). Artefacts included analogue and digital noise, as well as more generic terms such as buzz and hum. Spectral referred to specific types of noise such as white, pink or brown.

All of the respondents understood the concept of the soundscape, from either the natural or constructed perspective, but rarely both. One acoustician referenced

Schafer (1977), and a different acoustician referred to the importance of the point of listening, and range of time. Eighty-eight percent had encountered the term *soundscape*, with 43% defining it as a synthesized auditory environment, 33% as the auditory environment and 21% as the perceived auditory environment, which is an accepted definition in the acoustic ecology literature (Truax, 2001).

Definition of noise	Acoustics		Computing		Design		Combined	
	n	%	n	%	n	%	n	%
Preference	11	44%	14	56%	10	40%	35	47%
Artefacts	6	24%	15	60%	9	36%	30	40%
Spectral	5	20%	10	40%	6	24%	21	28%
Environment	4	16%	4	16%	4	16%	12	16%
Mechanical	2	8%	5	20%	4	16%	11	15%
Emotions	5	20%	3	12%	2	8%	10	13%
Information content	4	16%	4	16%	1	4%	9	12%
Type	3	12%	2	8%	4	16%	9	12%
Architectural acoustics	2	8%	4	16%	3	12%	9	12%
Transport	4	16%	2	8%	1	4%	7	9%
Dynamics	2	8%	1	4%	3	12%	6	8%
Human	3	12%			3	12%	6	8%
Perception			2	8%	3	12%	5	7%
Clarity	1	4%	2	8%	1	4%	4	5%
Source	1	4%	1	4%	1	4%	3	4%
Alerts					2	8%	2	3%

Table 5-7: Definitions of noise provided by the 75 respondents

In order to establish whether the three areas chosen for study: pre-existing sounds; hearing abilities of listeners and room acoustics, were important within the specified fields, respondents were asked to rank the importance of each area on a scale of 1-3 with 1 being the highest (see Table 5-8). Pre-existing sounds represented the background that all audio production has to be set against, from audience noise in a theatre, through to traffic noise when measuring reverberation times and office noise when interacting with an auditory interface. Hearing abilities of listeners referred to impairment or acuity. Room acoustics applied to the acoustic contribution of rooms in which performance, reproduction, recording or measurement takes place. Responses included all three factors being ranked as 1, through to each attribute being ranked individually, with only a single participant choosing to rank attributes using fractions. All three factors had a median rank of 2, suggesting that the factors are considered equally important within each field of work.

Importance of factors	Acoustics		Computing		Design		Combined	
	n	%	n	%	n	%	n	%
Pre-existing sounds	2	40%	2	36%	2	32%	2	36%
Hearing abilities of listeners	2	12%	2	64%	2	32%	2	25%
Room acoustics	2	36%	2	36%	2	32%	2	39%

Table 5-8: Importance of factors within each field of work, median rank and percentage of total responses

The quantitative and qualitative elements of sound events were frequently confused, with classifications or descriptions cropping up in both formal and informal sections of the questionnaire. The level of participants' relevant education was related to the number of quantitative methods used for measuring audio, and the use of formal methods for classifying sounds. Those with more formal training used a greater number of methods. Whilst this generated a large number of measurements of sound that had not been considered in the research up to this point, such as coverage, and clarity or intelligibility, this was not the case when it came to classifications. Sound pressure was the most commonly cited (55%), followed by frequency in Hz (40%) and amplitude in dB (40%) (see Table 5-9). When responses were combined, dynamics was the most common form of measurement, with 93% of respondents citing it. The designers and acousticians both referred to sound pressure level (SPL) the most, whereas the computing group were more concerned with amplitude.

Measuring audio	Acoustics		Computing		Design		Combined	
	n	%	n	%	n	%	n	%
Dynamics	24	96%	24	96%	22	88%	70	93%
Spectral	16	64%	16	64%	17	68%	49	65%
Clarity	8	32%	6	24%	7	28%	21	28%
Architectural acoustics	9	36%	7	28%	4	16%	20	27%
Temporal	5	20%	5	20%	7	28%	17	23%
Reproduction	5	20%	4	16%	6	24%	15	20%
Spatial	2	8%	5	20%	6	24%	13	17%
Hearing abilities	6	24%	2	8%			8	11%
Perceptual	1	4%	3	12%	1	4%	5	7%
Aesthetics	1	4%	3	12%			4	5%

Table 5-9: Attributes for measuring audio

There was little difference between the three professional groups in overall terms when referring to spectral attributes (65%). All three groups were interested in frequency in Hertz, only computing specialists and designers mentioned pitch.

Acousticians had a broader range for measuring spectra, such as modal build up and engine order levels. Clarity was the next most popular attribute (28%). Both acousticians and the designers measured clarity as percentage distortion. However, the computer practitioners did not use this criterion, they referred to colouration. Acousticians were more concerned with architectural acoustics, as well as having a more diverse range of measurements than the other two groups. All of the groups referred to reverberation time (RT) as being the most important, with one of the acousticians and two of the computing specialists mentioning impulse response. The acousticians referred to 13 different measurements whereas the designers referred to four: RT, excitement, isolation and behaviour. Temporal attributes were measured in seconds or milliseconds when applied to phase shifting. Reproduction was referred to in terms of power in watts and sampling rate, followed by coherency and bit depth, with the acousticians again having a slightly broader range of measurements. There was no consensus on spatial attributes, methods were informal such as panning or distance. Hearing abilities were mostly referred to by the acousticians with no instances from the designers, and only generic descriptions from the computing specialists: as in hearing loss and limitations. Perceptual and aesthetic attributes were only briefly alluded to.

Respondents employed a range of adjectives to describe sound events, specifically dynamics, spectral and aesthetic qualities (see Table 5-10). The most commonly referred to attribute for describing audio was dynamics (76%), which was expressed as either volume or loudness. Spectral descriptions were more varied, in terms of pitch, timbre or tone. Aesthetics mostly related to brightness, harshness or warmth, all of which referred to spectral aspects as cited by Katz (2002). Clarity was the first attribute for describing sound for which there was no consensus, the acousticians and computing specialists both cited sound quality and clarity, which were not referred to at all by the designers, who used terms such as dirty, unclear and overdriven. Architectural acoustics were mentioned by all three groups but without any commonalities, as were all of the remaining attributes.

Describing audio	Acoustics		Computing		Design		Combined	
	n	%	n	%	n	%	n	%
Dynamics	19	76%	21	84%	21	84%	61	81%
Spectral	14	56%	17	68%	13	52%	44	59%
Aesthetics	11	44%	12	48%	15	60%	38	51%
Clarity	14	56%	11	44%	11	44%	36	48%
Architectural acoustics	8	32%	8	32%	6	24%	22	29%
Perceptual	9	36%	5	20%	6	24%	20	27%
Spatial	9	36%	7	28%	4	16%	20	27%
Type	5	20%	5	20%	4	16%	14	19%
Temporal	2	8%	8	32%	3	12%	13	17%
Reproduction	3	12%	5	20%	5	20%	13	17%
Musical			5	20%	3	12%	8	11%
Interacting materials	3	12%	2	8%			5	7%
Onomatopoeia	3	12%			1	4%	4	5%
Hearing abilities	2	8%					2	3%

Table 5-10: Attributes for describing audio

The most common forms of visualising sound were spectral, dynamics and spatial (see Table 5-11). Spectral representations were the most commonly cited, with the most common being a spectrogram. The next was dynamics, with waveform being the only visualisation method selected by more than a single individual. Finally spatial attributes were represented, but without any consensus, with forms such as contour mapping and ray tracing. All of the other attributes were not used to any significant extent. There was limited reference to musical notation (7%), although it did figure much more prominently when referring to notating sound.

Graphically representing audio	Acoustics		Computing		Design		Combined	
	n	%	n	%	n	%	n	%
Spectral	20	80%	20	80%	17	68%	57	76%
Dynamics	12	48%	18	72%	15	60%	45	60%
Spatial	5	20%	7	28%	5	20%	17	23%
Sound type	1	4%	2	8%	4	16%	7	9%
Hearing abilities			5	20%	1	4%	6	8%
Reproduction			2	8%	3	12%	5	7%
Architectural acoustics	1	4%	3	12%	1	4%	5	7%
Music	1	4%	2	8%	2	8%	5	7%
Temporal			1	4%	2	8%	3	4%
Synthesis			1	4%	1	4%	2	3%
Linguistic					1	4%	1	1%

Table 5-11: Attributes used for graphically representing audio

Room acoustics were only referred to in any depth by the acousticians. Non-acousticians used terms such as reverberation time (RT) and frequency response, without specifying scales. Sound designers were the least concerned with room acoustics, but two specified the reproduction quality of listeners' audio hardware. When quantifying room acoustics only RT had a majority response across all three groups (65%). There was a broad range of measurements the most common being temporal (76%), with RT and decay rate being shared by all groups (see Table 5-12). Reflection had a more diverse set of measurements within the computing group (24) than either the acousticians (10) or the designers (16). With regards to spectral, only frequency response was shared, mostly attended to by the designers. Dynamics related to impulse response, but only by acousticians and computer practitioners.

Measuring	Acoustics			Computing			Design			Combined	
room acoustics	n	%		n	%		n	%		n	%
Temporal	20	80%		17	68%		20	80%		57	76%
Reflection	6	24%		13	52%		11	44%		30	40%
Spectral	7	28%		9	36%		10	40%		26	35%
Dynamics	10	40%		9	36%		5	20%		24	32%
Absorption	7	28%		8	32%		6	24%		21	28%
Clarity	10	40%		4	16%		4	16%		18	24%
Spatial	6	24%		4	16%		3	12%		13	17%
Room Type	1	4%		1	4%		1	4%		3	4%

Table 5-12: Attributes for measuring room acoustics

When describing room acoustics the most important attribute was reflections specifically dead/dry and reverberant/live (see Table 5-13). Spectral referred to bright, boomy and warmth. Clarity was used in terms of dull, muddy/muffled and clear. Finally dynamics produced a diverse group of responses, the only shared terms being by the acousticians and designers who referred to both noisy and quiet.

Describing	Acoustics			Computing			Design			Combined	
room acoustics	n	%		n	%		n	%		n	%
Reflections	13	52%		19	76%		21	84%		53	71%
Spectral	10	40%		13	52%		21	84%		44	59%
Clarity	9	36%		9	36%		11	44%		29	39%
Dynamics	7	28%		6	24%		5	20%		18	24%
Spatial	7	28%		4	16%		6	24%		17	23%
Temporal	3	12%		6	24%		3	12%		12	16%
Absorption	2	8%		3	12%		3	12%		8	11%
Room Type	1	4%		2	8%		2	8%		5	7%

Table 5-13: Attributes for describing room acoustics

When measuring hearing abilities the most common attribute was dynamics in terms of hearing level in dB, threshold and amplitude sensitivity (see Table 5-14). Spectral measurements of frequency sensitivity/response and hearing loss came next. There was no consensus on any other attributes apart from localization within the spatial attribute. Almost all of the measurements were concerned with hearing loss rather than acuity.

Measuring hearing abilities	**Acoustics**		**Computing**		**Design**		**Combined**	
	n	**%**	**n**	**%**	**n**	**%**	**n**	**%**
Dynamics	15	60%	13	52%	15	60%	43	57%
Spectral	12	48%	13	52%	12	48%	37	49%
Clarity	5	20%	6	24%	2	8%	13	17%
Spatial	4	16%	4	16%	3	12%	11	15%
Type of impairment	4	16%	3	12%	3	12%	10	13%
Temporal			6	24%			6	8%

Table 5-14: Attributes for measuring hearing abilities

When describing hearing abilities spectral attributes became more prominent, specifically frequency loss, and high frequency roll off/loss, followed by dynamics that were mostly described in terms of hearing loss in decibels (dBs) and sensitivity to level changes again in dBs (see Table 5-15). Types of impairment were applied predominantly to deafness and tinnitus. Clarity, spatial and temporal attributes were without any real consensus, beyond the ability to localize a sound source, which related to both the computing and design groups.

Describing hearing abilities	**Acoustics**		**Computing**		**Design**		**Combined**	
	n	**%**	**n**	**%**	**n**	**%**	**n**	**%**
Spectral	12	48%	13	52%	12	48%	37	49%
Dynamics	12	48%	11	44%	8	32%	31	41%
Type of impairment	11	44%	13	52%	7	28%	31	41%
Clarity	4	16%	4	16%	3	12%	11	15%
Spatial	2	8%	3	12%	3	12%	8	11%
Temporal	1	4%	3	12%	1	4%	5	7%

Table 5-15: Attributes for describing hearing abilities

The most common form of notation was musical notation. Other forms were both formal and informal, within the combined attributes, spectral pitch and frequency were the only two methods shared by all three groups, with timbre being used by the computing specialists and acousticians, but not by the designers, who were more interested in spectrograms/Fast Fourier analyses (see Table 5-16). Music notation was consistent across all three groups, both the computing specialists and the designers referred to a broader range such as non-western notations as well as piano rolls, tablature and chords. This was also the case when it came to the traditional comparison of amplitude versus time. Other methods were usually informal and unique to individual respondents such as intensity maps and relative levels. Duration was the only temporal attribute shared by all three groups, and all of the other attributes, with the exception of MIDI within programming language, were disparate.

Notating sound	Acoustics		Computing		Design		Combined	
	n	%	n	%	n	%	n	%
Spectral	18	72%	18	72%	18	72%	54	72%
Music	11	44%	19	76%	21	84%	51	68%
Dynamics	10	40%	12	48%	18	72%	40	53%
Temporal	5	20%	7	28%	6	24%	18	24%
Programming language	2	8%	11	44%	3	12%	16	21%
Linguistic	4	16%	4	16%	3	12%	11	15%
Spatial	3	12%	3	12%	5	20%	11	15%
Type	1	4%	3	12%	3	12%	7	9%
Clarity	2	8%	2	8%	1	4%	5	7%
Perceptual	1	4%	2	8%	2	8%	5	7%

Table 5-16: Attributes for notating sound

The classification of sounds was equally diverse. A few published methods were included by the respondents such as Smalley's spectro-morphology (1986), Schaeffer's typo-morphology (1966), Gaver's interacting materials (1993), Schafer's environmental method (1977), Wake and Asahi's verbal expressions (1998), along with the American National Standards Institute (ANSI) 1994 acoustical terminology. These responses represent a broad area, from the technical through to the aesthetic, but all were only cited singly, so are possibly not used as part of the standard practice across any of the three professional groups. The most popular method of classification was by sound type, which mostly fell into speech, music and non-speech/natural/artificial/everyday (see Table 5-17). Musical classifications fell mostly into the type of music or its instrumentation, again without consensus. Gaver's 1993 interacting materials was referred to by 5% of the respondents, and a further nine respondents detailed attributes contained within Gaver's taxonomy. A sound's artificiality was noted in terms of either being natural or mechanical/man made/artificial, again without consensus. The quantifiable attributes of sound such as temporal, dynamics, spectral clarity and spatial were rarely used to classify sound. The qualitative perceptual and aesthetic attributes were also rarely applied.

Classifying sound	Acoustics		Computing		Design		Combined	
	n	%	n	%	n	%	n	%
Sound Type	11	44%	14	56%	12	48%	37	49%
Music	2	8%	7	28%	5	20%	14	19%
Interacting Materials	4	16%	6	24%	4	16%	14	19%
Artificial	3	12%	7	28%	3	12%	13	17%
Temporal	4	16%	1	4%	8	32%	13	17%
Spectral	6	24%	4	16%	2	8%	12	16%
Environment	6	24%	2	8%	4	16%	12	16%
Interaction	3	12%	5	20%	4	16%	12	16%
Perceptual	4	16%	5	20%	3	12%	12	16%
Linguistic	4	16%	5	20%	2	8%	11	15%
Dynamics	4	16%	4	16%	3	12%	11	15%
Clarity	3	12%	2	8%	3	12%	8	11%
Spatial	1	4%	3	12%	3	12%	7	9%
Reproduction	1	4%	2	8%	3	12%	6	8%
Complexity	1	4%	1	4%	1	4%	3	4%
Aesthetics	1	4%	2	8%			3	4%
Room Acoustics					1	4%	1	1%

Table 5-17: Attributes for classifying sound

Participants were asked what parameters (characteristics) they would wish to be included if there were a technique to represent an auditory environment. This provided an insight into which attributes are currently under-represented, or difficult to capture. The first attribute was spatial which was mostly of interest to the computing specialists and the designers, this was broken down into location, direction and diffuseness (see Table 5-18).

Parameters	Acoustics		Computing		Design		Combined	
	n	%	n	%	n	%	n	%
Spatial	7	28%	15	60%	16	64%	38	51%
Dynamics	12	48%	6	24%	11	44%	29	39%
Architectural acoustics	6	24%	8	32%	13	52%	27	36%
Spectral	8	32%	8	32%	9	36%	25	33%
Temporal	7	28%	8	32%	9	36%	24	32%
Sound Type	4	16%	7	28%	7	28%	18	24%
Perception	4	16%	3	12%	3	12%	10	13%
Hearing abilities	1	4%	1	4%	1	4%	3	4%
All	1	4%	1	4%			2	3%
Clarity	1	4%	1	4%			2	3%
Emotions					2	8%	2	3%

Table 5-18: Parameters for representing an auditory environment

Dynamics was mostly related to either sound pressure level, power or perceived intensity. Architectural acoustics revolved around reverberation time and to a limited extent, absorption, with spectral relating mostly frequency followed by timbre. Time and duration represented the temporal attribute, with source and type being suggested under sound type. There was no agreement on the final attributes of: perception, hearing abilities, all attributes, clarity and emotions.

When respondents were asked which auditory environments they would be interested in capturing, acoustics was the most commonly cited. This was mostly made up of everyday sounds and music. The next, which were equally important, were natural environments and commercial interiors, closely followed by every environment (see Table 5-19). This data provides a set of environments to test the soundscape mapping tool with, as well as yielding information about which environments would be of interest to specific groups. The most common single environment, with a fairly equal response across all three groups was auditoria (21%).

Auditory	**Acoustics**		**Computing**		**Design**		**Combined**	
environments	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**
Acoustics	6	24%	6	24%	8	32%	20	27%
Natural	4	16%	6	24%	6	24%	16	21%
Commercial interiors	6	24%	5	20%	5	20%	16	21%
Every	3	12%	5	20%	7	28%	15	20%
Urban	4	16%	2	8%			9	12%
Domestic interiors	3	12%	4	16%	1	4%	8	11%
Virtual			4	16%	2	8%	6	8%
Commercial transport	4	16%	1	4%			5	7%
Religious	1	4%	2	8%	1	4%	4	5%
Anatomical	1	4%			3	12%	4	5%
Educational	4	16%					4	5%
Domestic transport	2	8%			1	4%	3	4%
Space travel	3	12%					3	4%
Sports			1	4%			1	1%
Commercial exteriors	1	4%					1	1%
Domestic exteriors	1	4%					1	1%

Table 5-19: Potential auditory environments for capture

Overall there was little overlap of terminology within the professional fields, except in the most general of terms. There was also little evidence of established methods to note, classify and visualise sound events, beyond those of waveform and spectrograph. There were specific exceptions within the acousticians' responses, but sound designers and computer practitioners evidenced little need for methods of visualisation.

Computing participants used the term sound event, whereas sound designers preferred the terms sound or audio, disassociating the sound from its source. The overall response to the research varied from not seeing its relevance, to requesting access to published results. An acoustic phonetician suggested that the proposed method would prove ideal for use within their field, which they felt that sound designers and engineers traditionally ignored.

The questionnaire elicited methods and terminologies that this group of audio professionals use when noting, classifying and visualising sounds. The survey has also suggested that concepts such as the soundscape and noise have limited standard accepted definitions, even within the same professional field. Attributes of sound that might be important to sound designers in order to describe sound have been identified.

5.2 Listener concurrent verbalisations

This section describes a study where listeners' concurrent verbalisations were used in order to provide an indication of key attributes for the perception of soundscapes and their relative importance. Concurrent verbalisations were first introduced as a technique by Duncker in 1945, and involve participants thinking aloud or verbalising what they are thinking of. This protocol has been widely used by the HCI community for usability evaluation studies, as concurrent verbalisation does not require participants to have any specialist training (Jensen, 2007). For this study participants were asked to describe what they were listening to. All of the verbal descriptions were recorded, transcribed, and then coded to provide attributes suitable for comparison with the data gleaned from the questionnaire completed by audio professionals as reported in the previous section.

The interviews reported in Chapter 4 concentrated on what Dubois, Guastavino and Raimbault (2006) refer to as representations in memory, where participants were not questioned about sound events that they had just heard, but about their recollections of sounds. This study concentrated on descriptions of what listeners were attending to, by eliciting concurrent verbalisations, from their working memory, which were subsequently analysed. Concurrent verbalisations can go further than just reflecting participants' working memories. Concurrent verbalisations can provide data about what participants are attending to (Turner & McGregor, 2004). Giles (2002) proposed that it was possible to elicit information about subjective experiences using content analysis of verbal accounts, a technique that is typically associated with grounded theory. Asking participants to describe experiences as they perceive them, and then analysing the data without recourse to published literature or established theories, can be described as a phenomenological approach (Merleau-Ponty, 1962 and Turner *et al.*, 2003).

Ericsson and Simon (1980) argued that using verbal protocols does not affect the primary task, although it might slow it down. Russo, Johnson and Stephens (1989) disagreed, citing problems with tasks that were working-memory intensive, which in their case reduced the accuracy of participants' mental arithmetic. Wilson and Schooler (1991) found that by verbally reflecting upon their decisions participants were less consistent. Conrad, Blair and Tracy (1999) suggested that this effect could be partially mitigated, by ensuring that those running the experiments remain detached, and only prompt participants to resume their description.

Detienne and Soloway (1990) referred to the capture of verbal protocols as a way of accessing participants processing strategies, which in the case of sound could provide information about the auditory attributes listeners are attending to, in order to identify a sound source. Vanderveer (1979) found that when asking listeners to describe what they heard from recordings, they predominantly referred to the action, object and location, with little attention being paid to any other attributes. Ballas and Howard (1987) extended this work, comparing environmental sounds to speech, and proposing that they could be regarded as linguistic. Ballas and Howard referred to bottom-up and top-down processes where a sound is given meaning by understanding the acoustical information and its context, as well as from prior experience and expectations. Ozcan and van Egmond (2005) concentrated on descriptions of product sounds, identifying what attributes listeners discerned, in order to create a lexicon for a wider range of auditory fields. They found that different amounts of recognition resulted in contrasting descriptions, which could be combined into 11 groups. A sound that was not recognised was described in onomatopoeic, psychoacoustical, temporal and emotional terms. Whereas, if a sound was recognised but not identified, a location where they had previously experienced the sound, the interacting materials and the sound type were communicated. If the source of a sound was identified, then its properties were conveyed using adjectives.

For the study detailed in this section, rather than describe sounds individually, as detailed above, participants were exposed to an entire soundfield, and asked to describe what they heard. Listeners' concurrent verbalisations were recorded so that descriptions could be transcribed, coded, a lexicon of terms used to describe sounds generated, and then classified. Coding was applied in order to understand what people were listening to in context, the lexicon provided the individual terms used to describe sounds, and the classification was used to derive specific attributes, and their relative importance.

Blauert and Jekosch (1997) highlighted the problem of a considerable reduction of information, in that listeners routinely referred to fewer than four parameters when describing sounds, whereas sound engineers required a high number of parameters in order to represent what is being heard. Guastavino and Dubois (2006) found that there was a 'lack of basic terms to describe soundscapes'. Descriptions of auditory environments were based on sources, actions, social activities, time and location, which they grouped into source events and background noise. The source events are what would be heard when everyday listening and the background noise would be musical listening. Only the background noise whose sources and actions could not be readily identified was described in spectral and temporal terms, although qualitative judgements were applied to both groups.

Ozcan and van Egmond (2005) found that the attributes which people used, when they described what they were listening to, varied according to their ability to identify the source. Carello, Wagman and Turvey (2005) stated that it is only when a source cannot be identified that sensory aspects are reported, moving from everyday to musical listening (Gaver, 1993). Guastavino (2007) reported that the sheer number of sound events within complex real-world auditory environments, make it difficult for listeners to identify individual sources.

Modifying the conditions in this study, so that the potential ease of identification varies, may provide a greater range of attributes. In addition this approach highlights the differences between listening to a real auditory environment and a recorded (virtual) auditory environment. The first condition was in the physical environment with full sight. Listeners knew where they were and were able to identify sources by looking at them. The second condition was again in the physical environment, but participants were wearing a blindfold, so that they could not immediately identify the sources of sounds, but knew where they were. The third condition was using the recording, participants were able to see that all of the sounds came from loudspeakers, but they did not know what the environment was, although they knew that it was artificial. The final condition again used the recording, but this time the participants were blindfolded prior to being let into the room, so that not only were they unaware of what the auditory environment was, they were also uninformed as to its artificial, recorded nature.

5.2.1 Method

The study involved 40 participants being asked to describe verbally what they could hear while listening to either a surround sound audio recording of, or the real, 500 seat computing centre illustrated in Figure 5-1. The computing centre is used for classes, individual study, entertainment, and exhibition purposes. Sound travels freely throughout the space. This complex auditory environment was chosen in order to generate descriptions about a wide variety of sound events.

Figure 5-1: 500 Seat, 8000 m^3, Jack Kilby Computing Centre

An eight-channel audio recording/replay system was employed in order to reproduce the soundfield of the Jack Kilby Computer Centre (JKCC, main computer lab at Edinburgh Napier University) during a weekday term time afternoon. The 30-minute recording involved eight identical omnidirectional tie-clip microphones, with subsequent speaker positioning matching the microphones in floor position and height (see Figure 5-2). Microphones were positioned in an ellipse at 1.5 m to emulate the majority of inhabitants' seated listening positions. Omnidirectional microphones were chosen to maximize the capture of natural reflections, as well as to ensure that nothing was off-axis as would be the case with directional microphones. Verbal permission to record the soundfield for 15 minutes was sought from all inhabitants in the immediate vicinity. Inhabitants were informed of the nature of study and assured that data would be anonymised to prevent individuals being identified.

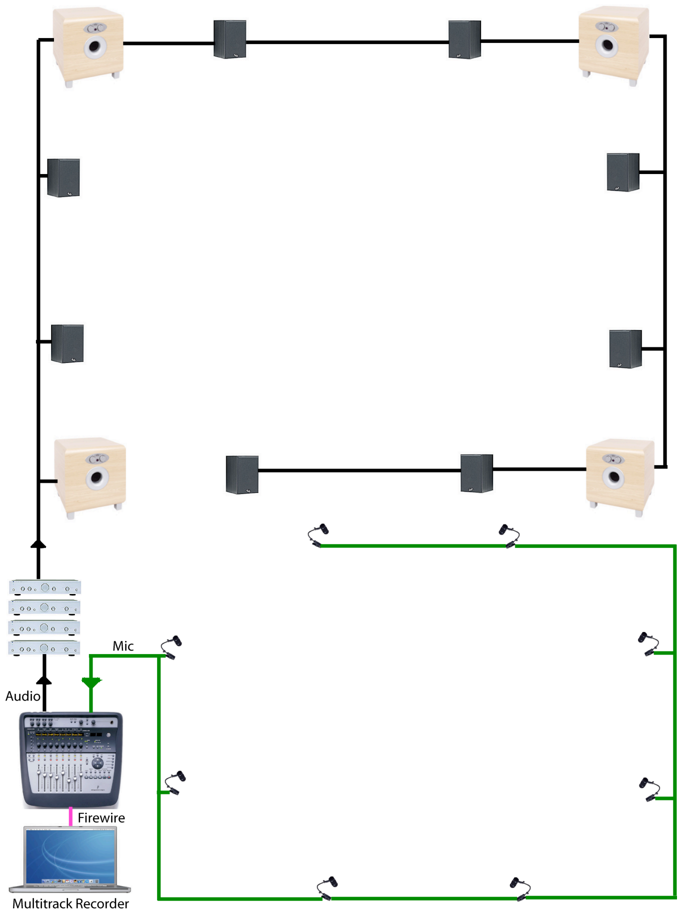


Figure 5-2: Representation of the recording/playback system. eight microphones were used for recording, eight speakers and four subwoofers were used for playback

The recording was made in a single pass onto eight separate channels, and a separate eight-channel microphone preamp was used to minimise distortion and ensure consistency in both dynamics and frequency. Each channel was recorded at 96kHz and 24 bits, which provided a theoretical dynamic range of 144 dB, and ensured that the full audible range was covered. Calibration between the physical soundfield and its subsequent reproduction was achieved with the aid of a sound pressure level (SPL) meter. The meter was set to the C scale and recorded an average of 48dBC.

For reproduction, four sub bass units supplemented eight compact monitors. Whilst bass transmission can normally be considered omnidirectional, the low SPL levels made accurate positioning of low frequency sounds, such as people walking on hollow resonant floors, difficult. The use of four sub bass units partially resolved this problem, achieving a more accurate representation than that normally associated with a 5.1 or 7.1 system, where a single sub bass loudspeaker is normally located in front of the listener. Freeman and Lessiter (2001) found that increased bass within a multichannel audio system improved the perceived naturalness. The sub bass units also compensated for the reduced frequency transmission range associated with compact monitors (see Figure 5-3).

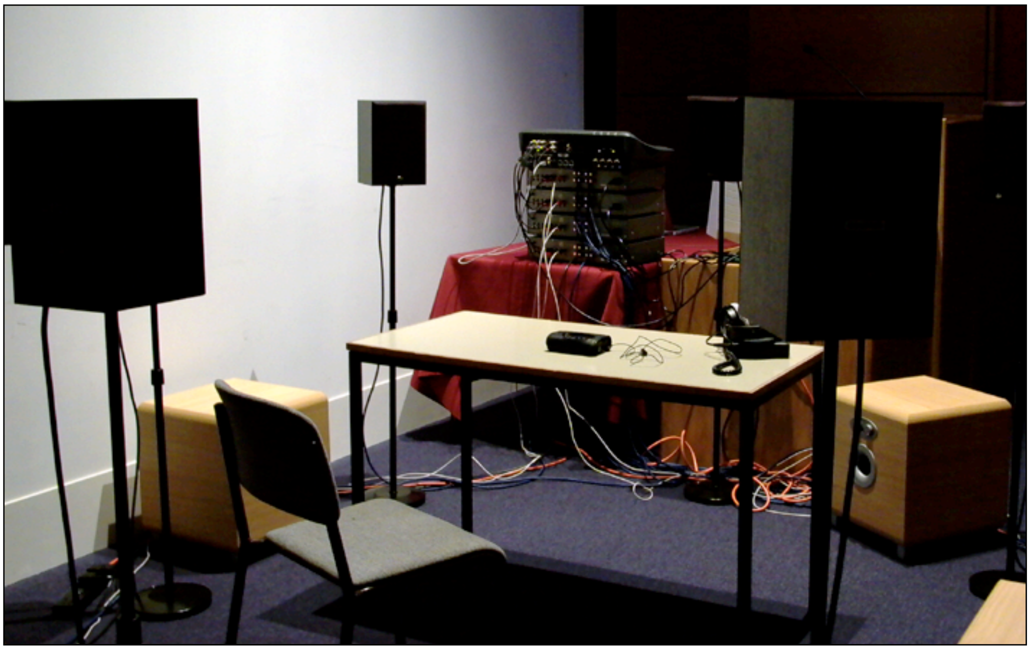


Figure 5-3: Surround sound reproduction apparatus

Forty participants were recruited from staff and students at Edinburgh Napier University. Participants varied with respect to age, sex and background. All participants took part in the study on a voluntary basis and had a good command of spoken English. Participants were informed of the nature of study and were assured that their responses would be anonymised to prevent individuals being identified. Verbal permission was obtained to make use of the results for this dissertation and associated publication. Participants were informed that the experiment would last approximately 15 minutes and that they could end the task at any point. Participants were also given the opportunity to ask questions after the experiment.

The 40 participants were divided into four groups according to the following conditions:

- Condition 1: while physically present in the JKCC for 15 minutes participants were asked to describe verbally what they could hear.
- Condition 2: participants were blindfolded while physically present in the JKCC for 15 minutes, and were asked to describe verbally what they could hear.
- Condition 3: participants were exposed to the recorded soundfield for 15 minutes. They were asked to describe verbally what they could hear.
- Condition 4: participants were blindfolded and exposed to the recorded soundfield for 15 minutes. They were asked to describe verbally what they could hear.

For conditions 3 and 4, the participants were guided into a room and seated at a table where they were asked to listen to the recording and describe what they could hear for 15 minutes. For condition 4 listeners were blindfolded and unaware that they would be seated in the midst of loudspeakers. After 15 minutes listeners were guided back out of the room, and their blindfold removed. At no stage during the experiment did listeners see the interior of the room or its contents.

The recording allowed half of the participants to experience an almost identical auditory environment, the sounds they generated themselves being the only variant.

The groups in conditions 1 and 2 each experienced a unique auditory environment, which extended the number of sound events that could be described by participants. The reasoning for four different groups was first to establish whether a surround sound recording could represent the real auditory environment, allowing repeatability. A second reason was to identify the effect of being able to see the sound sources in order to attribute meaning, and investigate how this affected descriptions.

All of the participants were asked to describe either the recorded or natural physical soundfield of the JKCC. Participants' descriptions were recorded using a stereo tie-clip microphone, onto a DAT recorder set to 48kHz 16 bit. This allowed an accurate stereo image in order to emulate the participant's listening experience with reference to their own voice, as well as a source for later transcription.

The recordings were transcribed and time-coded. Coding was conducted using the qualitative analysis software ATLAS.ti (2008). Codes were derived using a similar approach to Davenport, Higgins and Somerville (1998) and Turner, Davenport and Van De Walle (2004), that is a standard qualitative approach where transcriptions were read and re-read until recurring codes were identified. An open approach to qualitative coding was adopted (Goulding, 2002), where codes were generated through the first pass. These codes were then applied during a second pass. Dr Phil Turner from Edinburgh Napier University's School of Computing confirmed the codes through independent reading. No statistical test was applied in order to establish the level of agreement, as the codes were intended to be a starting point for the classification of the lexicon, rather than an end in themselves. Listing every term that had been used to describe a sound event created the lexicon. The lexicon not only provided a list for the identification of attributes that listeners used to describe sound, the lexicon could also be used to suggest criteria or a scale for each attribute.

5.2.2 Results

The verbal descriptions provided by the listeners were similar across all four conditions. Guastavino *et al.* (2005) also found that when they compared the experience of listening to a physical auditory environment, with stereo (2D) and surround sound (3D) recordings, the verbal descriptions of sound events were similar in all three conditions. A number of the participants in conditions 3 and 4, who were unaware of where the recording took place, started by trying to establish what the space was they were listening to. The listeners initially listed the individual sound events and then pieced the information together in order to establish the type of auditory environment. For example: ‘Again the same sense of people in the distance doing something... sitting, chatting but all very distant from me say oh... say fifteen, twenty, thirty feet it does still feel that I’m still in a large open space but indoors definitely indoors.’ The interpretation of the type of environment appeared to affect decisions about what sounds participants were listening to.

There was a wide variation in responses within the same condition. One participant, in the vision virtual condition, used short statements, and concentrated on identifying sources and actions, employing onomatopoeic terms when the source and action could not be identified:

> 00:00:03 to 00:00:05
> I can hear people talking
>
> 00:00:14 to 00:00:26
> I can hear someone typing
>
> 00:01:22 to 00:01:23
> I can hear printers
>
> 00:03:42 to 00:03:44
> I can hear a rustle noise
>
> 00:05:03 to 00:05:05
> I can hear foreign people speaking

A second participant provided a fuller description that addressed estimation, material, dynamics, location, gender, quantity, aesthetics, spectral, age and clarity, in addition to source, action and onomatopoeia, this suggested that their approach was much more considered:

> 00:00:03 to 00:00:51
> It s a sound that's like paper being slid across am... and also waves am there is some slight tapping very faint am... it sounds like lots and lots of people in the background perhaps walking along a woman's voice.. it's quite low am... it's like a hubbub... murmur... am and at the same time it sounds like wind rattling through and into a window and rattling blind... A plasticy sort of sound and you can hear a woman's and a man's voice occasionally above the hubbub... you hear odd voices coming
>
> 00:00:57 to 00:01:15
> Now there's a am a faint buzz in the background... the plastic rattling sound is getting more insistent... people coughing...
>
> 00:01:22 to 00:01:28
> More coughing lots of coughing... the volume seems to be going up quite a lot...
>
> 00:01:38 to 00:01:53
> There is a sharp small metallic /clinging sound and that's coming from the right hand side... where as the sound of the people and the coughing seems to be coming from the back left hand side behind me
>
> 00:01:58 to 00:02:01
> I can't hear what they are saying, a young woman's voice

Both participants referred to the sources and actions of what they were listening to, with further information being provided as they tried to interpret what they heard.

The qualitative codes generated through the first pass resulted in 45 individual codes, or descriptors, such as 'action', 'quality' and 'mass'. These codes applied during a second pass resulted in the identification of 5658 instances of sound events described by the participants, ranging from 1318 instances of 'source' down to 1 instance of 'privacy'. The data is summarised in Table 5-20, the column headed Sum records how often a sound event described by a participant was encoded with the descriptor.

Frequency is the number of participants who contributed descriptions that provided the code, which is shown as a Percentage of the total number of participants (40). The Mean, Median, Mode and Standard Deviation represent a typical response, with the Rank being calculated according to the highest Frequency and Sum.

Code	Sum	Frequency	Percent	Mean	Median	Mode	Standard Deviation	Rank
Source	1405	39	98%	36	28	18	28.6	1
Action	870	39	98%	22.3	16	18	20.3	2
Vocalisation	754	39	98%	19.3	12	8	18.6	3
Location	677	37	93%	18.3	13	13	22.6	4
Onomatopoeia	475	37	93%	12.8	8	8	12.8	5
Material	108	33	83%	3.27	2	1	3.28	6
Estimation	174	32	80%	5.44	5	1	4	7
Quantity	150	32	80%	4.69	4	3	2.92	8
Dynamics	153	31	78%	4.94	3	2	5.31	9
Structure	110	31	78%	3.55	2	1	3.03	10
Temporal	123	30	75%	4.1	4	1	3.03	11
Context	121	30	75%	4.03	3	2	3.6	12
Direction	100	30	75%	3.33	2	1	3.2	13
Clarity	77	29	73%	2.66	2	3	1.59	14
Content	81	27	68%	3	2	2	2.86	15
Gender	134	26	65%	5.15	2	1	6.6	16
Reflection	82	23	58%	3.57	3	1	2.43	17
Environment	76	22	55%	3.45	2	1	2.96	18
Unidentified	53	22	55%	2.41	2	1	1.76	19
Comparison	44	15	38%	2.93	2	1	3.43	20
Sequence	18	15	38%	1.2	1	1	0.56	21
Generic	36	14	35%	2.57	2	1	2.03	22
Pollution	24	13	33%	1.85	1	1	1.34	23
Nationality	27	12	30%	2.25	2	1	1.6	24
Aesthetics	20	12	30%	1.67	1	1	0.98	25
Acoustics	13	11	28%	1.18	1	1	0.4	26
Force	34	10	25%	3.4	2	1	3.69	27
Spectral	19	10	25%	1.9	1.5	1	1.1	28
Recipient	17	9	23%	1.89	1	1	1.05	29
Reproduction	16	8	20%	2	1.5	1	1.31	30
Differentiation	11	8	20%	1.38	1	1	0.52	31
Unusual	9	7	18%	1.29	1	1	0.76	32
Man-made	9	6	15%	1.5	1.5	1	0.55	33
Mass	7	6	15%	1.17	1	1	0.41	34
Emotion	11	5	13%	2.2	1	1	2.68	35
Participant	9	5	13%	1.8	1	1	1.1	36
Health	5	4	10%	1.25	1	1	0.5	37
Immersion	5	4	10%	1.25	1	1	0.5	37
Age	6	3	8%	2	1	1	1.73	39
Regulations	2	2	5%	1	1	1	0	40
Orientation	2	2	5%	1	1	1	0	40
Familiarity	1	1	3%	1	1			42
Interest	1	1	3%	1	1			42
Masking	1	1	3%	1	1			42
Privacy	1	1	3%	1	1			42

Table 5-20: Qualitative codes derived from participants' descriptions

A variety of codes arose, with the most prevalent being the source, or the 'sound of what?' (Metz, 1985). Sources varied from the vague 'somebody' to the more precise inclusion of gender and age in 'young woman' detailed by only 2 of the 40 participants. Vocalisations such as 'speech', 'conversation' and even 'cough' formed the largest detailed group, which corresponded with Cole's (1996) description of children's preference for speech over non-speech sounds. Nationality and accents were identified, together with content, which was mostly generic 'saying what's what' and 'asking a question'. Emotional content was not confined to purely speech, 'pens being clicked in frustration' as well as 'nervous juggling of coins in pocket', 4 of the participants commented on the poor health of some of the inhabitants of the environment.

> P 1: xxxxxx.txt – 1:3 (4:4) (Super)
> Codes: [dynamics] [vocalisation]
> More voices... sort of half shouting... laughing... more talking
>
> P10: xxxxxxx.txt – 10:21 (12:12) (Super)
> Codes: [content] [location] [recipient] [source] [vocalisation]
> the demonstrator is explaining java to somebody behind me...
>
> P24: xxxx.txt – 24:14 (46:46) (Super)
> Codes: [age] [gender] [source] [vocalisation]
> I can hear a young girl laughing
>
> P29: xxxxx.txt – 29:18 (19:19) (Super)
> Codes: [action] [emotion] [force] [source] [temporal]
> the continuing stamping down on the delete key as someone gets frustrated
>
> P33: xxxx.txt – 33:25 (57:57) (Super)
> Codes: [gender] [location] [nationality] [source] [vocalisation]
> And there is a French guy speaking just behind me or on my left hand side

When specifying the source of the sound event, most participants were confident, even when they described it in generic terms. A quarter of the respondents did come across sound events that they could not identify, but this represented a very small amount of the total sound events compared to those listeners felt could either estimated or identified. Comparisons were made, such as the air-conditioning being 'like the sea coming from behind me' or 'a moving airstrip around me', but the majority of sources were identified by single words. Materials where described as being 'metal' 'paper', 'plastic', 'velcro' or 'wood' with the mass of the objects described as either 'heavy' or 'large', but never light or small.

P 5: xxxxx.txt – 5:39 (85:85) (Super)
Codes: [source]
Mobile phone

P11: xxxxxx.txt – 11:8 (14:14) (Super)
Codes: [action] [comparison] [environment] [location] [temporal]
continuous wearing and movement and the background noise is almost fanlike...a... factory like... conveyor belt like, as it is a very busy environment

P12: xxxx.txt – 12:19 (46:47) (Super)
Codes: [action] [estimation] [mass] [material] [onomatopoeia] [source]
I can hear a creaking sound like... Mmmm... or a creaking knocking sound as though somebody has sat on a table and was heavy and the wood was creaking

P25: xxxx.txt – 25:36 (71:72) (Super)
Codes: [action] [estimation] [source]
people opening up pencil cases or something like that or CD cases on the top of the desk

P31: xxxx.txt – 31:9 (6:6) (Super)
Codes: [action] [source] [unidentified]
na I can't identify the office equipment that keeps... like something revolving... am lets see what other noises amm...

Actions that generated the sound event were then described such as 'typing', or the onomatopoeic 'tapping'. Individual sound events were generally described only once until the event varied or a lack of new sources became evident, at which point the temporal aspect of whether it was 'constant' or had just 'stopped' were detailed. This varied when applied to vocalisations, which were mentioned, mostly, whenever heard, even from the same source, further reinforcing an apparent predilection for human speech.

P 2: xxxx.txt – 2:55 (34:34) (Super)
Codes: [gender] [location] [source] [temporal] [vocalisation]
that was the guy right in front of me... still speaking constantly...

P16: xxxxxxxxx.txt – 16:73 (138:138) (Super)
Codes: [action]
Movement

P21: xxxxxxxx.txt – 21:35 (94:94) (Super)
Codes: [source] [temporal] [vocalisation]
People talkin' again, it sounds like, everyone starts and stopped doin' the same thing at the same time.

P30: xxxxxxx.txt – 30:17 (31:31) (Super)
Codes: [onomatopoeia] [source]
I can hear chairs squeaking ..

P34: xxxxx.txt – 34:25 (37:37) (Super)
Codes: [action] [onomatopoeia] [temporal]
and there is an occasional knock...

Physical properties such as dynamics and spectrum both featured, with dynamics, despite being mentioned the most, being predominantly confined to 'loud' which translated into the inferred force of the action such as 'hitting the keyboard hard'. Silence was only mentioned by its absence: 'It's a constant noise... there is no silence...', which as Cage (1973) discovered does not exist outside a vacuum, even in an anechoic chamber. Quiet sounds were rarely mentioned as being quiet, dynamics were mostly considered when they became 'loud'. Spectral referred mostly to voices with the limitations of 'deep' or 'low' and the less frequent 'high' or 'higher'.

P 3: xxxxxxxx.txt – 3:6 (16:16) (Super)
Codes: [dynamics] [temporal]
It's a constant noise... there is no silence...

P 9: xxxxxx.txt – 9:19 (42:42) (Super)
Codes: [dynamics] [reflection]
I am just aware of a high level of noise activity but nothing too distracting am... no this time of day it is usually quiet... quite busy... in here...

P18: xxxxxx.txt – 18:2 (9:9) (Super)
Codes: [estimation] [onomatopoeia] [source] [spectral]
There is a sort of rumble, which might be rain... Ah... low frequency content certainly, could be air-conditioning... could be rain...

P27: xxxxxxxxx.txt – 27:32 (37:37) (Super)
Codes: [dynamics] [source] [vocalisation]
People talking very loud

P39: xxxxxxx.txt – 39:60 (74:74) (Super)
Codes: [location] [spectral]
The elements of the background noise, they are changed from being such a low drone to being a higher pitched drone... it varies.

Clarity was referred to in terms of ‘distinct’ or ‘muffled’ with participants not being able to ‘make out’ the speech of the recordings, which a few found ‘annoying’. Differentiation between sound sources did occur, more by default rather than by considered identification. Quantities of sound sources were identified between one and four, otherwise it was a generic ‘few’ or ‘lots’. Only 23% of the participants referred to the receiver of a sound, in all cases the event was speech, with a single reference to masking ‘it drowns out the sound of people talking... well almost...’. Aesthetics were rarely mentioned, and were mostly negative such as ‘bland’, ‘drone’ and ‘monotonous’, with spectral aspects referred to as being ‘hard’ or ‘sharp’.

P 6: xxxxxxx.txt – 6:15 (34:34) (Super)
Codes: [aesthetics] [context] [source] [vocalisation]
Instructor’s voice from a tutorial or something... sounds very flat... a drone

P13: xxxxx.txt – 13:16 (47:47) (Super)
Codes: [clarity]
I can’t really make out any real words or comments but I’m certain that I could make out distinct words and sentences when I first sat down

P15: xxxxxx.txt – 15:29 (70:70) (Super)
Codes: [gender] [quantity] [recipient] [source] [vocalisation]
It sounds like a conversation between two... male individuals

P38: xxxxxx.txt – 38:16 (40:40) (Super)
Codes: [action] [differentiation] [source]
again the sounds of people typing into separate keyboards... different people.

P41: xxxxx.txt – 41:84 (235:235) (Super)
Codes: [clarity] [location] [source]
Someone’s, your man’s still on the phone, ahm... somewhere in my front left and quite near me, but can’t make out the conversation

The majority of sound source locations were described in terms of the relationship to the participant’s point of listening (POL). Spatial attributes were commonly detailed in terms of 'left', 'right', 'front' and 'back' with occasional generic references to distance, 'I'm starting to recognise the sounds constantly coming from the top right from my point of view somebody has just rolled over with their chair along rails in cluster one...'. A few participants specified height both in the physical environment and on the recording 'I'm getting some noise above me to the right...'. The recording had no height channel. Individuals were described as 'walking up and down steps' or 'walking by', or even moving from 'left to right'. Whilst descriptions were always generic, they illustrated awareness of moving objects rather than a static auditory environment, 'there is a bag of crisps flying around... it started on the front left and then went all the way to the back left...'.

> P 7: xxxx.txt – 7:21 (32:32) (Super)
> Codes: [direction] [source] [structure]
> Someone going to the stairs…
>
> P 8: xxxx.txt – 8:25 (64:64) (Super)
> Codes: [content] [direction] [gender] [location] [orientation] [source] [temporal] [vocalisation]
> Somebody with an Irish accent up to my right up the stairs... sounds like he might be coming towards me... nope must have turned to face me for a wee while
>
> P14: xxxxxx.txt – 14:24 (52:52) (Super)
> Codes: [direction] [location] [material] [onomatopoeia] [source]
> the rustle of paper to my left... very close now and behind me moving off to the right
>
> P23: xxxxxxx.txt – 23:6 (13:13) (Super)
> Codes: [action] [direction] [location] [source]
> I'm next to the stairs, so you can hear the people walking up and down as well.
>
> P36: xxxxxxx.txt – 36:6 (19:19) (Super)
> Codes: [action] [comparison] [estimation] [location]
> What is a sound on a... Diagonal in front and behind me... it sounds a bit like typing but not quite.

Context was occasionally described in some detail such as 'I can tell you that someone is pressing the key... and I can imagine that by the rhythm of their fingers when they press return or press space' or as a sequence of events 'checking of keys in their pocket in their left pocket... a checking of a mobile phone... turning it on probably picking up of a bag... of papers stuffing them in... zipping up the bag'. The environment was described in terms of its size, 'large' 'open plan space' with two participants guessing the original location and others going for either a computer lab or open plan office. When referring to the physical structure, participants detailed: 'door', 'floor', 'grating', 'rails', 'steps' with one participant who experienced the unidentified recording describing a 'high ceiling' with 'plaster walls'. Echoes were described when establishing the room size with sound 'pinging off the pillars'.

P 4: xxx.txt – 4:103 (85:85) (Super)
Codes: [action] [source] [structure]
jacket hitting off the wall..

P14: xxxxxx.txt – 14:38 (88:88) (Super)
Codes: [acoustics] [action] [material] [source] [structure] [temporal]
Occasionally I'm getting footsteps that sound as if they are on a hollow floor like a stage or a wooden surface away from me certainly on something with some resonance to it

P17: xxxxx.txt – 17:46 (110:110) (Super)
Codes: [action] [context] [estimation] [material] [source]
Moving paper or books out of the bag or into the bag

P20: xxxxxxxxxx.txt – 20:7 (15:15) (Super)
Codes: [content] [context] [source] [vocalisation]
Referring several okays and finishing base saying thank you so he has been communicating with somebody... and proceeding to an operation with this person... maybe transmitting information

P33: xxxx.txt – 33:16 (36:36) (Super)
Codes: [acoustics] [environment]
There is a fair amount of reverberation coming from the back so like... I presume there would be a big open space in my back...

P37: xxxxx.txt – 37:1 (5:5) (Super)
Codes: [environment]
a, sounds to me like in an office like in a large open plan office... ya...

Privacy was only considered by one participant ‘conversation private really...’, whereas pollution in terms of distraction and annoyance was more evident ‘it’s really quite annoying actually... I don’t particularly like this environment.’ Five of the participants referred to sounds that they generated themselves ‘I hear myself talking out loud...’ illustrating how listeners contribute to their own soundscapes. Immersion was detailed through comments such as ‘I’m really beginning to think that I am sitting in the office and not sitting in a dark room’ and ‘I think if I had eyes I would have turned around to have a look to see who it was’, were made by participants blindfolded listening to the recorded soundfield.

> P 2: xxxx.txt – 2:30 (22:22) (Super)
> Codes: [gender] [pollution] [source] [temporal] [vocalisation]
> the guy that is speaking non-stop is very distracting...
>
> P12: xxxx.txt – 12:45 (104:104) (Super)
> Codes: [environment] [immersion]
> Now I’m really beginning to think that I am sitting in the office and not sitting in a dark room
>
> P13: xxxxx.txt – 13:3 (11:11) (Super)
> Codes: [clarity] [immersion] [reproduction]
> Over time it is becoming very unclear as to whether people talking is coming through the speakers or whether they are actually here
>
> P22: xxxxx.txt – 22:6 (16:16) (Super)
> Codes: [action] [onomatopoeia] [pollution] [source]
> apart from the clicking of typing... am... keyboards which I find really really annoying...
>
> P28: xxxxxxxxx.txt – 28:4 (4:4) (Super)
> Codes: [action] [participant] [source] [temporal]
> me occasionally when I scratch my nose...

When comparing the responses from the four different conditions there are some differences. For participants who could see, the responses are similar, suggesting that the reproduction was comparable to the physical JKCC. The only notable difference were references to the environment. No reference was made to the environment by those physically present in the JKCC. When considering the effect of being blindfolded, there was a reduction in the number of sound events described. The number of instances of source, action, vocalisation, location and onomatopoeia mentioned by participants who were physically present in the JKCC, is in some cases more than double the number of events described by those in the recorded conditions. There was little difference in the number of respondents who mentioned terms contained within the codes, only the total sum. Participants in the recorded conditions appeared to be concentrating on working out what the environment was that they were listening to. Listeners in the physical JKCC appeared to be recounting what they could hear and matching it to what they had previously seen. Outside these two major differences everything else is similar (see Figure 5-4).

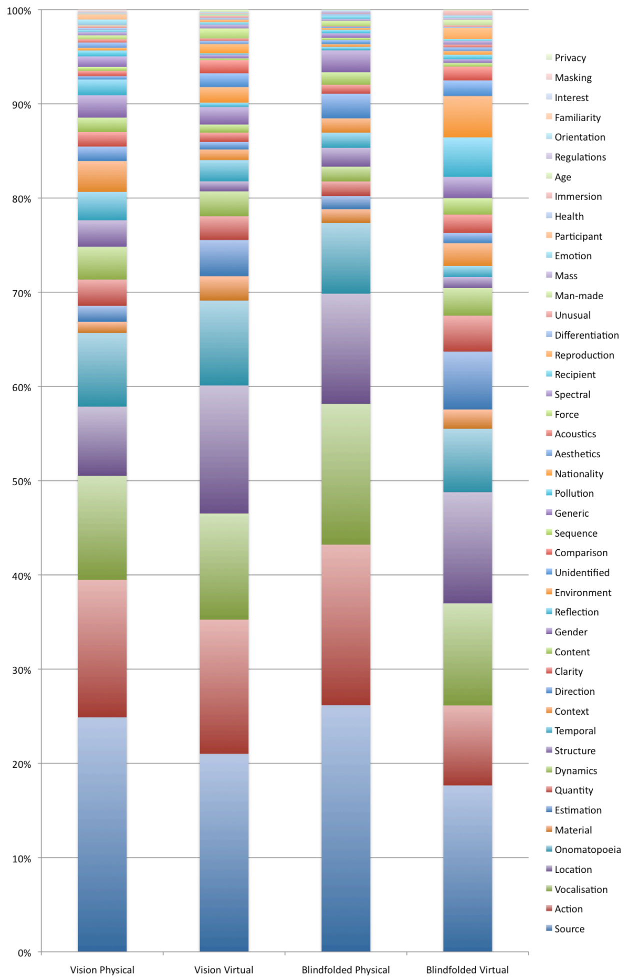


Figure 5-4: Relative frequency of codes as a function of environment

Once the codes had been finalised, a lexicon of individual words that were used by the participants in describing their auditory environments was classified. A first pass was made of the transcriptions, making note of individual words that were used to describe sound events as well as their frequency. No attempt at this stage was made to classify individual words other than ensuring that words such as ‘dropped’ referred to an action rather than a drop in volume or some other interpretation.

The sum, frequency and percentage response rate were then calculated for each experimental condition and then combined, with the results ranked as shown in Table 5-21. The original 45 codes were reduced to 15. A total of 6587 words were used with a frequency of 2702 unique descriptive words for each of the 40 participants, this resulted per word in a mean of 2.438, a median and mode of 1, and a standard deviation of 4.242. Codes ranged from source and actions, which were referred to by every participant, through to architectural acoustics, which were only mentioned by 20% of the participants.

Code	Sum	Unique words	Response	Mean	Median	Mode	SD	Rank
Source	2448	938	100%	2.6	1	1	4.6	1
Actions	2029	254	100%	2.5	1	1	4.3	2
Spatial	1093	312	88%	3.5	2	1	6.0	3
Dynamics	200	129	80%	1.6	1	1	1.1	4
Onomatopoeia	174	117	75%	1.3	1	1	1.0	5
Temporal	124	86	73%	1.4	1	1	1.1	6
Quantity	130	77	68%	1.7	1	1	1.1	7
Clarity	62	43	53%	1.4	1	1	1.5	8
Comparison	107	25	48%	4.3	2	1	5.7	9
Aesthetics	51	39	45%	1.3	1	1	0.8	10
Material	61	36	40%	1.7	1	1	1.2	11
Spectral	49	34	35%	1.4	1	1	0.9	12
Emotions	25	24	35%	1.0	1	1	0.2	13
Pollution	23	17	28%	1.4	1	1	0.8	14
Arch Acoustics	11	10	20%	1.1	1	1	0.3	15

Table 5-21: Classification of lexicon

Source and actions were the most common terms when describing the sound events, which the participants heard (see Figure 5-5). Source and action were both present in 100% of the responses, with source being mentioned more often than actions. Sources ranged from a single reference to a named individual, through to the more generic ‘bloke’, which retained gender and quantity, ‘somebody’ was applied the most for a single source and 'people' for sources that could not be separated. The use of generic sources such as 'something' were the most common, even by the group that could see what the sources were.

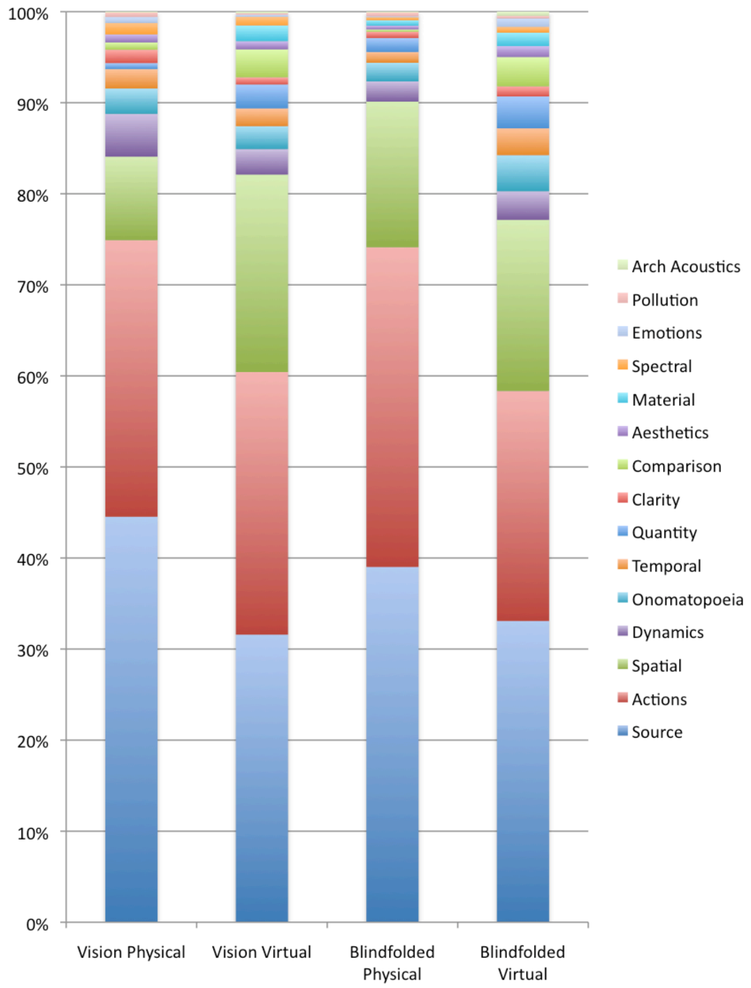


Figure 5-5: Relative frequency of words as a function of classified attributes and environment

Spatial attributes were the third most common attribute of the sound event mentioned, with an 88% response rate, most sound sources being located 'left', 'behind', 'right' and finally 'front'. Participants experienced sounds as coming from behind them almost twice as often as those coming from in front. This corresponds with Murch (2003) who theorised that when sounds above a certain level are heard from behind, then sounds from the front are suppressed as an innate response to a perceived ‘threat’, so sounds behind a listener are more perceptually salient due to inbuilt survival mechanisms. Spatial aspects were less important for those who could see and more important for those who could not see, with both blindfolded groups having a 100% response rate compared to 70% (sighted, present) and 80% (sighted, recorded) for the sighted participants.

Dynamics such as ‘loud’ and ‘louder’ were slightly more common than onomatopoeia: here the sighted groups mentioned them more than the blindfolded, although dynamics were only referred to a few times by each participant. There was a wide range of onomatopoeic words with ‘creaking’ being the most common, with the blindfolded groups having referred to onomatopoeic words more often than the sighted.

Quantity and temporal attributes were both generic. With ‘lots’ and ‘continuous’ occurring the most. The remaining attributes clarity, material, spectral, emotions, pollution and architectural acoustics were rarely referred to when compared to source and action, typically by a factor of approximately 40:1. However the results illustrate that some of the participants were aware of attributes associated with musical listening, as well as provide an insight into the terms used, which in the case of spectral were predominately ‘low’, ‘deep’ or ‘high’. Technical terms such as kHz had no place in listeners’ responses, with only a single participant referring once to frequency. Material was not mentioned at all by the sighted group within the physical environment, and the same group only referred to architectural acoustics once.

Only a single participant, who was listening to the recording blindfolded, mentioned all of the attributes, with two more listeners expressing 14 attributes. At the other end of the scale, one participant referred only to sources and actions, with a further two adding spatial characteristics. Otherwise participants used seven to eight of the attributes. Source was slightly more prevalent for all four experimental groups than actions.

The results might indicate the importance of source and action when describing sound, but they might also show that participants were aware of other attributes and had a broad vocabulary with which to describe them. Overall, responses varied in quantity and quality. The most basic was a series of sound events without sources or locations, ‘talking… walking… talking… talking… talking… walking…’. The other extreme provided detailed information about both the sound sources and their context, ‘Somebody is sitting in front of me and I can hear the typing quite clearly... he types quite strongly when he used the mouse I think ... the space on it.’

A set of attributes that this group of listeners used to describe what they were listening to has been identified. These attributes will be compared, in the next section, to the audio professionals’ responses in order to establish a set of attributes that are important to both designers and listeners.

5.3 Discussion

In order to establish a set of attributes for describing sound that are meaningful to both designers and listeners it was necessary to survey both groups. The designers provided information about attributes that they used to describe sound by completing a questionnaire. Listeners’ concurrent verbalisations were coded and classified in order to establish which attributes of sound were meaningful to them. The two studies presented in this chapter involved two different methods, and as such, comparisons have to be carried out with care.

When the attributes for describing sound are compared for the listeners and the audio professionals the most common attribute shared by both groups was dynamics (see Table 5-22). The listeners referred to dynamics in terms of high or low, whereas the professionals were more interested in the scale whether it was volume, loudness or level. Across all three professional groups dynamics had an 81% frequency response with ‘volume’ being the most common term with a 45% frequency overall.

Onomatopoeia, whilst having the third highest frequency in the concurrent verbalisations, only had a 5% overall frequency with the professionals.

Spectral terms were detailed in the same manner as dynamics, listeners again using high or low, with the professionals concerned with pitch, timbre, tone and frequency. Spectral terms had a 59% overall frequency by the professionals, with pitch having the third highest frequency of 17%. Whereas in the concurrent verbalisations spectral attributes were only mentioned by 10 of the participants with pitch only being referred to twice. The spatial attributes of a sound source were an important factor to the listeners with them detailing the source in reference to their own location. However, only 27% of the professionals referred to spatial aspects, with an even split between individual sounds and the soundfield as a whole.

Atrributes	Listeners	Professionals
Dynamics	80%	81%
Spectral	35%	59%
Aesthetics	45%	51%
Clarity	53%	48%
Architectural Acoustics	20%	29%
Spatial	88%	27%
Temporal	73%	17%
Onomatopoeia	75%	5%
Source	100%	0%
Actions	100%	0%
Quantity	68%	0%
Comparison	48%	0%
Material	40%	0%
Emotions	35%	0%
Pollution	28%	0%
Perceptual	0%	27%
Type	0%	19%
Reproduction	0%	17%
Musical	0%	7%
Hearing Abilities	0%	3%

Table 5-22: Comparison of attributes used by listeners and audio professionals

Fifty-one percent of the professional respondents referred to aesthetics, with brightness, harshness and warmth being the most common. Only 12 of the listeners referred to aesthetic aspects describing them typically as sharp or flat, nice or bad. Clarity was highlighted by 48% of the audio professionals, listeners were mainly concerned with making out speech, whereas professionals were more concerned with quality and definition.

In most areas the professionals applied a higher level of granularity, except when it came to sound source identification. Audio professionals' classification of type, which in itself only had a 19% frequency, was concerned with noise, artificial and natural. Listeners detailed the source more than any other factor, the only exception being when they found the source unidentifiable. While a few sound events were referred to as mechanical, there was no reference to a sound event being natural, and noise was referred to in terms of a distraction.

The audio professionals referred to noise as sounds with particular spectral properties or related to artefacts, or unwanted sounds. The professionals' approach contrasts with the way listeners predominately used noise to refer to an unidentified source. Some similarities were also noticed. Both groups predominately made aesthetic judgements in negative terms. This might be due to the environment that the listeners were asked to describe not offering much to be aesthetically enthusiastic about. On the other hand, the fact that professionals submitted a majority of negatively aesthetic terms might suggest that both listeners and designers might be more effective at communicating negative rather than positive experiences of auditory environments.

Both groups described room acoustics using similar terms, principally referring to reverberation or echo. No noticeable knowledge gap was visible in this area. Clarity judgements were consistently made on a binary scale in both groups. For example, the professionals used terms such as rough, smooth, transparent, muffled, dirty and clean. The listeners described events in a comparable fashion, without any moderating adverbs. The professionals referred to temporal aspects predominately in terms of pace and timing. Constancy of the sound event concerned the listeners, specifically whether a sound was continuous or intermittent.

Both emotions and pollution were referred to by the listeners, but not by the audio professionals. The lack of reference to pollution by the professionals might be due to the nature of their work. Practitioners commonly work in acoustic isolation: a sound designer for an interface might not have to routinely consider the auditory environment into which their work will be experienced. In contrast, listeners cannot easily isolate themselves to the same degree, even when using headphones.

Emotional responses are the mainstay in music and to a lesser extent sound design for the entertainment industry, but are rarely formally analysed, being confined to an individual's experience. Emotional content was mentioned by 35% of the listeners and included, to varying degrees all of the six basic emotions as identified by P. Ekman and Friesen (1986): surprise, anger, sadness, disgust, fear and happiness. The predominant terms related to happiness, followed by fear.

The data reported in this chapter addressed the first research question identifying attributes that are important to both sound designers and listeners when describing sound. While the approaches to data gathering and the populations were different the results can be used to inform the design of the soundscape mapping tool. The next chapter discusses the soundscape mapping tool and provides an illustration of its use during the evaluation of a simple in-car auditory display.

6 Soundscape mapping tool

This chapter describes the soundscape mapping tool (SMT), and provides an illustration of the SMT's use during the design and evaluation of an in-car auditory display. The results from the pro audio questionnaires and the concurrent verbalisations were employed to establish attributes for describing sound that are important to sound designers and listeners. This chapter addresses how a soundscape could be classified and visualised so that it is meaningful to a designer. In addition the chapter provides an example of how the SMT could be used by a designer to compare their intentions for a sound design with the experiences of listeners.

Watson and Sanderson (2007) stated that an auditory display's effectiveness at communicating information should be evaluated according to its context of use. The context of use could be studied by questioning listeners about their perceived soundscapes. Visualised results would then be passed to the designer for reference purposes. During the design process the designers might consider what they want listeners to experience, and in the process create their own soundscape map. Finally, listeners would experience the auditory elements in situ or within a simulated environment with new maps being created. A comparison of the maps could illustrate when designers' intentions and listeners experiences match, as well as highlight what impact new sound events have on pre-existing auditory environments.

This study illustrates the soundscape mapping tool in use during the design and evaluation of an in-car auditory display. This was an audio only device with no visual display. The auditory display was designed to convey information about the car's braking distance, vehicles overtaking in the driver's dead angle, and incoming emails. The study concentrated on the use of the SMT during the contextual design and evaluation of appropriate auditory cues. For this illustration a small petrol engine car, (Peugeot 205), travelling through a city centre, at rush hour, with speech radio was chosen. A moving car represented a different challenge from the office environments used in earlier studies reported in this dissertation.

6.1 Method

The SMT contained three distinct phases: capture, classification, and visualisation. Capture involved the researcher creating a schematic of the car, recording the soundfield, and noting the sound events. The designer and 10 listeners recruited from Edinburgh Napier University listened to the recordings and classified the sound events that they were aware of. The results were then visualised by the researcher in the form of 33 soundscape maps.

6.1.1 Participants

The designer was Dr Grégory Leplâtre, from Edinburgh Napier University's School of Computing. Dr Leplâtre specialised in the design and evaluation of non-speech sounds in mobile computing devices and had previously collaborated with the author.

Ten listeners were self-selected from staff and students within Edinburgh Napier University, and as such were a sample of convenience. Potential candidates were approached via e-mail. Each of the listeners was familiar with the inside of a car and with driving, and had no known hearing impairments. Participants varied with respect to age, sex and background. All participants took part in the study on a voluntary basis and had a good command of spoken English. Participants were informed of the nature of study and were assured that their responses would be anonymised to prevent individuals being identified. Verbal permission was obtained to make use of the results for this dissertation and associated publication. Participants were informed that the experiment would last between 45 - 60 minutes and that they could end the task at any point. Participants were also given the opportunity to ask questions after the experiment.

6.1.2 Apparatus and materials

The apparatus for this study included a car, multi-channel audio recording equipment and a multi-channel audio reproduction system. The materials generated for this study included a vehicle schematic, surround sound recordings with a list of the sound events, an auditory display with a description, a method of classification, a method of visualisation, and a set of listener guidelines.

Apparatus

The private car was a Peugeot 205 GR that belonged to the author. The vehicle was a five door hatchback constructed in 1989 and had a 1360 cc petrol engine with in excess of 133,000 miles on the odometer (Carfolio.com, 2010).

For the recording eight Audio-Technica AT803b omnidirectional condenser lavalier microphones were used in Audio-Technica AT8418 UniMount Instrument mounts (Audio-Technica, 2010a, Audio-Technica, 2010b). The microphones were set to a flat frequency response. Microphones signals were fed to four Sony TCD-D8 DAT (Digital Audio Tape) recorders set to record at 16 bit 48kHz (Sony, 1995). An Extech 407735 Dual Range Sound Lever Meter was used for calibration set to slow sensitivity peak dB C level (Extech Instruments, 2010). The DAT recordings were digitally transferred to a Pro Tools 7.4 LE system running at 24 bit 96 kHz (Digidesign, 2008).

For the reproduction eight Genelec 8030A bi-amplified monitors supported on microphone stands were connected to four Genelec 7070A active subwoofers (Genelec, 2010a, Genelec, 2010b). The bass roll-off was set to 85 Hz on the 8030As to complement the low pass filter that was set to 85 Hz on the 7070As. These self-powered loudspeakers were fed directly from an Avid 003 Rack interface that was connected to an Apple MacBook running Pro Tools LE 7.4 set to 24 bit 96 kHz (Avid, 2010, Apple, 2010, Digidesign, 2008).

Schematic

The dimensions of the car were measured in order to create an aerial schematic at a scale of 20:1 using Adobe's *Illustrator* software (Adobe, 2009b). A grid was overlaid on to the schematic, with each cell representing 20 cm by 20 cm. Three rows of additional cells were added around the perimeter, in order to allow the inclusion of sound events that occurred outside of the vehicle (see Figure 6-1). The grid was numbered in the same way as an ordnance survey map, beginning at the bottom left with 0 0 and finishing at the top right with 21 13. The car was orientated so that the front started at column 3 and the rear finished at column 18.

0 13	1 13	2 13	3 13	4 13	5 13	6 13	7 13	8 13	9 13	10 13	11 13	12 13	13 13	14 13	15 13	16 13	17 13	18 13	19 13	20 13	21 13
0 12	1 12	2 12	3 12	4 12	5 12	6 12	7 12	8 12	9 12	10 12	11 12	12 12	13 12	14 12	15 12	16 12	17 12	18 12	19 12	20 12	21 12
0 11	1 11	2 11	3 11	4 11	5 11	6 11	7 11	8 11	9 11	10 11	11 11	12 11	13 11	14 11	15 11	16 11	17 11	18 11	19 11	20 11	21 11
0 10	1 10	2 10	3 10	4 10	5 10	6 10	7 10	8 10	9 10	10 10	11 10	12 10	13 10	14 10	15 10	16 10	17 10	18 10	19 10	20 10	21 10
0 9	1 9	2 9	3 9	4 9	5 9	6 9	7 9	8 9	9 9	10 9	11 9	12 9	13 9	14 9	15 9	16 9	17 9	18 9	19 9	20 9	21 9
0 8	1 8	2 8	3 8	4 8	5 8	6 8	7 8	8 8	9 8	10 8	11 8	12 8	13 8	14 8	15 8	16 8	17 8	18 8	19 8	20 8	21 8
0 7	1 7	2 7	3 7	4 7	5 7	6 7	7 7	8 7	9 7	10 7	11 7	12 7	13 7	14 7	15 7	16 7	17 7	18 7	19 7	20 7	21 7
0 6	1 6	2 6	3 6	4 6	5 6	6 6	7 6	8 6	9 6	10 6	11 6	12 6	13 6	14 6	15 6	16 6	17 6	18 6	19 6	20 6	21 6
0 5	1 5	2 5	3 5	4 5	5 5	6 5	7 5	8 5	9 5	10 5	11 5	12 5	13 5	14 5	15 5	16 5	17 5	18 5	19 5	20 5	21 5
0 4	1 4	2 4	3 4	4 4	5 4	6 4	7 4	8 4	9 4	10 4	11 4	12 4	13 4	14 4	15 4	16 4	17 4	18 4	19 4	20 4	21 4
0 3	1 3	2 3	3 3	4 3	5 3	6 3	7 3	8 3	9 3	10 3	11 3	12 3	13 3	14 3	15 3	16 3	17 3	18 3	19 3	20 3	21 3
0 2	1 2	2 2	3 2	4 2	5 2	6 2	7 2	8 2	9 2	10 2	11 2	12 2	13 2	14 2	15 2	16 2	17 2	18 2	19 2	20 2	21 2
0 1	1 1	2 1	3 1	4 1	5 1	6 1	7 1	8 1	9 1	10 1	11 1	12 1	13 1	14 1	15 1	16 1	17 1	18 1	19 1	20 1	21 1
0 0	1 0	2 0	3 0	4 0	5 0	6 0	7 0	8 0	9 0	10 0	11 0	12 0	13 0	14 0	15 0	16 0	17 0	18 0	19 0	20 0	21 0

Figure 6-1: Aerial schematic of car with grid, red = bodywork, blue = seats

An A4 rotated version of the car schematic was created for the participants to refer to. The rotation made it easier for listeners to relate to their orientation within the car, negating the need to translate coordinates. The front and rear of the car was labelled to reduce confusion (see Appendix B).

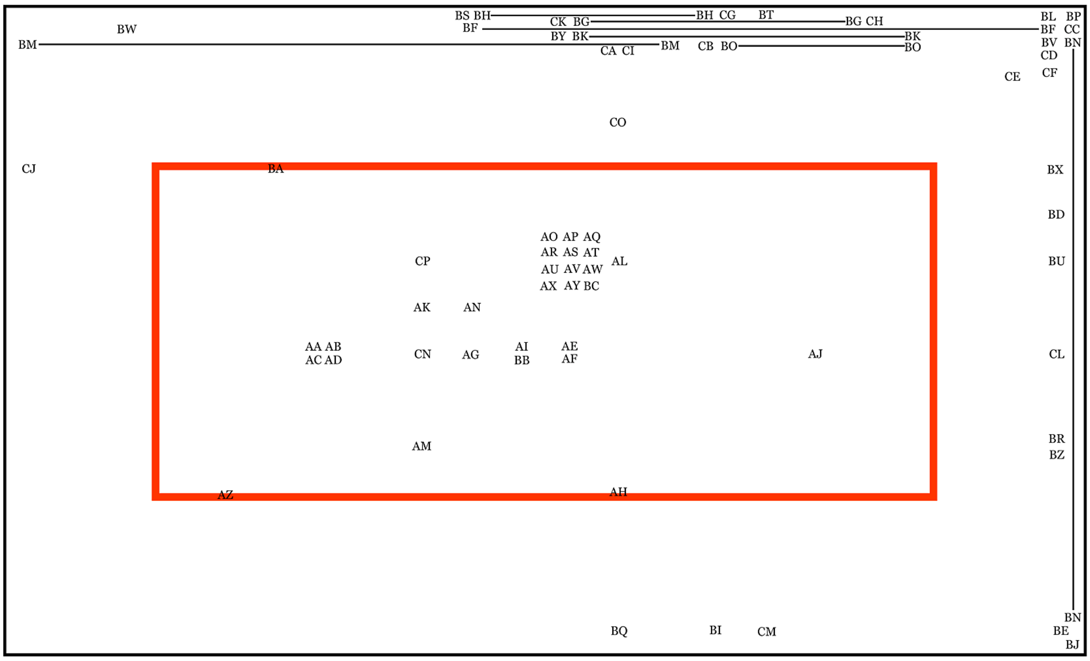
Surround sound recording

A 5 minute recording was made of the car driving through Edinburgh's city centre at rush hour, in order to provide a consistent soundfield for listeners to experience. Within the 5 minutes of audio recording that was transcribed by the author, 157 separate audible events were noted. The 157 sound events were identified as having been generated by 49 different sources. Grouping the sound events together according to their source and the event made it possible to reduce the total number of sound events to 65.

There were 28 audible sound events generated by the vehicle being studied, 28 passing vehicles, 5 people, a dog and some scaffolding. Within the car, the engine passed through different states (idling, accelerating, cruising and decelerating). Additional sound events were captured such as engaging and releasing the handbrake, changing gear and a wide range of vibrations. There were 11 distinct types of sound from the radio, these were split into speech, music and laughter. Outside the car 28 different vehicles were notated: 1 ambulance, 4 busses, 17 cars, 3 taxis, 1 truck and 2 vans. Sound events included a siren, vehicle passes, brake squeals, indicators and windscreen wiping. The remaining sound events included screaming, talking, rustling of clothes, barking and scaffolding being struck.

With regards to the spatial cues, all of the sound events associated with the car could be identified to specific points within the outline of the car, as shown in red in Figure 6-2. The majority of the passing vehicles were located on the driver's side, (top of the map), whilst most of the stationary vehicles were found to the rear of the car (right of the map). There were few sound events on the passenger's side and in front. The paucity of sound events on the passenger's side might be partially explained by the comparatively low volume level of sounds on the pavement, when compared to the much louder vehicles. The shortage of audible sound events at the front of the car is possibly due to masking associated with the car's engine, which was constantly running throughout the recording.

The recorded sound events were augmented with three sound events that formed a simple Auditory Display. Three versions of the audio recording were created for classification by the designer and the listeners. The first version was the recording from within the car (205, the car's model number), the second was the auditory display (AD) on its own, and the third was the augmented car recording (205 + AD).



Code	Event	Source	Code	Event	Source	Code	Event	Source
AA	Engine Idle	Engine	AX	Group Laughter	Radio	BU	Brake Squeal	Car 11
AB	Engine Accelerate	Engine	AY	Mobile Phone Interference	Radio	BV	Brake Squeal	Car 12
AC	Engine Cruise	Engine	AZ	Bump	Front Left Wheel	BW	Brake Squeal	Car 13
AD	Engine Decelerate	Engine	BA	Bump	Front Right Wheel	BX	Brake Squeal	Bus 3
AE	Handbrake Released	Handbrake	BB	Bumps	All wheels	BY	Brake Squeal	Bus 4
AF	Handbrake Engaged	Handbrake	BC	Clothes Rustle	Driver	BZ	Brake Squeal	Taxi 2
AG	Gear Change	Gear Stick	BD	Windscreen Wiper	Car 1	CA	Brake Squeal	Taxi 3
AH	Door Lock Release Rattle	Door Lock Release	BE	Wiper Blade	Windscreen Wiper	CB	Horn	Car 14
AI	Rattle	Sunroof	BF	Passing Car	Car 2	CC	Indicator	Car 8
AJ	Rattle	Parcel Shelf	BG	Car Pass	Car 3	CD	Indicator	Car 15
AK	Rattle	Dashboard	BH	Car Pass	Car 4	CE	Indicator	Car 16
AL	Seat Creak	Driver's Seat	BI	Car Pass	Car 5	CF	Siren	Ambulance
AM	Indicating	Indicator	BJ	Car Pass	Car 6	CG	Scaffolding Strike	Scaffolding
AN	Indicator Off	Indicator Switch	BK	Car Pass	Car 7	CH	Scaffolding Strike	Scaffolding
AO	Male 1 Speech	Radio	BL	Car Pass	Car 8	CI	Bark	Dog
AP	Male 2 Speech	Radio	BM	Passing Bus	Bus 1	CJ	Child Scream	Child
AQ	Female 1 Speech	Radio	BN	Bus Pass	Bus 2	CK	Child Scream	Child
AR	Female 2 Speech	Radio	BO	Van Pass	Van 1	CL	Female Speech	Woman
AS	Female 3 Speech	Radio	BP	Engine Rev	Car 9	CM	Male Speech	Man
AT	Male Chant	Radio	BQ	Engine Rev	Van 2	CN	Braking distance	Auditory Display
AU	Group Chant	Radio	BR	Engine Idle	Car 10	CO	Dead angle	Auditory Display
AV	Music 1	Radio	BS	Engine Idle	Taxi 1	CP	Message	Auditory Display
AW	Music 2	Radio	BT	Air Brakes	Truck			

Figure 6-2: Outline of car showing relative positions of notated sound events

The combined list of sound events was printed out on a single landscape A4 sheet in table form. Additional cells were included in the table to record the designer's and the listeners' responses when classifying each sound event (see Appendix C).

Auditory Display

The designer created a simple auditory display to communicate the braking distance, dead angle and receipt and response to an email. The designer provided the sound events as three separate audio files that he later imported into the sound recording choosing appropriate intervals, dynamic levels and spatial locations.

The designer also provided a description of each sound event within the auditory display for the listeners to refer to (see Appendix D). There were three sound events: braking distance, dead angle and message. The braking distance sound event was 'triggered when the car is getting too close to the vehicle in front.' The sound event had an 'increasing pitch [to] represent the diminishing distance between the two vehicles'. The dead angle was 'triggered when a vehicle overtaking the 205 (on the right) is in the driver's dead angle'. 'The intensity of the [dead angle] sound [event] increases as the car approaches the dead angle'. The message 'sound [event] represents a sequence of action[s] by the driver receiving a new email message'. 'The system notifies the driver of a new message. The driver presses as key to hear the subject. The system reads out the subject. The driver presses a key to hear the message. The system reads out the message body. The driver interrupts the system by pressing the delete key. The system confirms the message deletion. The system confirms that there are no other messages in the user's inbox.'

Classification

Both the designer and the listeners were provided with an A4 printed version of the attributes to aid them during the classification of the audio recordings (see Appendix F). The classification was generated based on a comparison of the findings from the audio professionals' survey and the listener concurrent verbalisations, as well as from the literature review. The classification differs from the one developed during the preliminary studies. All of the four measurements (spatial, temporal, spectral and dynamics) have been replaced by classifications so that the maps represent listeners' soundscapes rather than measured soundfields. The sound type, interacting materials and interactive functions have been simplified into type, material, interaction and content. The acoustical information has been replaced by clarity. Only emotions have been retained from the original information category and have been expanded.

Aesthetics were not part of the prototype and have been added. Listeners' responses from the concurrent verbalisations were used to identify attributes rather than the interviews conducted in the preliminary studies. There were two reasons, the first was the increased number of participants, 40 took part in the concurrent verbalisations, compared to only 18 for the interviews. The second reason was that the codes from the concurrent verbalisations were confirmed through independent reading, whereas another party had not confirmed the codes for the interviews.

The classification took the form of 11 distinct attributes, 10 of which had 3 states. HEAD acoustics (2006) recommend that, when conducting listening tests, no more than 12 dimensions be included in order to help prevent participants losing their concentration. Three point scales were chosen in order to keep the number of variables to a minimum, whilst still allowing participants to choose mid point values. Spatial attributes were noted using coordinates so that a wider range of responses was possible.

The first attribute was spatial (see Table 6-1). Both the audio professionals and listeners used spatial attributes when describing sound events (see Table 5-22). Audio professionals employed terms like envelopment and width, professionals were also concerned with direction and distance. When sound designers alter the spatiality of sounds they predominantly use two forms of manipulation: panning and reverberation. Panning can be used to control the left to right orientation of a sound event in relation to the listeners and reverberation can provide depth cues (Kerins, 2010, Marks, 2009, Rose, 2008). Listeners related whether a sound event was behind them, to their left, right or in front. Listeners also referred to their proximity to a sound event. Coordinates were chosen as the method for capturing spatial attributes, as this is the standard approach within cartography, which is considered an effective method for communicating spatial relationships (Monmonier, 1993).

For the sound type, choices were confined to *speech*, *music* or *sound effect*, with the last representing all sounds that are neither *speech* nor *music*. Only the audio professionals specified the type of sound. Within traditional media industries, the audio production pipeline is broken down according to dialogue, music and sound effects (K. Collins, 2008, Holman, 2005, Whittington, 2007). Listeners' concurrent verbalisations included some descriptions of sound events as either speech or music but they did not use the term sound effect.

Type	Category
Speech	Spoken language
Music	Performed composition
Sound effect	Audible events and actions
Material	Matter
Gas	Airborne
Liquid	Fluids
Solid	Objects
Interaction	Action
Impulsive	Explosion/drip/impact
Intermittent	Whooshing/splashing/scraping
Continuous	Blowing/flowing/rolling
Temporal	Duration
Short	Brief
Medium	Neither long nor short
Long	Extended
Spectral	Pitch
High	High pitch/frequency Treble
Mid	Medium pitch/frequency Alto
Low	Low pitch/frequency Bass
Dynamics	Volume/Loudness
Loud	High volume *Forte*
Medium	Medium volume/level
Soft	Quiet *Piano*
Content	Relevance
Informative	Relevant information
Neutral	Neither relevant nor irrelevant
Noise	Irrelevant/unwanted
Aesthetics	Beauty
Pleasing	Beautiful
Neutral	Mediocre
Displeasing	Ugly
Clarity	Quality
Clear	Easy to hear and comprehend
Neutral	Neither easy nor difficult to hear
Unclear	Difficult to hear and comprehend
Emotions	Feelings
Positive	Acceptance, Anticipation, Joy, Surprise
Neutral	No emotional content
Negative	Anger, Disgust, Fear, Sadness

Table 6-1: Sound event classification attributes and descriptions

Material relates to the substance which gives rise to the sound, either *gas*, *liquid* or *solid*, whilst the interaction specifies the nature of the sound's generation whether it was *impulsive*, *intermittent*, or *continuous*. The listeners used material as an attribute to describe sound events in the concurrent verbalisations, along with onomatopoeia. The audio professionals did not make use of material to describe sound events, but did make limited use of onomatopoeia. In the preliminary studies it was proposed that onomatopoeic descriptions of sound events could be represented by the material and interaction attributes of sound events in a similar manner to Gaver's 1993 interacting materials.

Temporal attributes reflect the total length of the sound event in terms of *short*, *medium* or *long*. Both listeners and the audio professionals used temporal attributes to describe sound events. For this study, start and end times were known for each sound event, but these measurements do not represent the perceived length, which can only be captured by querying listeners. Temporal attributes also relate to the rhythm and tempo of a sound event. Rhythm is partially represented by the interaction attribute in terms whether a sound event is *impulsive*, *intermittent* or *continuous*. Tempo is related to, but not the same as pitch, an object that is vibrating fast is perceived as having a high pitch, an object vibrating more slowly appears to have a correspondingly lower pitch. This effect can be heard in a Geiger counter where different tempi of clicks are experienced as different pitches (Neuhoff, 2004).

Spectral attributes apply to a sound event's pitch: *high*, *mid* and *low*. Both the listeners and the audio professionals used spectral attributes do describe sound. Spectral was the second most important attribute for the audio professionals with 59% of the respondents using it to describe sound. Listeners and professionals used the terms low and high, but only the professionals used mid. *High*, *mid* and *low* are terms universally used to label the equalizer section of recording consoles (B. Gibson, 2005)

Dynamics attributes pertain to a sound event's perceived intensity or volume: *loud*, *medium* or *soft*. This attribute was the most important for the audio professionals with 81% of the respondents specifying dynamics attributes. A similar percentage (80) of the listeners also specified dynamics when describing sound events. Both groups used the terms loud and soft when describing dynamics, neither used medium. *Loud* and *soft* are often used as descriptors in sound quality research (Björk, 1985, Guastavino, 2006, Susini, McAdams & Winsberg, 1999).

The content attributes are classified according to whether a sound event is *informative*, *neutral* or *noise*. In this case *noise* is defined as an unwanted or undesired sound, rather than unpleasant (Radomskij, 2007). Neither the listeners nor the professionals used content to describe sound, but according to Gaver (1986) establishing whether a sound is informative within an auditory interface has always been important. As the SMT is intended for use during the design and evaluation of auditory displays it was decided to include content as an attribute.

Barrass and Frauenberger (2009) referred to the importance of the balancing act between making an auditory display aesthetically pleasing and remaining informative. It has also been shown that a sound's aesthetics are integral to its functional effectiveness within an auditory display (Leplâtre & McGregor, 2004). When describing sound, both the listeners and the professionals highlighted aesthetic attributes. For this classification the terms were simplified to *pleasing*, *neutral* and *displeasing*, rather than the more commonly used terms by designers of harsh, warm, or bright. There is a general problem with describing aesthetics, even amongst professionals, which often requires designers to develop critical listening skills (Moylan, 2002).

Clarity applies to the intelligibility of a sound and for this classification is rated according to whether a sound event is *clear*, *neutral* or *unclear*. Almost half of the listeners and half of the audio professionals used clarity to describe sound. Both the professionals and the listeners used positive and negative terms to describe clarity.

The audio professionals were slightly more biased towards positive, whereas the listeners used a greater number of negative terms. *Clear* and *unclear* are often used as descriptors in sound quality research (Chesnokov & SooHoo, 1997, Gabrielsson & Lindström, 1985, Whitaker & Benson, 2002)

Emotions are considered in terms of *positive*, *neutral* or *negative*. Emotions were not identified by designers as a method of describing sound, but emotions were used by listeners. Listeners used positive and negative terms evenly to describe sound events. Both Thom (2003) and Johannsen (2004) argued that if a sound has been well designed, appropriate emotions should be evoked. Krebber *et al.* (2000) used what they termed the Exploration Associated Imagination of Sound Perception (AISP) to study the emotional impact of sounds within vehicles, where participants described what they heard without prompting. They found that this test prevented important attributes being overlooked, but required a considerable amount of experience to interpret results.

Visualisation

A form of visualisation was created, which took the form of a map, the key of which is shown in Figure 6-3. The key for the visualisation was provided as a printed A4 version for the designer to refer to when reviewing the soundscape maps (see Appendix G). Bertin's 1967 theory of cartographic communication was used to create the visualisation (Bertin, 1983). Bertin proposed that the visual variables of shape, size, value, orientation, hue, texture x and y coordinates could be applied to point, line and area symbols. Monmonier (1993) argued that Bertin's variables were also suitable for type, which could act as symbols.

The sound type was represented through either: a series of letters for *speech*, two quavers for *music*, or a loudspeaker symbol for *sound effect*. A loudspeaker symbol would be recognisable to sound designers who were familiar with audio circuit schematics (Talbot-Smith, 1999). In the preliminary studies three letters also represented *speech* and *music* was also displayed as ascending quavers.

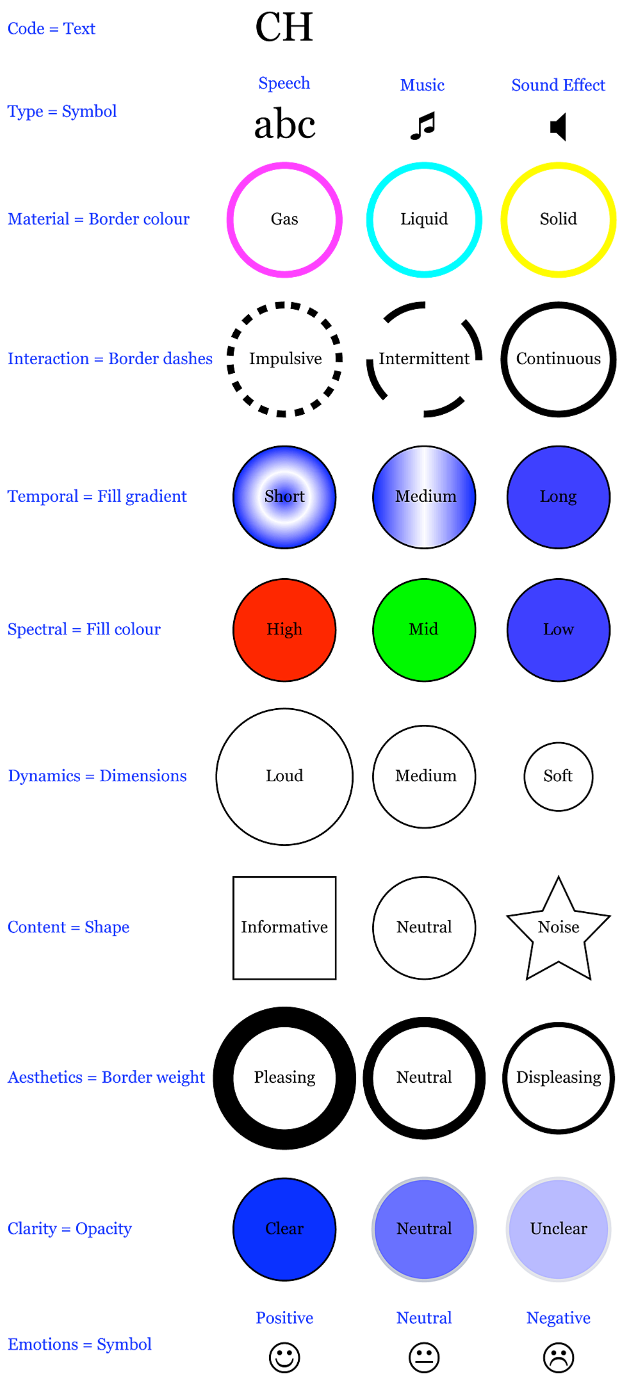


Figure 6-3: Visualisation key

The material attribute was illustrated through border colour. The phase state of elements (Gas, Liquid or Solid) can be visualized by colour coding the periodic table. Green is sometimes used for gas, blue for liquid and red for solid (Leach, 2009). Cyan Magenta and Yellow (CMY) were used in this study for material so that the colours were easy to differentiate from Red Green and Blue (RGB), which were applied to the spectral representation. *Gas* used magenta, *liquid* cyan and *solid* yellow, *liquid* has an association with water, which has a link with blue, so cyan was chosen for *liquid* in order to aid recognition. Border colour had been used to convey the material in the prototype method, but the colours were RGB rather than CMY.

The interaction was depicted using border dashes, *impulsive* had short dashes, whilst *intermittent* had longer, and therefore fewer dashes, whilst *continuous* was a single dash with no gaps. This approach was chosen to visually suggest the length of the sounds' interaction in a manner similar to a MIDI (Musical Instrument Digital Interface) piano roll editor. Piano roll editors are considered to be an intuitive way of viewing note lengths (Huber, 2007). Impulsive notes are displayed as short lines, intermittent notes are longer, and continuous notes are long lines. The interaction in the prototype tool was visualised in an identical manner with length and number of border dashes indicating the nature of the interaction.

Temporal attributes were shown using a fill gradient. A radial gradient was used to suggest a *short* event, which visually is associated with a droplet falling on to a liquid creating a surface gravity or circular wave. A *medium* event was portrayed with a linear gradient that is visually associated with longitudinal or plane waves. Both circular and plane waves are used in the field of acoustics to illustrate the propagation of sound, but neither forms of visualisation communicate the temporal length, only a waveform's presence (Crocker, 1998). A *long* event was a solid colour that implied that there was either no or minimal change. The gradient started with the spectral fill colour and then progressed to a pure white and then back to the original colour. Temporal attributes were not displayed within the prototype tool.

Fill colour was used for the spectral attribute, red was used for *high*, green for *mid* and blue for *low*. The choice of colour was based on an approach for pseudo-colouring topographical map by altitude (Hibbard, 2005). Red represents the regions above a central point, green is used for the middle range, and blue for areas below the mid point, this relationship corresponds to a colour's wavelength rather than its frequency. The analogy with height was adopted, as it is common for auditory professionals to consider pitch or frequency as the height of a sound (D. Gibson, 2005). D. Gibson also displayed different frequencies using colour, as did Matthews, Fong and Mankoff (2005). Fill colour was also used in the prototype tool, but was a continuous scale rather than a single colour. Within the prototype tool red represented high frequencies, green represented mid frequencies and blue was used for low frequencies.

Dynamics were illustrated using the scale of the shape, a *soft* sound was half the size of a *medium* one, and a *loud* sound event was 1.5 times the size of the *medium* and three times that of the *soft*. Servigne, Kang and Laurini (2000) proposed that the varying intensities of noise could be visualised by altering the radius of circles. D. Gibson (2005) also adopted this approach by indicating the volume of a sound in a mix by the object's size, with louder being larger than quieter. Scale was used in the prototype method to convey sound pressure level (SPL).

Three easily identifiable symbols were used for the content, informative used a square, whilst neutral was a circle and noise employed a star. The three distinct shapes do not share any stroke angles, making it easier to differentiate between them when sound events are overlapped on the map. Within the prototype tool, shape was used to visualise a sound event's interactive function, neutral was represented by a circle and noise was represented by star, a square was used to represent a relaxing sound event. Shape has been used to differentiate between the articulation of musical notes, legato = rounded, staccato = polygon (Friberg, 2004). Abstract shapes have also been applied to visualise phonemes that are not recognised by a phoneme recognition system, with high frequency sounds having spiky irregular forms (Levin & Lieberman, 2004). Circles have been regularly used to represent sound and can be found in a variety of visualisation schemes (Azar, Saleh & Al-Alaoui, 2007, Frecon *et al.*, 2004, Helyer, Woo & Veronesi, 2009). However, shape has not previously been used to convey sound content outwith semiotics.

Aesthetics were denoted by border weight, *pleasing* was represented with a thick line that was double the width of the *neutral* and four times the size of the *displeasing*. In terms of grouping sounds together in a mix, a full sound is often considered to be positive whilst a 'thin' sound is regarded as negative (Izhaki, 2008). Aesthetics did not form part of the original prototype tool reported in Chapter 4.

The clarity of a sound event was shown through the opacity of the shape, but not of the code or symbols. The codes and symbols were left unaffected in order to make the identification of sound events easier. *Clear* had an opacity of 100%, whilst *neutral* was 66% and *unclear* 33%, this related directly to the visual clarity of a sound event's representation. Opacity was used to display the acoustical information within the prototype tool with 100% opacity conveying *foreground* sound events and 33% indicating *background* sound events. Opacity has been used to a limited extent to communicate the volume or loudness of a sound event, but not to visualise the clarity of a sound event (Mathur, 2009, Radojevic & Turner, 2002 Thalmann & Mazzola, 2008).

Lastly graphical emoticons represented the emotions with a smile for *positive* emotions, a neutral expression for *neutral* and a frown for *negative*. As all of the codes were two letters an additional symbol located immediately below and next to the sound type was easily incorporated into the layout. Servigne *et al.* (1999) suggested that graphic seminology would be appropriate for displaying sounds, proposing that smiling faces overlaid onto a map could be used to display participant's preferences, a smile represented 'nice', a neutral expression 'neutral', and a frown for 'not so good'. Within the prototype a smile emoticon was used to represent *emotions* within the information category.

If a sound event was heard to move during the recording, then the start and end points were both marked, using the appropriate combination of symbols and colours. The two points were then joined with a line that had the properties of the appropriate material (colour), interaction (dashes) and clarity (opacity). This approach of indicating movement was also used in the prototype tool.

Guidelines

A set of guidelines were created by the author and printed on an A4 sheet for listeners to refer to at the start of each session (see Appendix E). The guidelines started by thanking the listeners for taking part. It then went on to state that the session was expected to take between 45 – 60 minutes. Listeners were then asked to feel free to leave at any time. The context of the work as part of an ongoing study into soundscape mapping was then explained. The document went to on to state that the data would be anonymised and used as part of the author's dissertation. The four different parts of the study were then described. The first part specified that listeners would be asked to listen to a series of alert sounds for a car auditory display, but would not be asked any questions afterwards. The following three questions were presented in random order. One of the parts specified that participants would be asked to listen to a 5 minute recording of a small car travelling through Edinburgh city centre, with a radio playing, and that listeners would be asked questions about what they heard. The second part stated that the participants would listen to a 5 minute recording of the audio interface only, after which they would be asked questions. The final part informed the listeners that they would be listening to a 5 minute recording of both the audio interface and the small car driving through Edinburgh city centre, and that they would be asked questions about what they had heard. The document finished by thanking the listeners for their help, and listeners were free to take the form for future reference.

6.1.3 Procedure

The procedure can be split into capture, design, classification, visualisation and designer's review of the SMT (see Figure 6-4). The capture stage involved the researcher creating a car schematic, recording the car soundfield and notating the soundfield. The schematic, recording and notation of soundfield were passed to the designer to aid the design of the Auditory Display. The Auditory Display, the car soundfield recording and the combined Auditory Display and car soundfield recording were then classified by the designer. The three audio recordings were then classified by the listeners. Next the researcher visualised in map form all of the participants' responses. Finally the designer reviewed the soundscape mapping tool.

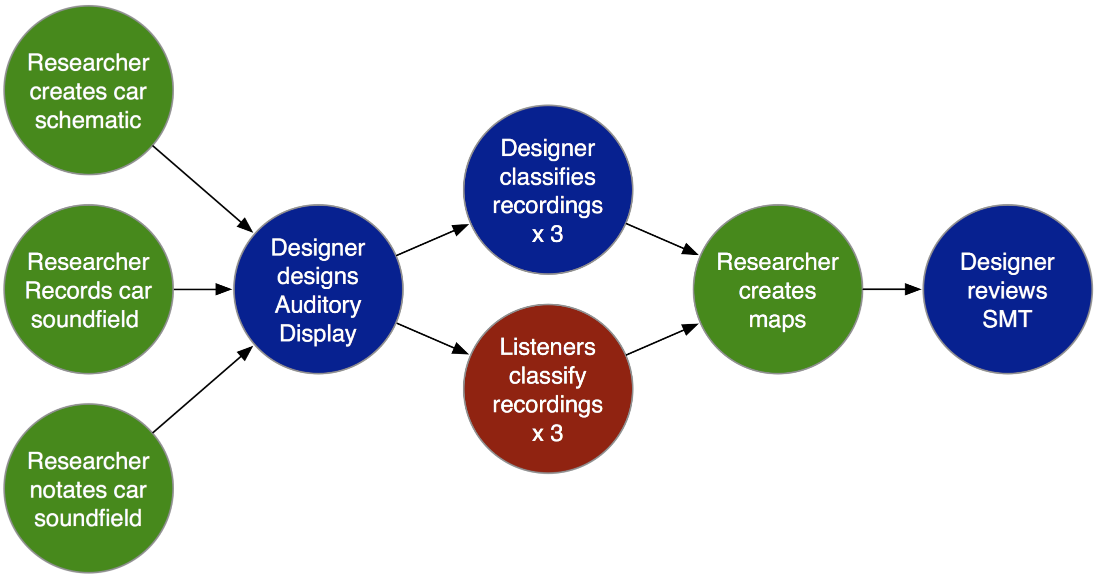


Figure 6-4: Procedure

Capture

The 5-minute recording was made using the custom eight-channel surround system and then augmented with the auditory display. Eight microphones were affixed in suspension mounts inside the car, at approximately head height (see Figure 6-5). Three microphones were placed equidistantly down each side, and two were mounted centrally, one from the rear view mirror, and one at the mid of the rear seat.

Figure 6-5: Microphone placement during setup of audio equipment, prior to final positioning and calibration, for surround sound recording

Calibration was achieved by a method borrowed from the film industry, the driver first read off the display of the SPL meter located on the passenger seat illustrating the slow sensitivity peak dB C level, and then the participant clapped their hands. The short peak acted as the starting point for the recording, allowing all eight tracks to be synchronised. A handclap by the driver completed the recording, confirming whether any of the tracks had drifted during the time period. Each track was transferred to the Digital Audio Workstation (DAW) to provide an auditory backdrop for the auditory display.

The source, action, start time, end time and location was identified for each sound by listening to the recording. As the researcher was the driver and the owner of the vehicle, it was relatively easy to identify discrete sound events. The location of each sound event was calculated using the perceived central point from the surround sound recording, and notated using x y coordinates according to the grid (see Table 6-2). If a sound event moved in relation to the car, the start and end points were documented. Start and end times of each sound event were also established from the recording, and were noted to the nearest second. In order to reduce the number of events that listeners had to classify, sound events that had the same source, action and location were grouped together and given alphabetical codes starting at AA.

Event	Source	Start	End	x	y	x	y
Engine Idle	Engine	00:00:00	00:00:21	6	6		
Male Speech	Radio	00:00:00	00:00:13	11	8		
Windscreen Wiper	Car	00:00:02		21	9		
Passing Car	Car	00:00:04	00:00:06	9	13	21	13
Siren	Ambulance	00:00:06	00:00:08	21	12		

Table 6-2: Example sound events

This list of sound events only represents what could be heard on the recording. Many more sounds were present, but were either masked or inaudible due to the method of capture. The transcription was made by listening to the multi-channel recording at normal levels, rather than over-amplifying to enhance barely audible sources. The decision to listen at normal levels was taken so that the levels replicated the conditions of the original journey, as well as the levels that listeners would experience. The process of transcribing the 5 minutes of recording took a full 8-hour day, with appropriate breaks to compensate for hearing fatigue.

All of the captured material was passed to the designer so that he could design the auditory display. The designer decided to limit the interface to only three sound events so as to minimise the requirement for memorisation by the listeners.

After creating the auditory display the designer overlaid the new sounds directly into the surround recording. This allowed the designer to control the level, incidence, duration and spatial location of the sound events. The augmented version of the surround sound recording was split into three versions by the author, so that all the timing and spatial cues remained consistent. The final three versions, each lasting 5 minutes were: Car only (205), Auditory Display only (AD), Car and Auditory Dsiplay (205+AD).

Classification

For reproduction each loudspeaker location corresponded to the equivalent position of the corresponding microphone during the recording. This ensured that all of the timings for the audio cues remained consistent, making for a more accurate spatial representation of the interior of the car. Each participant sat in the centre of eight compact loudspeakers and four sub bass units (see Figure 6-6).

Figure 6-6: Surround sound reproduction apparatus

Each listener was scheduled to take part individually. Listeners were asked to read the set of guidelines (see Appendix E). The sequence of the audio recordings was randomised to help reduce bias. After reading the guidelines participants were asked if they had any questions. They were then invited to listen to the three sounds created by the designer, as well as to refer to the printed descriptions (see Appendix D). The sound events for the auditory display were played back directly from a laptop using the internal speakers so that there was no association with the surround sound reproduction for either quality or spatial cues. Cues were triggered by the author and participants could listen as many times as they wished, which was on average three times per cue for the braking distance and dead angle, but only once for the message. Previewing the auditory display was necessary, so that listeners would be able to recognise the cues within the recordings, despite the effect of priming the listeners.

Listeners were questioned after each replay of the recording, rather than during, so that responses represented what they had attended to in context. After the first 5 minute sequence had finished playing, listeners were asked to turn over the rotated schematic of the car (see Appendix B). A printed A4 version of the classification was also provided for listeners to refer to (see Appendix F).

Individuals were first asked whether they were aware of a sound based on the list of sound events identified by the researcher from the recording (see Appendix C). If a listener was unaware of a sound event this was noted and the listener was then asked if they were aware of the next sound event on the list. If a listener was aware of a sound event they were asked to estimate the centre point of the sound source using the grid provided and then classify the sound event. This process continued until all of the sound events that the listeners were aware within the recording were classified. The procedure was identical for each of the three recordings (205, AD, 205+AD), and all of the 10 participants and the designer listened to all three recordings.

Once all of the responses had been captured, a combined response was created from the 10 participants so that the modal response could be generated. This was created using a spreadsheet, and was calculated using the mode. Combined coordinates were derived using a median rather than a mean, so as to reduce the effect of outliers skewing the data. If 50% or more of the listeners were aware of a sound event then the sound event was included in the combined map. Each classification was then obtained by calculating the most frequent occurrence, if an attribute had more than one mode then both values were recorded and subsequently visualised.

Visualisation

Each sound event was given a code as specified above, and the combination of shapes, colours and symbols were overlaid onto the grid according to the x y coordinates provided by the participants. If two or more sound events had identical coordinates then they were spaced evenly across the cell so that they remained visible. For ease of interpretation by the designer the grid, numbers and interior of the car were removed. The outline of the car was retained in order to provide some indication of orientation and scale. The spatial attributes were also displayed in this manner in the final version of the prototype tool.

6.2 Results

The results can be considered in terms of classification, visualisation and the designer's review of the soundscape mapping tool.

6.2.1 Classification

Participants listened to three different audio sequences in random order: the recording of the car (205), the auditory display (AD), and the car augmented with the auditory display (205+AD). Participants experienced both the car and the auditory display twice within the three recordings. The impact of the sequence in which the recordings were heard was reduced due to the random order of presentation. Listeners were aware of an average of 30% of the sound events, the highest recorded was 38% and the lowest 21%. An average of 25% (16) of all of the sound events from the car were heard by the participants the first time they heard the recording compared to 29% (19) for the second (see Table 6-3). With the auditory display, the average was 94% (2.82) for the first exposure, compared to 91% (2.73) for the second.

Participants had a high level of consistency with regards to sound events they were aware of, the average of 85% meant that only 15% varied between conditions.

Consistency within participant's responses was calculated by comparing each attribute and each sound event separately. If an attribute was classified identically for the same sound event within two separate conditions it was considered consistent.

	Car (205)				Auditory Display (AD)			
Participant	**1st**		**2nd**		**1st**		**2nd**	
Designer	18	28%	22	34%	3	100%	2	67%
P01	19	29%	28	43%	2	67%	3	100%
P02	21	32%	21	32%	3	100%	3	100%
P03	18	28%	19	29%	3	100%	3	100%
P04	22	34%	21	32%	3	100%	2	67%
P05	13	20%	17	26%	3	100%	3	100%
P06	15	23%	22	34%	2	67%	3	100%
P07	18	28%	16	25%	3	100%	2	67%
P08	10	15%	14	22%	3	100%	3	100%
P09	17	26%	15	23%	3	100%	3	100%
P10	9	14%	13	20%	3	100%	3	100%
Average	**16**	**25%**	**19**	**28%**	**2.82**	**94%**	**2.73**	**91%**

Table 6-3: Summary of participants awareness of sound events

When the participants are considered as a group there was a high level of awareness for the sounds associated with the car's engine and its handbrake. Whereas, the other sound events from the car, such as internal vibrations and indicating, went comparatively unnoticed, except when the vehicle passed over bumps in the road. On the radio the first male voice was discerned, whereas the second, and its chanting, was missed. Two out of three of the female voices, again on the radio, were identified, as was the interference from a mobile phone, but only one of the pieces of music was attended to. The group laughter was also generally missed, despite being the last thing that was present on the recording. Only two passing cars, and one passing bus were detected, which participants partially explained anecdotally as the overwhelming urge to listen to the conversation from the radio, even when they were experiencing the identical content for a second time. When listening to the three sound events from the auditory display all of the listeners were aware of all of the sounds. When the auditory display was listened to in context, then 4 out of the 10 listeners no longer recalled the braking distance cue. Even the designer was unaware of the braking distance cue, despite having added it into the recording himself.

Spatial accuracy was very much lower with an average consistency of only 23% (see Table 6-4). This was not as disparate as it first appeared, the average interval between the two instances was a max of only 2.56 cells, with the greatest average interval recorded by P02 at 4.65 cells and the lowest at 0.56 by P01. The designer had an average interval of 2.88 cells. When considering the average for the combined participants, the difference is 0.93 cells, this difference is possibly due to the decision to calculate the combined coordinates using the median rather than the mean. If the mean had been applied, this would have been closer to the participants' average cell interval of 2.56. Listeners appeared to find it hard to accurately recollect exactly where a sound originated from, but were more comfortable with its orientation in relation to their listening position, although there were the odd front to back errors. Problems with spatial discrimination are well documented, particularly when the source is not directly in front of the listener (Grantham, 1995).

Participant	**Aware**	**x y**	**Type**	**Material**	**Interaction**	**Temporal**	**Spectral**	**Dynamics**	**Content**	**Aesthetics**	**Clarity**	**Emotions**	**Sample**	**Mean**	**Classification only**
Designer	81%	6%	94%	100%	81%	50%	88%	50%	38%	94%	94%	88%	16	**72%**	**78%**
P01	65%	64%	100%	100%	50%	71%	100%	86%	64%	71%	79%	100%	14	**79%**	**82%**
P02	84%	17%	94%	89%	78%	83%	61%	61%	94%	100%	72%	78%	18	**76%**	**81%**
P03	93%	16%	89%	79%	53%	26%	95%	100%	84%	100%	63%	79%	19	**73%**	**77%**
P04	88%	10%	100%	80%	70%	80%	85%	90%	80%	50%	90%	65%	20	**74%**	**79%**
P05	88%	21%	100%	100%	93%	64%	93%	86%	100%	86%	79%	71%	14	**82%**	**87%**
P06	85%	13%	100%	94%	94%	50%	69%	94%	81%	94%	88%	75%	16	**78%**	**84%**
P07	84%	27%	100%	87%	73%	80%	73%	80%	80%	80%	87%	80%	15	**78%**	**82%**
P08	91%	8%	100%	92%	58%	75%	92%	75%	75%	75%	67%	92%	12	**75%**	**80%**
P09	85%	36%	100%	79%	57%	71%	71%	79%	50%	71%	93%	79%	14	**73%**	**75%**
P10	84%	22%	100%	67%	78%	67%	78%	67%	78%	56%	89%	44%	9	**69%**	**72%**
Mean	**85%**	**23%**	**98%**	**87%**	**70%**	**67%**	**82%**	**82%**	**79%**	**78%**	**81%**	**76%**	**15**	**76%**	**80%**
Combined	**85%**	**21%**	**100%**	**93%**	**57%**	**71%**	**79%**	**71%**	**86%**	**79%**	**86%**	**79%**	**14**	**76%**	**80%**

Table 6-4: Summary of consistency between individual classifications

For the classification as a whole there was an average consistency of 80% between individual attributes, the lowest being temporal at 67% and the highest type at 98%. The average of the 10 participants was also compared to the combined classification, which showed that apart from the interaction there was a good level of correspondence between the two sets of figures.

When classifying the type of sound, the majority of events were considered to be *sound effects* (73%) compared to *speech* (23%) and *music* (4%) (see Table 6-5). The consistency was 98%. The message from the auditory display, which used a synthesized voice, was sometimes considered to be a *sound effect,* and the chanting was occasionally referred to as *speech* rather than *music*.

	Type			Material			Interaction			Temporal			Spectral			Dynamics			Content			Aesthetics			Clarity			Emotions		
	Speech	Music	Sound Effect	Gas	Liquid	Solid	Impulsive	Intermittent	Continuous	Short	Medium	Long	High	Mid	Low	Loud	Medium	Soft	Informative	Neutral	Noise	Pleasing	Neutral	Displeasing	Clear	Neutral	Unclear	Positive	Neutral	Negative
Designer	11	0	35	36	0	9	8	25	12	21	20	4	9	30	8	1	32	13	27	15	3	8	34	3	29	16	0	7	36	2
	24%	0%	76%	80%	0%	20%	18%	56%	27%	47%	44%	9%	19%	64%	17%	2%	70%	28%	60%	33%	7%	18%	76%	7%	64%	36%	0%	16%	80%	4%
Participants	93	19	296	228	35	149	108	128	170	170	147	89	91	250	81	91	264	51	242	105	59	65	282	59	282	121	3	100	266	40
	23%	5%	73%	55%	8%	36%	27%	32%	42%	42%	36%	22%	22%	59%	19%	22%	65%	13%	60%	26%	15%	16%	69%	15%	69%	30%	1%	25%	66%	10%
Sum	104	19	331	264	35	158	116	153	182	191	167	93	100	280	89	92	296	64	269	120	62	73	316	62	311	137	3	107	302	42
	23%	4%	73%	58%	8%	35%	26%	34%	40%	42%	37%	21%	21%	60%	19%	20%	65%	14%	60%	27%	14%	16%	70%	14%	69%	30%	1%	24%	67%	9%

Table 6-5: Summary of classifications

The greater part of the sound events' materials were classified as *gas*; typically these included the passing cars and speech. *Liquid* was only applied to 8% of the sound events, and not at all by the designer. *Liquid* was occasionally used in conjunction with *speech*, but was never the mode for a single event. *Solid* was employed for 35% of the total sound events, and referred to objects such as the handbrake and the wheels passing over bumps in the road. The average consistency was high at 87%, inconsistencies occurring when classifying the engine idling, which was sometimes *gas* and sometimes *solid*.

Interaction had a lower level of consistency, with an average of 70%, with a participant averaging only 50%. The classifications were spread fairly evenly, more so than any other attribute. *Impulsive* sources like the handbrake and the braking distance alert made up 26%, whilst *intermittent* events, such as the seat creaking and the mobile phone interference contributed 34%. The largest group, *continuous*, which included the engine idling and passing cars, represented 40% of the responses.

Temporal responses were similar to interaction, but not directly related. *Impulsive* sounds were often classified as *short*, but this was less evident with *intermittent* and *medium* or *continuous* and *long*. *Short* sounds such as the mobile phone interference, and the wheel bumps made up 42% of the sound events. *Medium* contributed 37%, with the majority of the sounds originating from the engine. *Long* with half the percentage of *short* (21%) was applied to the male voice from the radio. After the spatial location, this category had the lowest average level of consistency (67%).

Part of the reason for both the temporal and interaction categories having lower levels of consistency might be due to a slight confusion with the nature of frequency, that is separating the rate of occurrence from the total length. Participants anecdotally reported that it took a while to distinguish between the temporal and interaction categories. The consistency may be improved by providing a training exercise prior to the main experiment such as examples of the different forms of interaction and length. The problem might be partially due to participants refining their responses the second time they heard an event, as they had a greater idea of how it related to the whole, and established for themselves more clearly the parameters of *short* versus *long*.

Spectral classifications had a higher level of average consistency at 82%. The majority of sound events were considered to be *mid* (60%), examples included all of the voices on the radio and the passing cars. *High* (21%) contained the braking distance alert and the driver's seat creaking, *low* (19%) included the bumps from the wheels and the engine accelerating and decelerating.

Dynamics had a similar weighting within the categories, with *medium* being applied to 65% of the total sound events, these included all of the auditory display cues, as well as the engine and passing cars. *Loud* represented 20% of the sound events, including the wheel bumps. *Soft* (14%) was applied to braking distance alert when it was heard in context. A preference for the mid value was also evident when classifying aesthetics and emotions. There was a high level of consistency at 82% with an overall bias towards the mid point.

Consistency, with regards to content was high amongst the participants (86%), but low with the designer (38%). Sixty percent of all of the sound events classified were considered *informative*, which included all of the auditory display sounds, and most of the sounds identified within the car. *Neutral* (27%) was applied to the second female voice on the radio as well as the music, again from the radio. *Noise* (14%) referred to the mobile phone interference and the seat creaking.

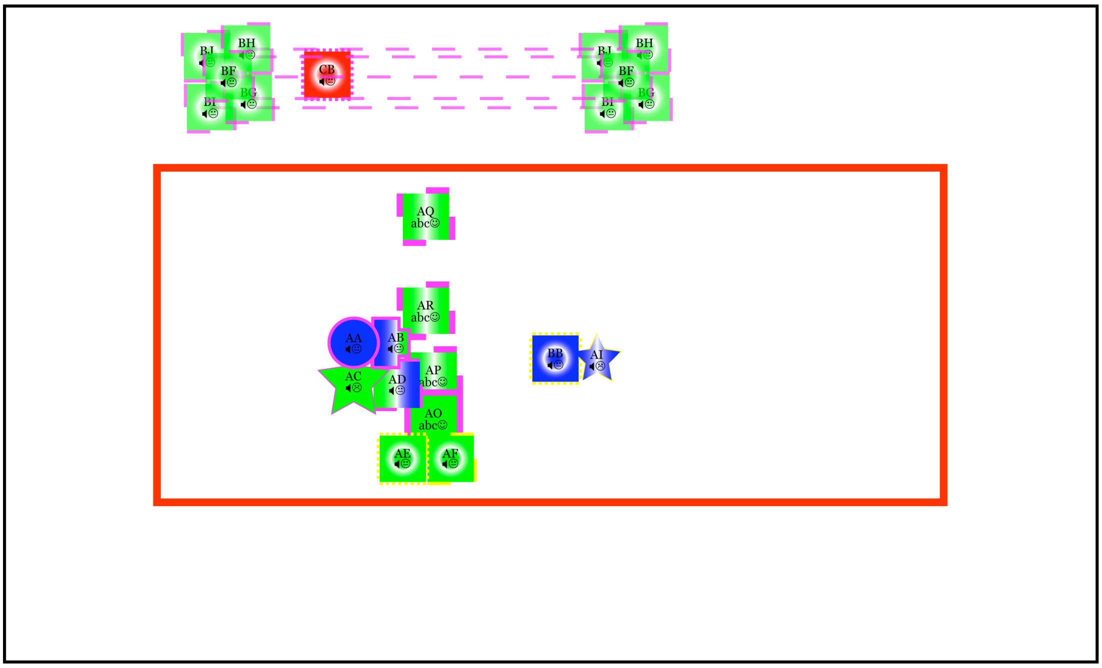
Aesthetics had a preponderance of sound events ranked as *neutral* (70%), which was also the highest single percentage for any category within a single attribute. *Neutral* sources included vehicular noises in general as well as the auditory display. *Pleasing*, which represented 16% of the total responses, included the second female voice and the first piece of music, both from the radio. *Displeasing* (14%) was used to classify the mobile phone interference and the bumps from the wheels. The low level of *displeasing* sound events may be context dependent, a greater percentage might be considered *displeasing* if the listener was in a park, rather than a simulated moving car in a city centre. The lack of *pleasing* sound events might be due to the lack of natural sound events that are normally considered agreeable (Anderson et al., 1983 and Kageyama, 1993). The level of consistency was comparatively high at 78%, which corresponds with the high level (70%) for a single category.

Clarity had a high average level of consistency at 81%, with *clear* being the most commonly chosen at 69%, the radio and the auditory display were the most obvious examples. *Neutral* represented a further 30%, with the engine idling and the driver's seat creaking. There were only three instances (1%) of sound events being classified as *unclear*, they were by three different participants and three different sound events: the second male voice and the male chant on the radio, as well as the message from the auditory display. This bias towards *clear* responses and lack of *unclear* may be due to the artificial nature of the experiment. As participants had no visual cues, they may have been paying greater attention to what they were listening to. Sound events that were *unclear* may have been ignored, and those that were constant were possibly habituated, and not consciously remembered (Sears & Jacko, 2008).

Emotions was the final attribute for classification, with an average 76% consistency. Responses were centred weighted with *neutral* representing 67%, this was uniformly applied to the auditory display and the car. *Positive* (24%) was used in reference to one of the female voices and occasionally for the music from the radio. *Negative* (9%) included the phone interference and sometimes the wheel bumps.

6.2.2 Visualization

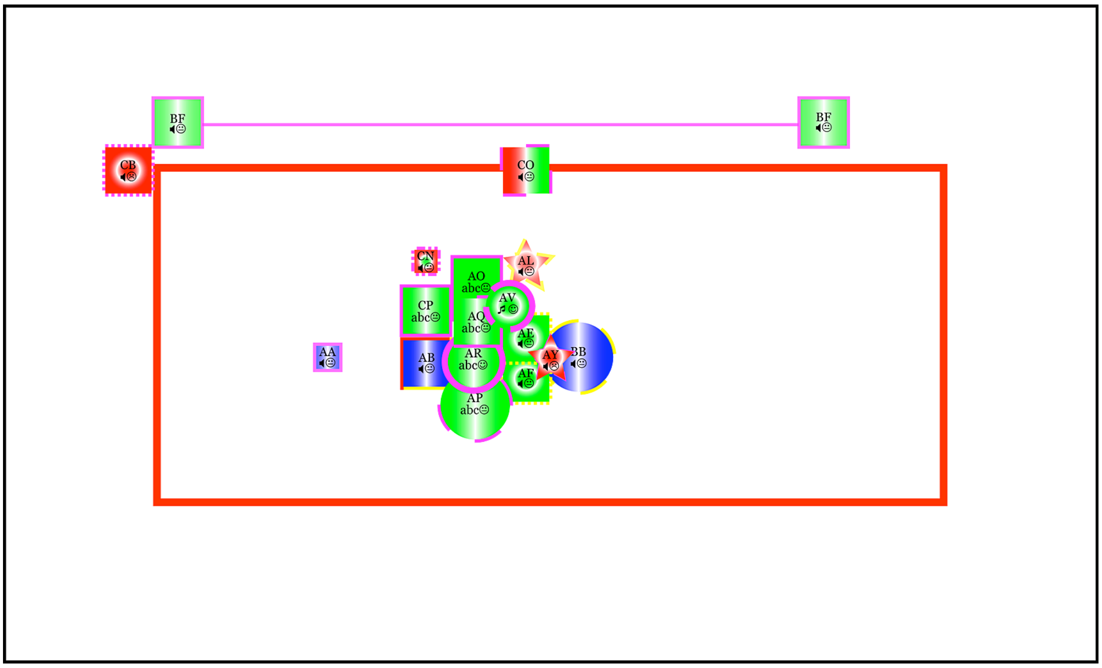
After all of the responses had been collated 36 maps were created using Adobe's *Illustrator* (2009b) software (see Figure 6-7). Each participant provided classifications for three maps: the car on its own (205), the isolated auditory display (AD), and the car augmented with the auditory display (205+AD). The combined classifications were also mapped in the same manner as the individuals' (see Figure 6-8). In addition, it was possible to create a fourth map that represented the auditory display as experienced in context, but isolated from the auditory backdrop.



Code	Event	Source	Code	Event	Source	Code	Event	Source
AA	Engine Idle	Engine	AI	Rattle	Sunroof	BF	Passing Car	Car 2
AB	Engine Accelerate	Engine	AO	Male 1 Speech	Radio	BG	Car Pass	Car 3
AC	Engine Cruise	Engine	AP	Male 2 Speech	Radio	BH	Car Pass	Car 4
AD	Engine Decelerate	Engine	AQ	Female 1 Speech	Radio	BI	Car Pass	Car 5
AE	Handbrake Released	Handbrake	AR	Female 2 Speech	Radio	BJ	Car Pass	Car 6
AF	Handbrake Engaged	Handbrake	BB	Bumps	All wheels	CB	Horn	Car 14

Figure 6-7: Visualisation of the Designer's soundscape for the car

Only sound events that participants stated that they were aware of were included in the maps. The first issue was when sound events occupied the same coordinates. If the clarity of one of the sound events was classified as being *neutral* or *unclear* then it was possible to overlap the sound events quite tightly. The partial opacity ensured that the relevant information was still clearly visible. However if all of the sound events that occupied the same grid were considered to be *clear*, and therefore opaque, then the amount of overlapping was minimal, as any area that was occluded was therefore no longer visible. Whilst this created problems with accurate positioning on the relevant coordinates, it did visually make it easier to see distinct *clear* sound events as they occupied a larger area. In contrast, clusters of *neutral* or *unclear* sound events were visually more complex due to their cluttered nature.



Code	Event	Source	Code	Event	Source	Code	Event	Source
AA	Engine Idle	Engine	AP	Male 2 Speech	Radio	BF	Passing Car	Car 2
AB	Engine Accelerate	Engine	AQ	Female 1 Speech	Radio	CB	Horn	Car 14
AE	Handbrake Released	Handbrake	AR	Female 2 Speech	Radio	CN	Braking distance	Auditory Display
AF	Handbrake Engaged	Handbrake	AV	Music 1	Radio	CO	Dead angle	Auditory Display
AL	Seat Creak	Driver's Seat	AY	Mobile Phone Interference	Radio	CP	Message	Auditory Display
AO	Male 1 Speech	Radio	BB	Bumps	All wheels			

Figure 6-8: Visualisation of soundscape for car and auditory display by combined participants

A simple solution to allow the inclusion of more sound events within a single grid would be to scale all of the attributes of the shapes down. This approach would make individual maps look both sparse as well as hinder legibility. Monmonier (1993) recommended that symbols are moved slightly apart to decrease the amount of overlap, and if this is not possible, then an inset at a larger scale could be used for crowded areas. The code and the type and emotions symbols were always kept at the same scale and opacity, which made sound events easier to locate and identify.

The maps show a propensity for awareness of sounds located in front and to a limited extent the sides. Sound events that were located to the side were normally moving, whilst those in front were almost always stationary. The use of CMY for borders and RGB for fills meant that any combination of colours, even a *continuous gas long high* sound event that had a continuous magenta border with a solid red fill was legible. Where this did not work as well as hoped was when a sound was classified as *displeasing*, the thin nature of the border width made it difficult to read the material and interaction. This could be partially rectified by increasing the overall scale of the borders, so that the thinnest is at least 2 points, which is currently the size of the neutral condition. Shape and size were easy to identify, even when partially occluded due to their symmetrical nature, which meant that the entire symbol does not have to be visible in order to identify its shape. Smaller *soft* sound events were layered on top of larger *loud* ones, and semi opaque *unclear* sounds appeared slightly washed out compared to the stronger colours of the *clear* ones.

6.2.3 Designer's review of SMT

The tool was found to be suitable for classifying and visualising a soundscape so that it was meaningful to a designer, but a number of issues were identified. The designer found the method of visualisation to be a quick way of interpreting the data. The principle of testing the auditory display within a surround sound recording was considered useful, and the technical quality was thought to be quite high. There were two areas in which the designer thought the realism could have been improved; the first was to include some height channels, and the second was to conduct the experiments in a room with improved acoustic treatment. Additional sounds, in order to confuse the listener, were also suggested. For temporally longer soundscapes it would be useful for listeners to make notes, and interruptions could also be used for longer experiments. But the designer considered that the correct method had been adopted for the current experiment, which used three separate durations of 5 minutes.

The designer requested a confidence rating for each individual icon, as well as an electronic version where information about how the values were derived was displayed in a side table, on mouse-over of the relevant icon. He also suggested giving designers a choice of classification scale, as sometimes being able to vary the number of choices is more appropriate. Some attributes might be improved with more categories such as spectral and dynamics, whilst others would suit less, for instance informative, where the neutral option could be dropped so that the decision is binary. The inclusion of spatialisation in the form of coordinates was deemed to be appropriate.

With regards to the classification, the type was believed to be useful, but the materials and interactions were not always relevant, especially with regards to *speech* and *music*. Though they were considered to be apposite when describing a *sound effect*. The designer found the temporal, spectral and dynamics attributes to be context dependent, but still relevant. The issue with the temporal attribute was that where a sound event could be considered to be *high* in relation to its source, such as a high note on a cello, which is essentially a bass instrument, or a high tone from a male voice which might be considered to be *low* pitched in overall terms. It was also suggested that practical examples such as a female voice for the *high* category might be more helpful than the current examples of 'high pitch/frequency treble'.

With respect to content, the need for *neutral* was queried and a request for a greater degree of granularity scale with possibly five or seven choices specifying the degree of information, such as moderately informative, informative, highly informative and so on. The term *noise* was considered to be ambiguous, *noise* could be considered as irrelevant and annoying. It was suggested that *noise* be changed to *uninformative* for consistency. The description was judged to be vague, as it could be relevant but unwanted, and could easily be improved by removing the term ‘unwanted’. This attribute was regarded as the most important for the purpose of interface evaluation.

Aesthetics were judged to be relevant, but like content, it would be more useful to have a larger scale. With regards to the descriptions, mediocre was considered to be *displeasing* rather than neutral, and it was felt that the *neutral* state did not require a description at all. Clarity was regarded as pertinent, and like type, material interaction spectral and dynamics had the correct number of choices, at three. Both the terms and descriptions were considered suitable. The classification of emotions could allow a greater degree of granularity, and the descriptors should be refined. Annoyance is not captured in the descriptor as a negative emotion, and it was queried as to whether ‘surprise’ and ‘anticipation’ were *positive* emotions.

Concern was raised about the possibility of aesthetics cancelling out the emotions. The designer wondered if it were possible for a *pleasing* sound to have an associated *negative* emotion. There was a tendency for *pleasing* sounds to be classified as *positive*, which occurred 78% of the time, with the remaining 22% being *neutral*. This was even larger for *neutral* aesthetics and *neutral* emotions at 82%, but was not the case with *displeasing* and *negative* emotions, which only coincided 51% of the time, 37% denoted as *displeasing* and *neutral*. Twelve percent of all of the responses were considered to be *displeasing* and *positive*, and four participants used this combination in three different conditions.

Almost all of the methods of visualising the attributes were regarded as effective, with two suggestions for changes. The first was to amend the gradient associated with the temporal attribute so that only a radial gradient was used and that its size varied according to the length of the event. A short event would have a smaller area where the gradient was applied, whilst a long event would have a correspondingly larger area. This would allow for a linear scale as well as addressing the issue of the linear gradient sometimes being difficult to see in conjunction with a low level of opacity. The spectral representation might also be changed from three distinct colours to a continuous scale, in order to allow a greater degree of granularity.

6.3 Discussion

When comparing maps it is possible to see what a participant or group are attending to, and how this differs from individual to individual. Figures 6-9 and 6-10 show the designer's map for the auditory display and the participants' combined responses in situ, with the car sound recording subtracted. The spatial cues have been identified, albeit with slight variation, the message and the braking distance alerts have remained in front of the driver, but are reversed, and the dead angle has been discerned as originating from the right, but not as far back as the designer intended.

The type has remained consistent for the braking distance and dead angle, both being considered to be *sound effects*, but the message has only been classified as *speech*, rather than a combination of *speech* and *sound effects*. This suggests that the sounds contained within the message may pass unnoticed. The material, which in this case was *gas*, remains constant, whereas the dead angle is perceived as being *intermittent* rather than *impulsive*. This shows that the dead angle is thought to be more of a whooshing sound rather than an explosion, which is also possibly due to a close association with the sound that a passing vehicle makes, this is also borne out through the alert being thought to be temporally *medium* in length rather than *short*.

The pitches for the two alerts were judged to be *high mid* rather than just *high* and the dynamics for the braking distance was considered to be *soft* but still *clear*. All of the events were classified as *informative* and aesthetically *neutral*, as well as emotionally *neutral*. It can be seen that the participants experienced the auditory display in context in a manner similar to the designer's intentions.

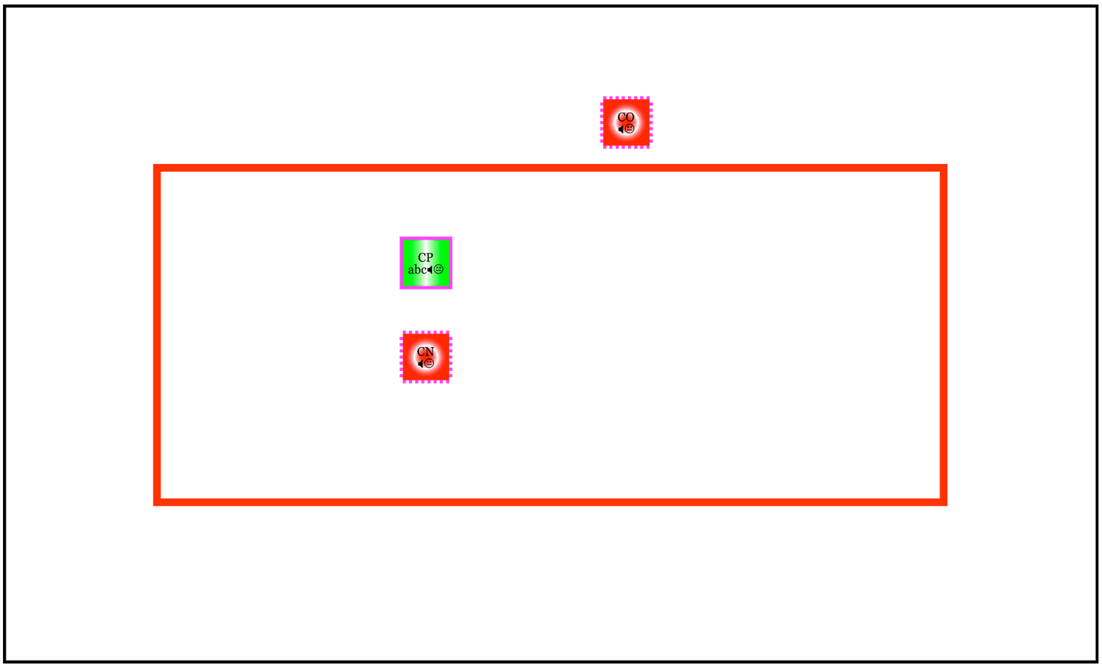


Figure 6-9: Designer's soundscape map for the auditory display (CN = Braking distance, CO = Dead angle, CP = Message)

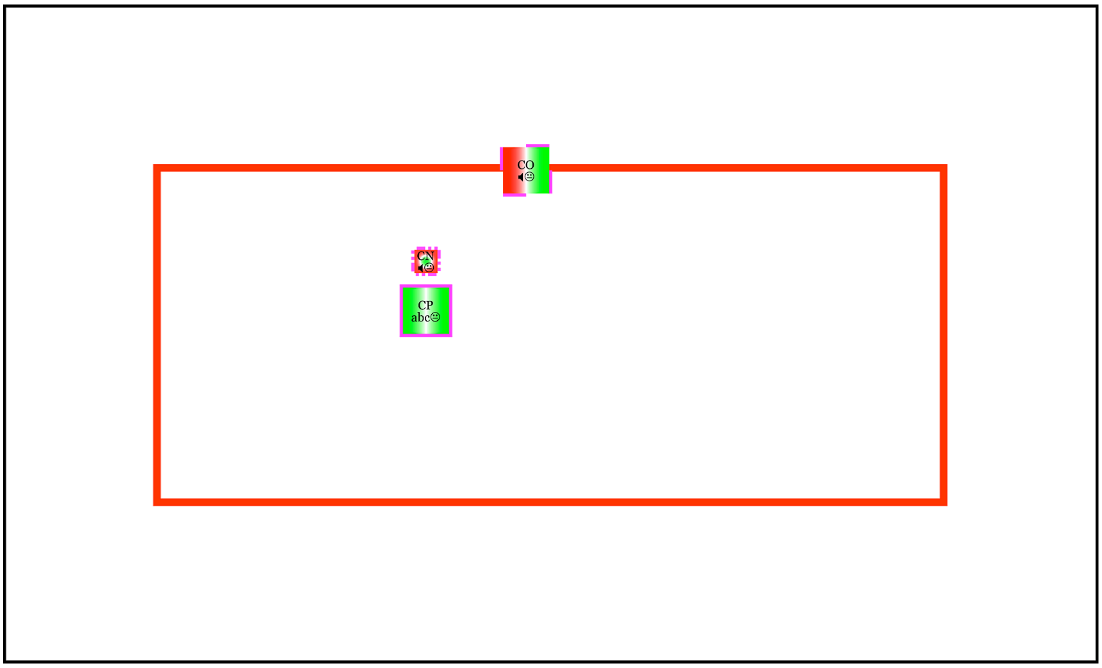


Figure 6-10: Combined participants' soundscape map for the auditory display in context with vehicle sound events subtracted (CN = Braking distance, CO = Dead angle, CP = Message)

Some of the changes suggested by the sound designer were made to the tool. Specifically the classification of *noise* was changed to *uninformative*, and the description of 'unwanted' removed. 'Mediocre' was removed from the description of the *neutral* category of aesthetics, and the descriptions for *positive* and *negative* emotions were changed from the eight basic emotions described by Plutchik (1980), to the six primary emotions formulated by Parrot (2001). The only modification made to the visualisation was to change the width of the radial gradient according to a sound event's temporal classification. Previously *short* used a radial gradient, whilst *medium* used a linear gradient, whilst *long* was a solid fill (see Figure 6-11).

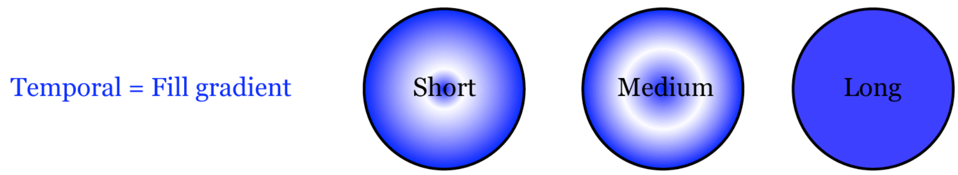


Figure 6-11: Updated visualisation key for temporal attribute

This chapter provided an illustration of a soundscape mapping tool that could be used to represent what attributes of sound people were listening to in visual form. It also showed that the tool could be used by a designer for the design and evaluation of sounds and an auditory environment. Whilst the attributes were derived from current professional practice and listener descriptions, and the procedure and visualisation were developed during preliminary studies, the tool still requires to be tested more extensively with a broader range of auditory environments.

The process of mapping has allowed a four dimensional auditory environment to be captured in two-dimensional form, allowing ease of comparison between a designer's intentions and listeners' experiences. It also represents the effect of listening rather than hearing, where it is clear what is being attended to, and what has become habituated or has been ignored. With the car it was evident that sounds emanating from beyond the rear of the vehicle fell into this latter category, whereas those in front of or immediately surrounding the driver fell into the former. The relevance of sounds were also shown so that unwanted elements such as mobile phone interference and the driver's seat creaking could be silenced or masked, but other sounds such as the engine idling or accelerating, and the handbrake being engaged and disengaged should remain clearly audible as they were considered informative.

The tool was found by participants to be relatively straightforward, if a little time consuming. The SMT could be applied as a form of notation for auditory environments without the need for a surround sound recording. The SMT could also be used as the basis for classifying a sound, listeners experiences could be combined in order to provide information about the shared experience, as well as an individual's interpretation. This is currently difficult using 'earwitness' accounts or concurrent verbalisations, where responses can range from simple lists of sound events, through to narratives, where elements such as comparisons, descriptions, emotional responses and semantics are shared, but are considered less useful for informing design (G. W. Coleman, 2008).

The work reported in this chapter addressed the second research question about how a soundscape could be classified and visualised so that it is meaningful to designers. In order to evaluate the tool's suitability as a method of comparing designers' intentions for a sound design with the experiences of listeners it is necessary to test the tool with a number of designers and a greater number of listeners. In the next chapter, 10 sound designs are mapped in order to provide examples of how the SMT could be used by designers.

7 Evaluation of soundscape mapping tool

The previous chapter illustrated the soundscape mapping tool using an in-car auditory display. To evaluate the soundscape mapping tool (SMT) the number of sound designs and listeners had to be expanded. This study provides additional examples of how the SMT can be used, from a range of fields within both traditional media and computing. Sound designers were asked to rate the importance of the attributes used in the classification, and to choose the most appropriate way to display the attributes. Designers were asked about their current methods for evaluating sound designs, as well as how they could use the SMT, and what they would change.

7.1 Method

Sound designers from different forms of media were approached to provide sound designs for evaluating the SMT. Both designers and listeners were asked to classify the designs using the same procedure. Responses from both groups were mapped and returned to the designers with a questionnaire about the SMT.

7.1.1 Participants

Three groups of participants took part in this study: 10 sound designers, 10 sound engineering students and 100 listeners. Within the sound designers group both professional practitioners and researchers were approached. The professional practitioners provided examples from traditional media and computer games, whilst the researchers supplied auditory displays. Participants were sought internationally in order to provide as broad a range as possible of sound designs. Requests for collaboration were posted on professional sound design forums, as well as via direct e-mail correspondence, with addresses gleaned from published papers. None of the designers were personally known to the author prior to the study. The sound designers provided designs from auditory displays, film, marketing, radio, sound art, and video games (see Table 7-1).

Design	Description	Sound events	Length
01	Auditory Display	32	02:09
02	Short Film	45	02:40
03	Soundscape Composition	16	00:31
04	Sonification	8	00:57
05	Simulation	19	07:26
06	Games Sound Effects	18	03:24
07	Radio Drama	14	00:42
08	Audiologos	14	00:20
09	Composition	26	01:30
10	Film Sound Effects	32	00:30

Table 7-1: Summary of sound designs

Ten undergraduate students who had studied sound engineering at Edinburgh Napier University took part in the pilot study. The 100 listeners were either staff or students at Edinburgh Napier University and formed a sample of convenience. None of the listeners had previously taken part in a listening study before.

7.1.2 Materials

In order to classify their sound designs the designers were provided with two Microsoft *Word* files, and two Adobe pdf (Portable Document Format) files. Designers listed and classified the sound events using the first *Word* file. This was laid out as a landscape table where descriptions of each sound event were inputted in the first column. All of the columns to the right of the description had drop-down form fields so that the designers could select appropriate attributes for each sound event (see Appendix F).

One of the pdf files was a grid so that the designers could estimate from where each sound event was emitted from (see Appendix I). The grid was laid out in the manner of an ordnance survey map with the numbering starting from the bottom left corner. For all of the stereo sound designs columns were used to represent panning and rows represented depth. The first column (0) was used for sound events that were panned hard left and the last column (21) was for sound events that were panned hard right. The first row (0) was for sound events that were perceived as being close to the listener. The last row (16) was for sound events that were perceived as being far from the listener. In the case of the surround sound design composition (Design 09) the listener was situated in the middle of the grid and the sound events were positioned according to their 360-degree orientation and depth.

The second pdf document comprised a list of attributes with descriptions that the designers used as a reference to classify each sound event within their designs (see Appendix J). The final document was an informed consent form that all participants were asked to complete (see Appendix K and Appendix L). This was used to provide written consent, and stated what was involved in order to take part, the purpose of the study and how the results would be anonymised so that participants could not be identified. The informed consent form was adapted from a template, modifications were approved by the supervisory team according to faculty ethics' guidelines.

A questionnaire was sent out to all of the designers with soundscape maps of their design, a visualisation key (see Appendix O) and anonymised data. The questionnaire was a Microsoft *Word* document that used a combination of drop down, check box and text form fields (see Appendix M). The questionnaire addressed three main areas: classification, visualisation and applications. Designers were asked to rate how important each attribute used in the soundscape mapping tool was "in order to compare sound designs with listeners' experiences". The five options were: 'Very Important', 'Important', 'Moderately Important', 'Of Little Importance' and 'Unimportant'.

The designers were then invited to "choose the most appropriate way to display the audio attributes used in the classification". A visual questionnaire approach was adapted from Bruseberg and McDonagh-Philp (2001), where each visualisation option was pictorially represented, and designers were asked to use a check box to indicate their choice. This was followed by their level of agreement with the statement that the "soundscape mapping tool allowed me to compare a sound design with the experience of listeners." The five options were: 'Strongly disagree', 'Disagree', 'Neither disagree nor agree', 'Agree' and 'Strongly agree'. The questionnaire concluded with four open ended questions about what methods they currently employed to evaluate sound designs, how they could use the tool, and suggestions for changes. The final question was a request for further comments.

According to Peat, Mellis & Williams (2002) it is necessary to pilot questionnaires in order to improve their internal validity. Ten undergraduate students who had studied sound engineering at Edinburgh Napier University were asked about the ambiguity or difficulty of the questions, along with the length of time it took to complete the above questionnaire (see Appendix N). Record was made of questions omitted by the respondents. Participants were asked to watch a Short Film (Design 02) and then refer to the sound designer's and the listeners' maps, along with a visualisation key and an explanation of the attributes. This was administered remotely via e-mail in order to replicate the method that was to be used by the sound designers. In order to help mitigate some of the problems highlighted in the pilot study the attributes table was expanded to include 'awareness' and 'spatial' cues (see Appendix J). The missing attributes ('awareness' and 'spatial') were also added to the visualisation key (see Appendix O).

7.1.3 Procedure

The procedure involved classification, visualisation and a survey (see Figure 7-1). The designers first classified their designs and forwarded them to the researcher for mapping. Listeners then classified the designs and the results were mapped by the researcher. Finally the maps were returned to the designers with a questionnaire.

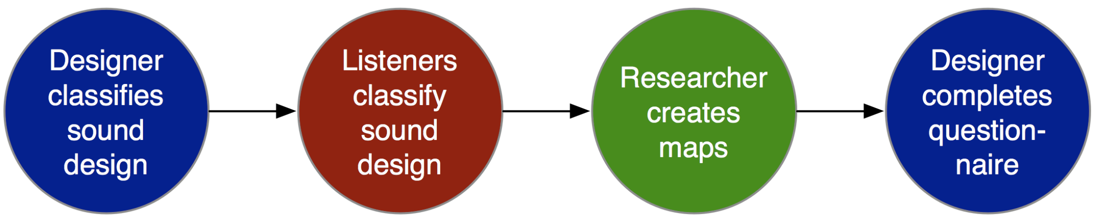


Figure 7-1: Procedure

All of the sound designers were located across Europe and the Americas and correspondence was conducted via e-mail. Designers were asked to supply a sound design that they would like to have evaluated. The choice of sound design was left entirely to the designer, and no guidance was given about length or complexity. There was no requirement to create a new design for the study and 7 of the 10 designers used existing works. The designers were asked to list all of the sound events using short descriptions that would be meaningful to non-designers. Designers were then asked to classify all of the sound events based on the specified attributes. After this a copy of the *Word* document and the sound design were sent to the author so that ten listeners could experience the design.

The participants were randomly assigned to each sound design, until 10 participants had listened to each design. The tests were conducted in a quiet office with stereo loudspeaker reproduction, except for design 9 (Composition) which required a surround sound system located in an isolated acoustically untreated room. For six of the sound designs participants were asked first to listen to the complete sound design and then classify, in random order, the sound events. For the other four sound designs participants were played short sections and asked to classify specified sound events based upon what they had just heard, and for two (04 and 09) they had the ability to repeat the sound as often as they wished due to the abstract and complex nature of the designs. The decision as to which approach was adopted was left up to the sound designer as the SMT can be used in both ways.

The author conducted questioning about the attributes of each sound event verbally, with the listeners having access to the grid and the list of attributes, but not the list of sound events. The responses from the listener were collated and the mode for each attribute was calculated for individual sound events. The author then visualized the results into two different maps for both the designer's intentions and the listeners' experiences. When the maps were returned to the designer the self-administered questionnaire was included.

7.2 Results

The results are split into examples of soundscape maps (20) and questionnaire responses (10). Ten sound designs were submitted, each of which were mapped according to the sound designer's intentions and the listeners' experiences.

7.2.1 Design 01: Auditory Display

The first design was for an Auditory Display, the sound events had been designed for a large manufacturer of electrical appliances, and were used in a variety of their products (see Table 7-2). Four of the 32 sound events were for indicating 'power on' of electrical appliances. One of the sound events was theme music for all of the company's appliances. The remaining 27 sound events were all designed for smartphones, which Töyssy and Helenius (2006, p.110) define as "mobile phone[s] that includes software that a user is able to modify and update." The sound events included auditory icons, earcons, a spearcon and music.

Code	Description	Code	Description	Code	Description
AA	Low Battery	AL	C_WashingMachine_PowerOn	AW	Touch02
AB	Navigation Warning	AM	C_ThemeMusic	AX	Failure
AC	No Key Area	AN	Air-Refresh for GUI	AY	Notification
AD	No Signal	AO	PND Booting Sound	AZ	Success
AE	Power Off	AP	Message Song	BA	Low Battery Alert
AF	Power On	AQ	Message Music	BB	I'm calling you_Man
AG	Routing Completion	AR	Message Music 2	BC	I'm calling you_Woman
AH	System Warning	AS	Message Knocking	BD	Ringtone1
AI	C_Buzzer_PowerOn	AT	Phone Power On	BE	Ringtone2
AJ	C_Refrigerator_PowerOn	AU	Phone Power Off	BF	Warning_Spearcon
AK	C_Airconditioner_PowerOn	AV	Touch01		

Table 7-2: Sound events (Design 01: Auditory Display)

No spatial cues were recorded by the designer, as the sound events were tested in isolation rather than within products. As all the spatial cues were missing the sound events were visualised in rows and columns rather than map form (see Figure 7-2). The designer made limited use of the material and interaction attributes, recording this information for only 9 of the 32 sound events. However, all of the other attributes were applied. On the designer's map 23 of the borders are continuous black to denote that the designer made no choice for the material and interaction, it was necessary to include a border to convey the aesthetics (border width).

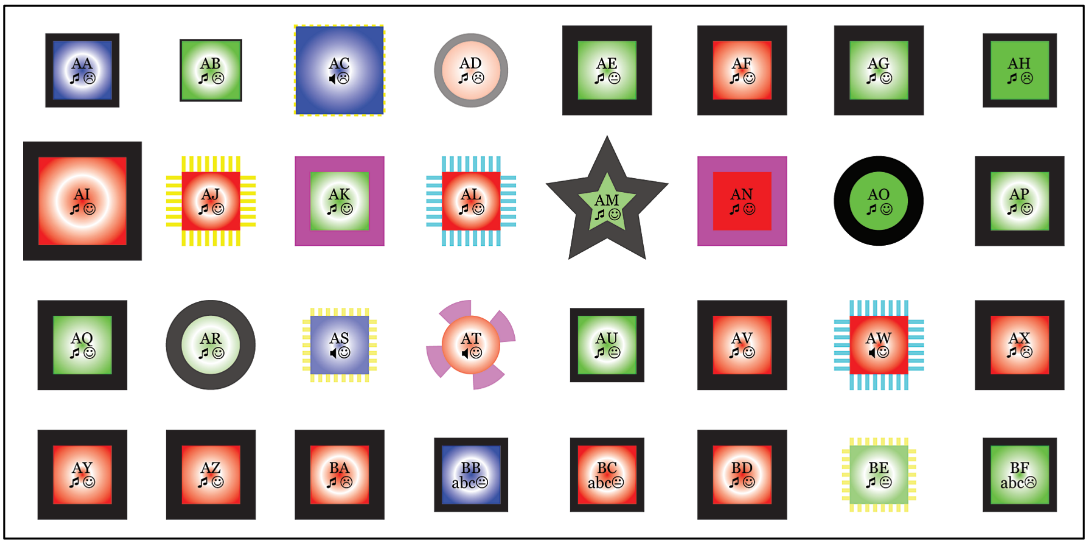


Figure 7-2: Designer's soundscape map (Design 01: Auditory Display)

The designer classified 25 of the sound events as *music* (quavers), 4 as *sound effects* (loudspeaker) and 3 as *speech* (text). *Music* was applied to earcons as well as the theme music, *sound effect* was selected for the auditory icons and *speech* was chosen for the spearcon, as well as the male and female voices. The designer classified 27 of the sound events as *informative* (square), 4 as *neutral* (circle) and 1 as *uninformative* (star). The *uninformative* sound event was the theme music (AM), and the 4 *neutral* sound events were all designed indicate interaction on a smartphone (AD, AO, AR & AT). All of the sound events, except for two (AC and AI), were classified as having *medium* dynamics. This was displayed using dimension, which in the case of *medium* was 30 pixels, two were considered to be *loud*, which increased the size to 45 pixels. None of the sound events were *soft*, which would have been 15 pixels.

Listeners included all of the material and interaction attributes, which meant that all of the borders are coloured, with 21 having dashes (see Figure 7-3). The listeners attributed 3 sound events as *music*, 20 as *sound effects*, 8 as *speech* and 1 as both *music* and *sound effect*. The listeners classified 22 as *informative*. A further four sound events were classified by the listeners as *informative* and either *neutral* (circle) or *uninformative* (star), the two that were *informative* and *neutral* (AM and AV), are shown as half square and half circle, and the two that were considered to be *informative* and *uninformative* (AT and AX), are visualised as half square and half star. The contradiction about the content for the sound event for the 'phone power on' (AT) might not be considered a problem, as the designer classified it as *neutral*. Whereas, the 'failure' sound event (AX) was classified as *informative* by the designer. This might be due to the listeners perceiving the sound event as *speech* rather than *music*, as well as their considering the sound event's emotional attributes to be *positive/ neutral* rather than *negative.* However, both the listeners and the designer found the sound event to be *clear*, so clarity could not have been a contributing factor.

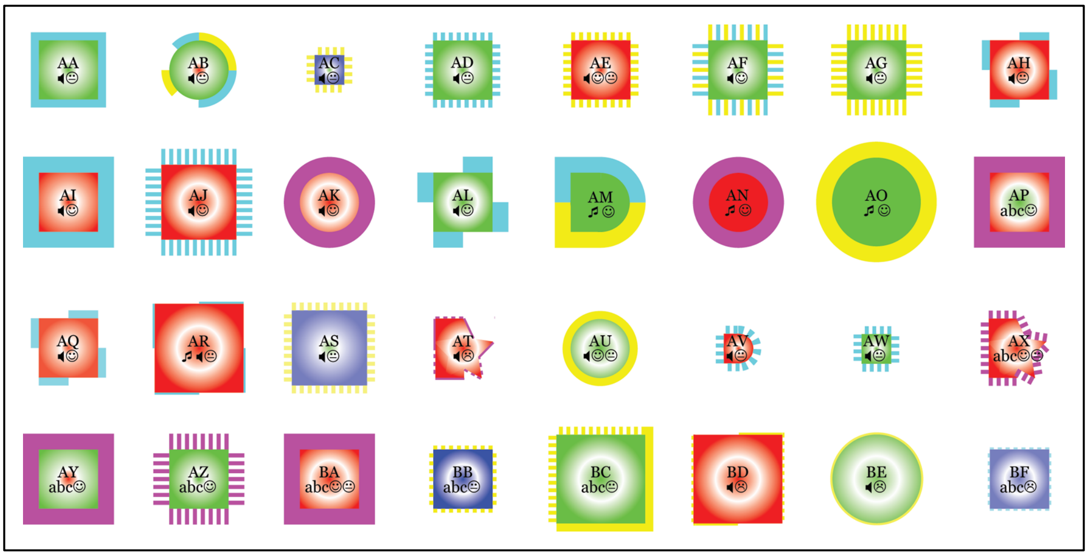


Figure 7-3: Listeners' soundscape map (Design 01: Auditory Display)

The listeners' responses for dynamics were slightly more varied than the designer's, 22 sound events were *medium*, 5 were *loud*, 3 *soft*, and 2 were both *loud* and *medium*, which meant that the diameter of the object was 38 pixels. Fewer of the sound events were considered to be *pleasing* by the listeners (13) when compared to the designer (21), but more were classified as either *clear* (28) or *clear/neutral* (3).

As an Auditory Display the sound design might be regarded as successful, the majority of the sound events were considered by the listeners to be *informative* and *clear*. The listeners considered sound events that were thought to be *displeasing* by the designer *neutral*. However, three of the sound events that were considered to be *pleasing* by the designer were found to be *displeasing* by the listeners (AR, AT and BD). These three sound events along with the two that were found to be partially *uninformative* (AT and AX) might benefit from further review by the designer.

7.2.2 Design 02: Short Film

The second sound design was for an independent Short Film that had both music and sound effects, but no scripted dialogue. The 2 minute 40 second film had 45 sound events that included ambiences, vocalisations, spot effects and music (see Table 7-3).

Code	Description	Code	Description	Code	Description
AA	Room ambience	AP	Low/mid modulated ambience	BE	Weird multi-gun layered sweeps
AB	Synth pad ambience	AQ	Steps on wet concrete	BF	Low frequency boom closing of the scene
AC	Printer like sound	AR	Punch	BG	Layered distorted pad sound
AD	Typing	AS	Outdoors ambience busy street	BH	Distorted scream
AE	Typing confirmation beeps	AT	Crow	BI	Distorted electric guitar
AF	Drum stick instrumental	AU	They are over there vocal	BJ	Processed reverse scream
AG	Bells instrumental	AV	Helicopter	BK	Breathing wearing mask
AH	Leather clothing	AW	Dub music	BL	Touch of the mask
AI	Floor creak	AX	Distorted evolving pad	BM	Low frequency ambient pad
AJ	Steps on wooden floor	AY	Gas release	BN	Small bells sound
AK	Noise sweep	AZ	Distorted woosh	BO	Drum stick
AL	Monitor flash noise	BA	Gun loading	BP	Touch mask
AM	Low frequency kick	BB	Gun shot	BQ	Breathing
AN	Short aa vocal	BC	Multilayered distorted scream	BR	Pulling the mask off
AO	Outdoor ambience	BD	Weird branches coming out of mouth	BS	Big gong sound

Table 7-3: Sound events (Design 02: Short Film)

The designer made use of all of the attributes to classify the sound events (see Figure 7-4). Ambiences such as ‘synth pad ambience’ (AB) and ‘low frequency ambient pad’ (BM) were generally classified as *uninformative* by the designer. Spot effects such as ‘typing’ (AD) and ‘gun shot’ were predominantly *informative*. The designer considered the vocalisations such as the ‘short aa vocal’ (AN) and the ‘multilayered distorted scream’ to be more evenly spread between *informative*, *neutral* and *uninformative*. All of the music was regarded as *uninformative*.

The most obvious difference between the designer’s and listeners’ experiences was the number of sound events that the listeners were *unaware* of. Only 23 out of the 45 sound events were recalled by the listeners, this does not appear to be due solely to the length of the clip, as even the last sound (BS: Big gong sound) went unnoticed by some, and the omissions were equally placed throughout the clip (see Figure 7-5).

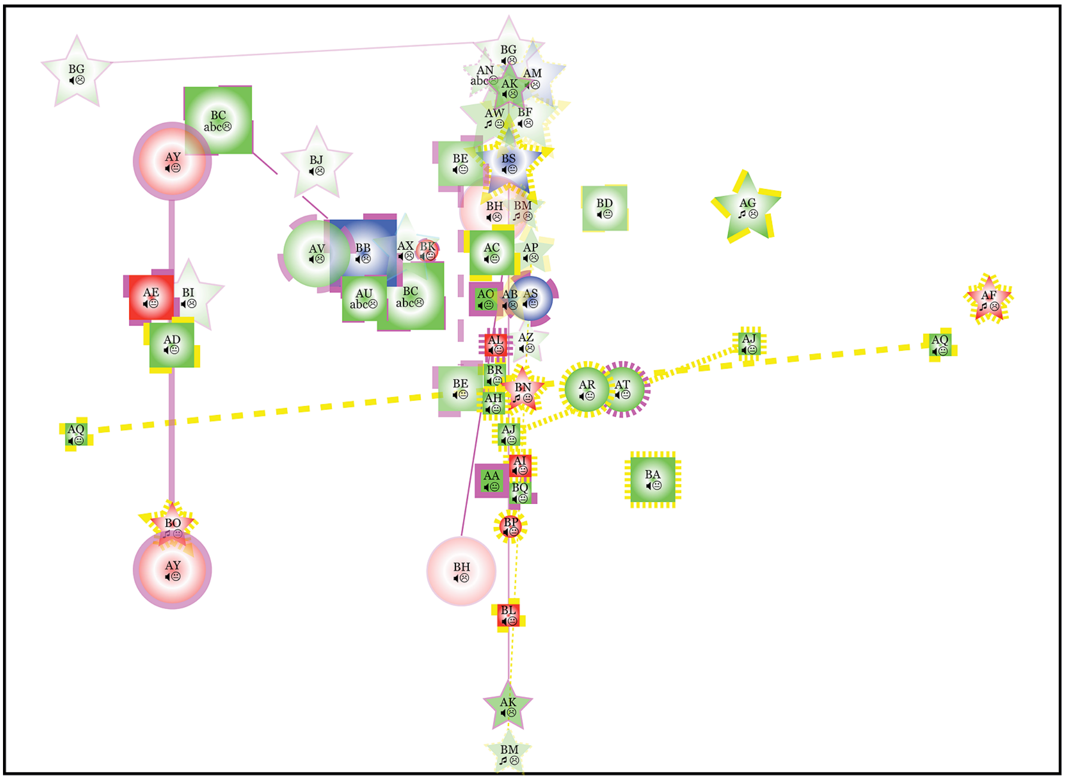


Figure 7-4: Designer’s soundscape map (Design 02: Short Film)

Atmospheric musical sound events like ‘synth pads’ and ‘drum sticks’ (AB, AF, AK, AM, AY, AZ, BG, BM, BN, BO and BS) were not recalled, as were sound events that were associated with strong visual design such as ‘weird branches coming out of mouth’ or ‘weird multi-gun layered sweeps’ (BD and BE). Twelve of the 22 sound events that the listeners were *unaware* of were classified as *uninformative* by the designer, 7 were *informative* and 3 were *neutral*. All of the sound events that listeners were *aware* of were classified by the designer as either *informative* (18) or *neutral* (5). Dynamics does not seem to have had an effect on the lack of awareness of sound events, as 6 were classified as *loud*, 13 were *medium* and only 3 were *soft*. Clarity (opacity) appears to have even less of an effect on listeners’ awareness of sound events, as 10 sound events were *clear*, 5 were *neutral*, and 7 were *unclear*.

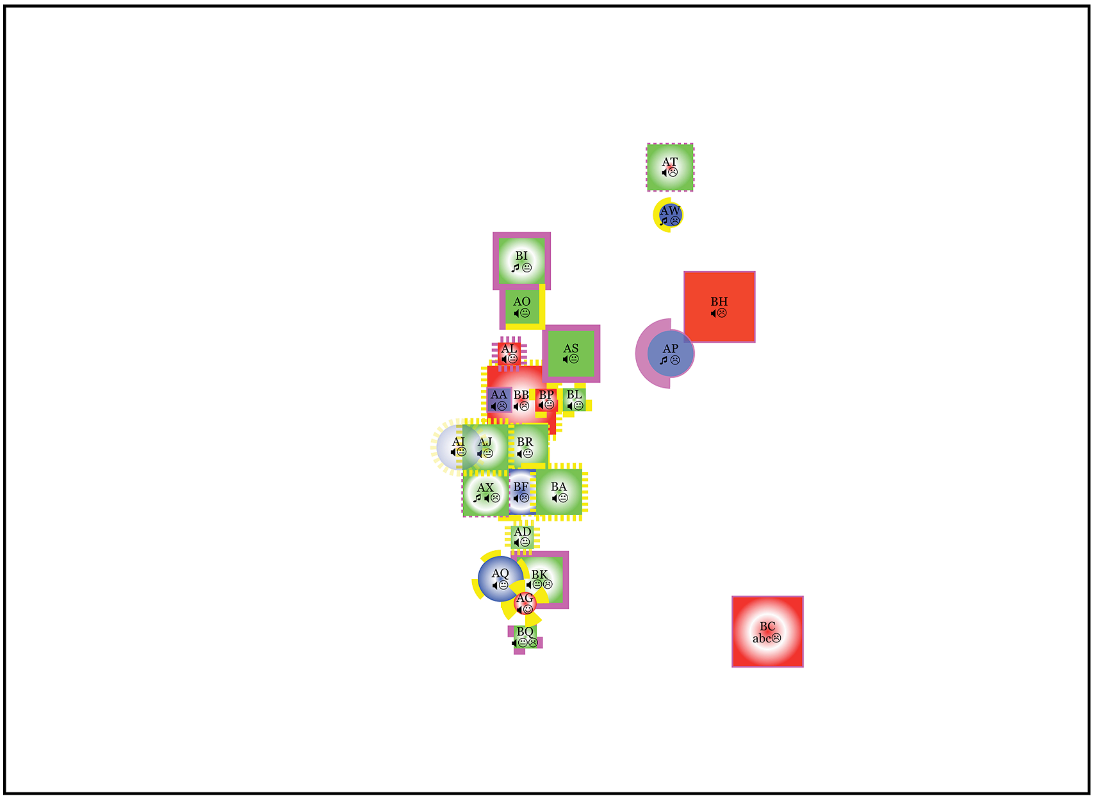


Figure 7-5: Listeners soundscape map (Design 02: Short Film)

The listeners were *unaware* of 8 out of the 12 sound events that were considered to be both *displeasing* and *negative* by the designer. The sound design did not contain any sound events considered to be either aesthetically *pleasing* or emotionally *positive* by the designer, although the listeners considered the 'bells instrumental' (AG) to be both *pleasing* and *positive*. The motion of sounds, either through panning left to right or through depth cues such as reverb, were not recalled by the listeners. The designer made use of these cues for sound events that represented movement such as footsteps, as well as for enhancing natural decays with distorted screams and pads. The majority of the sound events were located in the centre of the x-axis, and this corresponded with the listeners' recollections. The designer had used the full potential depth, whereas listeners described a slightly shallower range.

The designer considered the majority of the sound events to be spectrally *mid* (29), with 11 being *high* and only 5 being *low*. Listeners were aware of four out of five of the low sounds (blue fill), although there was little evidence for an increased awareness of *high* (red fill) or *mid* sound events (green fill). Greater use might have been made of the potential frequency range within the design.

Just under half of the sound events experienced by the listeners were considered to have the same dynamics (11), as classified by the designer, the remainder were evenly split between being softer (6) or louder (6). Events such as footsteps and breathing were considered to be louder, whereas music such as the ‘bells instrumental’ and the ‘dub music’ were softer. Sound events closely associated with the narrative, as in the case of the ‘distorted scream’ and the ‘gun shot’ were clearly heard and recalled, but sound events which supported visible actions that were not important to the narrative were not recalled, as in the case of ‘typing confirmation beeps’ and ‘leather clothing’ (AE and AH).

7.2.3 Design 03: Soundscape Composition

The third design was for a Soundscape Composition, this took the form of a 30 second composed piece made up of music and sound effects. A soundscape composition differs from a soundscape in that it is a constructed composition intended to convey its ‘environmental context’ (Truax, 2001). The composition consisted of a person playing a flute beside a stream, with birds and some disturbance (see Table 7-4). The design has a short narrative of a musician who starts to play a happy piece that changes to a sad piece, we also hear the flautist hit a rock that bounces away, as well as stand on a leaf.

Code	Description	Code	Description	Code	Description
AA	Water trickling	AF	Type 3 Bird	AK	Slight leaf crunch
AB	Flute music A	AG	Odd chirp	AL	Big leaf crunch
AC	Flute music B	AH	Loud chirp	AN	Swirling disturbance (Airplane)
AD	Type 1 Bird	AI	Rock first hit	AO	Big disturbance (Airplane)
AE	Type 2 Bird	AJ	Rock bounce	AP	Far big disturbance (Airplane)

Table 7-4: Sound events (Design 03: Soundscape Composition)

The designer took a sparse, spatially dispersed approach to the design, which was created for this study (see Figure 7-6). Only a single sound event (AJ) moved, and full use was made of the width and depth cues. The designer included 16 sound events and classified 12 as *aware*. The four sound events that the designer was *unaware* of were an ’odd chirp’, a ‘slight leaf crunch’ made by the flautist’s boot and a ‘far air’ and ‘swirling disturbance’ made by an airplane. Only three of the sound events were regarded as *informative* (AA, AB and AC). The designer also only considered three of the sound events to be *pleasing* (AA, AB and AD), three as *neutral* (AE, AO and AP) and six as *displeasing* (AC, AF, AH, AI, AJ and AL).

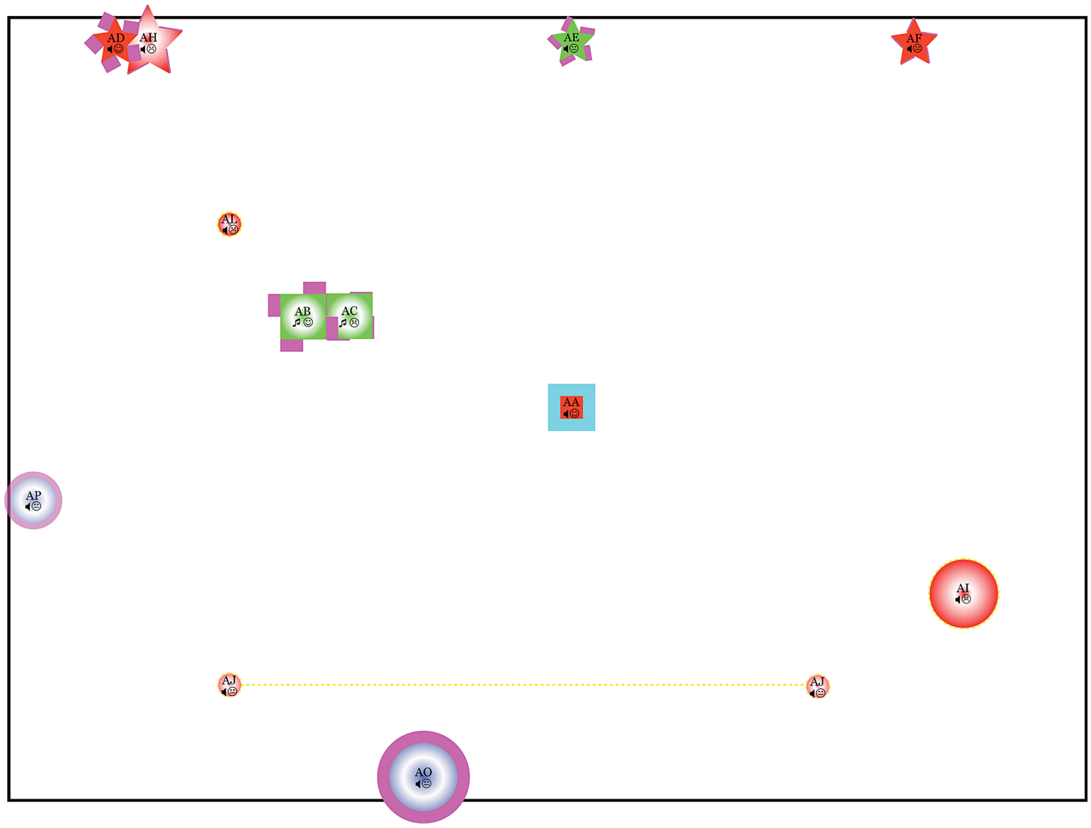


Figure 7-6: Designer’s soundscape map (Design 03: Soundscape Composition)

The designer made full use of the material, interaction, spectral, temporal and dynamics attribute scales. *Gas* (AB), *liquid* (AA) and *solid* (AL) sound events were included, along with *impulsive* (AI), *intermittent* (AC) and *continuous* (AO). Sound events were spread spectrally across *high* (AF), *mid* (AE) and *low* (AP), as well as being temporally *short* (AJ), *medium* (AC) and *long* (AD). The designer also made full use of dynamics, including *loud* (AI), *medium* (AE) and *soft* (AA). The designer adopted an approach used in the music industry when mixing sounds together, ensuring that all of the available attributes of sound events are spread across a wide range of values so that everything can be clearly heard (Owsinski, 1999).

The listeners were *aware* of nine of the sound events (see Figure 7-7). Listeners combined the 'flute music' into a single sound event, as they reported anecdotally that they could not tell where the first piece (AB) finished and the second (AC) started. The designer had separated the two sections as he considered the first piece to be *pleasing* and *positive*, and the second to be *displeasing* and *negative*. A similar grouping problem occurred with the third type of bird (AF) that was not identified as being distinct from either the first (AD) or second type (AE), and also with the 'airplane disturbance' (AM, AN, AO and AP).

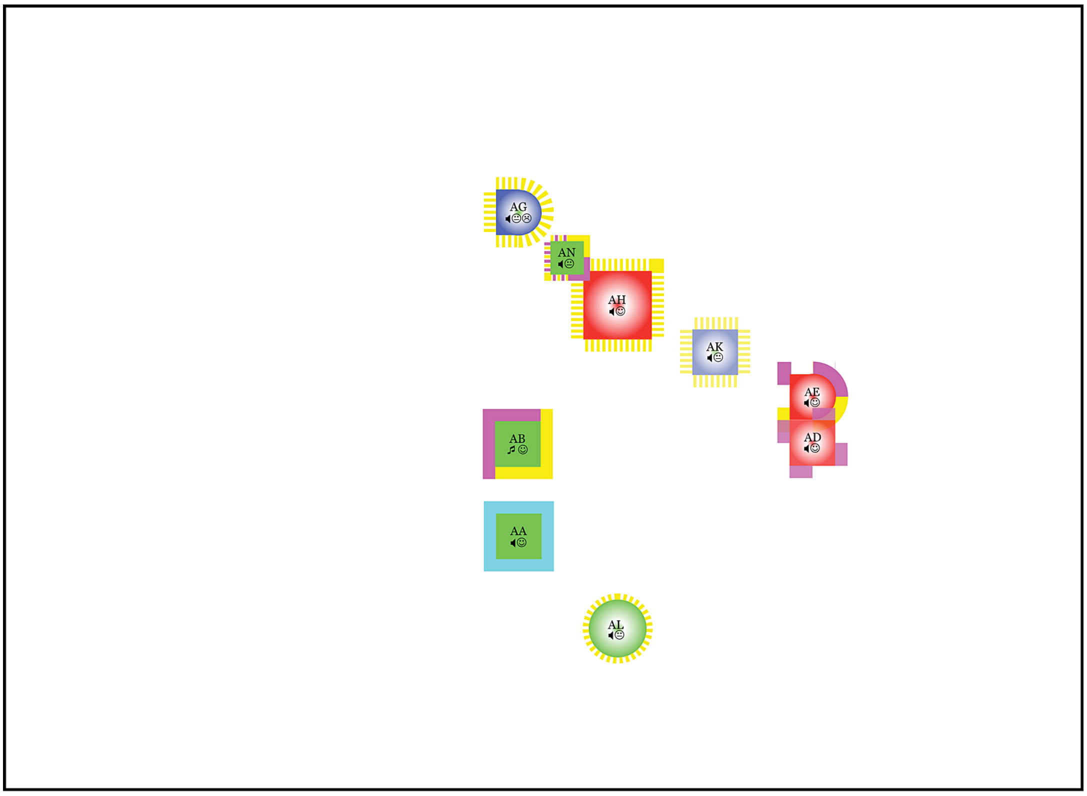


Figure 7-7: Listeners' soundscape map (Design 03: Soundscape Composition)

Spatially, the listeners grouped the sound events closer together than the designer did. The listeners experienced both a narrower width and a shallower depth. Six sound events were thought to be *informative* by the listeners (AA, AB, AD, AH, AK and AN), they considered two sound events to be both *informative* and *neutral* (AE and AG). The listeners also found all of the sound events that they were *aware* of to be either *pleasing* (6) or *neutral* (3), as visualised by border width, 8 point representing *pleasing*, 4 point for *neutral* and 1 point for *displeasing*.

Listeners disagreed about the material for the ‘flute music A’ (AB), the ‘type 2 bird’ (AE), and the ‘swirling disturbance’ (AN), all of which were classified as both *gas* and *solid*. The designer considered AB and AE to both be *gas*, and he was *unaware* of AN. There was also disagreement between the listeners over the interaction for AE and AN, both of which were regarded as *intermittent* and *continuous*. The designer classified AE as *intermittent*. There was no confusion regarding the *impulsive* sound events such as the ‘odd’ and ‘loud’ chirps (AG & AH). The listeners found the Soundscape Composition to be more *informative*, *pleasing* and *positive* than the designer. They grouped some sound events together and only missed the events associated with the ‘rock hit’ and ‘rock bounce’ (AI & AJ).

7.2.4 Design 04: Sonification

The fourth design was for a sonification, an acceleration trace from a four man coxless rowing team was sonified using a continuous tone, which varied in pitch, rising as the acceleration of the skiff (coxless four) increased. The technique was used to monitor the various stages of a rowing cycle in order to help improve the athletes’ rowing technique (see Table 7-5).

Code	Description	Code	Description	Code	Description
AA	Acceleration trace	AD	Front catch	AG	Front reversal
AB	Drive-phase	AE	Mid catch	AH	Back reversal
AC	Recovery-phase	AF	End catch		

Table 7-5: Sound events (Design 04: Sonification)

The sonification is to help the athletes minimise the forward reversal when the oars enters the water as well as the back reversal when the oars leave the water. Minimising the reversals helps to increase the speed of the boat (British Rowing, 2010). Parameter mapping sonification was used as it could allow athletes to hear short events such as the reversals that were difficult to see either in real-time or from video recordings.

All of the sound events were panned hard left as part of a stereo track within the video file. Depth cues were included for all of the sound events except for the ‘acceleration trace’ (AA). Slight panning was applied to the ‘drive phase’ (AB), ‘recovery phase’ (AC) and the ‘back reversal’ (AH). The designer classified all of the sound events as *sound effects*, *gas*, *informative*, aesthetically *neutral*, *clear* and emotionally *neutral* (see Figure 7-8).

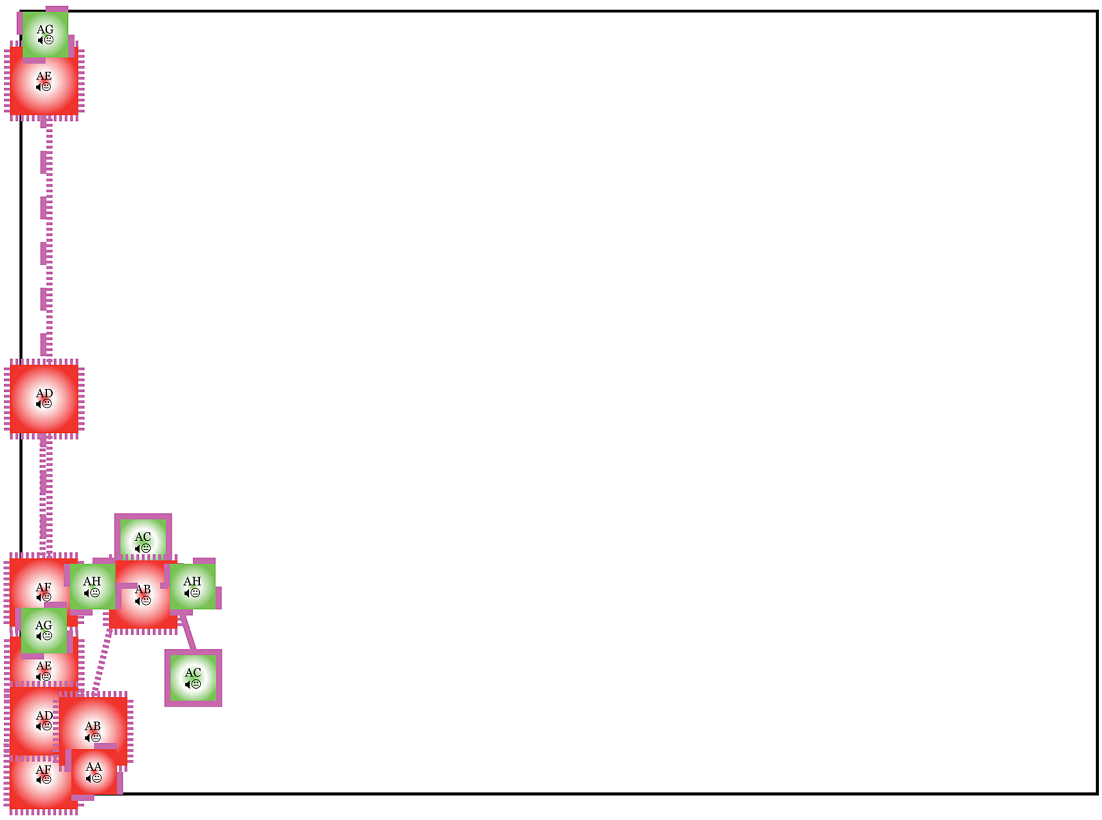


Figure 7-8: Designer’s soundscape map (Design 04: Sonification)

Interactions varied from *impulsive* (AB, AD, AE & AF) to *intermittent* (AA, AG & AH), with a single instance of *continuous* (AC). The designer considered all of the *impulsive* and *intermittent* sound events to be temporally *short* and the *continuous* sound event (AC) to be *medium*. Spectrally all of the sound events were either *high* or *mid*, with all of the *impulsive* sound events being *high*, and all but one of the *intermittent* being *mid*. The single *continuous* sound event was also *mid*. All of the *loud* sound events were classified by the designer as *impulsive*, whereas those with *medium* dynamics were *intermittent* or *continuous*.

Participants watched a 56 second video of the rowers with synchronised sonification, but without the original rowing soundtrack. As the sonification conveyed complex actions, listeners were allowed to refer to the video as often as they wished. The designer provided written explanations of the terms used in the sonification for listeners to refer to (see Table 7-6). None of the participants had any experience of rowing, so all of the listeners were equally unfamiliar with the terminology used.

Acceleration	sine-tone related to the pitch. It represents the acceleration-trace of a rowing boat at stroke rate 38 strokes per minute (=average stroke frequency in a rowing race).
Drive phase	This is the phase from the catch to the extraction
Recovery-phase	This is the phase from the extraction to the catch.
Front catch	where the oar-blade is placed in the water and the rower applies pressure to the oars at the front
Mid catch	where the oar-blade is placed in the water and the rower applies pressure to the oars at the mid
End catch	where the oar-blade is placed in the water and the rower applies pressure to the oars at the end
Front reversal	the rower applies pressure during the mid-drive
Back reversal	the rower applies pressure during the end-drive

Table 7-6: Explanation of terms used for Sonification provided by the designer

The listeners were aware of all eight sound events, and considered all but one (AE) to be *informative* (see Figure 7-9). The 'mid-catch' was classified as being both *informative* and *neutral* by the listeners, who were not trained athletes, so possibly found the short duration of the 'mid catch' difficult to distinguish, from the 'front' and 'end catch'. Spatially all of the listeners also experienced the sound events as being panned hard left, but to a lesser extent than the designer. They also did not consider the sound events to be as close, or experience the variation of depth cues (AB, AC, AD, AE, AF and AG) as identified by the designer.

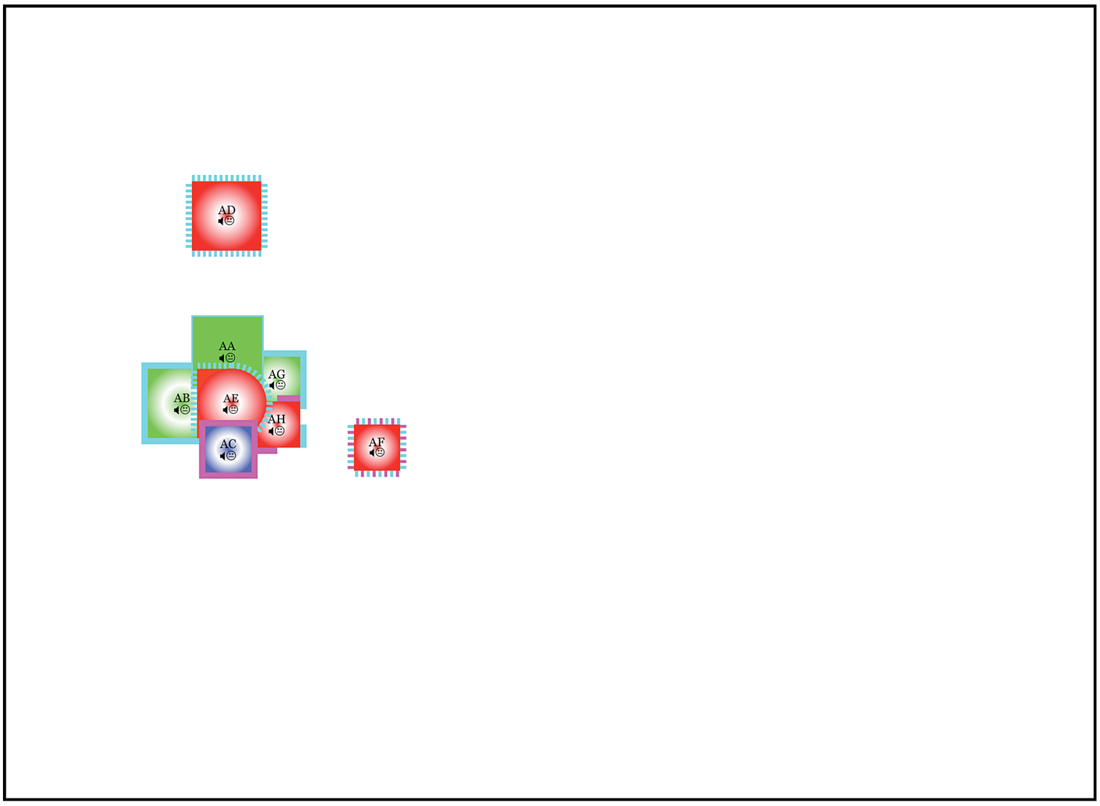


Figure 7-9: Listeners' soundscape map (Design 04: Sonification)

The listeners grouped the materials of the sound events into one *gas* (magenta border), five *liquid* (cyan border), one *solid* (yellow border), and two as both *liquid* and *solid*. Listeners experienced a greater range of spectral attributes than the designer, listeners considered four sound events to be *high*, three to be *mid* and one (AC) to be *low*. This might be considered positive, as it indicated that listeners were able to separate out the differences in pitch slightly more than the designer. There was agreement between the designer and the listeners about clarity and emotions, all of the sound events were found to be both *clear* and emotionally *neutral*. The designer considered each sound event to be aesthetically *neutral*, and the listeners agreed with all but one, the 'acceleration trace' (AA), which was *displeasing*.

The sonification could be considered as successful, as almost all of the sound events were considered *informative* and listeners were able to distinguish between the differences in pitch. The range of pitch variation could be increased so that it extended into the *low* range and some form of panning might be considered, if only to move the sound events into the centre of the stereo field.

7.2.5 Design 05: Simulation

A 7 minute and 25 second simulation of the soundfield of a multimedia laboratory and its immediate environment was created using a non-linear sequencing model called GeoGraphy (Valle, Lombardo and Schirosa, 2009). GeoGraphy had previously been tested by comparing simulations with recordings from real environments and asking listeners to identify which was which. The designer, unlike others in this study, decided to group multiple sound events together in order to minimise the total required for classification. Some sound events represented simple actions like 'shifting of a chair', whilst others were complicated constructs, such as 'sounds emitted by a key-holder while someone is walking in the passage, other noises, some steps, aeration ducts' (see Table 7-7). This grouping of sounds reflects the nature of the software used, which triggers zones made up of atmospheres, events and sound objects grouped together. Each sound event is a single zone, and the descriptions represent the sounds that were used to create the zone.

Code	Description	Code	Description	Code	Description
AA	It represents the opening of the front door	AH	It consists in the tonic of the bottom part of the laboratory, it coincides with ducts of the thermic station	AO	Some people of the staff is talking with some students
AB	It consists in the tonic (keynote) of the laboratory, it coincides with the wheels of personal computers.	AI	The sounds produced by a wash-basin, in the bathroom	AP	A member of the staff is speaking with a student
AC	The sound concerning the use of a pc: keyboard and mouse.	AJ	The sounds produced by the use of a towel, in the bathroom	AQ	Sounds emitted by a key-holder while someone is walking in the passage, other noises, some steps, aeration ducts
AD	Sounds emitted by some people that walk and speak.	AK	Someone that throw a towel in a waste basket	AR	Some steps in the passage, sounds emitted by aeration ducts
AE	Sounds emitted during a copy of a document with the machine	AL	It consists in the tonic of the laboratory which are near some servers, it coincides with wheels	AS	Shifting of a chair
AF	It is produced by drumming fingers on a desk	AM	The emission sounds of a television: a woman's voice		
AG	Sounds emitted in an auditorium, where is projected a movie	AN	It consists in the tonic of the laboratory, without a frequent use of pc; it coincides with aeration ducts and wheels of personal computers		

Table 7-7: Sound events (Design 05: Simulation)

The designer considered the nine of the sound events to be *informative*, five *neutral* and five *uninformative* (see Figure 7-10). Sound events such as the photocopier (AE) and the film (AG) were *informative*, the sounds of people's actions such as drying their hands (AJ) or footsteps (AR) were *neutral*, and room tones (AL & AN) were *uninformative*. Almost all of the speech was considered to be *pleasing* by the designer (AD, AG, AO & AP), however, the 'woman's voice' emanating from a television (AM) was rated as having *neutral* aesthetics and *negative* emotions.

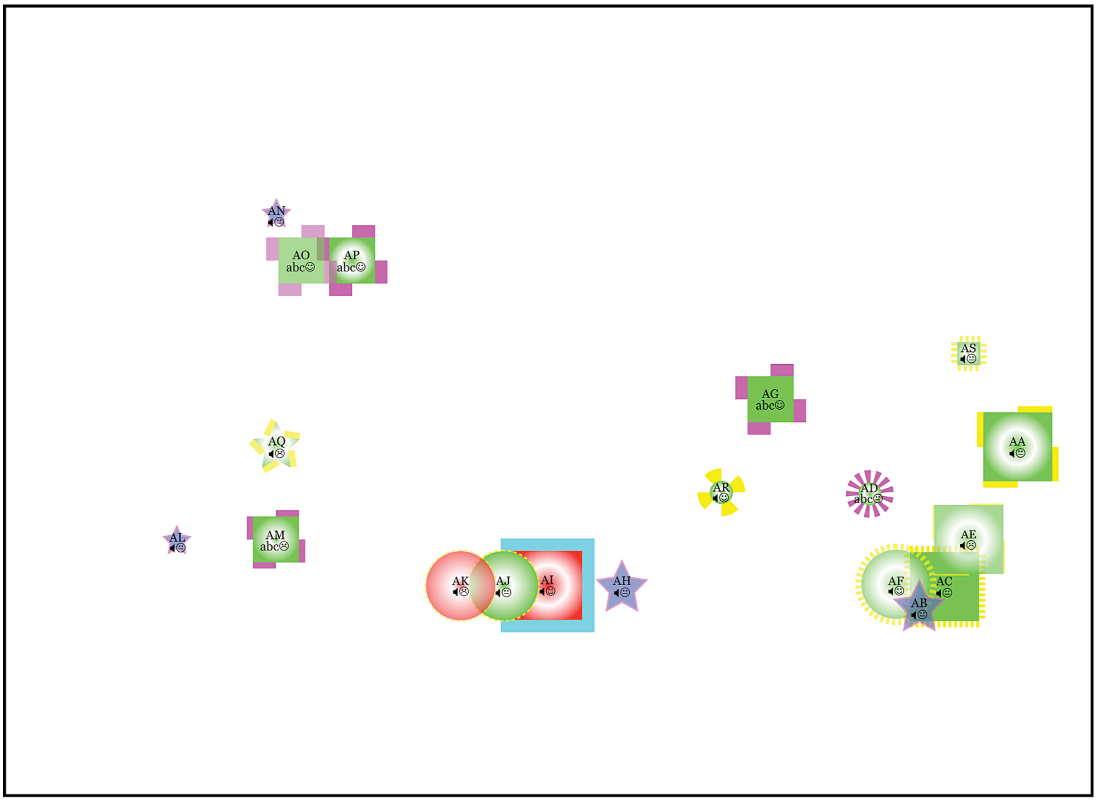


Figure 7-10: Designer's soundscape map (Design 05: Simulation)

For this study listeners were questioned after listening to the entire sound file, without referring back to it. Despite the long duration of the file the listeners stated that they were *aware* of 17 of the 19 sound events (see Figure 7-11). Listeners were *unaware* of two of the sounds associated with the washing of hands (AJ & AK). The designer classified the sound events as being spatially dispersed, while the listeners experienced them closer together. This was possibly due to the way in which the sound events were required to be recalled. It might be expected that the spatial cues be recalled more accurately if the file was shorter, or if listeners made note of their impressions whilst listening. Referring to notes may alter the levels of awareness of specific sound events, and possibly other attributes such as content. The listeners thought that 14 sound events were *informative*, 1 *neutral* and 2 as both *informative* and *neutral*. Listeners may be trying to make sense of what they were listening to, and constructing a narrative in order to understand the sequence of sound events. This might be attributed to the number of sound events that the listeners found to be *clear* (15), which contrasts with the designer who rated 9 of them as *clear* and the remaining 10 as *unclear*.

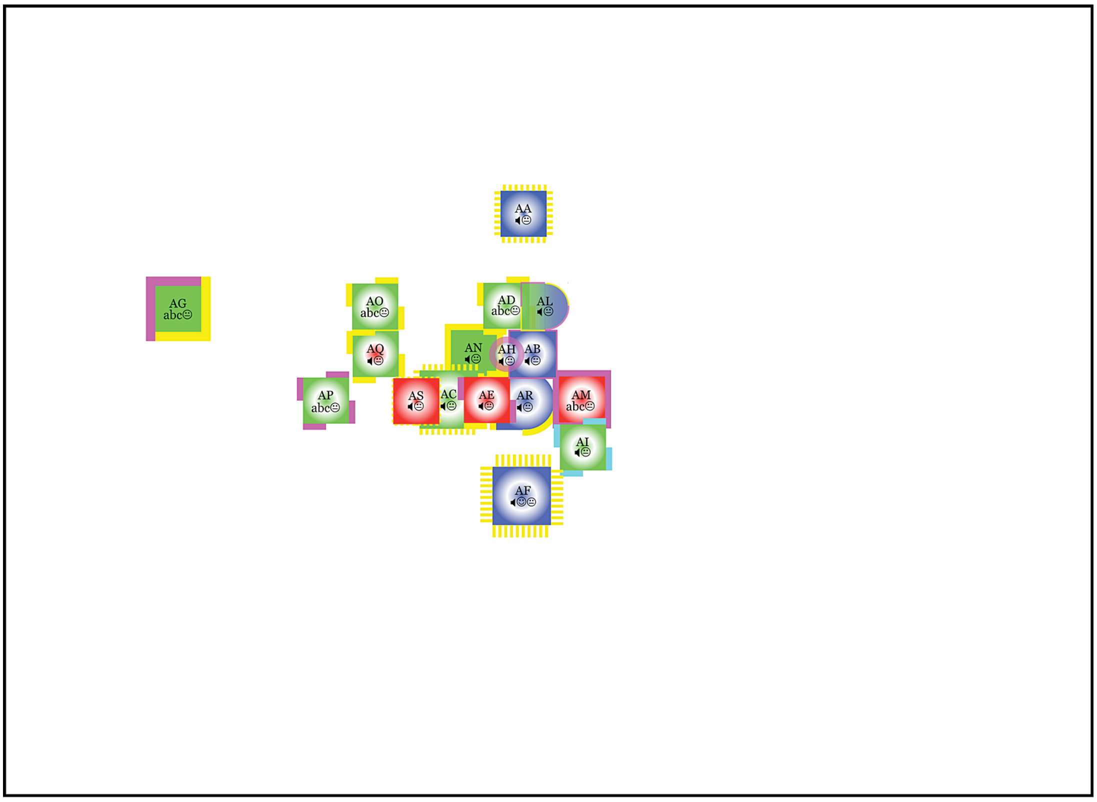


Figure 7-11: Listeners' soundscape map (Design 05: Simulation)

The sound types were consistent, with all of the sound events being categorized identically by the both the listeners and the designer. Four of the sound events were considered to be *speech*, with the remainder (15) being *sound effect*, there were no instances of *music*. The designer considered the sound events to have a greater difference in dynamics, 7 were *loud*, 7 *medium* and 5 *soft*. The listeners found 3 to be *loud/medium* (AF and AR), 1 *soft* (AH) and the remainder (16) *medium*. This might suggest that the variation in dynamics is too subtle, and that a greater difference needs to be applied in order to convey the range intended by the designer.

The designer regarded both the aesthetics and the emotions to be more varied than the listeners considered them to be. The aesthetics were split almost evenly between the 3 options by the designer, whereas 11 of the 17 sound events were aesthetically *neutral* for the listeners. This was even more pronounced for the emotions with 16 out of the 17 sound events judged to be *neutral* by the listeners, compared to 9 by the designer. As a simulation the sound design might be considered successful, listeners were aware of almost all of the sound events, and considered the majority of them to be *informative* and *clear*. If the intention was to convey the subtleties of the aesthetics and the emotions then the design has been less successful, as the majority were rated as *neutral* for both attributes.

7.2.6 Design 06: Games Sound Effects

The sixth design utilised sound effects used for a commercially released console video game. All of the sound events were part of a game company's sound library, for designers to use in the construction of games. Eight separate audio files were included, the shortest was less than 1 second long and the longest 1 minute and 19 seconds. Half of the files were single sound events, which were all recordings of a female voice speaking single words, the remaining four were atmospheric constructs with between three to five sound events each (see Table 7-8). Sound events included dogs, 'birds', 'water', a 'kiss' and a 'hit'. The constructs suggested a limited narrative which listeners were free to interpret, being informed only of the component sounds.

Code	Description	Code	Description	Code	Description
AA	Female voice 'Bye' (4051)	AG	Water (95532)	AM	Birds (95545)
AB	Female voice 'Hello' (4053)	AH	Dog growl (95533)	AN	Waterfall (95545)
AC	Female voice 'Tomorrow' (4056)	AI	Dog barking (95533)	AO	Voice (95545)
AD	Female voice 'Tonight' (4057)	AJ	Water (95533)	AP	Birds soft (95552)
AE	Dog growl (95532)	AK	Kiss (95533)	AQ	Birds high and loud (95552)
AF	Barking dog (95532)	AL	Hit (95533)	AR	Waterfall (95552)

Table 7-8: Sound events (Design 06: Games Sound Effects)

The designer considered all of the 18 sound events to be *informative* (see Figure 7-12). All of the sound events were either *speech* or *sound effects*. Full use was made of the range of the remaining attributes. For the material attribute *gas* was predominantly used to classify the voices, most of the 'birds' and some of the dogs. *Liquid* was consistently chosen for the 'water', and *solid* was applied to the 'kiss', 'hit' and some of the dog sounds. There was increased consistency for the interaction classification. The designer used *continuous* to classify only the water sounds (AG, AJ, AN & AR), all of the birds were *intermittent* (AM, AP & AQ), and all of the voices (AE, AB, AC, AD &AO) were *impulsive*. Only the dog sounds were inconsistent, being either *impulsive* or *intermittent* (AE, AF, AH & AI). The majority (10) of the sound events were temporally *short*, only 3 were *medium* (AC, AD & AQ), and 5 were *long*. Atmospheric *sound effects* such as the waterfall tended to be temporally *long*, whereas *speech* was either *short* or *medium*.

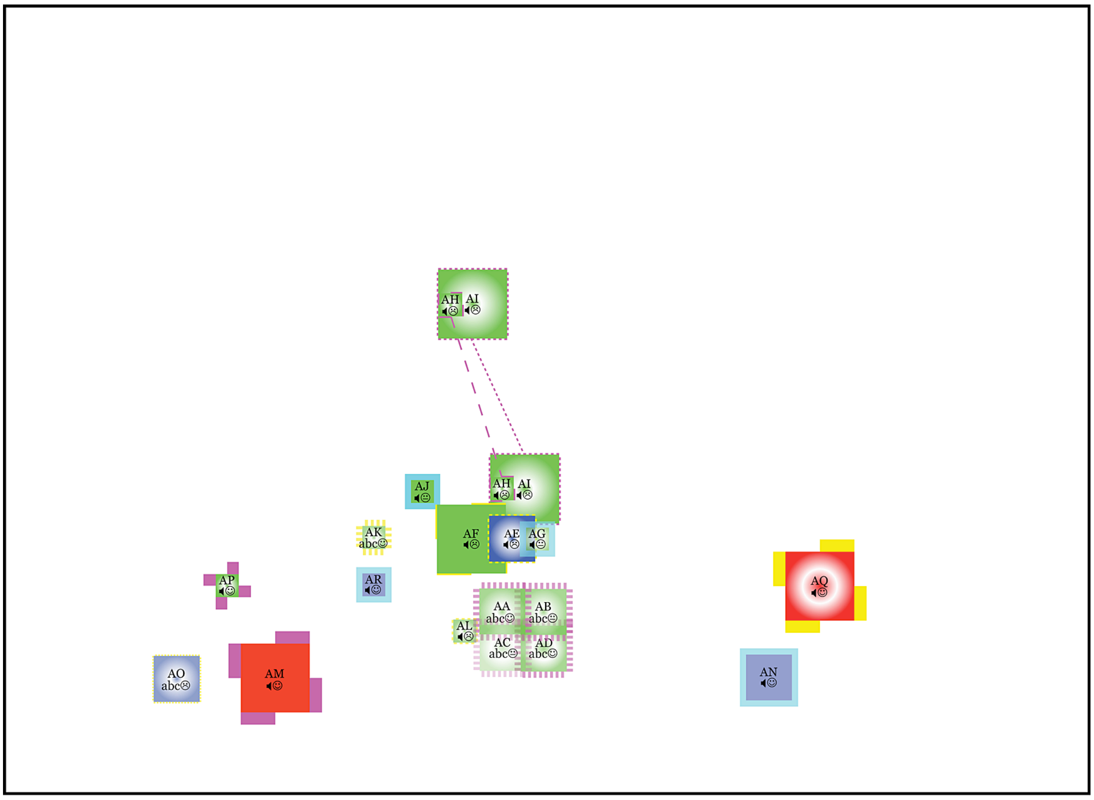


Figure 7-12: Designer's soundscape map (Design 06: Games Sound Effects)

The order in which the listeners experienced the audio files was randomised, and listeners were questioned after each file was played. This meant that they were aware of all of the sound events, replicating a game sound designer auditioning a library to select appropriate sounds for inclusion within a game (Rogers, 2010). All of the sound events were grouped together in a single map for ease of comparison (see Figure 7-13).

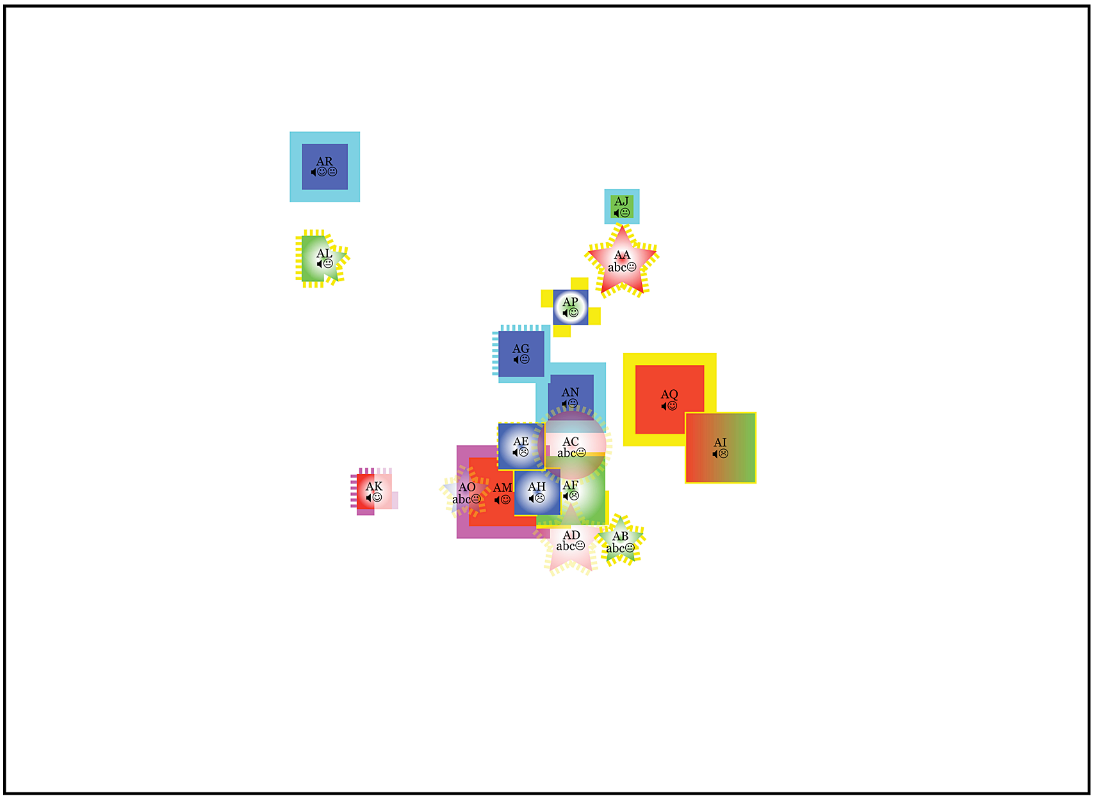


Figure 7-13: Listeners' soundscape map (Design 06: Games Sound Effects)

Spatially the designer considered the sound events to be spread slightly wider across the stereo field when compared to the listeners' responses. In contrast the listeners reported a slightly greater range of depth, without experiencing the movement of the dog barking and growl (AH and AI). One of the sound events (AR) was *soft* and close for the designer, but for the listeners it was *medium* and far.

The type was consistent for both the designer and the listeners, only a single sound event differed. The 'kiss' (AK) was classified as *speech* by the designer and as a *sound effect* by the listeners. There was greater contrast with spectral attributes, the designer considered 12 out of the 18 sound events to be *mid*, compared to the listeners who found the same number to be *high* or *low*. The designer weighted the sounds between *medium* and *soft*, whereas the listeners experienced *medium* and *loud*. The aesthetics had a greater level of concurrence, with half of the sound events being *neutral* for both groups, but the emotions were more diverse for the designer than the listeners.

The listeners rated only 12 of the 18 sound events as being *informative*. Four were found to be *uninformative*, 1 was *neutral* and 1 was both *informative* and *uninformative* (AL), illustrating that there were contradictory responses from the listeners for this sound event. Each of the sound events classified by the listeners as *uninformative*, as well as the single *neutral* sound event, were *speech*. Three of these were also *unclear*, whilst the remaining two were *clear*. The designer regarded only one of the sound events as *unclear* (AC).

When considering the sound design as a whole then the *sound effects* are successful, they are informative and convey the required emotions accurately. There is a difference for the two groups with regards to the *speech*. The emotions are not conveyed, being consistently considered as *neutral* by the listeners, as well as predominantly *uninformative*. However, the designer judged them to be both *informative*, conveying either *positive* or *negative* emotions. This is perhaps due to a problem with the dialogue delivery rather than the sound design.

7.2.7 Design 07: Radio Drama

The seventh design was a 42 second section from a radio classic-crime drama. The designer had created all of the sound effects manually using physical props, which he layered on top of a previously recorded dialogue track (see Table 7-9). Manual sound effects are a popular way of creating radio sound effects as they are easily synchronised with actions and dialogue (Mott, 2009). The scene covers the arrival of a safecracker to a home run by crooks posing as aristocrats. The criminals need some papers that have been locked in a desk safe in their house. The action includes a crook sending off a henchman (AB), a doorbell ringing (AA), a butler answering a door (AE & AF), a safecracker coming in (AF & AH), trying the safe (AI), and then being greeted by the crooks/aristocrats (AL & AN).

Code	Description	Code	Description	Code	Description
AA	Ringing bell (doorbell)	AF	Sid's voice	AK	Woman's footsteps enters
AB	Butler's Voice	AG	Door closes	AL	Woman's voice
AC	Girl's voice	AH	Sid's footsteps	AM	Chetwood enters
AD	Butler Footsteps	AI	Safe Door jiggled	AN	Chetwood's voice
AE	Door opens	AJ	Tools put down		

Table 7-9: Sound events (Design 07: Radio Drama)

There were five speaking characters and all of the nine remaining sound events were *sound effects*. The files were recorded in mono, so there were no panning cues and the designer considered there to be minimal depth cues as well (see Figure 7-14). The appearance of panning in the designer's map is an artefact of the mapping process to limit the amount of visual overlapping. Mono compatibility is an important issue for broadcast listeners, as summing stereo signals can mean that sounds are artificially loud if they are panned to the centre compared to left or right (Newell, 2008).

The majority, nine, of the sound events were temporally *short*, with five being *medium* and none classified as *long*. The *sound effects* were predominantly *short* (AE, AG, AH, AI, AJ, AK & AM) whereas the *speech* was more evenly split between *short* (AC & AL) and *long* (AB, AC & AN). Twelve of the sound events were spectrally *mid*, with three *high* (AA, AC & AJ) with none being *low*. It is important in radio production for the content to be predominantly in the *mid* range, as small radios often cannot reproduce bass efficiently (Rose, 2009). Radio is often listened to at a distance, or even in the next room and all of the information content has to be in the *mid* range in order for it to remain clearly audible (Katz, 2002). Confining most of the content in radio to *mid* frequencies ensures intelligibility, as listeners' are more sensitive to this range of frequencies (Reese, Gross & Gross, 2006).

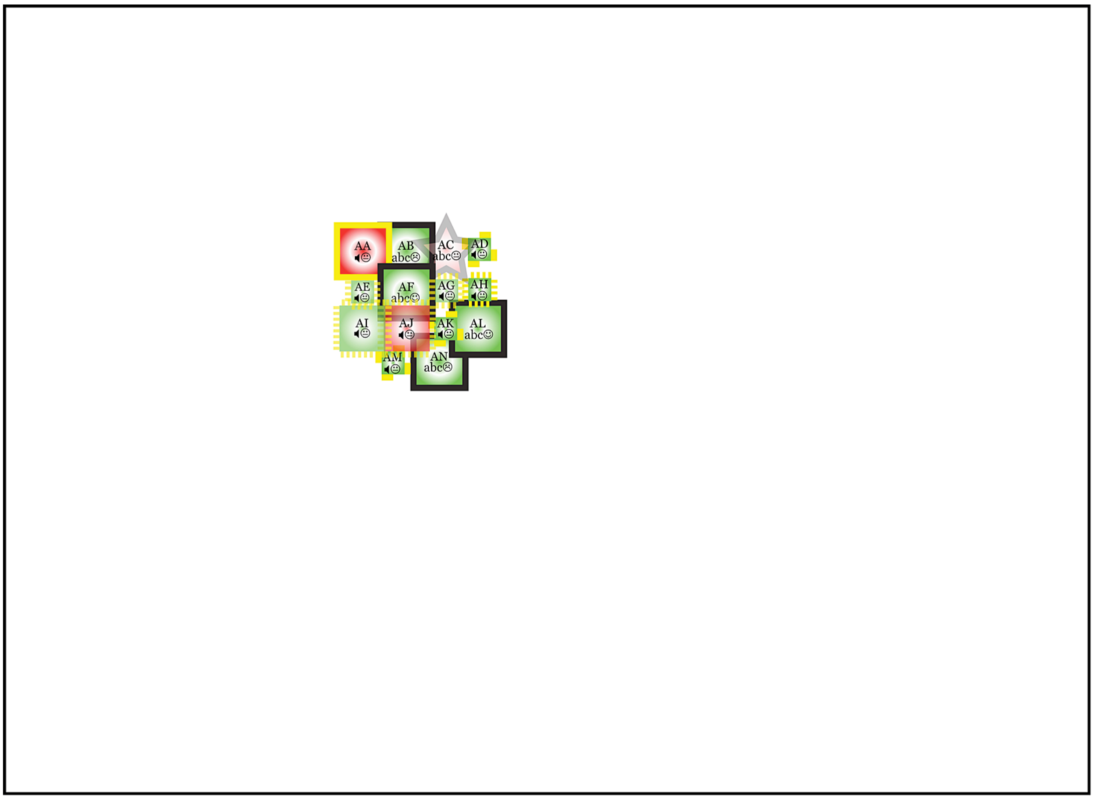


Figure 7-14: Designer's soundscape map (Design 07: Radio Drama)

The designer requested that listeners be played the dialogue only version prior to the one with the sound effects, so that listeners had a greater understanding of the narrative. Without being told that the file was in mono, the listeners also judged the sound events to be in the centre of the stereo field (see Figure 7-15). Listeners did experience a greater sense of depth than the designer, which could be thought of as a positive effect as this reflects the action within the drama, where, for example, characters enter a room and congregate around a desk.

All of the sound events, except one, were judged to be *informative* by both the listeners and the designer. The 'girl's voice' (AC) was classified as *uninformative* and *unclear* by the designer, the listeners were *unaware* of it. The designer omitted to classify the material or interaction for the *speech*, which the listeners almost consistently considered to be *solid* and *continuous*. The designers and listeners also concurred with the aesthetics, which were uniformly classified as *neutral*.

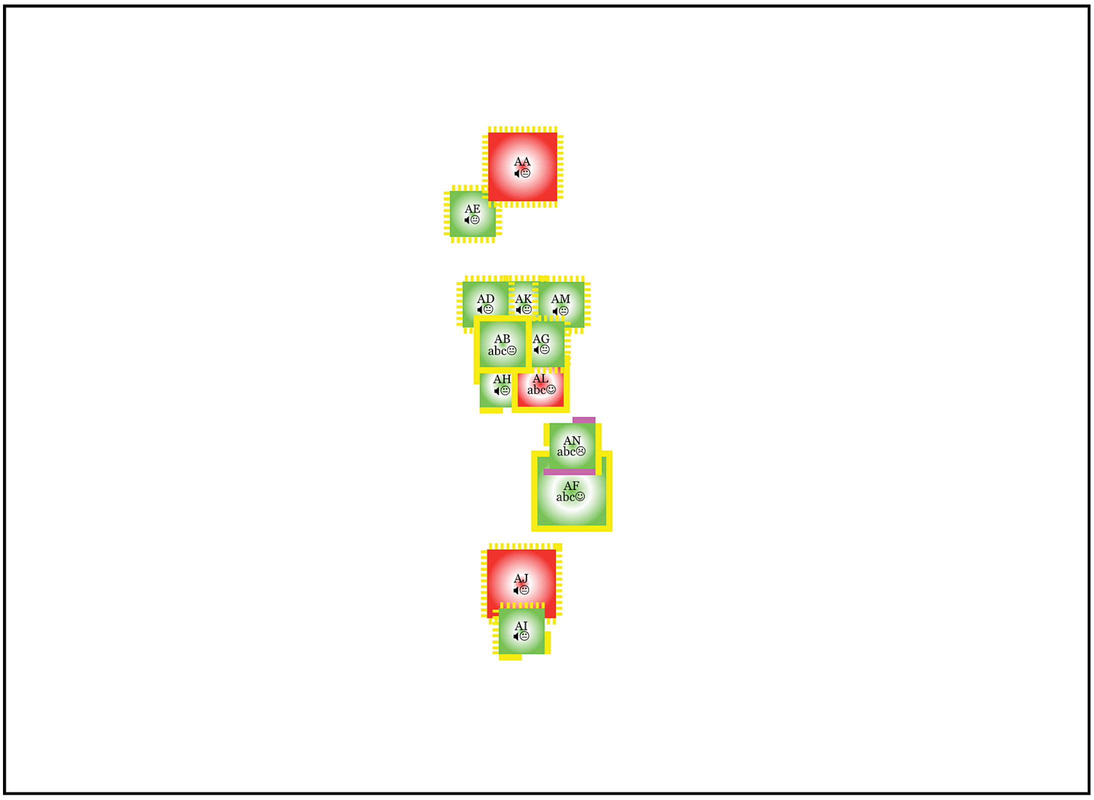


Figure 7-15: Listeners' soundscape map (Design 07: Radio Drama)

The listeners found all of the sound events to be *clear*, whereas the designer thought that the sounds associated with the doors (AE, AG, AI) and the tools (AJ) were *neutral*. As mentioned above the only *unclear* sound event went unnoticed by the listeners. Both the designer and the listeners only attributed emotions to the *speech*, all of the *sound effects* were classified as *neutral*. The two characters considered to be conveying *positive* emotions by the designer, Sid (AF) and the Woman (AL), were also judged to be *positive* by the listeners. Only one of the two characters' voices that the designer considered to be *negative* (AB and AN) was regarded as being *negative* by the listeners (AN). All of the sound events were considered to be *soft* or *medium* by the designer, whereas they were either *loud* or *medium* for the listeners. This is possibly due to a lack of calibration between the designer's equipment and the reproduction equipment used for the study, the listeners might have been listening to a higher reproduction levels than the designer.

The sound design might be thought of as being successful as all of the sound events, which the listeners were *aware* of, were considered *informative* by the designer. In addition there was a greater sense of depth perceived by the audience than intended by the designer without any perceived loss of clarity or dynamics.

7.2.8 Design 08: Audiologos

Design 8 consisted of a series of Audiologos. Audiologos are a type of sonic branding and are sometimes referred to as sonic logos. Audiologos are commonly short musical phrases played on a single instrument, occasionally with sound effects and/or dialogue (Jackson, 2003). Four different Audiologos were trialled, varying in length from 4 to 6 seconds (see Table 7-10). *Speech* and *music* were present in all four audiologos, but *sound effects* were only included in three. The first audiologo was named *Classico* and consisted of a discordant door, some plucked strings and the company's name (xxxxx). The second audiologo was called *Folcklore* and had 'birds', a 'classical guitar', a drum and the company's name. The third audiologo (*Piano*) was the simplest containing only a 'piano' and the company name. The last audiologo (*Piano2*) was the most complicated with a 'door', 'voice', 'wood knocks' 'drum' and 'piano'. It is common to present a client with a number of audiologos so that they can conduct listener tests to identify which is most effective. Alternatively, different versions of audiologos are created for different target markets (ibid.).

Code	Description	Code	Description	Code	Description
AA	Wooden Country side door (clasico)	AF	Leather bass drum (Folcklore)	AK	Voice "xxxxx" (Piano2)
AB	Plucked Strings (clasico)	AG	Voice "xxxxx" (Folcklore)	AL	Wood knocks (Piano2)
AC	Voice (clasico)	AH	Piano (Piano)	AM	Leather bass drum (Piano2)
AD	Birds (Folcklore)	AI	Voice "xxxxx" (Piano)	AN	Piano (Piano2)
AE	Classical guitar (Folcklore)	AJ	Door (Piano2)		

Table 7-10: Sound events (Design 08: Audiologos)

The designer supplied mono files, which is common for this industry. Audiologos typically form part of an advert or some other promotional material, where listeners are often at a large distance from the reproduction source, which means that any stereo output is summed to mono by the time it reaches the listeners ears (Huber & Runstein, 2010). Designing the logos in mono prevents any phase issues from being introduced, which can affect clarity, and therefore the impact of the audiologo. The sound events were displayed on a single map for ease of comparison (see Figure 7-16). The designer made use of spatial depth cues which were not experienced by the listeners to the same degree, who reported differences in the stereo width. This effect was not a recall artefact, as questioning was conducted after each audiologo had been played.

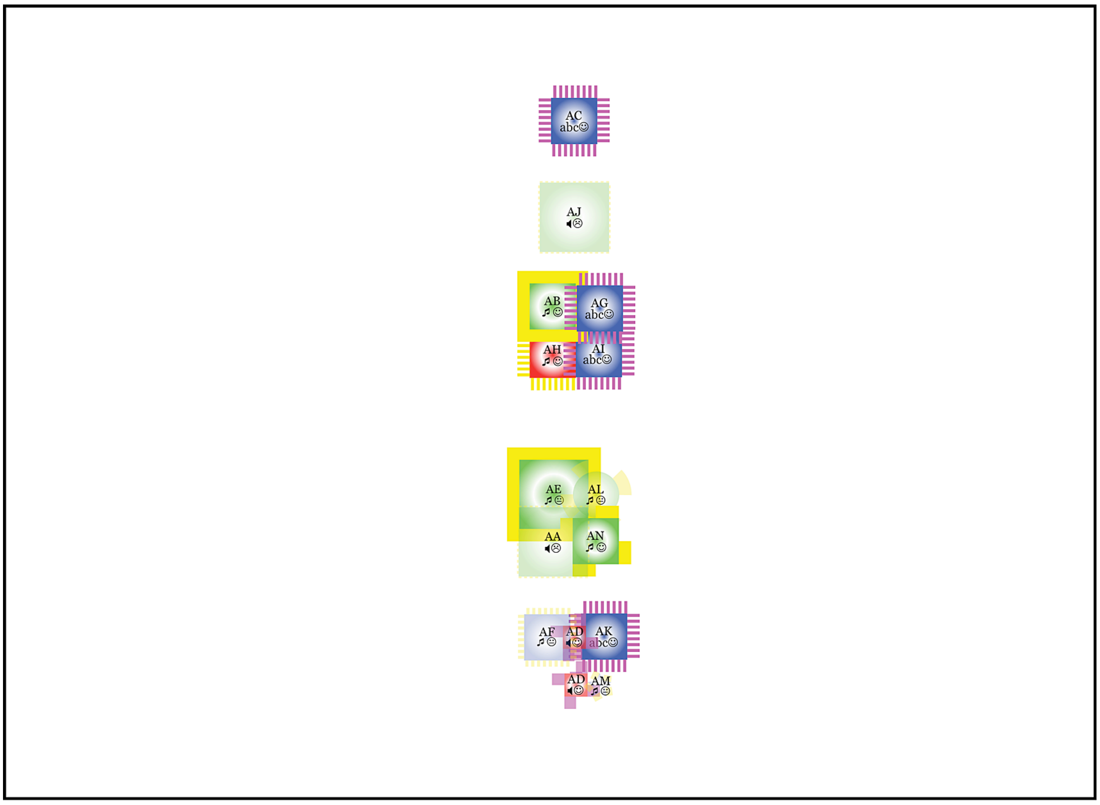


Figure 7-16: Designer's soundscape map (Design 08: Audiologos)

All of the sound events were either *solid* or *gas* and were spectrally spaced to include *high*, *mid* and *low*. The sound events also occupied the full dynamic range from *loud*, through *medium* to *soft*. The predominant considerations for the designer were aesthetics, clarity and emotions. The designer considered all but 4 of the 14 sound events to be *pleasing*, in contrast only 5 were found to be *pleasing* by the listeners. This was reversed when it came to clarity, all but one of the sound events (AL) were considered *clear* by the listeners, whereas the designer classified only eight as *clear*, one as *neutral*, and five as *unclear* (see Figure 7-17). The designer judged the majority, eight of the sound events to be *positive*, this differed from the listeners who thought that half, seven were *neutral*.

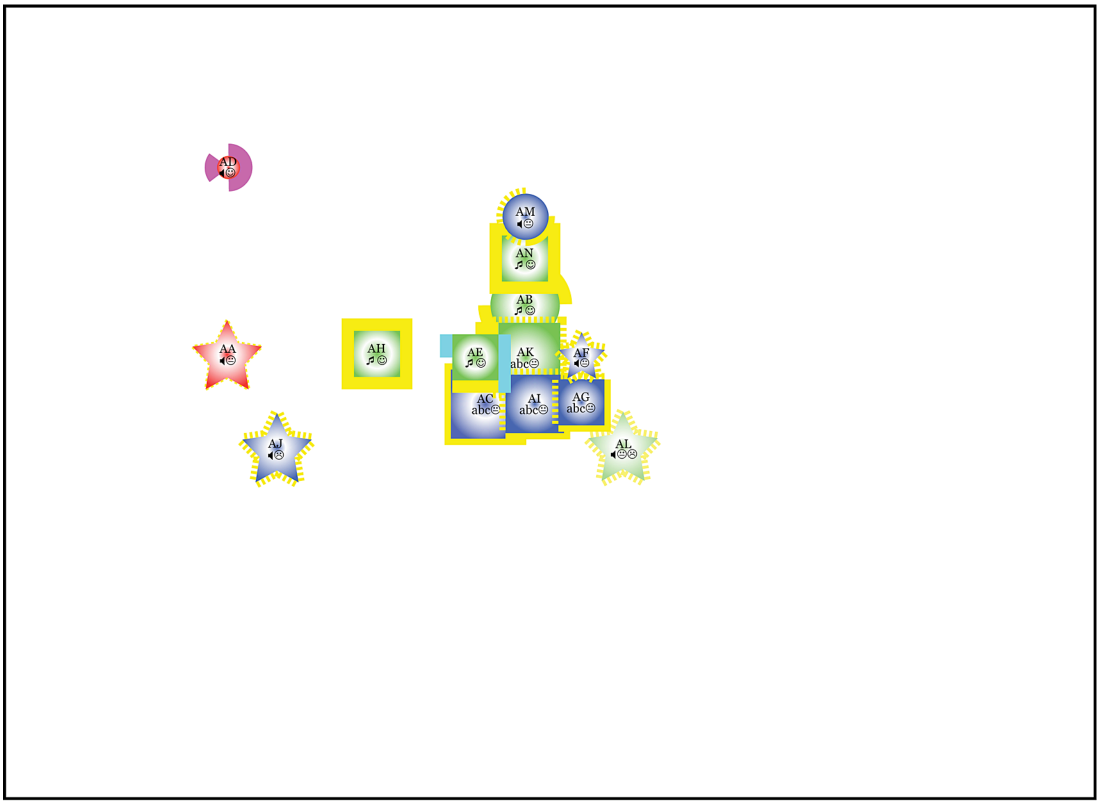


Figure 7-17: Listeners' soundscape map (Design 08: Audiologos)

Listeners were not familiar with the company represented by the audiologos, so were able to judge them purely on what was presented. A key part of the sound design in each audiologo was the voice stating the company name. The designer considered them to be *pleasing* and *positive*, this was not the case with the listeners, who rated them as aesthetically and emotionally *neutral*. As this was an identical sound event for each audiologo it could be suggested that the context of the sound event had no effect on the voice and that it might be worth considering a new recording in order to emphasise the desired *pleasing* and *positive* response. This might have something to do with the *impulsive* nature of the voice as the listeners found the *intermittent* and *continuous* sound events, generally, to be *positive*. The listeners consistently found the musical instrument which played the code to be *pleasing*, *clear* and *positive*, and all but the ‘plucked strings’ (AB) to be *informative* as well. It is important that the music in an audiologo is appropriate as customers have an increased likelihood of purchase if they like the music associated with a brand (North, Hargraves & McKendrick, 1999). The designer included two sound events (AA and AJ), both associated with doors, which he considered to be *displeasing*, *unclear* and *negative*. Both were considered by the listeners to be *clear* and only one (AA) was *displeasing*, whilst the other (AJ) was *negative*.

If a decision were to be made about the most appropriate audiologo based on the results of this study, then the ‘piano’ version should be considered. Both sound events that made up the audiologo were thought of as *informative* and *clear* by the listeners, and there were no *uninformative*, *displeasing* or *negative* sound events as there were in the other three audiologos.

7.2.9 Design 09: Composition

An abstract composition was chosen in order to find out whether the tool could be used for visualising complex conceptual soundscapes that did not have easily identifiable sound sources. The Composition used sounds like ‘ice cascade’, ‘bass rumble’ and ‘ripping detritus drop’ within complex patterns in order to create an immersive environment (see Table 7-11). The 1 minute and 30 seconds of a longer piece (09:42) was chosen, this design differed from all of the other mono/ stereo designs, in that the reproduction utilized a custom eight-channel surround sound system.

Code	Description	Code	Description	Code	Description
AA	(0-5") opening rotary ice cascade	AJ	(24-52") background ambience	AS	(1'04-1'06) rotary rip gesture
AB	(5") match swipe	AK	(24"-43") ice cascade	AT	(1'06) thump
AC	(6-8") intermediary ice cascade	AL	(40-45") water hiss	AU	(1'06-1'10) ice scrape judders spectropen
AD	(8") juddering anacrusis	AM	(45-47") flick-back gesture	AV	(1'11) ice scrape deep judders
AE	(9") ice click gesture	AN	(51-52") transition-click	AW	(1'12-1'20) ice busy-texture
AF	(9-20") rotary ice cascade to fragment	AO	(52"-1'02) rotary cascade	AX	(1'22) ripping detritus-drop
AG	(9-12") intermittent ice clunk activity	AP	(1'-1'02) tearing anacrusis to thump	AY	(1'24) ripping detritus-drop2
AH	(20") closing scrape-glide	AQ	(1'-1'32) bass rumble	AZ	(1'27-1'30) ripping detritus-drop3 to detritus
AI	(23-24") tearing swipe open gest	AR	(1'-1'32) intermediary ice cascade		

Table 7-11: Sound events (Design 09: Composition)

Unlike all of the other soundscape maps, where the listening position was located at the bottom centre of the map, the listener was this time located in the absolute centre (see Figure 7-18). The piece made extensive use of panning, with the majority of the sound events being located in front of the listener. There were also instances where the depth cues varied as well (AE, AH, AI, AZ).

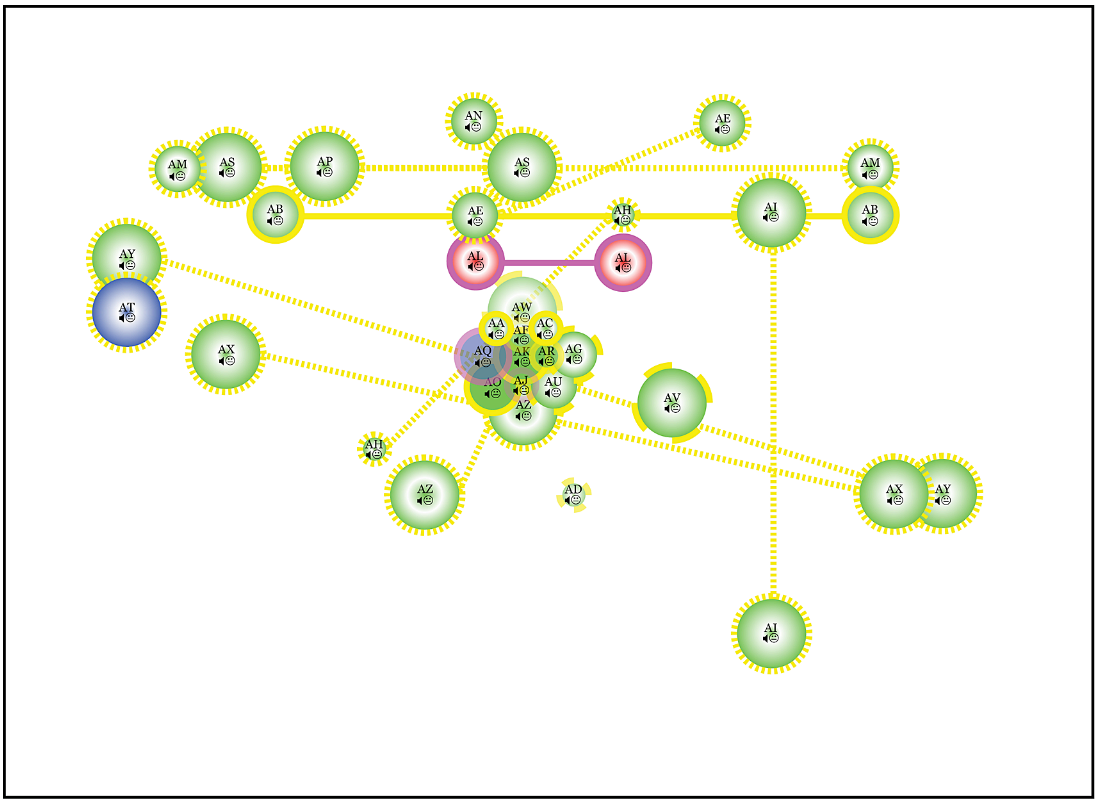


Figure 7-18: Designer's soundscape map (Design 09: Composition)

The designer classified 11 out of the 26 sound events as *impulsive*, with 5 being *intermittent* and 10 *continuous*. *Impulsive* included 'tearing' (AI) and 'ripping' (AX, AY & AZ) sound events. *Intermittent* sound events consisted mostly of 'judders' (AD, AU & AV). 'Ice cascade[s]' (AA, AC, AF, AK, AO & AR) formed the predominant sound event for *continuous*. Half (13) of the sound events were temporally *short*, 8 were *medium* and only 5 were *long*. The *short* sound effects were mostly 'click[s]' (AE & AN), 'thumps' (AP & AT) and 'judders' (AD, AU & AV). *Medium* sound events incorporated 'cascade[s]' (AA, AC & AF), 'hiss' (AL) and 'texture' (AW). All of which lasted somewhere between 3 to 12 seconds in length. *Long* also included 'cascade[s]' (AK, AO & AR) as well as 'ambience' (AJ) and 'rumble' (AQ). *Long* sound events varied from 11 to 33 seconds. The composition began by alternating between mostly dynamically *soft* and *medium* sound events (0 – 39 s), and then progressed to a section with predominantly *medium* dynamics (40 – 59 s). The piece then alternated between *loud*, *medium* and *soft* (60 – 70 s) before finishing with almost only *loud* sound events (71 – 90 s).

Due to the abstract nature of the Composition the designer requested that participants be allowed to listen to the Composition as often as they wished. When classification took place an example of each sound was played in context to aid identification. The appropriate section of the piece was repeated until the participant was confident that they were classifying the correct sound. This meant that the listeners were *aware* of every sound event. Whilst the listeners were able to identify the panning cues they did not experience the depth cues to the same extent as the designer (see Figure 7-19). The listeners also thought that some of the stationary sound events had panning.

Both the designer and the listeners considered all of the sound events to be *sound effects* rather than *music*. The designer classified the majority of the sound events as *solid*, whereas the listeners distributed the material attributes to be more evenly distributed between *gas*, *liquid* and *solid*. The listeners also reported a wider range of difference for the spectral attributes, specifically more *high* sound events. The response for clarity was very similar for both the designer and the listeners, with the 20 sound events being *clear* in the designer's case and 21 for the listeners.

The biggest difference was that the designer considered all of the sound events to have *neutral* content, *neutral* aesthetics and *neutral* emotions. This did not correspond with the listeners' experiences, who rated only 9 out of the 26 sound events as having *neutral* content, *neutral* aesthetics and *neutral* emotions. It might be suggested that although the designer intended the Composition to have no informative, aesthetic or emotional resonance, the listeners still experienced these attributes. This may be a desire on the part of listeners to attribute meaning and express preferences when presented with an abstract composition.

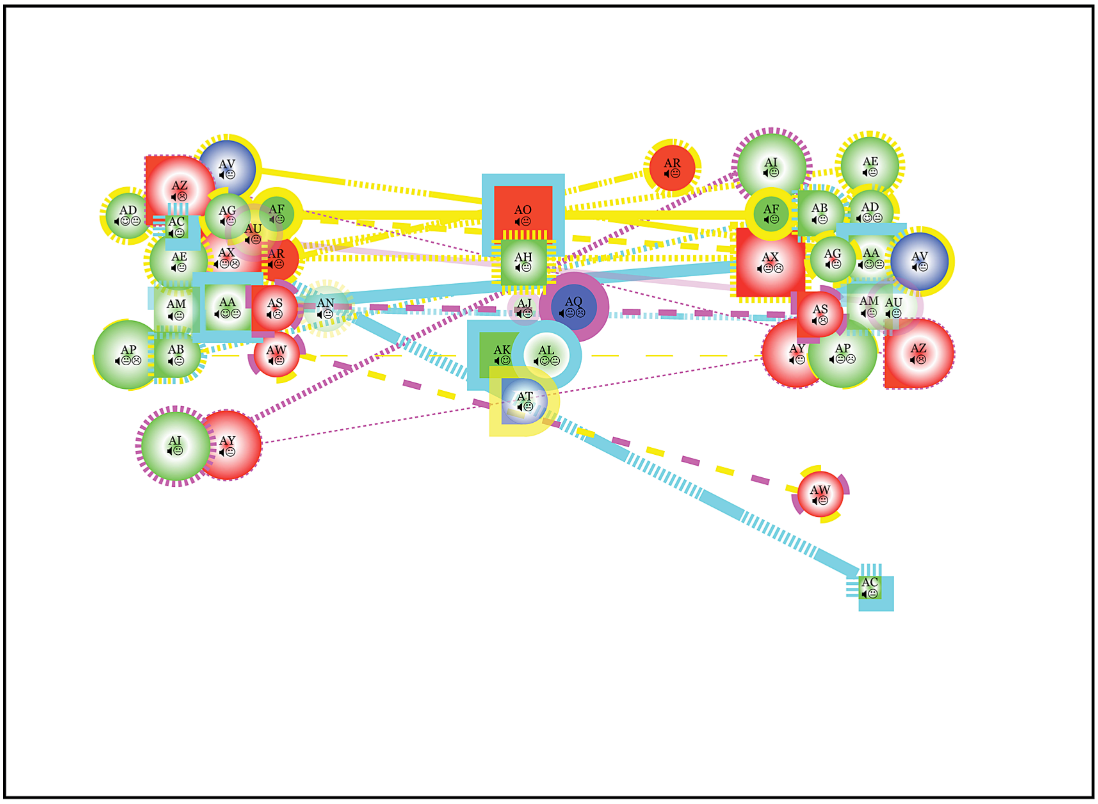


Figure 7-19: Listeners' soundscape map (Design 09: Composition)

The designer intended the composition to be an environment of total immersion, where the listener experiences the sonic activity from within the sound itself, and as such it might be suggested to have partially achieved these aims. Witmer and Singer (1998) defined immersion as the perception of being "enveloped by, included in and interacting with an environment that provides a continuous stream of stimuli and experiences" (p.227). Listeners experienced the sound events moving around them and engaged with them finding some of the sound events to be *informative*, *pleasing* or *positive* throughout the 1 minute and 30 seconds.

7.2.10 Design 10: Film Sound Effects

Film Sound Effects formed the basis of the tenth design. A new 30 second audio only sequence was created by a sound designer from the film industry in order to trial the tool. The sequence contained 32 sound effects all of which came from the designer's sound effects library and contained vehicles, as well as sound events associated with media production. The design also included a few musical elements (see Table 7-12).

Code	Description	Code	Description	Code	Description
AA	Film Camera	AL	Radio Static	AW	Quick Synth Hit
AB	Sci Fi Hit	AM	Vintage TV Knob Turns	AX	Alien Synth Pad
AC	Synth Pad-Crystal-like	AN	Processed Reverse Cymbal	AY	Building Transitional Whoosh
AD	Fast Whoosh	AO	Train Horn-Entry	AZ	Synth with eerie movement
AE	Processed Reverse Cymbal	AP	Train-On Tracks-Fast	BA	Shaker-Modulating
AF	Metal Switches-Clicks	AQ	Train Horn-Doppler-Exit	BB	Stretching In and Out Synth Transition
AG	Processed Cymbal Scrape	AR	Processed Cymbal Scrape-Transition	BC	Synth Whoosh-Modulating
AH	Jet-Entry	AS	Hit-Ding	BD	Chirping beeps 1
AI	Jet-Engine Fires	AT	Eerie Synth Pad	BE	Chirping beeps 2
AJ	Jet-Exit	AU	Guitar-Powerchord	BF	Chirping beeps 3 (x4)
AK	Poof-Flash	AV	Sliding Synth Whoosh		

Table 7-12: Sound events (Design 10: Film Sound Effects)

The designer made full use of the range of panning, and utilized almost the full range of depth cues (see Figure 7-20). In terms of the type of sound event only the 'guitar powerchord' (AU) was considered to be *music* by the designer, all of the other sound events were rated as *sound effects*, there were no instances of *speech*. The designer used bass sparingly as only 3 out of 32 sound events were classified as *low* (AB, AI & AK). The spectrally *high* and *mid* sound events were almost equally matched in number with 15 *high* and 14 *low*. Some sound events moved from *high* to *low* as part of the Doppler effect (AH, AJ, AO, AP). The sounds were included as two separate sound events when they were split by a third such as AI and AQ, if the Doppler effect was uninterrupted then the sound was a single sound event such as AD and AV. The sequence started and ended with what the designer considered to be emotionally *positive* sound events (AA & BF). The three other *positive* sound events were associated with entry of vehicles, specifically a jet and a train (AH, AI & AO). Twenty-one of the sound events were classified as emotionally *neutral* and 6 sound events were *negative*, three of which were ‘synth[s]’ (AC, AT, AX).

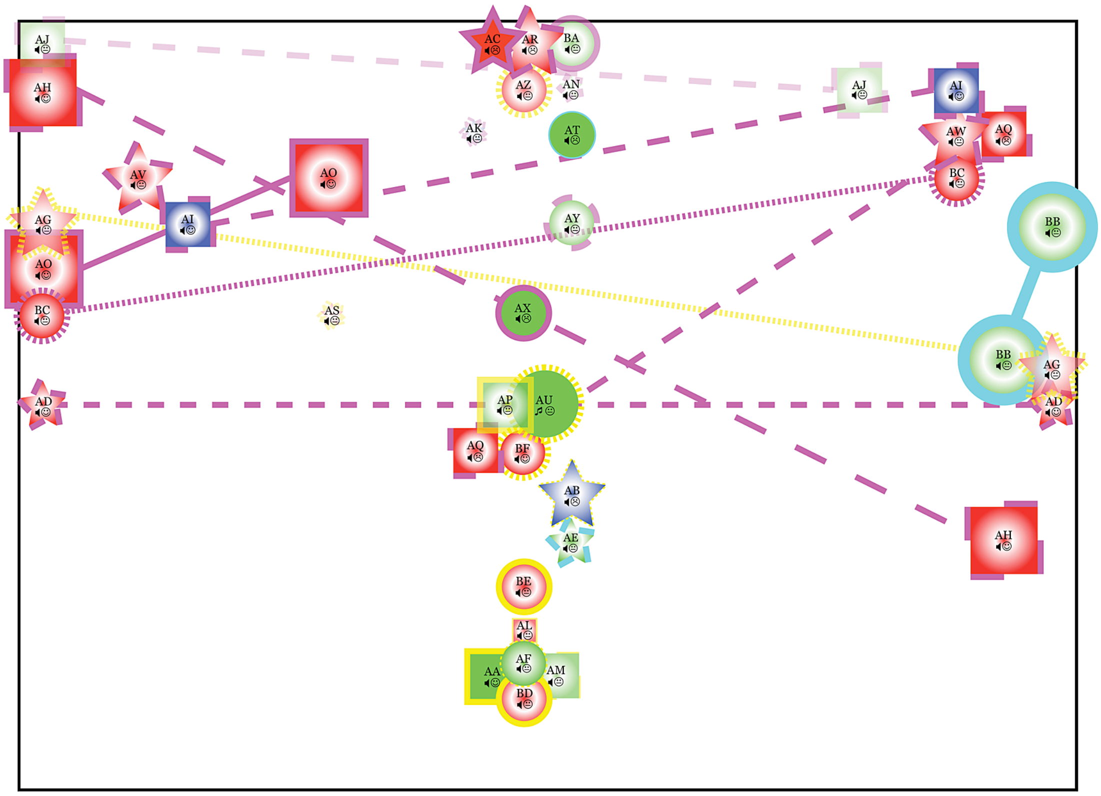


Figure 7-20: Designer’s soundscape map (Design 10: Film Sound Effects)

The difference in the number of sound events recalled by the listeners (18) might be partly due to the high number of sound events (32) required to be recalled in 30 seconds. The designer intended that the listeners should be *unaware* of three of the sound events (AN, AS and AZ), the majority of the remaining sound events were associated with synth pads and cymbal scrapes and were classified as either *uninformative* or *neutral* by the designer. There were three exceptions (AA, AL and AM) which may have gone unrecognised as they all were associated with vintage media production and might not have been easily identified and recalled by the listeners. The listeners were able to identify the full extent of the panning but not the depth cues, which were slightly truncated (see Figure 7-21).

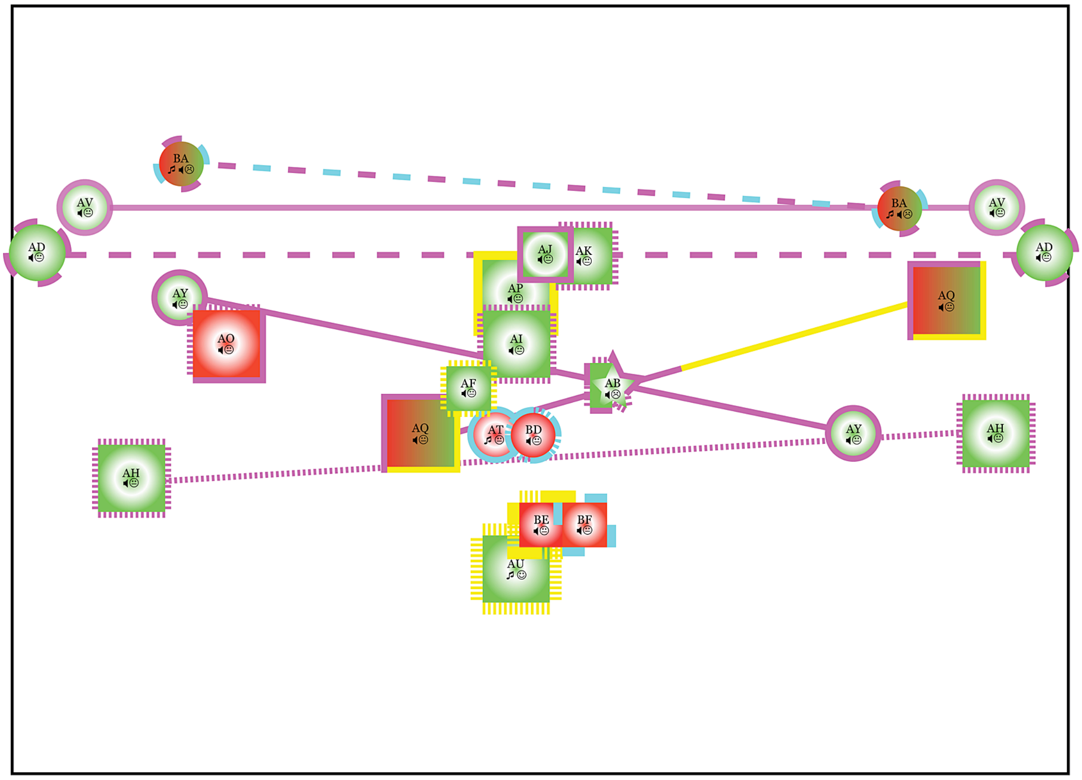


Figure 7-21: Listeners' soundscape map (Design 10: Film Sound Effects)

The listeners considered two of the synth sounds (AT and BA) to be *music*, and a third synth (BA) was classified as *music/sound effect*. Otherwise the listeners found the remaining 15 sound events that they were aware of to be *sound effects*. There was a high level of agreement with regards to material (border colour), 12 out of the 18 were rated identically with the designer. A further two sound events were rated as *gas* and either *liquid* or *solid* by the listeners, with the designer recording them both as *gas* (AQ and BA). There was little correspondence between the interaction attributes for the designer and the listeners (border dashes), only a third of the sound events matched. Sometimes, when the designer considered the sound event to be *intermittent* the listeners regarded it as *continuous* (AJ, AQ, AV and AY), on other occasions the listeners chose *impulsive* (AH and AI).

The listeners found the majority of the sounds that they were aware of to be *informative*, the designer considered them to be more evenly distributed between *informative*, *neutral* and *uninformative*. The listeners rated all but one of the sound events as *clear*, the single exception was both *clear* and *neutral* (AV). The designer considered the majority (21) to be *clear*. Only a single sound event was classified as *pleasing* by the designer, a 'synth transition' (BB) which listeners were *unaware* of. None of the sound events were regarded as *displeasing* by the listeners. Both the designer and the listeners rated the majority of the emotions attributes to be *neutral*.

In general terms the listeners were *aware* of the majority (18) of the sound events considering the number contained in a short period of time (30 seconds). Listeners were *unaware* of all of the sound events that the designer classified as *soft* (AK, AL, AN & AS). However, it did not follow that listeners were *aware* of each *loud* sound event. Listeners did not regard any of the sound events as being *soft*, and found more of the sound events to be *informative* than the designer. More could have been made of the aesthetics and the emotions, but in film sound design there is still an emphasis on the technical rather than the aesthetic (Holman, 2010). The emotions in a film soundtrack are often left to the music, as the score can suggest memories that have positive or negative associations (Plantinga, 2009).

7.2.11 Designers' survey

All ten of the designers completed the questionnaire; none of the questions were omitted. Seven out of the 12 audio attributes were considered to be either important or very important by six of the designers (see Table 7-13). A further four attributes were rated as important or very important by five of the designers, and only a single attribute (*interaction*) was rated as being either important or very important by fewer than half of the designers. This suggests that *awareness*, *spatial cues*, *type*, *dynamics*, *content*, *clarity* and *emotions* should be chosen as the reduced set of attributes, with *material*, *temporal*, *spectral* and *aesthetics* being offered as optional attributes, should designers wish to customise the SMT.

	Unimportant	Of Little Importance	Moderately Important	Important	Very Important	Important/Very Important	%
Awareness		2	1	3	4	**7**	**70%**
Spatial Cues		1		4	5	**9**	**90%**
Type			1	6	3	**9**	**90%**
Material		3	2	3	2	**5**	**50%**
Interaction		3	3	2	2	**4**	**40%**
Temporal		1	4	3	2	**5**	**50%**
Spectral		2	3	3	2	**5**	**50%**
Dynamics			2	6	2	**8**	**80%**
Content			3	4	3	**7**	**70%**
Aesthetics		2	3	1	4	**5**	**50%**
Clarity		1	2	2	5	**7**	**70%**
Emotions		1	3		6	**6**	**60%**

Table 7-13: Designers' ratings of importance of audio attributes

The second part of the questionnaire asked designers for their preferred choice of displaying each audio attribute. Seven out of the 12 attributes had a single method of display chosen by the majority of the sound designers. Two of the methods of visualisation were chosen by all 10 of the designers, these were the position on grid for the *spatial* cues and symbols for the *type* (see Table 7-14).

	Inclusion of object	Position on grid	Symbols	Border colour	Border dashes	Fill gradient	Fill colour	Dimension	Shape	Border weight	Opacity	Emoticon		%
Awareness	9				1								**Inclusion of object**	**90%**
Spatial Cues		10											**Position on grid**	**100%**
Type			10										**Symbols**	**100%**
Material			1	4	2		1		2				**Border colour**	**40%**
Interaction			1		7	2							**Border dashes**	**70%**
Temporal			1			3	1	3		2			**Fill gradient/Dimension**	**30%**
Spectral				2		3	3	1		1			**Fill gradient/Fill colour**	**30%**
Dynamics			1			1	1	4		2	1		**Dimension**	**40%**
Content					1		2		6	1			**Shape**	**60%**
Aesthetics			1			1	1			4		3	**Border weight**	**40%**
Clarity					2		1				7		**Opacity**	**70%**
Emotions			1									9	**Emoticon**	**90%**

Table 7-14: Designers' choice of display of audio attributes

A further two visualisation methods were chosen by 9 out of 10 designers: inclusion of object for *awareness* and emoticons for *emotions*. Opacity for *clarity*, border dashes for *interaction* and shape for *content* were also chosen by more than half of the sound designers. There was no clear single choice of display for the remaining attributes: *material*, *temporal*, *spectral*, *dynamics* and *aesthetics*.

The designers were asked to state their level of agreement with the statement that the soundscape mapping tool allowed them to compare a sound design with the experience of listeners. All ten found that they either agreed (6) or strongly agreed (4). This result must be approached with caution as all of the designers self-selected to take part in the study, and as such are possibly more likely to agree with the statement (Stanton, 1998).

Responses to the question about how they (the designers) currently evaluate their sound designs varied from 'expert rating (via questionnaires)' (D04) to 'asking them [listeners] for their input' (D03). Four of the designers evaluate their work themselves by breaking it down into different 'parameters' (D06) or 'criteria' (D08) such as 'pitch volume and dynamics' (D06) or whether it 'is enhancing the picture' (D10). Three designers 'aurally' (D09) evaluate the work by asking for listeners' impressions, trying to establish informally whether they 'perceive what ... was intended' (D07). This allows the designers to be 'more critical about my [their] work' (D02). One designer used 'expert rating' (D04) with questionnaires and two designers had not had their work evaluated prior to this study (D01 and D05).

When answering the question about how they could use the SMT the designers stated that it could used to 'provide an overview' (D01), and as an 'artefact to communicate and discuss design ideas' (D02). They also asserted that it could be employed to 'allow others to give controlled but honest interpretation' (D03) as well as to 'check all kinds of feelings and reactions' (D06). The designers highlighted its 'value as an analytical tool' (D09) potentially extending existing 'questionnaires' (D04). The maps were considered useful for 'evaluation' (D05) establishing if spatial cues are ‘accurately perceived by listeners’ (D07). Finally the designers thought that the tool could be used ‘when choosing the final design’ (D08) and to see ‘how much more [successful] the sound design makes the final product’ (D10).

Changes that were proposed for the SMT predominantly related to reducing ‘the number of variables’ (D01) in order to make it ‘much more “user friendly”’ (D10). One designer found it ‘complete’ (D10) and did not want anything changed, four wished to reduce the number of attributes as ‘there are too many’ (D04) and some were considered ‘useless’ (D07). Two of the designers requested animation along a ‘timeline’ (D02) so that ‘temporal aspects’ (D02) such as ‘complex spatial gestures’ (D09) and ‘relationships between sonic elements’ (D02) could be captured, as this information would be ‘both informative and compelling’ (D09). One designer recommended a ‘z-axis’ to ‘improve accuracy of interpretations’ (D03), whilst another suggested confining the results to being ‘simply numeric’ as they found the visualisation to be ‘complicated and unnatural’ (D08). The descriptions for the *spectral* and *aesthetics* attributes ‘wasn’t very clear’ (D05) and it would be useful ‘if listeners can guess or be aware of the intended meaning of the sounds’ (D01) prior to informing them of what the sound event was for classification.

In the additional comments section the designers were generally supportive. There were comments about the lack of any ‘fixed taxonomy for sounds’ and that there were not any ‘standardized evaluation tools for sound design’ (D01). The maps were considered ‘useful’ for ‘comparison’ but should be used in ‘combination with discussion’ as it was ‘missing the wealth of information that can derive out of discussion with listeners’ (D02). One designer would ‘love to see a full blown product for sound designer’s to use freely’ (D03) and another thought that it was an ‘interesting tool’ (D04). ‘Some instructions to create symbols for the visualisation’ were proposed (D05), but this should not prove necessary, as the intention is to automate the process using software. The issue of ‘artificial representation’ of ‘stereo’ was raised, as it does not represent ‘the true aural landscape’, this was an artefact of the form of media delivery. It was considered as ‘valuable as an exercise... and certainly as a listening tool’ (D09) and ‘gave me another dimension in the thinking of how my work is perceived’ (D10). A single designer (D06) did not have any additional comments.

7.3 Discussion

Having increased the number of participants from 1 designer and 10 listeners in the previous study (in-car auditory display) to 10 designers and 100 listeners in this study a number of issues became apparent. The task of recalling all of the sound events was generally difficult for the listeners, even when the sequences were quite short. However, in a few cases they were noticeably accurate. Designers chose whether they wished the listeners to be able to listen only once or multiple times. When the listeners were able to listen repeatedly they were *aware* of every sound event, which made asking listeners about their awareness of a sound event redundant. It might be effective to question the listeners after they had listened to the sound file for the first time about which sound events they were aware of, and then allow them to listen to each sound event in context in order to improve the accuracy of their responses.

Informally, listeners reported that the short descriptions provided by the designers were easy to follow for all but two of the designs. Two sets of descriptions that were considered difficult by the listeners were the Sonification and the Composition. This did not present a problem as the listeners were allowed to listen to the designs as often as they wished so they were eventually able to successfully identify the sound events in order to classify them. The short descriptions of the sound events within the designs ranged from single words such as ‘Helicopter’ within the Short Film through to “It consists in the tonic of the laboratory, without a frequent use of PC; it coincides with aeration ducts and wheels of personal computers” in the Simulation.

All of the designers specified actions within the short descriptions, but not for every sound event. Actions included: ‘loading’ (Short Film), ‘swipe’ (Composition) and ‘turns’ (Film Sound Effects). Each designer identified the source of sound events, except for the Sonification. Sources included: ‘airplane’ (Soundscape Composition), ‘tools’ (Radio Drama) and ‘piano’ (Audiologos). Six of the designers referred to the meaning or content of the sound event, such as ‘low battery alert’ (Auditory Display), ‘recovery phase’ (Sonification) and ‘hello’ (Games Sound Effects). The type of sound was also specified by six designers and included: ‘music’ (Auditory Display, Short Film and Soundscape Composition), ‘ambience’ (Short Film and Composition) and ‘spearcon’ (Auditory Display). Four designers included information about the dynamics of the sound event within the descriptor, examples included: ‘loud’ (Soundscape Composition, Games Sound Effects), ‘soft’ (Games Sound Effects) and ‘building’ (Film Sound Effects). Spectral descriptors were used by four designers, samples comprise: ‘low frequency’ (Short Film), ‘high’ (Games Sound Effects) and ‘bass’ (Composition). Four designers included references to spatial cues, they included: ‘outdoors’ (Short Film), ‘distant’ (Soundscape Composition) and interior locations (Simulation). Temporal references were also included in the short descriptions of the sound events by four designers, terms included: ‘short’ (Short Film), ‘opening’ (Composition) and ‘fast’ (Film Sound Effects). Onomatopoeia was used by four designers, examples comprise: ‘beeps’ (Short Film and Film Sound Effects), ‘whoosh’ (Short Film and Film Sound Effects) and ‘hiss’ (Composition). Three designers included information about the material of the sound source: ‘leather’ (Short Film and Audiologos), ‘wood’ (Short Film and Audiologos) and ‘metal’ (Film Sound Effects). A single designer specified clarity in terms of ‘distortion’ (Short Film) and the designer of the Simulation included information about quantity, differentiating between ‘some students’ and ‘a student’.

7.3.1 Awareness

In order to gain an insight into the suitability of the attributes for describing the different sound designs, a comparison of the results is required. When comparing all of 10 of the designs there were a total of 224 sound events contained within 20:09 minutes of audio, which gave an average length for each design of 2:00 minutes and 22 sound events. The listeners were aware of 178 of the sound events, which represented a level of 79% awareness. Five of the designs, such as the Games Sound Effects, had levels of 100% awareness, the Short Film had the lowest level of listener awareness (51%). The reason for the low level of awareness is possibly due to the large number of sound events (45), which was the highest number for any design, but might also be due to the short average interval between sound events of only 3 seconds. The Film Sound Effects and Soundscape Composition had similar level of awareness (both 56%) with an average interval of 1 second or less. Listeners tended to be *unaware* of sound events that did not have a recognisable source such as the 'synth ambience' in the Short Film or the 'stretching in and out synth transition' in the Film Sound Effects. Sound events that the designers considered *displeasing*, such as the bathroom sounds in the Simulation or the 'weird branches coming out of the mouth' in the Short Film, also went unnoticed by the listeners. A further factor might be that the sound event was regarded as *uninformative* by the designer, for example the 'girl's voice' in the Radio Drama and the 'rock first hit' in the Soundscape Composition.

7.3.2 Spatial

Spatial cues were used by 9 out of 10 designers, only the Auditory Display omitted this attribute. The designers considered that 75% of the sound events were static and 25% had some motion (panning and/or depth). The listeners thought that 89% of the sound events were static and 11% had some motion. The design with the greatest percentage of motion was the Sonification with 88%, almost all of which was depth, two of the designs (Simulation and Radio Drama) had no motion. Listeners only perceived two of the designs as having motion (Composition and Film Sound Effects). The listeners thought that the design with the greatest amount of motion was the Composition with 69% of the sound events being regarded as having motion.

The designer of the Composition considered 42% of the sound events to have motion.

For stationary sound events the designers used almost the entire x axis (panning) stopping only 1 cell short at 20 (maximum = 21), the entire y axis (depth) was used (16). The listeners experienced slightly less panning for the static sound events with an x axis range of 3 to 16, and a y axis extending from 3 to 13. The designers recorded an average spatial attributes value for the static sound events of 8, 7, which was slightly to left of centre and in the middle for depth. The listeners experienced an average spatial attribute of 9, 8 for the static sound event, which represents the same depth but a panning slightly (1 cell) closer to the centre. For sound events that had motion the entire x and y axes were used by the designers. In contrast the listeners experienced the entire range of panning (0 – 21) but a lesser amount of depth (4 – 13).

7.3.3 Type

Within the type attribute, *speech* was predominantly used to classify identifiable words or phrases by both the designers and the listeners, such as 'I'm calling you (Man)' in the Auditory Display or the 'Butler's voice' in the Radio Drama. *Music* was for the most part chosen when there was a clearly identifiable melody such as the 'dub music' in the Short Film or the 'flute music A' in the Soundscape Composition. *Sound effect* was used for a wide range of sound events, examples include the 'birds' in the Audiologos and the 'recovery phase' in the Sonification.

When classifying the type of sound events the designers considered 12% to be *speech*, which was slightly lower than the listeners level of 15% (see Table 7-15). The designers classified 19% of the sound events as *music*, in contrast the listeners only considered 9% of the sound events to be *music*. This was mostly due to the listeners only classifying four of the sound events in the Auditory Display as *music*, whereas the designer considered 25 sound events to be *music*. *Sound effects* formed the largest group for both the designers and the listeners at 70% and 77%

respectively, this difference was predominantly down to responses about the Auditory Display. Twenty-one of the sound events classified as *music* by the designer were classified as *sound effects* by the listeners. The designers considered that the Composition and the Sonification were completely made up of *sound effects*. All of the other designs except the Audiologos were regarded by the designers as having the majority of sound events as *sound effects*. Half of the sound events within the Audiologos were classified as *music* by the designer, the remaining seven were split, almost equally, between *speech* and *sound effect*. Only 2% of the sound events were classified by the listeners as being multiple type (*music* and *sound event*), there were single instances in three designs (Auditory Display, Short Film, Film Sound Effects). The designers classified the first instance in the Auditory Display as *music* and the second two as *sound effects*.

		Designers		**Listeners**	
		n	**%**	**n**	**%**
Type	Speech	26	12%	27	15%
	Music	41	19%	16	9%
	Sound effect	153	70%	140	77%
Material	Gas	81	42%	55	28%
	Liquid	13	7%	38	19%
	Solid	98	51%	103	53%
Interaction	Impulsive	74	39%	80	39%
	Intermittent	70	36%	51	25%
	Continuous	48	25%	75	36%
Temporal	Short	106	48%	89	46%
	Medium	77	35%	67	35%
	Long	37	17%	36	19%
Spectral	High	69	31%	61	31%
	Mid	121	55%	100	51%
	Low	30	14%	35	18%
Dynamics	Loud	55	25%	52	26%
	Medium	118	54%	123	62%
	Soft	47	21%	22	11%
Content	Informative	118	54%	132	69%
	Neutral	61	28%	47	25%
	Uninformative	41	19%	12	6%
Aesthetics	Pleasing	44	20%	46	24%
	Neutral	133	60%	115	61%
	Displeasing	43	20%	27	14%
Clarity	Clear	146	66%	165	88%
	Neutral	45	20%	13	7%
	Unclear	29	13%	9	5%
Emotions	Positive	51	23%	40	21%
	Neutral	118	54%	125	64%
	Negative	51	23%	29	15%

Table 7-15: Summary of Designers' and Listeners' application of attributes

7.3.4 Material

When classifying the material attribute of sound events *gas* was often chosen for sound events that involved the movement of air as in ‘it consists in the tonic (keynote) of the laboratory, it coincides with the wheels of personal computers’ in the Simulation, or the ‘jet-entry’ in the Film Sound Effects. *Liquid* was predominantly selected for sound events such as the ‘waterfall’ in the Games Sound Effects and ‘water trickling’ in the Soundscape Composition.

The designers regarded the material attributes for the sound events to be 51% *solid*, 42% *gas* and only 7% *liquid*. The listeners reported a similar figure for *solid* (53%) but less for *gas* (28%) and a greater figure for *liquid* (19%). Two of the designers considered their designs to have only one form of material, *solid* in the case of the Radio Drama and *gas* for the Sonification. The listeners’ responses did not concur with the Sonification designer, as 70% of the sound events were considered to be *liquid*. The listeners’ experiences concurred closely with the designer of the Radio Drama, only a single sound event varied, ‘Chetwood’s voice’ was considered to be both *gas* and *solid* by the listeners. There were a greater number of multiple classifications by the listeners for the material than the type. Ten percent of the sound events had multiple classifications, and this occurred in all but one of the designs (Games Sound Effects). Eleven out of 18 of the multiple classifications were *gas* and *solid*, with 7 remaining being split between *gas* and *liquid* (3) and *liquid* and *solid* (4).

7.3.5 Interaction

When classifying the interaction attributes *impulsive* was primarily used for percussive type sound events, such as ‘message knocking’ in the Auditory Display, or ‘it is produced by drumming fingers on a desk’ in the Simulation. *Intermittent* was chosen when sounds had a percussive element but had an underlying sustained element beneath it, as in the ‘dog growl’ in the Games Sound Effects or the ‘juddering anacrusis’ in the Composition. *Continuous* was applied when there was a sustained sound event without any obvious percussive elements, examples include the ‘room ambience’ in the Short Film, and the ‘background ambience’ in the Composition.

Designers made full use of the range of interaction attributes, 39% were *impulsive*, 36% were *intermittent* and 25% were *continuous*. The designers of the Auditory Display and the Audiologos made greater use of *impulsive* sound events. *Continuous* sound events were only used once in both the Sonification (‘acceleration trace’) and the Radio Drama (‘ringing bell’). The listeners reported the same level of *impulsive* sound events as the designers, but considered 25% to be *intermittent* and 36% *continuous*. Fifteen percent of the sound events had multiple classifications and this occurred in seven of the sound designs, only the Sonification, Simulation and the Games Sound Effects were excepted. The multiple classifications were somewhat evenly split between all three combinations, with slightly more *impulsive* and *intermittent*, and there was a single instance of all three descriptors being applied, which was for the ‘Sci-Fi Hit’ in the Film Sound Effects design.

7.3.6 Temporal

Within the temporal attributes *short* was chosen for brief non-repeating sound events such as the ‘poof flash’ in the Film Sound Effects or the ‘voice “xxxxx”’ in the Audiologos. *Medium* was used when a sound event was of indeterminate length, neither *short* nor *long*, examples include ‘Sid’s voice’ in the Radio Drama’ and ‘sounds emitted by a key-holder while someone is walking in the passage, other noises, some steps, aeration ducts’ in the Simulation. *Long* was applied to extended uninterrupted sound events, such as the ‘theme music’ in the Auditory Display’, or the ‘water’ in the Games Sound Effects.

The designers did not apply the temporal attributes equally. Greater use was made of *short* (48%) sound events than *medium* (35%) or *long* (17%). Seven out of the eight sound events in the Sonification were regarded as *short* by the designer, and three of the designers considered that there were no *long* sound events in their designs

(Sonification, Radio Drama and Audiologos). The listeners classified the sound events similarly to the designers, with 46% as *short*, 35% *medium* and 19% *long*. Fourteen of the 178 sound events that the listeners were aware of had multiple classifications, and these occurred in five of the designs (Short Film, Radio Drama, Audiologos, Composition, Film Sound Effects). Nine of the multiple classifications were *short* and *medium*, five were *medium* and *long*.

7.3.7 Spectral

The spectral attribute *high* was commonly applied to bright percussive sound events both natural and man-made, these included 'loud chirp' in the Soundscape Composition and the 'front catch' in the Sonification. *Mid* was chosen for sound events that fell between *high* and *low* as well as for sound events that had broadband spectral content. Examples of broadband content include the 'distorted evolving pad' in the Short Film, and the 'jet engine fires in the Film Sound Effects. *Low* was selected for obvious bass content, as in the 'bass rumble' in the Composition and the 'leather bass drum' in the Audiologos.

Designers predominantly used *mid* for classifying the spectral attributes of the sound events in their designs, 55% of all of the sound events were *mid*, with 31% *high* and only 14% *low*. The designers of two of the designs (Sonification, Radio Drama) did not consider their content to have any *low* sound events at all. Only one designer made extended use of *low*, this was for the Audiologos and they represented 43% of the sound events, which was equal to the amount of *mid* sound events. The designer of the Composition classified 88% of the sound events as *mid*, and four other sound designers had the majority of their sound events as *mid* (Short Film, Simulation, Games Sound Effects and Radio Drama). The listeners experienced a similar percentage of *high*, *mid* and *low* spectral attributes. *High* represented 31%, which was identical to the designers' classification. *Mid* was 51% which was 4% lower than the designers' responses, and *low* was 18% which was 4% higher than the designers' ratings. Multiple classifications by the listeners occurred with 10% of the sound events and in six of the designs (Auditory Display, Short Film, Simulation, Games Sound Effects, Composition and Film Sound Effects). The majority (10) of the multiple classifications were *high* and *mid*, 7 were *mid* and *low*, and 1 was both *high* and *low* (‘floor creak’, Short Film).

7.3.8 Dynamics

When dynamics attributes were applied *loud* was often used for short prominent sound events such as the ‘gun shot’ in the Short Film, or the ‘mid catch’ in the Sonification. *Medium* was chosen for moderate intensity sound events that provided context for a further action, examples include ‘gun loading’ in the Short Film and ‘safe door jiggled’ in the Radio Drama. *Soft* was used to classify gentle sound events that formed an auditory backdrop, examples include the ‘background ambience’ in the Composition and ‘birds’ in the Audiologos.

As in the case of the spectral attributes designers mainly applied the *medium* value of the dynamics attributes. *Medium* was used for 54% of the sound events, *loud* for 25% and *soft* for 21%. The designer for the Auditory Display considered that 94% (30) of the sound events were *medium* and that there were no *soft* sound events. The designer of the Sonification also did not include any *soft* sound events but split the eight sound events equally between *loud* and *medium*. The Radio Drama did not contain any *high* sound events, but both *medium* (8 sound events) and *soft* (6) were included. The designers of the Short Film, Simulation, Games Sound Effects and Composition applied the spectral attributes more evenly across the range. The listeners also reported that the majority of sound events were *medium*. *Medium* represented 62% of the sound events, which was 8% higher than the designers reported. *Loud* accounted for 26% of the sound events and *soft* for 11%, from the listeners’ experiences. Multiple classifications accounted for 10% of the sound events, and this occurred in eight of the designs. *Low* and *medium* occurred 11 times, *medium* and *soft* occurred 6 times. There was a single instance of *loud*, *medium* and *soft* in the Soundscape Composition for the ‘flute music’.

7.3.9 Content

The most obvious examples of *informative* sound events for the content attribute were those associated with warnings such as the ‘low battery alert in the Auditory Display or the ‘Ringing bell (doorbell)’ in the Radio Drama. *Neutral* was applied to sound events that were neither regarded as necessary nor unnecessary to comprehend the sound design. Examples include the ‘chirping beeps 1’ in the Film Sound Effects and the ‘big leaf crunch’ in the Soundscape Composition. *Uninformative* was retained for those sound events which were considered unnecessary as in the ‘leather bass drum’ in the Auditory Display or the voices in the Games Sound Effects, the latter of which was only *uninformative* from the listeners’ perspective.

The majority (54%) of the content of the sound events were rated as *informative* by the designers. *Neutral* accounted for 28% of the responses and *uninformative* 19% amongst the designers. The designer for the Sonification consider all 8 sound events to be *informative*, as did the designer for the Games Sound Effects (18 sound events). In contrast, the designer of the Composition classified all 26 of the sound events as *neutral*, and the Radio Drama designer did not regard any of the sound events as having *neutral* content, 13 were *informative* and 1 was *uninformative*. The designers for the Short Film, Soundscape Composition, Simulation and Film Sound Effects had responses more evenly spread across all three options. As might be expected no designer considered the majority of the sound events for their design to be *uninformative* as they could choose, for the most part, which sound events to include. The listeners experienced a greater percentage of the sound events as *informative* (69%) than the designers did. The listeners reported a similar level of *neutral* sound events to the designers (25%), but a much lower level of *uninformative* (6%) sound events, which were all within four sound designs (Auditory Display, Games Sound Effects, Audiologos and Film Sound Effects). The listeners provided multiple content classifications for 9% of the sound events, 11 of which were *informative* and *neutral*, and four were *informative* and *uninformative*.

7.3.10 Aesthetics

Within the aesthetics attributes *pleasing* was predominantly applied to *positive* sound events that came from an acoustic source, examples include the 'birds high and loud' in the Games Sound Effects and the 'classical guitar' in the Audiologos. *Displeasing* was often chosen for sound events that had negative associations such as a 'dog growl' in the Games Sound Effects or the 'distorted scream' in the Short Film. *Neutral* was used for abstract sound events that had no physical analogue, such as the 'back reversal' in the Sonification or the 'ripping detritus-drop' in the Composition.

The designers considered that the majority (60%) of the sound events were aesthetically *neutral*. The designers classified 20% of the sound events as *positive* and 20% as *negative*. Three sound designers rated all of the sound events within their designs as aesthetically *neutral* (Sonification, Radio Drama and Composition). The designers of the Auditory Display and Audio logos regarded the majority of the sound events as *pleasing*, whereas the designers of the Soundscape Composition and the Simulation thought that the majority of the sound events in their designs were *displeasing*. The listeners reported a similar percentage of sound events as being aesthetically *neutral* 61% when compared to the designers (60%). They considered more sound events to be *pleasing* (24%) than the designers did, and fewer to be *displeasing* (14%). Only 6% of the sound events within six designs had multiple classifications from the listeners, five were *pleasing* and *neutral*, four were *neutral* and *displeasing* and one was both *pleasing* and *displeasing* ('low/mid modulated ambience' in the Short Film).

7.3.11 Clarity

Within the clarity attribute *clear* was applied to often explicit sound events that were foreground in the designs, examples include 'the emission sounds of a television: a woman's voice' from the Simulation and the 'woman's voice' from the Radio Drama. *Unclear* was used for sound events that whilst still audible were difficult to discern as in the 'female voice "Tomorrow"' from the Games Sound Effects and the 'background ambience' in the Composition. *Neutral* sound events were those which were regarded as neither *clear* nor *unclear*, examples comprise 'rock bounce' from the Soundscape Composition for the designers and 'warning spearcon' from the Auditory Display for the listeners.

In terms of the clarity attribute the designers considered 66% of the sound events within their designs to be *clear*, 20% were *neutral* and 13% *unclear*. Only the designer of the Sonification classified all of the eight sound events in their design as *clear*. The designers of the Games Sound Effects and the Simulation regarded approximately half the sound events in their designs as *neutral*. All of the other seven designers specified that the majority of the sound events within their designs were *clear*. The listeners experienced a greater percentage of the sound events as being *clear* (88%) than the designers reported. The listeners considered 7% of the sound events to have *neutral* clarity and 5% to be *unclear*. All of the sound events for the Sonification and the Radio Drama were classified as *clear* by the listeners and only the Short Film, Games Sound Effects and Composition were considered to have any *unclear* sound events. Only 5%, nine, of the sound events in half of the designs had double classifications made by the listeners, seven were *clear* and *neutral*, two were *clear* and *unclear* ('low/modulated ambience' in the Short Film, 'kiss' in the Games Sound Effects).

7.3.12 Emotions

In terms of the emotions attributes *positive* was applied when a sound events with obvious affirmative associations such as the 'kiss' in the Games Sound Effects or 'success' in the Auditory Display. *Neutral* was used when the sound events were abstract, examples include the 'drive phase' from the Sonification and the 'building transitional whoosh' in the Film Sound Effects. *Negative* denoted sound events that were designed to have an unpleasant effect, these included the 'door' in the Audio logos and 'Chetwood's voice' in the Radio Drama.

When considering the emotions attribute the designers again used *neutral* most frequently (54%), both *positive* and *negative* accounted for 23% each of the sound events. The designers of the Sonification and the Composition classified all of the sound events within their designs as *neutral*. The designer of the Short Film did not consider any of the sound events to be *positive*, but the other seven designers made use of *positive*, *neutral* and *negative* sound events in their designs. The majority of the sound events were *positive* for the Auditory Display and Audiologos designers, but for all of the other eight designers the majority of the sound events were *neutral*. The listeners found that 64% of the sound events were *neutral*, 21% were *positive* and 15% were *negative*. The listeners agreed with the designer of the Sonification and classified all eight of the sound events as *neutral*. There was only a single design (Soundscape Composition) that had the majority of the sound events classified as *positive* by the listeners. Although the Auditory Display had just under half of the sound events classified as *positive* (44%). The Short Film had the largest percentage of *negative* sound events for the listeners at 44% (11). Only 9% of the sound events had double classifications made by the listeners within seven designs, nine sound events were *positive* and *neutral*, and seven were *neutral* and *negative*.

In general terms the designers' responses were weighted towards the middle value in six of the attributes (Spatial, Spectral, Dynamics, Aesthetics and Emotions) (see Table 7-15). Only one of the attributes (material) had a value of under 10% (*liquid*, 7%) according to the designers. With regard to the listeners, three of the attributes had responses that fell below 10%: *music* (9%), *uninformative* (6%), *neutral* clarity (7%) and *unclear* (5%). For six of the attributes the internal ranking of responses was consistent between the designers and listeners, although the percentages differed. The four attributes that did not have consistent ranking between the two groups were type, interaction, aesthetics and emotions although the majority response was always the same. In conclusion then, all of the attributes were used by both the designers and the listeners, and as such appear to be suitable for describing soundscapes. What has not been established is the designers' opinions about the importance of the attributes and their preferred methods of visualisation, which is discussed in the following section.

7.3.13 Survey

A reduced set of attributes (7) has been suggested by the designers (see Table 7-16), along with appropriate methods of display. All but one of the attributes (*interaction*) were considered to be either 'important' or 'very important' by at least half of the designers, with seven attributes being selected by the majority. It is proposed that a standard set of the seven most popular attributes is offered. Designers might like to select from a further four attributes (*material*, *temporal*, *spectral* and *aesthetics*) should they wish to elicit this additional information, as this information was still considered important by half of the designers. This was not as clear with the preferred visualisations, with only 7 out of the 12 attributes had a single form of display chosen by the majority of designers.

Awareness	Cognizance
Aware	Conscious of sound event
Unaware	Not conscious of sound event
Spatial cues	Position
x axis	panning/left-right
y axis	depth/reverb/close-far
Type	Category
Speech	Spoken language
Music	Performed composition
Sound effect	Audible events and actions
Dynamics	Volume/Loudness
Loud	High volume *Forte*
Medium	Medium volume *Mezzo*
Soft	Quiet *Piano*
Content	Relevance
Informative	Relevant information
Neutral	Neither relevant nor irrelevant
Uninformative	Irrelevant information
Clarity	Quality
Clear	Easy to hear and comprehend
Neutral	Neither easy nor difficult to hear
Unclear	Difficult to hear and comprehend
Emotions	Feelings
Positive	Joy, Love, Surprise
Neutral	No emotional content
Negative	Anger, Fear, Sadness

Table 7-16: Possible reduced set of attributes

All of the designers agreed that the SMT allowed them to compare a sound design with the experience of listeners. There was no suggestion as to any omissions, although two of the designers suggested that the SMT could be used either to extend existing questionnaires or as a starting point for further discussions with listeners.

The methods of visualisation that have been removed due to the reduced number of attributes are either related to the border (colour dashes and weight), or the fill (colour and gradient). It is necessary to have a fill colour in order to be able to perceive the clarity (opacity) of the sound event. Different colours could be chosen for each sound event so that interpretation could be improved, but the different hues might make the clarity (opacity) difficult to interpret. This was not a problem previously as all three hues representing the spectral attributes (red, green and blue) had the same levels of brightness. A black border has also to be retained so that when sound events overlap shapes are still identifiable.

All but one (*dynamics*) of the forms of display, which remain, was chosen by the majority of the designers (see Table 7-14). 'Shape' was chosen for *dynamics* by 4 out of 10 of the designers, but 2 designers chose 'border weight' and the remaining 4 designers chose either 'symbols', 'fill gradient', 'fill colour' or 'opacity'. Further studies will have to be conducted into the most appropriate way of visualising the dynamics attributes.

This chapter reported an evaluation of the soundscape mapping tool that involved 10 designers and 100 listeners. The study addressed the third research question by providing examples of how the SMT could be used by designers to compare their intentions for a sound design with the experiences of listeners. Designers proposed a number of modifications, the most important being the reduction in the number of attributes. Ten sound designs were mapped and the tool was found to be suitable for comparing sound designs with listeners' experiences by all 10 of the designers who took part. In the next chapter (8) the research aim and questions are reviewed, strengths and limitations of the thesis are discussed, and further work is identified.

8 Conclusions and indication of further work

This final chapter begins by restating the research aim and questions, and then looks at how each part of the thesis has addressed them. Strengths and limitations are discussed, and the chapter concludes with further work.

8.1 Answering the research questions

The research aim was to develop and trial a method that can allow sound designers to represent their designs and listeners to communicate what they are attending to. In order to achieve this aim the following three research questions had to be addressed. The first question related to identifying what attributes are important to both sound designers and listeners when describing sound. The second question addressed how a soundscape could be classified and visualised so that it was meaningful to designers. The final question investigated how soundscape mapping could be used by designers to compare their intentions for a design with the experiences of listeners.

The first question addressed the identification of attributes of sounds. In order to compare the different experiences of a sound design then both designers and listeners have to use the same attributes to communicate what they are listening to. It was necessary to identify attributes for describing sound that were understandable to listeners so that their experiences could be captured. It was also important to establish attributes that allowed designers to represent their sound designs.

Chapter 2 reviewed published methods of classifying sound from a listener's perspective. Chapter 3 was concerned with approaches to sound design in both traditional media and computing, which provided an insight into the attributes required by designers. In a preliminary study in Chapter 4 interviews with office inhabitants were used to elicit responses about which attributes were important to listeners when describing sound. A survey completed by 75 audio professionals furnished a list of attributes from contemporary practice. The concurrent verbalisations reported in Chapter 5 gave a corresponding list from listeners. These two sets of attributes (designers and listeners) were then combined to create the first version of the soundscape mapping tool (Chapter 6). After evaluating the tool in Chapter 7, it was found that the 12 attributes conveyed too much information. However, the designers did not suggest that any attributes were missing, and all but one of the attributes were considered to be important by at least half of the designers.

The second question was concerned with the formalisation of attributes for describing sound into a classification that could be visualised in a meaningful form for sound designers. Appropriate scales for each attribute had to be established along with effective methods of visualising each attribute. A published but untested method was chosen and a visualisation was created in Chapter 4. This method was then extended using published work based on the findings of the literature review and the office inhabitant interviews together with a new visualisation. This was tested in the shared office environment demonstrating its effectiveness at capturing listeners' experiences. Having established a way of visually comparing listening experiences, sound designers current methods of classifying and visualising sound were sought in a survey completed by 75 audio professionals in Chapter 5. The responses from the questionnaire were used to make the first version of the soundscape mapping tool, which was used by a single designer for the design, and evaluation of an in-car Auditory Display in Chapter 6. After minor modifications suggested by the designer of the Auditory Display the SMT was used by 10 sound designers and all of them agreed that the tool allowed them to compare a sound design with the experiences of listeners in Chapter 7. Some changes were suggested such as the reduction in the number of attributes, but the methods of displaying the attributes for the remaining attributes were judged to be appropriate. It was suggested that a way of displaying only the differences between experiences was mapped and a method was created but not tested.

The third question addressed how soundscape mapping could be used to compare listening experiences. It was important to survey the ways that sound is designed for different forms of media. A range of sound designs had to be mapped in order to identify and resolve some of the procedural problems associated with the capture of designers' and listeners' responses. Methods of designing sound for different forms of media were surveyed in Chapter 3 in order to identify different approaches to sound design. The preliminary studies in Chapter 4 were used to identify some of the practical problems with classifying and visualising a soundscape. The soundscape mapping tool was used by 11 designers to compare their intentions for a sound design with the experiences of listeners. A version was used for the design and evaluation of a simple in-car auditory display and a few minor modifications were suggested for the attributes and the visualisation. After some of the modifications had been made, 10 designers from traditional media and computing used the SMT to compare their design with the experiences of listeners. The SMT was capable of visualising all 10 of the designs, which ranged from an Auditory Display through to a Composition. Designers found that the map not only provided them with a way of visualising the design but also could be used as a form of communication with other designers; they also thought that it could be used during the design phase to check their progress. It was also suggested that the tool could be used to help choose between designs.

8.2 Strengths and limitations of the tool

Any method, which attempts to capture the experience of inhabiting a soundscape, will have issues with granularity; a balance must be achieved with gathering sufficient data, and overwhelming the participants with queries. Only limited time periods can be studied, as there are necessary time constraints for listeners, not only for availability, but also for issues of fatigue. This makes it difficult to choose representative intervals, which has the knock on effect of potentially missing sound events, due to the short durations that can be studied.

In order to capture the in-car auditory display in context the soundscape mapping tool used surround sound recording equipment, which was expensive, time consuming and required specialist knowledge to operate. This allowed an element of repeatability, in that multiple listeners could be exposed to the same recording at separate times, with minimal impact on the original environment. A recording, even in surround sound, can only capture some of the experience of inhabiting an auditory environment, as the choice of microphones and their placement, along with the reproduction environment colour the sound. This approach does however make it easy to augment the environment, as the necessary reproduction equipment is already part of the process. When studying sound designs out of context, it was comparatively simple to use stereo reproduction systems such as loudspeakers or headphones, but this potentially introduces other problems about levels of listening and colouration of room acoustics and reproduction equipment.

The SMT did capture what people were listening to in a form from which the modal response could be easily calculated. Participants who used the updated tool did not feel any desire to add any additional information. However there are other attributes that contribute to the experience of inhabiting a soundscape; Gabrielsson and Sjögren (1979) refer to the feeling of space, whether it is wide and open or narrow and closed. Hellström (1998) suggests that the origin of a sound is important whether it is man made or natural, whilst Hedfors and Berg (2002) propose that, amongst other characteristics, a soundscape's 'devotional' or 'trivial' nature should be considered, in terms of its importance to the individual inhabiting it. There are a wide variety of attributes that could be included in order to capture the unique experience of inhabiting a soundscape, but for this research it was decided to limit them to those that were considered relevant to both listeners and designers.

G. W. Coleman (2008) highlights the distrust that designers have for non experts' descriptions of auditory environments. Audio professionals spend a considerable amount of time learning to shift between critical and natural listening. The soundscape mapping tool places an emphasis on the need for the designer's classification of the soundscape in conjunction with listeners', so that a clear comparison can be made as to the difference between the two groups. The classification itself was based on the principle of a common language, having been derived from a lexicon generated from descriptions used by participants to describe what they were listening to, and a questionnaire where audio professionals were asked for terms that they used to describe sounds. This meant that the resultant terms would be meaningful to both groups.

A simple comparison of the designer's soundscape map of the pre-existing environment with the listeners' illustrates where similarities and differences lie. Cross referencing what participants were aware of, with all of the recorded sound events, highlights what was being attended to, and what was ignored. The classification provides information about what the perceived events sound like, how relevant they are, whether they are pleasing, clear and what, if any, their emotional impact was. This informs the designer what is favourable and what is considered to be neutral or unfavourable.

The simplest form of evaluation is a comparison between what the designer intended and what the listeners experienced. This can either be done in isolation, or in context with the pre-existing environment subtracted. Both approaches could show where the classifications matched and where they differed. Auditory alerts that the designer considered to be clear and informative, but which are classified by listeners as unclear and uninformative, could be said to be unsuccessful. It is also possible to subtract the augmentation from the results in order to study the impact upon the pre-existing environments, such as a sound event no longer being attended to, or changing from clear to unclear.

Auditory environments can also be evaluated on their own, in order to establish where they could be improved. Acoustic treatments, and removal or relocation of sound generating objects can be mapped in the same manner as augmented audio.

Temporary or permanent solutions can be developed and their effects monitored. Separate recordings may be made before and after the changes, or participants can be questioned in the physical location. Providing that the two time periods under study are similar, differences can be highlighted, such as traffic becoming soft and speech becoming clear.

In a similar manner the impact upon the experience of an auditory display or environment whilst a listener is wearing single ear-pieces, passive or active noise cancelling headphones could also be evaluated. Increasingly individuals working in the emergency services have to wear ear-pieces whilst performing their normal tasks, from driving, through to interviewing witnesses, this partially isolates them from their auditory environment, as well as often distracting them. The extent of this isolation and the level of distraction can be shown through the changes in classification and the number and types of sound events which they are aware of when experiencing the auditory environment with and without the ear-pieces, which can also be tested when they are silent and with appropriate content. In a similar manner it is possible to test how the soundscape alters when using noise cancelling technologies. Anecdotal accounts have referred to how much more annoying speech and other mid to high-pitched sounds become when the background ambient noise has been reduced.

8.3 Further work

Although the 10 sound designers who take part in the final study agreed that the soundscape mapping tool allowed them to compare a sound design with the experiences of listeners, further work has been identified. Work will have to be conducted on the internal validity of each attribute, as well as on how to calculate the modal response. In order to improve the internal validity, attributes will be tested individually in order to establish the correct scale and provide easily understood descriptors. Initial studies will be done using pure tones, with the complexity of the sound events increasing as scales are established until a consistent response can be obtained, this approach is suitable for four out of seven of the attributes (awareness, spatial, dynamics and clarity), but for the remaining three (type, content and emotions) studies will have to begin with complex sound events.

Of the seven attributes that are retained six had a single method of visualisation chosen by the majority of the designers (see Figure 8-1). There was only a low level of agreement for the visualisation of the dynamics attribute, varying the dimensions of the shape was only chosen by 4 out of 10 of the designers. Alternative methods of displaying this attribute will be sought and then trialled with designers. Choices might include dimensions, colour, symbols, gradients, border weight and hatching. After a reduced set of preferred methods of visualisation has been identified, the procedure will be repeated with the inclusion of all of the other attributes.

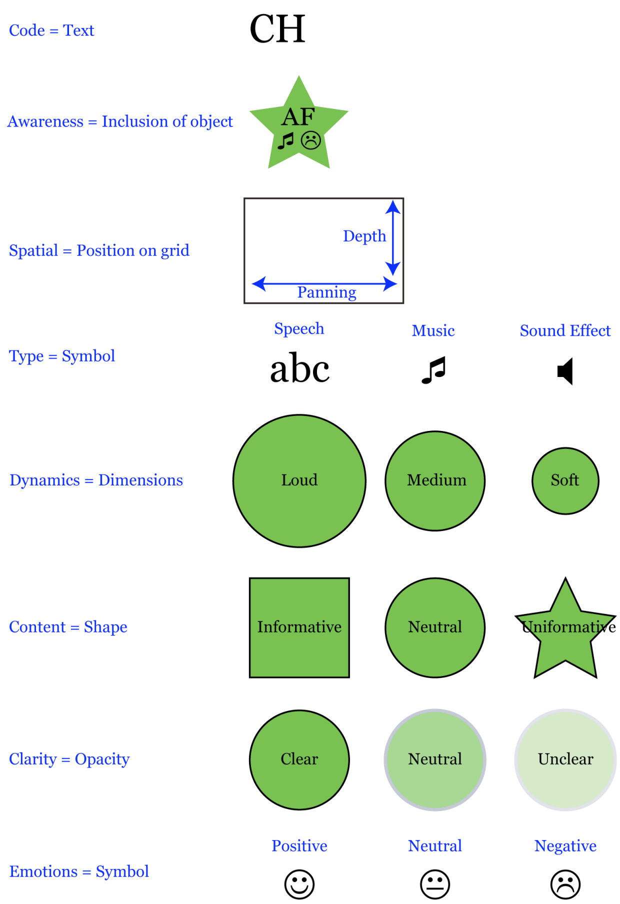
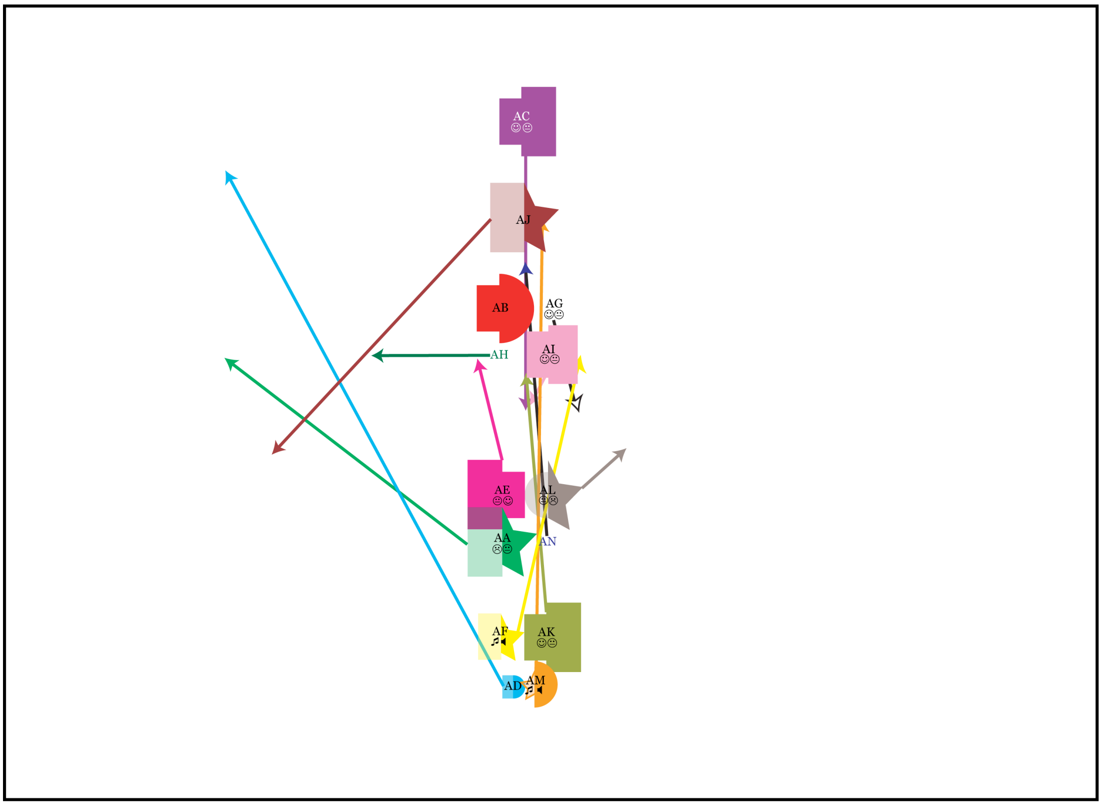


Figure 8-1: Possible reduced set of visualisations

Studies will also be conducted to establish the most appropriate length and level of complexity for listeners to experience. There was not an easily identifiable relationship between length and complexity of a sound design with regards to the level of awareness of sound events. A balance between reliance on listeners' memories and the level of awareness needs to be made. During the final study designers were given the choice of whether participants could listen once or repeatedly. Listening only once appeared to provide more information about levels of awareness, but meant that sound events were either forgotten or possibly inaccurately recalled. Allowing listeners to listen repeatedly meant that listeners were aware of every sound event, and were able to describe relatively accurately what they were listening to. A study will be designed to investigate the relationship between complexity and length of time. This experiment will also have to consider the length of time between listening to the sound event and questioning, especially if there is a long list of sound events, as this may have an impact as well. Randomizing the order of questioning can help mitigate this effect, but it should still be investigated.

A method of highlighting only the differences between a sound designers' intentions and the listeners' experiences within a single map was suggested by one of the designers. A possible way of showing the differences on a single map might be to overlay the listeners' responses onto the designer's, but omit all of the attributes that are identical, whilst still retaining the code in order to indicate the sound event's presence (see Figure 8-2). In this example each sound event has been given a different colour in order to aid identification. The sound events are located on the grid according to the designer's responses and the arrows indicate the positions according to the listeners. The designer's response is on the left hand side of each object, with the listeners' on the right.

If the two classifications are identical then only the letters denoting the sound event will be visible, if the listeners have indicated that they experienced it from a different 'spatial' location then an arrow will mark the difference. In the case of AH (Piano [Piano]) all of the attributes are identical except for 'spatial', so the code is visible with a green arrow showing that the listeners considered it be panned further to the left. With AM (Leather bass drum [Piano2]) there were a greater number of differences, the small semi-opaque half-star with the quavers on the left of the orange object indicate that the designer considered the sound event to be *music*, *soft*, *uninformative* and *unclear*. The larger opaque semi-circle with the loudspeaker symbol on the right of the orange object show that the listeners thought AM to be a *sound effect* with *medium* dynamics, *neutral* content and *clear*, they also thought that it was spatially further back, in a similar location to AJ.



Code	Description	Code	Description	Code	Description
AA	Wooden Country side door (clasico)	AF	Leather bass drum (Folcklore)	AK	Voice "xxxxx" (Piano2)
AB	Plucked Strings (clasico)	AG	Voice "xxxxx" (Folcklore)	AL	Wood knocks (Piano2)
AC	Voice (clasico)	AH	Piano (Piano)	AM	Leather bass drum (Piano2)
AD	Birds (Folcklore)	AI	Voice "xxxxx" (Piano)	AN	Piano (Piano2)
AE	Classical guitar (Folcklore)	AJ	Door (Piano2)		

Figure 8-2: Soundscape map for the Audiologos indicating differences between Sound Designer's and Listeners' responses.

The same designer also suggested classifying the design as a whole, so that separate designs might be easily compared. This could either be achieved by asking the listeners and the designer to classify the complete sound design during the original questioning, or it could be derived in the same manner as the listeners' combined response. The median could be used to calculate the spatial values and the mode for each of the remaining six attributes: awareness, type, dynamics, content, clarity and emotions.

The visualisation showing only the differences between the designer's and listeners' responses also requires to be trialled. Initially maps will be generated for all of the designs in the final study (Chapter 7) and responses will be sought from the original designers. If the concept is thought to be successful, then different methods of visualising differences in designs will be explored and trialled with designers.

Once all of the above work has been conducted, a software version of the tool will be developed. At present all of the visualisation is created by hand and depending on the number of sources within an environment a map can take anywhere from 15 minutes to 4 hours to create. The visualisation process would be speeded up through automation. This could be done in two ways data could be combined in similar manner to SPSS (Statistical Package for the Social Sciences) (IBM, 2010) and then modelled accordingly after an appropriate auditory backdrop has been generated. This could be extended so that designers import sounds directly and then drag a visualisation directly onto the map, and then apply their own classification. AB comparisons and subtractions could then be easily made. The designer could also have access to the original multi-channel recording, as well as being able to generate a soundfield from the audio files that they imported, in a similar manner to a digital audio workstation, with the main difference that dynamics and panning were controlled directly from the map rather than from a traditional simulated mixer. The output could then be experienced via headphones using Head Related Transfer

Functions (HRTFs) or from multiple loudspeakers, in a manner similar to Auralisation.

Three immediate applications have arisen from this research; the first is the evaluation of an auditory display for commercial vehicles. The current auditory displays under study in commercial vehicles have warnings with different levels of immanency, from non-urgent such as a new SMS (Short Message Service) message through to highest urgency where the driver needs to stop the vehicle as there is an oil leak, or similar. The responses from the co-driver as well as the driver can be sought, as there are also warning alerts for when the driver is not paying sufficient attention to their driving. The driver's experiences could be compared to the designer's intentions. Sound events that are not part of the auditory display but are still important to the driver can be identified, such as cars passing or the handbrake release. Note can be made to make sure that none of the auditory display sounds either mask them or sound similar. If warning sounds are designed to convey *negative* emotions and the listeners experience it as *neutral*, then sound events that were considered to be *negative* by the listeners can be studied in order to try and identify what makes a sound event *negative*.

A second application is for the evaluation of a simulated virtual audio environment. As with all forms of audio reproduction there is a balance to be achieved between realism and artifice. The level of awareness might indicate which sound events are perceived and the level of clarity could be used to identify which sound events can be heard clearly. If there is little agreement about a particular attribute, such as when a sound event has been classified as being both *informative* and *neutral*, issues such as lack of clarity and *soft* dynamics might suggest possible causes. The spatial cues might indicate if the sound events are experienced as being located in the correct coordinates, especially if the listener moves around the environment. The amount of emotional involvement in either *positive* or *negative* terms can also be inferred. The granularity of the sound events may be scaled up or down so that individual or combined elements could be classified. The effect of simplifying a simulation could also be tested. Only sound events considered *informative* by the designers could be included, or *uninformative* sound events might be introduced at low levels to test what impact this has upon the listeners' responses.

A third application for the SMT could be a study of the impact of auditory displays within hospitals. Persson, Waye and Ryherd (2008) highlight the negative impact of the sound environment in hospitals. International standards have been set (ISO, 2006) but Pope (2008) believes that further studies should be conducted about the effect of the sounds on patients and staff. In 2009 a registered nurse was suspended for 18 months partially due to having silenced a child's heart monitor (Nursery and Midwifery Council, 2009). The SMT could be used to test listeners' responses to new and existing machinery. Combinations of technologies can change according to patient diagnoses as well as with shared occupancy. Simulated environments could be mapped to highlight problems and successes. In hospitals staff and patients experiences could be captured. *Spatial* cues could be studied to test if it is possible to accurately identify where an alarm is located, thereby associating the warning with the correct patient. Sounds which are *informative* to staff, could be emotionally *negative* to patients, possibly causing stress. The SMT might call attention to sound events that staff might be *unaware* of through habituation, as well as sound events that are *uninformative* or *unclear*.

8.4 Conclusions

A significant shift in design practice will have to take place prior to any method of soundscape mapping being adopted. At present sounds are evaluated either by the designer or a listening jury, predominantly out of context. Contextual studies are concerned more with issues of speech intelligibility or noise pollution rather than inhabitants' unique perspectives, excepting the field of acoustic ecology, which by its very nature does not wish to add any sound events, to what is often classified as an acoustically crowded 'lo-fi' soundscape. Designers never have the luxury of creating a new auditory environment; each new sound has to coexist with the pre-existing auditory environment. Even a listener within an anechoic chamber experiences sound from their own central nervous system, whilst experiencing the supposed silence (Cage, 1973).

Moore (1997) referred to the difficulty of attaining accurate measurements of loudness, let alone pitch, reminding us of sound perception's qualitative nature. The fact that we stop to listen takes the event out of context, and as Porteous and Mastin (1985) found, due to the variability of individual perceptions, any form of classification is difficult to achieve. J. S. Brown and Duguid (2000) drew attention to the 'stuff around the edges', context, background, history, common knowledge and social resources. If these aspects are considered, then at least a few of the gaps could be filled taking us closer to a method of effectively mapping soundscapes. A greater understanding of how listeners would like to affect their auditory community, and what they would like to eliminate or control, might allow designers to produce effective and aesthetically pleasing sounds and auditory environments.

The 10 designers who took part in the final evaluation agreed that the soundscape mapping tool allowed them to compare a sound design with the experiences of listeners. This suggests that the research aim of developing and trialling a method that allows sound designers to represent their designs and listeners to communicate what they are attending to has been achieved. In addition, the three research questions have been addressed. A list of attributes that are important to both sound designers and listeners when describing sound have been identified. A method of classifying and visualising soundscapes that is meaningful has been created. And finally examples of how the soundscape mapping tool could be used by the designers to compare their intentions for a sound design with the experiences of listeners have been provided.

References

Abdulkarim, D., & Abu-Obied, N. (2001). The Sonic Image of the Environment - Characterizing and Evaluating Soundscapes. In J. L. Drever (Ed.), *Sound practice: The 1st UKISC conference on sound culture and environments* (pp. 1-10). Dartington, UK: UK and Ireland Soundscape Community.

Adams, E. (2009). *Fundamentals of Game Design* (2nd ed.). Berkerley, CA: New Riders.

Adobe. (2009a). Animation software, multimedia software: Adobe Flash CS4 Professional Retrieved April, 30, 2009, from http://www.adobe.com/products/flash/

Adobe. (2009b). Computer drawing: Adobe Illustrator CS4 Retrieved April, 30, 2009, from http://www.adobe.com/products/illustrator/

Al'tman, Y. A., Bekhterev, N. N., Vaitulevich, S. F., & Nikitin, N. I. (2004). The Role of Phase Changes in Sound Signals in Localization of Sound Sources. *Neuroscience and Behavioral Physiology, 34*(8), 765-771.

Alexander, C. (1979). *The Timeless Way of Building*. New York: Oxford University Press.

Alexanderson, P., & Tollmar, K. (2006). Being and Mixing - Designing Interactive Soundscapes. In A. Mørch, K. Morgan, T. Bratteteig, G. Ghosh & D. Svanaes (Eds.), *Proceedings of the 4th Nordic conference on Human-computer interaction: changing roles* (pp. 252-261). New York: ACM.

Alkin, G. (1989). *Sound Techniques for Video & TV* (2nd ed.). London: Focal Press.

Alten, S. R. (2011). *Audio in Media* (9th ed.). Boston, MA: Wadsworth.

Altman, R. (1992). The Material Heterogeneity of Recorded Sound. In R. Altman (Ed.), *Sound Theory/Sound Practice* (pp. 15-31). New York: Routledge.

Alty, J. L. (1995). Can we use music in computer-human communication? In M. A. R. Kirby, A. J. Dix & J. E. Finlay (Eds.), *People and Computers X. Proceedings of HCI 95* (pp. 409-423). Cambridge, UK: Cambridge University Press.

Amazon. (2009). Kindle 2: Amazon's New Wireless Reading Device Retrieved February 25, 2009, from http://www.amazon.com/Kindle-Amazons-Wireless-Reading-Generation/dp/B00154JDAI

Ament, V. T. (2009). *The Foley Grail: The Art of Performing Sound for Film, Games, and Animation*. London: Focal Press.

American National Standards Institute. (1994). Acoustical Terminology. New York: American National Standards Institute.

Amphoux, P. (1997). L'identite sonore des villes Europeennes (Vol. 1). Grenoble/Lausanne: Cresson/IREC.

Anderson, L. M., Mulligan, B. E., Goodman, L. S., & Regen, H. Z. (1983). Effects of Sounds on Preferences for Outdoor Settings. *Environment and Behaviour, 15*(5), 539-566.

Appia, A. (1918). *Music and the Art of Theatre* (R. W. Corrigan & M. D. Dirks, Trans.). Coral Gables, FL: University of Miami Press.

Apple. (2010). MacBook Retrieved December, 4, 2010, from http://www.apple.com/uk/macbook/

Arias, E., Eden, H., Fischer, G., Gorman, A., & Scharff, E. (2000). Transcending the individual human mind: Creating shared understanding through collaborative design. *ACM Transactions on Computer Human-Interaction, 7*(1), 84-113.

Asakawa, C., Takagi, H., Ino, S., & Ifukube, T. (2003). *Maximum listening speeds for the blind.* Paper presented at the Proceedings of the 9th International Conference on Auditory Display (ICAD2003), Boston, USA.

Ashcraft, M. H. (2006). *Cognition* (4th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

ATLAS.ti. (2008). ATLAS.ti - The Knowledge Workbench: Topic Overview Retrieved March 03, 2008, from http://www.atlasti.com/product.html

Audio-Technica. (2010a). AT803b Retrieved December, 4, 2010, from http://eu.audio-technica.com/en/products/product.asp?catID=1&subID=6&prodID=555

Audio-Technica. (2010b). AT8418 Retrieved December, 4, 2010, from http://eu.audio-technica.com/en/products/product.asp?catID=6&subID=49&prodID=3354

Augoyard, J.-F. (1998). The Cricket Effect: Which tools for research on sonic urban ambiences? In H. Karlsson (Ed.), *Stockholm, Hey Listen!* (pp. 116-125). Stockholm: The Royal Swedish Academy of Music.

Avid. (2010). Avid | 003 Rack Retrieved December, 4, 2010, from http://www.avid.com/US/products/003-Rack

Azar, J., Saleh, H. A., & Al-Alaoui, M. A. (2007). Sound Visualization for the Hearing Impaired. *International Journal of Emerging Technologies in Learning (iJET), 2*(1).

Babbie, E. (1990). *Survey Research Methods* (2nd ed.). Belmont, CA: Wadsworth Publishing Company.

Ballas, J. A., & Howard Jr, J. H. (1987). Interpreting the Language of Environmental Sounds. *Environment and Behaviour, 19*(1), 91-114.

Ballou, G. M. (2002). Fundamentals and Units of Measurement. In G. M. Ballou (Ed.), *Handbook for Sound Engineers* (3rd ed., pp. 1491-1530). Boston: Focal Press.

Banham, M. (1995). *The Cambridge guide to theatre*. Cambridge: Cambridge University Press.

Barker, L., & Watson, K. (2000). *Listen up: how to improve relationships, reduce stress, and be more productive by using the power of listening*. New York: St. Martin's Press.

Barrass, S. (2003). Sonification Design Patterns. In E. Brazil & B. Shinn-Cunningham (Eds.), *Proceedings of the 9th International Conference on Auditory Display* (pp. 170-175). Boston: Boston University.

Barrass, S., & Frauenberger, C. (2009). A coummunal map of design in auditory display *Proceedings of the 15 International Conference on Auditory Display, Copenhagen, Denmark, May 18-22, 2009* Copenhagen, Denmark: ICAD.

Barrass, S., & Kramer, G. (1999). Using sonification. *Multimedia Systems, 7*(1), 23-31.

Barthes, R. (1977). *Image-Music-Text* (S. Heath, Trans.). New York: Hill and Wang.

Beaman, J. (2006). *Programme Making for Radio*. London: Routledge.

Beauchamp, R. (2005). *Designing Sound for Animation*. Oxford: Focal Press.

Bech, S. (1992). Selection and Training of Subjects for Listening Tests on Sound-Reproducing Equipment. *Journal of the Audio Engineering Society, 40*(7/8), 590 - 610.

Bech, S., & Zacharov, N. (2006). *Perceptual Audio Evaluation*. Chichester, West Sussex: Wiley.

Beck, J., & Grajeda, T. (2008). *Lowering the boom: critical studies in film sound*. Champaign, IL: University of Illinois Press.

Begault, D. R. (1994). *3-D Sound for Virtual Reality and Multimedia*. Boston: AP Professional.

Berg, R. E., & Stork, D. G. (2005). *The Physics of Sound* (3rd ed.). London: Pearonson Prentice Hall.

Bernstein, D. (1997). Creating an Interactive Audio Environment Retrieved February, 20, 2008, from http://www.gamasutra.com/view/feature/3238/creating_an_interactive_audio_.php

Berque, A. (1990). *Médiance: de milieux en paysages*. Montpellier: Reclus.

Bertin, J. (1983). *Semiology of graphics: Diagrams, networks, maps* (W. J. Berg, Trans.). Madison, WI: University of Wisconsin Press.

Björk, E. A. (1985). The perceived quality of natural sounds. *Acustica, 57*, 185–188.

Blattner, M. M., Greenberg, R. M., & Kamegai, M. (1992). Listening to Turbulence: an Example of Scientific Audiolization. In M. M. Blattner & R. M. Dannenberg (Eds.), *Multimedia Interface Design* (pp. 87-102). New York: ACM Press.

Blattner, M. M., Sumikawa, D. A., & Greenberg, R. M. (1989). Earcons and Icons: Their Structure and Common Design Principles. *Human-Computer Interaction, 4*(1), 11-44.

Blauert, J., & Jekosch, U. (1997). Sound Quality Evaluation - A Multi-Layered Problem. *Acustica, Acta Acustica, 83*, 747 - 753.

Blesser, B., & Salter, L. (2007). *Spaces speak, are you listening?: experiencing aural architecture*. London: MIT Press.

Bly, S. (1982). Presenting information in sound *Conference on Human Factors in Computing Systems* (pp. 371-375). New York: ACM.

Boer, J. (2003). *Game Audio Programming*. Hingham, MA: Charles River Media.

Bohme, G. (2000). Acoustic Atmospheres. *Soundscape, 1*(1), 14-18.

Bohme, G. (2004). Atmosfere acustiche. Un contributo all'estetica ecologica. In A. Colimberti (Ed.), *Ecologia della musica: Saggi sul paesaggio sonoro*. Roma: Donzelli.

Bordwell, D., & Thompson, K. (1985). Fundamental Aesthetics of Sound in the Cinema. In E. Weis & J. Belton (Eds.), *Film Sound: Theory and Practice* (pp. 181 - 199). New York: Columbia University Press.

Borgmann, A. (1999). *Holding On To Reality: The Nature of Information at the Turn of the Millenium*. Chicago: The University of Chicago Press.

Bose. (2010). Modeler Sound System Software Retrieved January, 12, 2010, from http://pro.bose.com/ProController?url=/pro/products/modeler.jsp

Boyd, A. (2003). When Worlds Collide: Sound and Music in Film and Games Retrieved February, 21, 2009, from http://www.gamasutra.com/view/feature/2901/when_worlds_collide_sound_and_.php

Bracewell, J. L. (1993). *Sound Design in the Theatre*. Upper Saddle River, NJ: Prentice Hall.

Brandon, A. (2005). *Audio for Games*. Berkeley, CA: New Riders.

Brashears-Morlet, S., Santucci, M., & Morlet, T. G. (2006). *Musicians and Hearing Loss: Comparison to a non-noise exposed population Incidence and Risk Factors*. Presentation. Louisiana State University, Sensaphonics Hearing Conservation & Alfred I. duPont Hospital for Children. New Orleans.

Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge, MA: MIT Press.

Breinbjerg, M. (2005). The Aesthetic Experience of Sound: Staging of Auditory Spaces in 3D computer games. *Aesthetics of Play: A conference on Computer Game Aesthetics* Retrieved March, 30, 2009, from http://www.aestheticsofplay.org/breinbjerg.php

Brewster, S. A. (1994). *Providing a Structured Method for Integrating Non-Speech Audio into Human-Computer Interfaces.* PhD, University of York, York.

Brewster, S. A. (2008). Nonspeech auditory output. In A. Sears & J. Jacko (Eds.), *The Human Computer Interaction Handbook* (2nd ed., pp. 247-264).

Brewster, S. A., Wright, P. C., & Edwards, A. D. N. (1994). A Detailed Investigation into the Effectiveness of Earcons. . In K. Gregory (Ed.), *Auditory Display, Sonification, Audification and Auditory Interfaces* (pp. 471-498). Reading, MA: Addison-Wesley.

Bridgett, R. (2008). Dynamic range: subtlety and silence in video game sound. In K. Collins (Ed.), *From Pac-Man to Pop Music* (pp. 127-133). Aldershot, Hampshire: Ashgate.

British Rowing. (2010). The perfect stroke Retrieved November, 29, 2010, from http://www.britishrowing.org/upload/files/CoachingTraining/PerfectStrokePoster.pdf

Brock, J. W. (1950). *A study of the use of sound effects in Elizabethan Drama.* PhD, Northwestern University, Evanston.

Brown, J. S., & Duguid, P. (2000). *The Social Life of Information*. Boston, MA: Harvard Business School Press.

Brown, M. H., & Hershberger, J. (1992). Color and Sound in Algorithm Animation. *Computer, 25*(12), 52-63.

Brown, P. (2002a). Fundamentals of Audio and Acoustics. In G. M. Ballou (Ed.), *Handbook for Sound Engineers* (pp. 3-21). Boston: Focal Press.

Brown, P. (2002b). Test and measurement. In G. M. Ballou (Ed.), *Handbook for Sound Engineers* (pp. 1463-1489). Boston: Focal Press.

Bruseberg, A., & McDonagh-Philp, D. (2001). New product development by eliciting user experience and aspirations. *International Journal of Human-Computer Studies, 55*(4), 435 - 452.

Buhler, J., Neumeyer, D., & Deemer, R. (2009). *Hearing the movies: music and sound in film history*. Oxford: Oxford University Press.

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

Burris-Meyer, H., Mallory, V., & Goodfriend, L. S. (1979). *Sound in the Theatre*. New York: Theatre Arts Books.

Burwell, C. (2003). Composing for the Coen Brothers. In L. Sider, D. Freeman & J. Sider (Eds.), *Soundscape: The School of Sound Lectures 1998 - 2001* (pp. 195 - 208). London: Wallflower Press.

Bussemakers, M. P., & de Haan, A. (2000). *When it Sounds like a Duck and it Looks like a Dog… Auditory icons vs. Earcons in Multimedia Environments.* Paper presented at the International Conference on Auditory Display, Atlanta, Georgia.

Butler, J. G. (2007). *Television: critical methods and applications* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Buxton, W. A. S. (1989). Introduction to this special issue on nonspeech audio. *Human-Computer Interaction, 4*, 1-9.

Buxton, W. A. S., Baecker, R. M., & Arnott, J. (1985). *A holistic approach to user interface design*. Unpublished manuscript.

Cage, J. (1973). *Silence: Lectures and Writings*. Middletown, CT: Wesleyan University Press.

Cancellaro, J. (2006). *Exploring Sound Design for Interactive Media*. Clifton Park, New York: Thomson Delmar Learning.

Cano, P., Koppenberger, M., Celma, O., Herrera, P., & Tarasov, V. (2004). *Sound Effects Taxonomy Management in Production Environments*. Paper presented at the 25th AES International Conference. Metadata for Audio, London.

Card, S., Mackinlay, J. D., & Shneiderman, B. (1999). *Information visualization: Using vision to think*. San Francisco: Morgan Kaufmann Publishers.

Carello, C., Wagman, J. B., & Turvey, M. T. (2005). Acoustic Specification of Object Properties. In J. Anderson & B. Anderson (Eds.), *Moving Image Theory: Ecological Considerations* (pp. 79-104). Carbondale, IL: Southern Illinois Press.

Carfolio.com. (2010). Peugeot 205 1.4i car technical specifications 1.4 litre (1360 cc) Inline 76 PS, 5 Speed manual Retrieved December, 5, 2010, from http://www.carfolio.com/specifications/models/car/?car=36179

Carlile, S. (2002). An overview of auditory dimensions that can be used to represent information. In R. Nakatsu & H. Kawahara (Eds.), *ICAD 2002* (pp. 3 - 8). Kyoto: Advanced Telecommunications Research Institute (ATR).

Chandler, R. (2005). Organizing And Formatting Game Dialogue Retrieved February, 16, 2010, from http://www.gamasutra.com/features/20051118/chandler_01.shtml

Cheenne, D. J. (2002). Modeling and Auralization. In G. M. Ballou (Ed.), *Handbook for Sound Engineers*. Boston: Focal Press.

Chesnokov, A., & SooHoo, L. (1997). *Subjective and Objective Evaluation of Listening Rooms Acoustics*. Paper presented at the 102nd AES Convention, Munich, Germany.

Childs, G. W. (2007). *Creating Music and Sound for Games*. Boston, MA: Thomson Course Technology.

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

Cohen, J. (1993). "Kirk here:": using genre sounds to monitor background activity. In S. Ashlund, K. Mullet, A. Henderson, E. Hollnagel & T. White (Eds.), *INTERACT '93 and CHI '93 conference companion on Human factors in computing systems* (pp. 63-64). Amsterdam: ACM.

Cohen, J. (1994). Out to Lunch: Further Adventures Monitoring Background Activity. In G. Kramer & S. Smith (Eds.), *Proceedings of the Second International Conference on Auditory Display* (pp. 15-20). Santa Fe, NM: Santa Fe Institute.

Cole, M. (1996). *Cultural Psychology: A Once and Future Discipline*. Cambridge, MA: The Belknap Press of Harvard University Press.

Coleman, G. W. (2008). *The Sonic Mapping Tool.* PhD, University of Dundee.

Coleman, G. W., Macaulay, C., & Newell, A. F. (2008). Sonic mapping: towards engaging the user in the design of sound for computerized artifacts *5th Nordic conference on Human-computer interaction: building bridges* (pp. 83-92). Lund, Sweden: ACM.

Coleman , P. (2004). *Basics: A Beginner's Guide to Stage Sound*. Great Shelford, UK: Entertainment Technology Press.

Collins, K. (2007). An Introduction to the Participatory and Non-Linear Aspects of Video Games Audio. In S. Hawkins & J. Richardson (Eds.), *Essays on Sound and Vision* (pp. 263-298). Helsinki: Helsinki University Press.

Collins, K. (2008). *Game Sound: An Introduction to the History, Theory, and Practice of Video Game Music and Sound Design*. Cambridge, MA: The MIT Press.

Collins, M. (2003). *A professional guide to audio plug-ins and virtual instruments*. Burlington, MA: Focal Press.

Collison, D. (2008). *The Sound of the Theatre*. Eastbourne: Professional Sound and Lighting Association (PLASA).

Connelly, D. W. (2005). *Digital Radio Production*. London: McGraw-Hill.

Conrad, F., Blair, J., & Tracy, E. (1999). Verbal Reports are Data! A Theoretical Approach to Cognitive Interviews *Proceedings of the Federal Committee on Statistical Methodology Research Conference.* Washington, DC: Federal Committee on Statistical Methodology.

COST. (2008). Memorandum of Understanding for the implementation of a European Concerted Research Action designated as COST Action TD0804: Soundscape of European Cities and Landscapes (pp. 34). Brussels: European Cooperation in the field of Scientific and Technical Research

Crisell, A. (1994). *Understanding Radio* (2nd ed.). London: Routledge.

Crocker, M. J. (1998). Introduction. In M. J. Crocker (Ed.), *Handbook of Acoustics* (pp. 3-20). Hoboken, NJ: J. Wiley.

Crook, T. (1999). *Radio Drama: Theory and Practice*. Abingdon, Oxon: Routledge.

Dakic, V. (2007). *Sound Design for Film and Television*. Norderstedt, GmbH: Druck und Bindung: Books on Demand GmbH.

Davenport, E., Higgins, M., & Somerville, I. (2000). Narratives of new media in Scottish households: the evolution of a framework of inquiry. *Journal of the American Society for Information Science, 51*(10), 900-912.

Davies, W. J., Cain, R., Carlyle, A., Hall, D. A., Hume, K. I., & Plack, C. J. (2009). The positive soundscape project: a synthesis of results from many disciplines *Internoise 2009, 23-26 August 2009,* . Ottawa, Canada: Internoise.

de Fossard, E. (2005). *Writing and Producing Radio Dramas*. London: Sage Publications.

Delage, B. (1998). On sound design. In H. Karlsson (Ed.), *Stockholm, Hey Listen!* (pp. 67-73). Stockholm: The Royal Swedish Academy of Music.

Detienne, F., & Soloway, E. (1990). An Empirically-Derived Control Structure for the Process of Program Understanding. *International Journal of Man-Machine Studies, 33*(3), 323-342.

Devereux, P. (2001). *Stone Age Soundtracks: The Acoustic Archaeology of Ancient Sites*. London: Vega.

Digiano, C. J., & Baecker, R. M. (1992). Program Auralisation: Sound Enhancements to the Programming Environment. *Proceedings of Graphics Interface '92*, 44-52.

Doane, M. A. (1985). Ideology and Practice of Sound Editing and Mixing. In E. Weis & J. Belton (Eds.), *Film Sound: Theory and Practice* (pp. 54 - 62). New York: Columbia University Press.

Dolby. (2009). Dolby Axon - Realistic & Interactive Voice Communication for Video Games Retrieved July, 25 2009, from http://www.dolby.com/professional/game_development/technologies/dolby-axon.html

Dombois, F. (2001). Using Audification in Planetary Seismology *International Conference on Auditory Display*. Espoo, Finland.

Dubois, D., Guastavino, C., & Raimbault, M. (2006). A Cognitive Approach to Urban Soundscapes: Using Verbal Data to Access Everyday Life Auditory Categories. *Acta Acustica united with Acustica, 92*(865-874).

Duncker, K. (1945). On problem solving. In J. F. Dashiell (Ed.), *Psychological Monographs* (Vol. 58, pp. 1-114). Washington, D.C.: The American Psychological Association.

Eargle, J. (2001). *The Microphone Book*. Oxford: Focal Press.

Eargle, J., & Foreman, C. (2002). *Audio engineering for sound reinforcement*. Milwaukee, WI: Hal Leonard Corporation.

Edworthy, J. (1998). Does sound help us to work better with machines? A commentray on Rauterberg's paper 'About the importance of auditory alarms during the operation of a plant simulator'. *Interacting with Computers, 10*, 401-409.

Edworthy, J., Loxley, S., Geelhoed, E., & Dennis, I. (1989). The perceived urgency of auditory warnings. *The Proceedings of the Institute of Acoustics, 11*(5), 73-80.

Eide, E., Fernandez, R., Hoory, R., Hamza, W., Kons, Z., Picheny, M., et al. (2006). The IBM Submission to the 2006 Blizzard Text-to-Speech Challenge *Interspeech 2006*. Pittsburgh, Pennsylvania.

Ekman, I. (2005). *Meaningful Noise: Understanding Sound Effects in Computer Games.* Paper presented at the Digital Arts and Cultures (DAC), Copenhagen.

Ekman, I., & Lankoski, P. (2009). Hair-Raising Entertainment: Emotions, Sound, and Structure in Silent Hill 2 and Fatal Frame. In B. Perron (Ed.), *Horror video games: essays on the fusion of fear and play* (pp. 181-199). Jefferson, NC: McFarland.

Ekman, P., & Friesen, W. V. (1986). A new pan-cultural facial expression of emotion. *Motivation and Emotion, 10*(2), 159-168.

Engelen, H. (1998). Sounds in Consumer Products. In H. Karlsson (Ed.), *Stockholm, Hey Listen!* (pp. 65-66). Stockholm: The Royal Swedish Academy of Music.

EPSRC. (2006). The Positive Soundscape Project: A re-evaluation of environmental sound Retrieved February, 27, 2010, from http://gow.epsrc.ac.uk/ViewGrant.aspx?GrantRef=EP/E011624/1

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review, 87*(3), 215-251.

Esslin, M. (1987). *The Field of Drama: How the Signs of Drama Create Meaning on Stage and Screen*. London: Methuen.

Everest, F. A. (2001). *Master Handbook of Acoustics* (4th ed.). New York: McGraw-Hill.

Everest, F. A. (2002). Psychoacoustics. In G. M. Ballou (Ed.), *Handbook for Sound Engineers* (pp. 23-40). Boston: Focal Press.

Everest, F. A., & Pohlmann, K. C. (2009). *Master Handbook of Acoustics* (5th ed.). London: McGraw Hill.

Extech Instruments. (2010). 407735: Dual Range Sound Lever Meter Retrieved December, 4, 2010, from http://www.extech.com/instruments/product.asp?catid=18&prodid=237

Fagerlönn, J., & Liljedahl, M. (2009). *AWESOME Sound Design Tool: A Web Based Utility that invites End Users into the Audio Design Process.* Paper presented at the 15th International Conference on Auditory Display, Copenhagen, Denmark.

Ferrington, G. (1994). Keep Your Ear-Lids Open. *The Journal of Visual Literacy, 14*(2), 51 - 61.

Finelli, P. (2002). *Sound for Stage*. Cambridge: Entertainment Technology Press.

Fitch, W. T., & Kramer, G. (1994). Sonifying the Body Electric: Superiority of an Auditory over a Visual Display in a Complex, Multivariate System. In G. Kramer (Ed.), *Auditory Display, Sonification, Audification and Auditory Interfaces* (pp. 307-325). Reading, MA: Addison-Wesley.

Fleming, C. (2002). *The Radio Handbook* (2nd ed.). London: Routledge.

Flickinger, R. C. (1926). *The Greek Theater and Its Drama* (3rd ed.). Chicago: University of Chicago Press.

Flowers, J. H. (2005). Thirteen Years of reflection on auditory graphing: Promises, pitfalls and potential new directions *International Conference on Auditory Display* (pp. 406-409). Limerick, Ireland.

Foreman, C. (2002). Sound System Design. In G. M. Ballou (Ed.), *Handbook for Sound Engineers* (pp. 1115-1213). Oxford: Focal Press.

Francioni, J. M., Albright, L., & Jackson, J. A. (1991). Debugging parallel programs using sound. *ACM SIGPLAN Notices, 26*(12), 68-75.

Frantti, G. E., & Leverault, L. A. (1965). Auditory Discrimination of Seismic Signals from Earthquakes and Explosions. *Bulletin of the Seismological Society of America, 55*, 1-25.

Frecon, E., Stahl, O., Soderberg, J., & Wallberg, A. (2004). *Visualising Sound Perception in a Submarine: A Museum Installation.* Paper presented at the Eighth IEEE International Symposium on Distributed Simulation and Real-Time Applications, 2004. DS-RT 2004.

Freeman, J., & Lessiter, J. (2001). Here, There and Everywhere: The effects of multichannel audio on presence *Proceedings of the 2001 International Conference on Auditory Display* (pp. 231-234). Espoo, Finland: ICAD.

Friberg, A. (2004). *A fuzzy analyzer of emotional expression in music performance and body motion.* Paper presented at the Music and Music Science, Stockholm.

Friberg, J., & Gardenfors, D. (2004). *Audio games: New Perspectives on game audio.* Paper presented at the Advances in Computer Entertainment Technology, Singapore.

Gabrielsson, A., & Lindström, B. (1985). Perceived Sound Quality of High-Fidelity Loudspeakers. *JAES, 33*(1/2), 33-53.

Gabrielsson, A., & Sjögren, H. (1979). Perceived sound quality of sound-reproducing systems. *Journal of the Acoustical Society of America, 65*(4), 1019-1033.

Gaver, W. W. (1986). Auditory Icons: Using Sound in Computer Interfaces. *Human-Computer Interaction*(2), 167-177.

Gaver, W. W. (1993). What in the World do we Hear? *Ecological Psychology, 5*(1), 1-29.

Gaver, W. W. (1997). Auditory Interfaces. In R. M. Baecker, J. Grudin, W. A. S. Buxton & S. Greenberg (Eds.), *Readings in Human Computer Interaction* (2nd ed., pp. 1003-1041). San Francisco: Morgan Kauffmann Publishers Inc.

Gelfand, S., A. (1998). *Hearing: an Introduction to Psychological and Physiological Acoustics*. New York: Marcel Dekker, Inc.

Genelec. (2010a). Genelec 8030A - Powerful Bi-Amplified Nearfield Monitor System Retrieved December, 4, 2010, from http://www.genelec.com/products/2-way-monitors/8030a/

Genelec. (2010b). Genelec 7070A Active Subwoofer is a Powerful Bass Monitoring Tool Retrieved December, 4, 2010, from http://www.genelec.com/products/subwoofers/7070a/

Giaccardi, E., Eden, H., & Fischer, G. (2006). The Silence of the Lands. *Proceedings of the New Heritage Forum*, 94-114.

Gibson, B. (2005). *The S.M.A.R.T. Guide to Mixers, Signal Processors, Microphones and More*. Boston, MA: Thomson Course Technology PTR.

Gibson, D. (2005). *The Art of Mixing: A Visual Guide to Recording Engineering and Production* (2nd ed.). Boston: Artist Pro Publishing.

Giles, D. C. (2002). *Advanced research methods in psychology*. Abbingdon, Oxon: Routledge.

Gillam, R. (2006). *Performance and Drama in Ancient Egypt*. London: Gerald Duckworth & Co Ltd.

Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should We Trust Web-Based Studies? A Comparative Analysis of Six Preconceptions About Internet Questionnaires. *American Psychologist, 59*(2), 93-104.

Goulding, C. (2002). *Grounded Theory: A Practical Guide for Management, Business and Market Researchers*. London: Sage.

Graham, R. (1999). Use of auditory icons as emergency warnings: evaluation within a vehicle collision avoidance application. *Ergonomics, 42*(9), 1233-1248.

Granö, J. G. (1929). Reine Geographie. *Acta Geographica, 2*, 1-202.

Granö, J. G. (1997). *Pure Geography* (M. Hicks, Trans.). Baltimore: The Johns Hopkins University Press.

Grantham, W. D. (1995). Spatial Hearing and Related Phenomena. In B. C. J. Moore (Ed.), *Hearing* (2nd ed., pp. 297-345). London: Academic Press.

Gray, P. S., Williamson, J. B., Karp, D. A., & Dalphin, J. R. (2007). *The Research Imagination: An Introduction to Qualitative and Quantitative methods*. Cambridge: Cambridge University Press.

Griffiths, G. (2009). Subtitles: Increasing Game Accessibility, Comprehension Retrieved February, 16, 2010, from http://www.gamasutra.com/view/feature/3922/subtitles_increasing_game_.php

Grimshaw, M. (2008). *The Acoustic Ecology of the First-Person Shooter: The Player Experience of Sound in the First-Person Shooter Computer Game*. Saarbrucken: VDM Verlag Dr. Muller.

Guastavino, C. (2006). The Ideal Urban Soundscape: Investigating the Sound Quality of French Cities. *Acta Acustica United with Acustica, 92*, 945-951.

Guastavino, C. (2007). Categorization of Environmental Sounds. *Canadian Journal of Experimental Psychology 61*(1), 54-63

Guastavino, C., & Dubois, D. (2006). From language and concepts to acoustics: How do people cognitively process soundscapes? In C. Burroughs & G. Maling (Eds.), *Proceedings of Inter-noise 2006, the 35th International Congress and Exposition on Noise Control Engineering*. Honolulu, Hawaii: Institute of Noise Control Engineering of the USA.

Guastavino, C., Katz, B. F. G., Polack, J.-D., Levitin, D. J., & Dubois, D. (2005). Ecological Validity of Soundscape Reproduction. *Acta Acustica United with Acustica, 91*(333-341).

Guerraz, A., & Lemordant. (2008). Indeterminate adaptive audio for games on mobiles. In K. Collins (Ed.), *From Pac-Man to Pop Music: Interactive Audio in Games and New Media* (pp. 55-72). Aldershot, Hampshire: Ashgate.

Handel, S. (1989). *Listening: An Introduction to the Perception of Auditory Events*. Cambridge, MA: MIT Press.

Hannan, M. (2003). *The Australian guide to careers in music*. Sydney, NSW: UNSW Press.

Hartmann, W. M. (1997). *Signals, sound, and sensation*. New York: Springer.

Haselgrave, C. M. (1995). Auditory environment and noise assessment. In J. R. Wilson & E. N. Corlett (Eds.), *Evaluation of human work: a practical ergonomics methodology* (2nd ed., pp. 506 - 540). London: Taylor & Francis.

Haughton, P. M. (2002). *Acoustics for audiologists*. London: Academic Press.

Hayward, C. (1994). Listening to the Earth Sing. In G. Kramer (Ed.), *Auditory Display: Sonification, Audification, and Auditory Interfaces* (pp. 369-404). Reading, MA: Addison-Wesley.

HEAD Acoustics. (2006). Conducting Listening Tests Retrieved August 7, 2009, from http://www.head-acoustics.de/downloads/eng/application_notes/Hoertests_07_2006_E.pdf

Heckl, M., & Müller, H. A. (1994). *Taschenbuch der Technischen Akustik*. Berlin: Springer Verlag.

Hedfors, P. (2003). *Site Soundscapes: landscape architecture in the light of sound.* Ph.D., Swedish University of Agricultural Sciences, Uppsala.

Hedfors, P., & Berg, P. G. (2002). Site Interpretation by Skilled Listeners - methods for communicating soundscapes in landscape architecture and planning. In H. Jarviluoma & G. Wagstaff (Eds.), *Soundscape Studies and Methods* (pp. 91 - 114). Helsinki: The Finnish Society for Ethnomusicology, Department of Art, Literature and Music, University of Turku.

Hellström, B. (1998). The Voice of Place: A Case-study of the Soundscape of the City Quarter of Klara, Stockholm. In R. M. Schafer & H. Jarviluoma (Eds.), *Yearbook of Soundscape Studies 'Northern Soundscapes', Vol. 1, 1998* (Vol. 1, pp. 25-42). Tampere: University of Tampere, Department of Folk Tradition.

Hellström, B. (2003). *Noise Design: Architectural Modelling and the Aesthetics of Urban Acoustic Space*. Goteborg, Sweden: Bo Ejeby Forlag.

Helyer, N., Woo, D., & Veronesi, F. (2009). The Sonic Nomadic: Exploring Mobile Surround-Sound Interactions. *IEEE Multimedia, 16*(2), 12-15.

Henkelmann, C. (2007). Improving the Aesthetic Quality of Realtime Motion Data Sonification. Bonn: University of Bonn.

Hermann, T. (2008). Taxonomy and Definitions for Sonification and Auditory Display *International Conference on Auditory Display*. Paris.

Hermann, T., & Hunt, A. (2005). An Introduction to Interactive Sonification. *IEEE Multimedia, 12*(2), 20-24.

Hermann, T., & Ritter, H. (1999). Listen to your Data: Model-Based Sonification for Data Analysis. In G. E. Lasker (Ed.), *Advances in intelligent computing and multimedia systems* (Vol. 08, pp. 189-194). Baden-Baden, Germany: International Institute for Advanced Studies in System research and cybernetics.

Hibbard, B. (2005). Vis5D, Cave5D, and VisAD. In C. D. Hansen & C. R. Johnson (Eds.), *The Visualization Handbook* (pp. 673-688). Oxford: Elsevier Butterworth-Heinemann.

Holland, S., & Morse, D. R. (2002). Audio GPS: spatial audio in a minimal attention interface. *Personal and Ubiquitous Computing, 6*(4), 253-259.

Holman, T. (2000). *5.1 Surround Sound Up and Running*. Oxford, UK: Focal Press.

Holman, T. (2002). *Sound for Film and Television* (2nd ed.). Boston: Focal Press.

Holman, T. (2005). *Sound for Digital Video*. Oxford: Focal Press.

Holman, T. (2010). *Sound for Film and Television* (3rd ed.). Oxford: Focal Press.

Houix, O., Lemaitre, G., Misdariis, N., Susini, P., Franinovic, K., Hug, D., et al. (2006). Closing the Loop of Sound Evaluation and Design (CLOSED), Deliverable 4.1 Everyday sound classification: Sound perception, interaction and synthesis, Part 1 State of the Art (pp. 52). Paris: IRCAM.

Houtsma, A. J. M. (1995). Pitch perception. In B. C. J. Moore (Ed.), *Hearing* (pp. 267-295). London: Academic Press Ltd.

Howard, D. M., & Angus, J. (2006). *Acoustics and Psychoacoustics* (3rd ed.). London: Focal Press.

Huber, D. (2007). *The MIDI Manual: a Practical Guide to MIDI in the Project Studio*. Oxford: Focal Press.

Huber, D. M., & Runstein, R. E. (2010). *Modern Recording Techniques* (7th ed.). Oxford: Focal Press.

IBM. (2010). IBM SPSS Modeler - Data mining, text mining, predictive analysis Retrieved Oct, 31, 2010, from http://www.spss.com/software/modeler/

ISO. (2006). IEC 60601-1-8:2006 *Medical electrical equipment -- Part 1-8: General requirements for basic safety and essential performance -- Collateral standard: General requirements, tests and guidance for alarm systems in medical electrical equipment and medical electrical systems*. Geneva: International Organization for Standardization.

ISO. (2010). TC43/SC1: Noise Retrieved February, 27, 2010, from http://www.iso.org/iso/standards_development/technical_committees/other_bodies/iso_technical_committee.htm?commid=48474

Izhaki, R. (2008). *Mixing Audio: Concepts, Practices and Tools*. London: Focal Press.

Jackson, D. M. (2003). *Sonic Branding*. Basingstoke: Palgrave Macmillan.

James, F. (1998). Lessons from developing audio HTML interfaces *Proceedings of the third international ACM conference on Assistive technologies* (pp. 27-34). New York: ACM.

Jensen, J. J. (2007). Evaluating in a Healthcare Setting: A Comparison Between Concurrent and Retrospective Verbalisation. In J. A. Jacko (Ed.), *Human-computer Interaction: Interaction design and usability* (pp. 508-516). Berlin: Springer.

Jeon, J.-H., Kook, C., & Jang, G.-S. (2006). Classification of Preferred Sounds by Climatic Elements & Field Application of SAFRS. In T. Imada, K. Hiramatsu & K. Torigoe (Eds.), *The West Meets the East in Acoustic Ecology* (pp. 238 - 249). Hirosaki: Japanese Association for Sound Ecology and Hirosaki University International Music Centre.

Johannsen, G. (2004). Auditory Displays in Human–Machine Interfaces. *Proceedings of the IEEE, 92*(4), 742-758.

Jones, D. (2002). Small Room Acoustics. In G. M. Ballou (Ed.), *Handbook for Sound Engineers* (pp. 89-107). Boston: Focal Press.

Jones, D., & Szymanski, J. (2002). Acoustical Treatment for Indoor Areas. In G. M. Ballou (Ed.), *Handbook for Sound Engineers* (pp. 73-88). Boston, MA: Focal Press.

Jordan, P. W. (2000). *Designing Pleasurable Products: An Introduction to the New Human Factors* London: Taylor & Francis.

Jorgensen, K. (2006). On the Functional Aspects of Computer Game Audio *Audio Mostly* (pp. 48-52). Pitea, Sweden.

Jorgensen, K. (2008). Left in the dark: playing with the sound turned off. In K. Collins (Ed.), *From Pac-Man to Pop Music* (pp. 163-176). Aldershot, Hampshire: Ashgate.

Kageyama, T. (1993). Individual Differences in Preference for Environmental Sounds. *Perceptual and Motor Skills, 76*(1), 279 - 284.

Katz, B. (2002). *Mastering Audio: The Art and the Science*. Burlington, MA: Focal Press.

Katz, B. (2007). *Mastering Audio: The Art and the Science* (2nd ed.). Oxford: Focal Press.

Kaye, D., & Lebrecht, J. (2000). *Sound and Music for the Theatre: The Art and Technique of Design* (2nd ed.). Oxford: Focal Press.

Keith, M. C. (2002). Turn On ... Tune In: The Rise and Demise of Commercial Underground Radio. In M. Hilmes & J. Loviglio (Eds.), *Radio Reader: Essays in the Cultural History of Radio* (pp. pp. 389 - 404). New York: Routledge.

Keith, M. C. (2007). *The Radio Station: Broadcast, Satellite and Internet*. Oxford: Focal Press.

Kerins, M. (2010). *Beyond Dolby: Cinema in the Digital Sound Age*. Bloomington, IN: Indiana University Press.

Klæboe, R., Engelien, E., & Steinnes, M. (2006). Context sensitive noise impact mapping. *Applied Acoustics, 67*, 620-642.

Kleiner, M., Dalenbäck, Bengt-Inge, & Svensson, P. (1993). Auralization-An Overview. *Journal of the Audio Engineering Society, 41*(11), 861-875.

Kornfeld, A.-L. (2007). *SOUNDSLIKE: Online-PPGIS kommuniziert akustische Raumperspektive. Über Potentiale und Visualisierung akustischer Rauminformationen im Internet. Sound-Mapping des Schanzenviertels in Hamburg.* Master thesis, University of Hamburg, Hamburg.

Kramer, G. (1994). An Introduction to Auditory Display. In G. Kramer (Ed.), *Auditory Display: Sonification, Audification, and Auditory Interfaces* (pp. 1-77). Reading, MA: Addison-Wesley.

Kramer, G., Walker, B., Bonebright, T., Cook, P., Flowers, J., Miner, N., et al. (1999). Sonification Report: Status of the Field and Research Agenda: ICAD.

Krebber, W., Adams, A., Chourad, N., Genuit, K., Hempel, T., von Hofe, R., et al. (2000). Objective Evaluation of Interior Car Sound – the OBELICS Project *DAGA 2000, Fortschritte der Akustik* (pp. 186-187). Oldenburg, Germany: DAGA.

Krokstad, A., Strøm, S., & Sørsdal, S. (1968). Calculating the acoustical room response by the use of a ray tracing technique. *Journal of Sound and Vibration, 8*(1), 118-125.

Kruth, P., & Stobart, H. (Eds.). (1999). *Sound*. Cambridge: Cambridge University Press.

Larsson, P. (2010). Tools for Designing Emotional Auditory Driver Vehicle Interfaces. In S. Ystad, M. Aramaki, R. Kronland-Martinet & K. Jensen (Eds.), *Auditory Display: 6th International Symposium, CMMR/ICAD 2009, Copenhagen, Denmark, May 18-22, 2009, Revised Papers* (pp. 1-11). Berlin: Springer.

Laurel, B. (1991). *Computers as Theatre*. Menlo Park, CA: Addison Wesley.

Lazar, J., Feng, J. H., & Hochheiser, H. (2009). *Research Methods in Human-Computer Interaction*. Chichester: John Wiley & Sons.

Leach, M. R. (2009). Phase State: Solid, Liquid, Gas at 20°C & 700°C Retrieved December, 7, 2010, from http://www.meta-synthesis.com/webbook/35_pt/pt_database.php?PT_id=135

Lecky-Thompson, G. W. (2002). *Infinite Game Universe: Level Design, Terrain, and Sound*. Hingham, MA: Charles River Media.

Lemmens, P. M. C. (2005). Using the major and minor mode to create affectively-charged earcons *International Conference on Auditory Display* (pp. 205-211). Limerick, Ireland.

Leonard, J. A. (2001). *Theatre Sound*. London: A & C Black.

Leplâtre, G. (2002). *The Design and Evaluation of non-Speech Sounds to Support Navigation in Restricted Display Devices.* PhD, University of Glasgow.

Leplâtre, G., & Brewster, S. A. (2000). Designing non-speech sounds to support navigation in mobile phone menus *International Conference on Auditory Display*. Atlanta, Georgia.

Leplâtre, G., & McGregor, I. (2004). How to Tackle Auditory Interface Aesthetics? Discussion and Case Study In S. Barrass & P. Vickers (Eds.), *Proceedings of ICAD 04-Tenth Meeting of the International Conference on Auditory Display*. Sydney, Australia: International Community for Auditory Display (ICAD).

Levin, G., & Lieberman, Z. (2004). *In-situ speech visualization in real-time interactive installation and performance*. Paper presented at the Proceedings of the 3rd international symposium on Non-photorealistic animation and rendering, Annecy, France.

Lucas, P. A. (1994). *An evaluation of the communicative ability of auditory icons and earcons.* Paper presented at the International Conference on Auditory Display, Santa Fe, New Mexico.

Lukin, A., & Todd, J. (2006). *Adaptive Time-Frequency Resolution for Analysis and Processing of Audio*. Paper presented at the AES 120th Convention, Paris, France.

Lynch, K. (1960). *The image of the city*. Cambridge, MA: The MIT Press.

Lyver, D. (1999). *Basics of Video Sound* (2nd ed.). Oxford, UK: Focal Press.

Macaulay, C., Benyon, D., & Crerar, A. (1998). Voices in the forest: Sounds, soundscapes and interface design. In N. Dahlback (Ed.), *Exploring Navigation: Towards a Framework for Design and Evaluation in Electronic Spaces, SICS Technical Report T98:01*. Stockholm: SICS.

Macaulay, C., & Crerar, A. (1998). 'Observing' the Workplace Soundscape: Ethnography and Auditory Interface Design *ICAD 98, International Conference on Auditory Display*. University of Glasgow, Glasgow.

Madell, J. R., & Flexer, C. (2008). *Pediatric Audiology: Diagnosis, Technology and Management*. New York: Thieme.

Mahnke, F. H. (1996). *Color, Environment and Human Response - an Interdisciplinary Understanding of Color and Its Use as a Beneficial Element in the Design of the Architectural Environment.* New York: Van Nostrand Reinhold.

Marks, A. (2001). *The Complete Guide to Game Audio*. Berkeley, CA: CMP Books.

Marks, A. (2009). *The Complete Guide to Game Audio: For Composers, Musicians, Sound Designers, and Games Developers* (2nd ed.). Berkeley, CA: Focal Press.

Marshall, L. G. (2009). Acoustical Design: Places for Listening. In W. J. Cavanaugh, G. C. Tocci & J. A. Wilkes (Eds.), *Architectural Acoustics: Principles and Practice*. Hoboken, NJ: John Wiley & Sons.

Mason, R. (2002). *Elicitation and measurement of auditory spatial attributes in reproduced sound.* PhD, University of Surrey, Guildford, Surrey. Retrieved from http://www.surrey.ac.uk/soundrec/pdf/masonthesisf.pdf

Mathur, P. (2009). *Visualizing Remote Voice Conversations: Uses from Artifacts to Archival.* Masters, University of Illinois, Urbana, IL.

Matthews, T., Fong, J., & Mankoff, J. (2005). Visualizing Non-Speech Sounds for the Deaf *Proceedings of the 7th international ACM SIGACCESS conference on Computers and accessibility* (pp. 52-59). New York: ACM.

McGookin, D. K., & Brewster, S. A. (2004). Understanding concurrent earcons: applying auditory scene analysis principles to concurrent earcon recognition. *ACM Transactions on Applied Perception, 1*(2), 130-155.

McLeish, R. (2005). *Radio Production*. Oxford: Focal Press.

McWhinnie, D. (1959). *The Art of Radio*. London: Faber & Faber.

Merleau-Ponty, M. (1962). *Phenomenology of Perception* (C. Smith, Trans.). London: Routledge.

Metric Halo. (2008). Metric Halo SpectraFoo Retrieved April, 29, 2009, from http://www.mhlabs.com/metric_halo/products/foo/

Metz, C. (1985). Aural Objects. In E. Weis & J. Belton (Eds.), *Film Sound: Theory and Practice* (pp. 154 - 161). New York: Columbia University Press.

Miller, W. C. (1931). The Illusion of Reality in Sound Pictures. In L. Cowan (Ed.), *Recording Sound for Motion Pictures* (pp. 210-217). London: McGraw-Hill Book Company.

Misra, A., Cook, P. R., & Wang, G. (2006). *Musical Tapestry: Re-composing Natural Sounds.* Paper presented at the International Computer Music Conference (ICMC), New Orleans.

Misra, R. P., & Ramesh, A. (1989). *Fundamentals of cartography*. New-Delhi: Concept Publishing Company.

Monmonier, M. (1993). *Mapping it out: Expository Cartography for the Humanities and Social Sciences*. London: University of Chicago Press.

Moore, B. C. J. (1997). *An Introduction to the Psychology of Hearing* (4th ed.). London: Academic Press.

Mott, R. L. (1993). *Radio Sound Effects: Who Did It, and How, in the Era of Live Broadcasting*. London: McFarland.

Mott, R. L. (2009). *The audio theater guide: vocal acting, writing, sound effects and directing for a listening audience*. Jefferson, NC: McFarland & Company.

Moylan, W. (2002). *The Art of Recording: Understanding and Crafting the Mix*. London: Focal Press.

Murch, W. (2003). Touch of Silence. In L. Sider, D. Freeman & J. Sider (Eds.), *Soundscape: The School of Sound Lectures 1998-2001* (pp. 83-102). London: Wallflower Press.

Mynatt, E. D. (1994). Designing with Auditory Icons: How Well Do We Identify Auditory Cues? *CHI '94* (pp. 269-270). Boston, MA.

Napier, F. (1962). *Noises Off: A Handbook of Sound Effects* (4th ed.). London: J. Garnet Miller Ltd.

Navcon. (2007). Noise Contour Map Retrieved February, 23, 2007, from http://www.navcon.com/SPLANContourmap.htm

Nees, M. A., & Walker, B. N. (2009). Auditory Interfaces and Sonification. In C. Stephanidis (Ed.), *The Universal Access Handbook* (pp. 507-521). New York: CRC Press.

Nelson, R. (2007). Quality TV Drama: Estimations and Influences Through Time and Space. In J. McCabe & K. Akass (Eds.), *Quality TV: contemporary American television and beyond*. London: I.B. Tauris.

Neuhoff, J. G. (2004). *Ecological psychoacoustics*. London: Elsevier.

Newell, P. (2008). *Recording Studio Design* (2nd ed.). Oxford: Focal Press.

Newman, R. (2009). *Cinematic game secrets for creative directors and producers: inspired techniques from industry legends*. Oxford: Focal Press.

Nisbett, A. (2003). *The sound studio: audio techniques for radio, television, film and recording*. London: Focal Press.

Norberg-Schulz, C. (1976). The Phenomenon of Place. *Architectural Association Quarterly, 8*(4), 3 - 10.

North, A. C., Hargreaves, D. J., & McKendrick, J. (1999). The effect of music on in-store wine selections. *Journal of Applied Psychology*(84), 271-276.

Nursery and Midwifery Council. (2009). A Substantive Hearing of the Conduct and Competence Panel Retrieved June 15, 2009, from http://www.nmc-uk.org/aDisplayDocument.aspx?DocumentID=5926

Ohlson, B. (1976). Sound fields and sonic landscapes in rural environments. *Fennia*(148), 33 - 45.

Ong, W. (1967). *The Presence of the Word: Some Prolegomena for Cultural and Religious History*. New York: Simon & Schuster.

Owsinski, B. (1999). *The Mixing Engineer's Handbook*. Vallejo, CA: Mix Books.

Ozcan, E., & van Egmond, R. (2005). Characterizing Descriptions of Product Sounds. In D. Brock (Ed.), *Proceedings of ICAD 05* (pp. 55-60). Limerick: ICAD.

Palladino, D. K., & Walker, B. N. (2007). *Learning Rates for Auditory Menus Enhanced with Spearcons Versus Earcons.* Paper presented at the Proceedings of the 13th International Conference on Auditory Display (ICAD2007), Montreal, Canada.

Patterson, R. D. (1982). *Guidelines for Auditory Warning Systems on Civil Aircraft*. London: Civil Aviation Authority.

Paul, L. J. (2008). An introduction to granular synthesis in video games. In K. Collins (Ed.), *From Pac-Man to Pop Music* (pp. 135-149). Aldershot, Hampshire: Ashgate.

Peat, J., Mellis, C., & Williams, K. (2002). *Health science research: a handbook of quantitative methods*. London: Sage.

Persson Waye, K., & Ryherd, E. (2008). Relating the hospital sound environment to occupant psychological and physiological response *Acoustics '08*. Paris: Acoustical Society of America (ASA).

Peterson, A. P. G., & Gross, E. E. (1972). *Handbook of Noise Measurement* (7th ed.). Concord, MA: General Radio.

Pierce, A. D. (1981). *Acoustics: An Introduction to Its Physical Principles and Applications*. New York: McGraw-Hill Book Company.

Plack, C. J., & Carlyon, R. P. (1995). Loudness Perception and Intensity Coding. In B. C. J. Moore (Ed.), *Hearing* (2nd ed., pp. 123-160). London: Academic Press.

Plantinga, C. R. (2009). *Moving viewers: American film and the spectator's experience*. London: University of California Press.

Plutchik, R. (1980). A general psychoevolutionary theory of emotion. In R. Plutchik & H. Kellerman (Eds.), *Emotion: Theory, research, and experience: Vol. 1. Theories of emotion* (pp. 3-33). New York: Academic.

Pope, D. S. (2008). Hospital Noise: Risk to Health and Safety. *Perspectives on Hearing and Hearing Disorders: Research and Diagnostics, 12*(2), 40-47. doi: 10.1044/hhd12.2.40

Porteous, J. D., & Mastin, J. F. (1985). Soundscape. *Journal of Architectural Planning Research, 2*(3), 169-186.

Purcell, J. (2007). *Dialogue Editing for Motion Pictures: A Guide to the Invisible Art*. London: Focal Press.

Rabbitt, P. (1990). Mild hearing loss can cause apparent memory failures which increase with age and reduce with IQ. *Acta Oto-Laryngologica. Supplementum, 476*, 167-175.

Radojevic, M. D., & Turner, R. (2002). *Spatial Forms Generated by Music – The Case Study*. Paper presented at the GA2002 Generative Art and Design Conference, Milan.

Radomskij, P. (2007). Measurement of noise. In L. Luxon & L. Prasher (Eds.), *Noise and Its Effects* (pp. 13 - 43). Chichester, UK: John Wiley and Sons Ltd.

Raimbault, M., & Dubois, D. (2005). Urban soundscapes: Experiences and knowledge. *Cities, 22*(5), 339-350.

Ramakrishnan, C., & Greenwood, S. (2009). *Entropy sonification.* Paper presented at the Proceedings of the 15th International Conference on Auditory Display (ICAD2009), Copenhagen, Denmark.

Reese, D. E., Gross, L. S., & Gross, B. (2006). *Radio Production Worktext: Studio and Equipment* (5th ed.). London: Focal Press.

Renkus-Heinz. (2007). EASE 4.1 Retrieved February, 23, 2007, from http://www.renkus-heinz.com/ease/index.html

Reynolds, R. A., Woods, R., & Baker, J. D. (2007). *Handbook of Research on Electronic Surveys and Measurements*. London: Idea Group Reference.

Rice, T. (2001). Soundselves: An Acoustemology of Sound and Self within the Edinburgh Royal Infirmary *Sound Practice* (pp. 87-92). Dartington, UK: UK and Ireland Soundscape Community.

Riley, M., Byrne, W., Finke, M., Khudanpur, S., Ljolje, A., McDonough, J., et al. (1999). Stochastic pronunciation modelling from hand-labelled phonetic corpora. *Speech Communication, 29*, 209-224.

RNID. (2007). Facts and figures about deafness Retrieved January, 8, 2007, from http://www.rnid.org.uk/mediacentre/key_rnid_facts/facts_and_figures_about_deafness_and_tinnitus/

Roads, C. (1996). *The Computer Music Tutorial*. Cambridge, MA: The MIT Press.

Robare, P., & Forlizzi, J. (2009). Sound in Computing: A Short History. *Interactions, XVI*(1), 62-65.

Röber, N., & Masuch, M. (2004). Interacting with Sound: An Interaction Paradigm for Virtual Auditory Worlds *ICAD 04 Tenth Meeting of the International Conference on Auditory Display*. Sydney.

Rodaway, P. (1994). *Sensuous Geographies: Body, Sense and Place*. London: Routledge.

Rogers, S. (2010). *Level Up!: The Guide to Great Video Game Design*. Chichester: John Wiley & Sons.

Rose, J. (2009). *Audio Postproduction for Film and Video* (2nd ed.). London: Focal Press.

Rossing, T. D. (2007). *Springer handbook of acoustics*. New York: Springer.

Rossing, T. D., Moore, F. R., & Wheeler, P. A. (2002). *The science of sound* (3rd ed.). San Francisco: Addison Wesley.

Rumsey, F. (2001). *Spatial Audio*. Oxford: Focal Press.

Russo, J. E., Johnson, E. J., & Stephens, D. L. (1989). The validity of verbal protocols. *Memory and Cognition, 17*(6), 759-769.

Sanger, G. A. (2004). *The Fat Man on Game Audio: Tasty Morsels of Sonic Goodness*. Indianapolis, IN: New Riders.

Scaletti, C., & Craig, A. B. (1991). Using sound to extract meaning from complex data. *Proceedings of the 1991 SPIE/SPSE Symposium on Electronic Imaging Science and Technology, 1459*, 207-219.

Schaeffer, P. (1966). *Traite des objets musicaux*. Paris: Seuil.

Schafer, R. M. (1977). *The Tuning of the World*. Toronto: McClelland and Stewart Limited.

Schafer, R. M. (1993). *Voices of Tyranny: Temples of Silence*. Indian River, Ontario: Arcana Editions.

Schafer, R. M. (1994). *The soundscape: Our sonic environment and the tuning of the world*. Rochester, Vt: Destiny Books.

Schiewe, J., & Kornfeld, A.-L. (2009). Framework and Potential Implementations of Urban Sound Cartography *12th AGILE International Conference on Geographic Information Science*.

Schleicher, R., Sundaram, S., & Seebode, J. (2010). *Assessing audio clips on affective and semantic level to improve general applicability.* Paper presented at the Fortschritte der Akustik - DAGA 2010, Berlin.

Schoenborn, C. A., & Heyman, K. (2008). *Health Disparities Among Adults With Hearing Loss: United States, 2000-2006*. Centers for Disease Control and Prevention Retrieved from http://www.cdc.gov/nchs/data/hestat/hearing00-06/hearing00-06.pdf.

Schwartz, T. (1973). *The Responsive Chord*. Garden City, New York: Anchor Books.

Sears, A., & Jacko, J. A. (2008). *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications* (2nd ed.). London: CRC Press.

Sergi, G. (2004). *The Dolby era: Film sound in contemporary Hollywood*. Manchester, UK: Manchester University Press.

Servigne, S., Kang, M.-A., & Laurini, R. (2000). GIS for Urban Soundscape: From Static Maps to Animated Cartography. *2nd International Conference on Decision Making in Urban and Civil Engineering*, 437-448.

Servigne, S., Laurini, R., Kang, M.-A., & Li, K. J. (1999). First Specifications of an Information System for Urban Soundscape *IEEE International Conference on Multimedia Computing and Systems (ICMCS'99), 2*, 262-266.

Sieveking, L. (1934). *The Stuff of Radio*. London: Cassell and Company Ltd.

Sikora, C. A., Roberts, L., & Murray, L. T. (1995). Musical vs. Real World Feedback Signals *CHI '95 Mosaic of Creativity* (pp. 220-221). Denver, Colorado.

Smalley, D. (1986). Spectro-morphology and Structuring Processes. In S. Emmerson (Ed.), *The Language of Electroacoustic Music* (pp. 61-93). New York: Harwood Academic Publishers.

Smith, S., Bergeron, R. D., & Grinstein, G. G. (1990). Stereophonic and surface sound generation for exploratory data analysis *SIGCHI conference on Human factors in computing systems: Empowering people* (pp. 125-132). Seattle, WA: ACM.

Soderholm, M. (1998). Listening Test as a Tool in Sound Quality Work: Applied to Vacuum Cleaners *The Marcus Wallenberg Laboratory for Sound and Vibration Research, Department of Vehicle Engineering*. Stockholm: Royal Institute of Technology.

Sonnenschein, D. (2001). *Sound Design: The Expressive Power of Music, Voice and Sound Effects in Cinema*. Studio City, CA: Michael Wise Productions.

Sony. (1995). TCD-D8 Digital Audio Tape-corder Operating Instructions Retrieved December, 4, 2010, from http://www.docs.sony.com/release/TCDD8.pdf

Southworth, M. (1969). The Sonic Environment of Cities. *Environment and Behaviour, 1*(1), 49-70.

Speeth, S. D. (1961). Seismometer Sounds. *Journal of the Acoustical Society of America, 33*(7), 909-916.

Stanton, J. M. (1998). An Empirical Assessment of Data Collection Using the Internet. *Personnel Psychology, 51*(3), 709-725. doi: 10.1111/j.1744-6570.1998.tb00259.x

Starkey, G. (2004). *Radio in Context*. New York: Palgrave MacMillan.

Steiglitz, K. (1996). *A Digital Signal Processing Primer, with Applications to Digital Audio and Computer Music*. Menlo Park, CA: Addison-Wesley Publishing Company.

Sterne, J. (2003). *The Audible Past: Cultural Origins of Sound Reproduction*. Durham, NC: Duke University Press.

Stockburger, A. (2003). *The Game Environment from an Auditive Perspective.* Paper presented at the Level Up: Digital Games Research Conference, Utrecht.

Stratoudakis, C., & Papadimitriou, K. (2007). A Dynamic Interface for the Audio-Visual Reconstruction of Soundscape, Based on the Mapping of its Properties *Proceedings SMC'07, 4th Sound and Music Computing Conference*, 185-191.

Sumikawa, D. A. (1985). *Guidelines for the integration of audio cues into computer interfaces.* Masters, University of California, Davis, CA.

Sundaram, S., & Narayanan, S. (2008). *Classification of Sound Clips by Two Schemed: Using Onomatopoeia and Semantic Labels.* Paper presented at the 2008 IEEE International Conference on Multimedia and Expo, Hanover, Germany.

Susini, P., McAdams, S., & Winsberg, S. (1999). A Multidimensional Technique for Sound Quality Assessment. *Acustica, Acta Acustica, 86*, 650-656.

Szendy, P. (2008). *Listen: a History of our Ears* (C. Mandell, Trans.). Bronx, NY: Fordham University Press.

Talbot-Smith, M. (1999). *Audio engineer's reference book* (2nd ed.). Oxford: Focal press.

Thalmann, F., & Mazzola, G. (2008). *The BigBang Rubette: Gestural Music Composition with Rubato Composer.* Paper presented at the IMCMC 2008 International Computer Music Conference, Belfast.

Thom, R. (2003). Designing a Movie for Sound. In L. Sider, D. Freeman & J. Sider (Eds.), *Soundscape: The School of Sound Lectures 1998-2001* (pp. 121-137). London: Wallflower Press.

Thomas, R. K. (2000). The Function of the Soundscape *International Organization of Scenographers, Theatre Architects and Technicians Scenography Symposium*. Bregenz, Austria: International Organization of Scenographers, Theatre Architects and Technicians.

Thomson, P. (1992). *Shakespeare's Theatre* (2nd ed.). Oxford: Routledge.

THX. (2010). THX Certified Studio Specifications Retrieved January, 12, 2010, from http://thx.com/studio/specs.html

Toole, F. E. (2006). Loudspeakers and Rooms for Sound Reproduction—A Scientific Review. *JAES, 54*(6), 451-476.

Torigoe, K. (2002). A City Traced by Soundscape. In H. Jarviluoma & G. Wagstaff (Eds.), *Soundscape Studies and Methods* (pp. 39-57). Helsinki: Finnish Society for Ethnomusicology; Department of Art, Music and Literature.

Töyssy, S., & Helenius, M. (2006). About malicious software in smartphones. *Journal in Computer Virology, 2*(2), 109-119.

Truax, B. (Ed.). (1978). *Handbook for Acoustic Ecology*. Burnaby, Canada: Aesthetic Research Centre.

Truax, B. (2001). *Acoustic Communication* (2nd ed.). Norwood: Ablex Publishing Corporation.

Turcan, P., & Wasson, M. (2004). *Fundamentals of Audio and Video Programming for Games*. Redmond, WA: Microsoft Press.

Turner, P., Davenport, E., & Van de Walle, G. (2004). Familiarity as Changing Perception. In D. J. Reed, G. Baxter & M. Blythe (Eds.), *Twelfth European Conference on Cognitive Ergonomics* (pp. 171-179). York: European Association of Cognitive Ergonomics (EACE).

Turner, P., & McGregor, I. (2004). Telling People About Virtual Places: A Qualitative Perspective. *Seventh International Workshop: Presence 2004*, 332 - 339.

Turner, P., McGregor, I. P., Turner, S., & Carroll, F. (2003). Evaluating Soundscapes as a means of creating a sense of place *2003 International Conference on Auditory Display*. Boston, MA: ICAD.

Valle, A., Lombardo, V., & Schirosa, M. (2009). A Graph-based System for the Dynamic Generation of Soundscapes. In M. Aramaki, R. Kronland-Martinet, S. Ystad & K. Jensen (Eds.), *Proceedings of the 15th International Conference on Auditory Display*. Copenhagen, Denmark: ICAD.

Vanderveer, N. J. (1979). *Ecological acoustics: human perception of environmental sounds.* PhD, Cornell University, Ithaca, NY.

Vickers, P. (1999). *CAITLIN: Implementation of a Musical Program Auralisation System to Study the Effects on Debugging Tasks as Performed by Novice Pascal Programmers.* PhD, Loughborough University.

Vickers, P. (2004). External Auditory Representations of Programs: Past, Present, and Future - An Aesthetic Perspective *International Conference on Auditory Display*. Sydney, Australia.

Vickers, P., & Hogg, B. (2006). *Sonification abstraite/sonification concrete: An'aesthetic perspective space'for classifying auditory displays in the ars musica domain.* Paper presented at the 12th International Conference on Auditory Display, London.

Viers, R. (2008). *The Sound Effects Bible: How to create and record Hollywood style sound effects*. Studio City, CA: Michael Wiese Productions.

Waaser, C. M. (1976). *Sound and music for the theatre*. New York: Richards Rosen Press.

Wake, S., & Asahi, T. (1998). Sound Retrieval with Intuitive Verbal Expressions *ICAD 98*. Glasgow: Department of Computer Science, The University of Glasgow.

Walker, B. N., Nance, A., & Lindsay, J. (2006). *Spearcons: speech-based earcons improve navigation performance in auditory menus.* Paper presented at the Proceedings of the 12th International Conference on Auditory Display (ICAD2006), London, UK.

Walne, G. (1990). *Sound for the Theatre*. London: A & C Black.

Warusfel, O., & Eckel, G. (2004). LISTEN: Augmenting everyday environments through interactive soundscapes. *Virtual Reality for Public Consumption, IEEE Virtual Reality 2004 Workshop, Chicago IL, 27*, 268–275.

Watson, M. O., & Sanderson, P. (2007). Designing for Attention With Sound: Challenges and Extensions to Ecological Interface Design. *Human Factors,, 49*(2), 331-346.

Werne, F. (1987). *Den Osynliga artkitekturen (The Invisible Architecture)*. Lund: Vinga Press.

Westerkamp, H. (2000). Editorial. *Soundscape: The Journal of Acoustic Ecology, 1*(1), 3-4.

Whitaker, J., & Benson, B. (2002). *Standard Handbook of Audio and Radio engineering* (2nd ed.). Columbus, OH: McGraw-Hill.

White, G., & Louie, G. (2005). *The audio dictionary*. Seattle, WA: University of Washington Press.

Whittington, W. (2007). *Sound Design and Science Fiction*. Austin, TX: University of Texas Press.

Wilde, M. D. (2004). *Audio Programming for Interactive Games*. London: Focal Press.

Wilson, T. D., & Schooler, J. W. (1991). Thinking too much: introspection can reduce the quality of preferences and decisions. *Journal of Personality and Social Psychology, 60*(2), 181-192.

Wishart, T. (1994). *Audible Design: A Plain and Easy Introduction to Practical Sound Composition*. York: Orpheus the Pantomime.

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments, 7*(3), 225 - 240.

Wollman, E. L. (2006). *The theater will rock: a history of the rock musical : from Hair to Hedwig*. Ann Arbor, MI: University of Michigan Press.

Wyatt, H., & Amyes, T. (2005). *Audio Post Production for Television and Film: An introduction to technology and techniques* (3rd ed.). Oxford: Focal Press.

Yewdall, D. L. (1999). *Practical Art of Motion Picture Sound* Boston: Focal Press.

Zaza, T. (1991). *Audio Design: Sound Recording Techniques for Film and Video*. Englewood Cliffs, NJ: Prentice Hall.

Zwicker, E., & Fastl, H. (1999). *Psychoacoustics: Facts and Models* (2nd ed.). Berlin: Springer.

Appendix A: Pro audio questionnaire

Audio questionnaire i.mcgregor@napier.ac.uk

Introduction

We are very grateful for your input to the early stages of this research into the possibility of creating a computer-based system for notating, classifying and visualising auditory environments. We aim to gain a greater understanding of the auditory contexts in which system designers currently operate, helping to ensure that any new sounds introduced by auditory interface designers are both audible and appropriate, for both the listeners and the shared auditory environment. At the present we are only considering the physical properties of sound.

The purpose of this questionnaire is to find out what methods and terminology audio professionals currently use when notating, classifying and representing sounds.

Many thanks for your help.

Iain McGregor
School of Computing, Napier University
10 Colinton Road, Edinburgh, EH10 5DT
0131 455 2788
i.mcgregor@napier.ac.uk

Your background

1. **What is your job title?**

2. **What are your main responsibilities?**

3. **What is your highest qualification relevant to sound?**

4. **Have you ever experienced any hearing loss, if yes how would you describe it?**

Audio questionnaire i.mcgregor@napier.ac.uk

5. If textbooks or journals are relevant to your work, please list the ones you would recommend?

6. Please rank the importance of the following three factors within your work.

Factors	**Rank** (1=most important, 3= least important)
Pre-existing sounds	
Room acoustics	
Hearing abilities of users	

7. Please give your definitions of noise and rank them according to relevance to your field?

Definition of noise	**Rank** (1= most important, 5= least important)
e.g. Unwanted sound	
a)	
b)	
c)	
d)	
e)	

8. Have you encountered the term soundscape?

If yes, please define your understanding of it?
If no, what does it suggest to you?

Quantitative Measurements

In this section we are interested in discovering the terminology you use when measuring audio, room acoustics and hearing abilities. Please give up to 10 examples of terms you use and rank them according to frequency of usage.

9. What terms are you aware of for measuring audio?

Terms for measuring audio	**Rank** (1=most freq used, 10=least freq used)
e.g. amplitude in Decibels	
a)	
b)	
c)	
d)	
e)	
f)	
g)	
h)	
i)	
j)	

10. What terms are you aware of for measuring room acoustics?

Terms for measuring room acoustics	**Rank** (1=most freq used, 10=least freq used)
e.g. Reverb Time in seconds	
a)	
b)	
c)	
d)	

e)	
f)	
g)	
h)	
i)	
j)	

11. What terms are you aware of for measuring the hearing abilities of listeners?

Terms for measuring hearing abilities	**Rank** (1=most freq used, 10=least freq used)
e.g. Hearing Level in decibels	
a)	
b)	
c)	
d)	
e)	
f)	
g)	
h)	
i)	
j)	

Qualitative Measurements

In this section we are interested in discovering the terms you use for describing audio, room acoustics and hearing abilities (these may include words used in the quantitative section above). Please give up to 10 examples and rank their importance according to frequency of usage.

12. What terms are you aware of for describing audio?

Terms used to describe audio	**Rank** (1=most freq used, 10=least freq used)
e.g. Volume (soft to loud)	
a)	
b)	
c)	
d)	
e)	
f)	
g)	
h)	
i)	
j)	

13. What terms are you aware of for describing room acoustics?

Terms used to describe room acoustics	**Rank** (1=most freq used, 10=least freq used)
e.g. Bright (proportionately large number of high frequency reflections)	
a)	
b)	
c)	
d)	
e)	

f)	
g)	
h)	
i)	
j)	

14. What terms are you aware of for describing the hearing abilities of listeners?

Terms used to describe hearing abilities	**Rank** (1=most freq used, 10=least freq used)
e.g. Frequency Loss (1 – 4 kHz)	
a)	
b)	
c)	
d)	
e)	
f)	
g)	
h)	
i)	
j)	

Notation, Classification and Visualisation

In this section we are interested in discovering the methods you use for notating, classifying and graphically representing sounds. We are interested in both formal and informal methods. Please give up to 10 examples of each and rank their importance according to frequency of use.

15. What methods are you aware of for notating sounds?

Formal method of notating sounds	**Rank** (1=most freq used, 10=least freq used)
e.g. Musical notation	
a)	
b)	
c)	
d)	
e)	
Informal method of notating sounds	**Rank** (1=most freq used, 10=least freq used)
e.g. pitch (Low, Medium, High)	
f)	
g)	
h)	
i)	
j)	

16. What methods are you aware of for classifying sounds?

Formal method of classifying sounds	**Rank** (1=most freq used, 10=least freq used)
e.g. Gaver: Vibrating (Impacts/Scraping/Others), Aerodynamic (Explosions/Continuous), Liquid (Dripping/Splashing).	
a)	
b)	
c)	
d)	

e)	
Informal method of classifying sounds	**Rank** (1=most freq used, 10=least freq used)
e.g. Speech/Music/Other Known/Other Unknown	
f)	
g)	
h)	
i)	
j)	

17. What graphical methods are you aware of for representing sounds?

Formal method of graphically representing sounds?	**Rank** (1=most freq used, 10=least freq used)
e.g. Spectrogram (time v. frequency)	
a)	
b)	
c)	
d)	
e)	
Informal method of graphically representing sounds	**Rank** (1=most freq used, 10=least freq used)
e.g. colour (blue = low frequency, red = high frequency)	
f)	
g)	
h)	
i)	
j)	

Audio questionnaire i.mcgregor@napier.ac.uk

Final Questions

18. If there were a technique to represent an auditory environment, what parameters (characteristics) would you want or expect to be included?

19. What auditory environments would you be interested in capturing?

20. Would you like to elaborate further upon any of your answers, or make any points that might be interesting to follow-up?

Appendix B: Rotated car schematic

Front of Car

0 0	0 1	0 2	0 3	0 4	0 5	0 6	0 7	0 8	0 9	0 10	0 11	0 12	0 13
1 0	1 1	1 2	1 3	1 4	1 5	1 6	1 7	1 8	1 9	1 10	1 11	1 12	1 13
2 0	2 1	2 2	2 3	2 4	2 5	2 6	2 7	2 8	2 9	2 10	2 11	2 12	2 13
3 0	3 1	3 2	3 3	3 4	3 5	3 6	3 7	3 8	3 9	3 10	3 11	3 12	3 13
4 0	4 1	4 2	4 3	4 4	4 5	4 6	4 7	4 8	4 9	4 10	4 11	4 12	4 13
5 0	5 1	5 2	5 3	5 4	5 5	5 6	5 7	5 8	5 9	5 10	5 11	5 12	5 13
6 0	6 1	6 2	6 3	6 4	6 5	6 6	6 7	6 8	6 9	6 10	6 11	6 12	6 13
7 0	7 1	7 2	7 3	7 4	7 5	7 6	7 7	7 8	7 9	7 10	7 11	7 12	7 13
8 0	8 1	8 2	8 3	8 4	8 5	8 6	8 7	8 8	8 9	8 10	8 11	8 12	8 13
9 0	9 1	9 2	9 3	9 4	9 5	9 6	9 7	9 8	9 9	9 10	9 11	9 12	9 13
10 0	10 1	10 2	10 3	10 4	10 5	10 6	10 7	10 8	10 9	10 10	10 11	10 12	10 13
11 0	11 1	11 2	11 3	11 4	11 5	11 6	11 7	11 8	11 9	11 10	11 11	11 12	11 13
12 0	12 1	12 2	12 3	12 4	12 5	12 6	12 7	12 8	12 9	12 10	12 11	12 12	12 13
13 0	13 1	13 2	13 3	13 4	13 5	13 6	13 7	13 8	13 9	13 10	13 11	13 12	13 13
14 0	14 1	14 2	14 3	14 4	14 5	14 6	14 7	14 8	14 9	14 10	14 11	14 12	14 13
15 0	15 1	15 2	15 3	15 4	15 5	15 6	15 7	15 8	15 9	15 10	15 11	15 12	15 13
16 0	16 1	16 2	16 3	16 4	16 5	16 6	16 7	16 8	16 9	16 10	16 11	16 12	16 13
17 0	17 1	17 2	17 3	17 4	17 5	17 6	17 7	17 8	17 9	17 10	17 11	17 12	17 13
18 0	18 1	18 2	18 3	18 4	18 5	18 6	18 7	18 8	18 9	18 10	18 11	18 12	18 13
19 0	19 1	19 2	19 3	19 4	19 5	19 6	19 7	19 8	19 9	19 10	19 11	19 12	19 13
20 0	20 1	20 2	20 3	20 4	20 5	20 6	20 7	20 8	20 9	20 10	20 11	20 12	20 13
21 0	21 1	21 2	21 3	21 4	21 5	21 6	21 7	21 8	21 9	21 10	21 11	21 12	21 13

Rear of Car

Appendix C: List of sound events for classification

Code	Event	Source	Unaware	x y	Type	Material	Interaction	Temporal	Spectral	Dynamics	Content	Aesthetics	Clarity	Emotions
AA	Engine Idle	Engine												
AB	Engine Accelerate	Engine												
AC	Engine Cruise	Engine												
AD	Engine Decelerate	Engine												
AE	Handbrake Released	Handbrake												
AF	Handbrake Engaged	Handbrake												
AG	Gear Change	Gear Stick												
AH	Door Lock Release Rattle	Door Lock Release												
AI	Rattle	Sunroof												
AJ	Rattle	Parcel Shelf												
AK	Rattle	Dashboard												
AL	Seat Creak	Driver's Seat												
AM	Indicating	Indicator												
AN	Indicator Off	Indicator Switch												
AO	Male 1 Speech	Radio												
AP	Male 2 Speech	Radio												
AQ	Female 1 Speech	Radio												
AR	Female 2 Speech	Radio												
AS	Female 3 Speech	Radio												
AT	Male Chant	Radio												
AU	Group Chant	Radio												
AV	Music 1	Radio												
AW	Music 2	Radio												
AX	Group Laughter	Radio												
AY	Mobile Phone Interference	Radio												
AZ	Bump	Front Left Wheel												
BA	Bump	Front Right Wheel												
BB	Bumps	All wheels												
BC	Clothes Rustle	Driver												
BD	Windscreen Wiper	Car 1												
BE	Wiper Blade	Windscreen Wiper												
BF	Passing Car	Car 2												
BG	Car Pass	Car 3												
BH	Car Pass	Car 4												
BI	Car Pass	Car 5												
BJ	Car Pass	Car 6												
BK	Car Pass	Car 7												
BL	Car Pass	Car 8												
BM	Passing Bus	Bus 1												
BN	Bus Pass	Bus 2												
BO	Van Pass	Van 1												
BP	Engine Rev	Car 9												
BQ	Engine Rev	Van 2												
BR	Engine Idle	Car 10												
BS	Engine Idle	Taxi 1												
BT	Air Brakes	Truck												
BU	Brake Squeal	Car 11												
BV	Brake Squeal	Car 12												
BW	Brake Squeal	Car 13												
BX	Brake Squeal	Bus 3												
BY	Brake Squeal	Bus 4												
BZ	Brake Squeal	Taxi 2												
CA	Brake Squeal	Taxi 3												
CB	Horn	Car 14												
CC	Indicator	Car 8												
CD	Indicator	Car 15												
CE	Indicator	Car 16												
CF	Siren	Ambulance												
CG	Scaffolding Strike	Scaffolding												
CH	Scaffolding Strike	Scaffolding												
CI	Bark	Dog												
CJ	Child Scream	Child												
CK	Child Scream	Child												
CL	Female Speech	Woman												
CM	Male Speech	Man												

Appendix D: Designer's description of Auditory Display

Auditory display

1. Braking distance:
    a. Triggered when the car is getting too close to a vehicle in front.
    b. The increasing pitch and intensity represent the diminishing distance between the two vehicles.

2. Dead angle:
    a. Triggered when a vehicle overtaking the 205 (on the right) is in the driver's dead angle.
    b. The intensity of the sound increases as the car approaches the dead angle.

3. Message:
    a. This sound represents a sequence of action by the driver receiving a new email message:
        i. The system notifies the driver of a new message
        ii. The driver presses a key to hear the subject
        iii. The system reads out the subject
        iv. The driver presses a key to hear the message
        v. The system reads out the message body
        vi. The driver interrupts the system by pressing the delete key
        vii. The system confirms the message deletion
        viii. The system confirms that there are no other messages in the user's inbox

Appendix E: Guidelines provided to participants

Thank you for taking part in this experiment, which will take 45 - 60 minutes to complete. Please feel free to leave at any time.

The experiment is part of an ongoing study into soundscape mapping, where individual's auditory environments are mapped in order to understand what people are listening to.

The data will be anonymised and used for publication in Iain McGregor's PhD thesis, which is due for submission in August 2009.

There are four parts to the experiment.

1. A recording of a series of alert sounds for a car auditory interface for you to listen to. There are no questions to answer at this stage.

2. A five minute recording of a small car driving through Edinburgh city centre, with a radio playing. After you have finished listening you will be asked questions about what you heard.

3. A five minute recording of the audio interface only. After you have finished listening you will be asked questions about what you heard.

4. A five minute recording of both the audio interface and the small car driving through Edinburgh city centre. After you have finished listening you will be asked questions about what you heard.

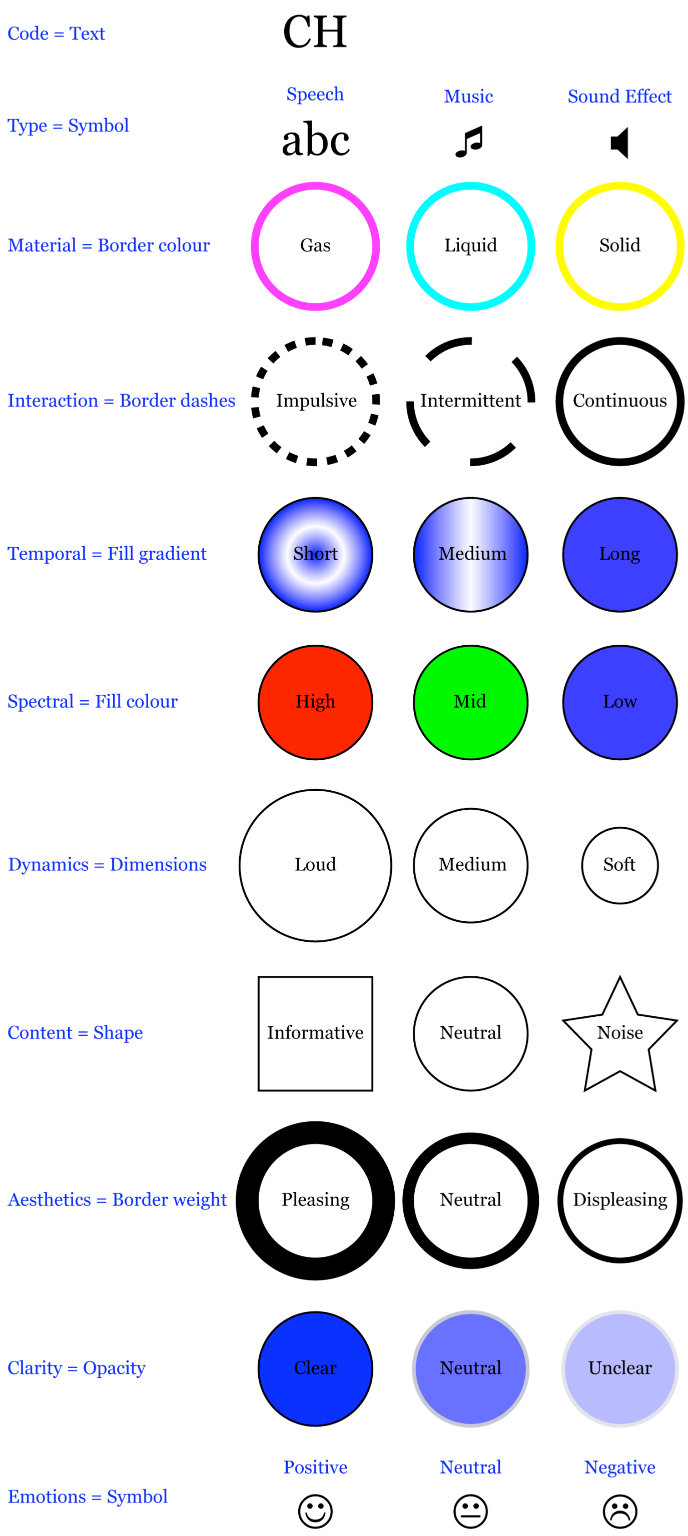
Thank you very much for your help, it is really appreciated.

Iain McGregor
School of Computing, Edinburgh Napier University

Appendix F: Sound event classification reference

Type	Category
Speech	Spoken language
Music	Performed composition
Sound effect	Audible events and actions
Material	Matter
Gas	Airborne
Liquid	Fluids
Solid	Objects
Interaction	Action
Impulsive	Explosion/drip/impact
Intermittent	Whooshing/splashing/scraping
Continuous	Blowing/flowing/rolling
Temporal	Duration
Short	Brief
Medium	Neither long nor short
Long	Extended
Spectral	Pitch
High	High pitch/frequency Treble
Mid	Medium pitch/frequency Alto
Low	Low pitch/frequency Bass
Dynamics	Volume/Loudness
Loud	High volume *Forte*
Medium	Medium volume/level
Soft	Quiet *Piano*
Content	Relevance
Informative	Relevant information
Neutral	Neither relevant nor irrelevant
Noise	Irrelevant/unwanted
Aesthetics	Beauty
Pleasing	Beautiful
Neutral	Mediocre
Displeasing	Ugly
Clarity	Quality
Clear	Easy to hear and comprehend
Neutral	Neither easy nor difficult to hear
Unclear	Difficult to hear and comprehend
Emotions	Feelings
Positive	Acceptance, Anticipation, Joy, Surprise
Neutral	No emotional content
Negative	Anger, Disgust, Fear, Sadness

Appendix G: Visualisation key



Appendix H: Sound events' attributes form

Sound event attributes

	Start		Finish												
Short description of sound	**x**	**y**	**x**	**y**	**Awareness**	**Type**	**Material**	**Interaction**	**Temporal**	**Spectral**	**Dynamics**	**Content**	**Aesthetics**	**Clarity**	**Emotions**

A

Appendix I: Spatial attributes' grid

0 16	1 16	2 16	3 16	4 16	5 16	6 16	7 16	8 16	9 16	10 16	11 16	12 16	13 16	14 16	15 16	16 16	17 16	18 16	19 16	20 16	21 16
0 15	1 15	2 15	3 15	4 15	5 15	6 15	7 15	8 15	9 15	10 15	11 15	12 15	13 15	14 15	15 15	16 15	17 15	18 15	19 15	20 15	21 15
0 14	1 14	2 14	3 14	4 14	5 14	6 14	7 14	8 14	9 14	10 14	11 14	12 14	13 14	14 14	15 14	16 14	17 14	18 14	19 14	20 14	21 14
0 13	1 13	2 13	3 13	4 13	5 13	6 13	7 13	8 13	9 13	10 13	11 13	12 13	13 13	14 13	15 13	16 13	17 13	18 13	19 13	20 13	21 13
0 12	1 12	2 12	3 12	4 12	5 12	6 12	7 12	8 12	9 12	10 12	11 12	12 12	13 12	14 12	15 12	16 12	17 12	18 12	19 12	20 12	21 12
0 11	1 11	2 11	3 11	4 11	5 11	6 11	7 11	8 11	9 11	10 11	11 11	12 11	13 11	14 11	15 11	16 11	17 11	18 11	19 11	20 11	21 11
0 10	1 10	2 10	3 10	4 10	5 10	6 10	7 10	8 10	9 10	10 10	11 10	12 10	13 10	14 10	15 10	16 10	17 10	18 10	19 10	20 10	21 10
0 9	1 9	2 9	3 9	4 9	5 9	6 9	7 9	8 9	9 9	10 9	11 9	12 9	13 9	14 9	15 9	16 9	17 9	18 9	19 9	20 9	21 9
0 8	1 8	2 8	3 8	4 8	5 8	6 8	7 8	8 8	9 8	10 8	11 8	12 8	13 8	14 8	15 8	16 8	17 8	18 8	19 8	20 8	21 8
0 7	1 7	2 7	3 7	4 7	5 7	6 7	7 7	8 7	9 7	10 7	11 7	12 7	13 7	14 7	15 7	16 7	17 7	18 7	19 7	20 7	21 7
0 6	1 6	2 6	3 6	4 6	5 6	6 6	7 6	8 6	9 6	10 6	11 6	12 6	13 6	14 6	15 6	16 6	17 6	18 6	19 6	20 6	21 6
0 5	1 5	2 5	3 5	4 5	5 5	6 5	7 5	8 5	9 5	10 5	11 5	12 5	13 5	14 5	15 5	16 5	17 5	18 5	19 5	20 5	21 5
0 4	1 4	2 4	3 4	4 4	5 4	6 4	7 4	8 4	9 4	10 4	11 4	12 4	13 4	14 4	15 4	16 4	17 4	18 4	19 4	20 4	21 4
0 3	1 3	2 3	3 3	4 3	5 3	6 3	7 3	8 3	9 3	10 3	11 3	12 3	13 3	14 3	15 3	16 3	17 3	18 3	19 3	20 3	21 3
0 2	1 2	2 2	3 2	4 2	5 2	6 2	7 2	8 2	9 2	10 2	11 2	12 2	13 2	14 2	15 2	16 2	17 2	18 2	19 2	20 2	21 2
0 1	1 1	2 1	3 1	4 1	5 1	6 1	7 1	8 1	9 1	10 1	11 1	12 1	13 1	14 1	15 1	16 1	17 1	18 1	19 1	20 1	21 1
0 0	1 0	2 0	3 0	4 0	5 0	6 0	7 0	8 0	9 0	10 0	11 0	12 0	13 0	14 0	15 0	16 0	17 0	18 0	19 0	20 0	21 0

Appendix J: Sound event classification reference

Awareness	Cognizance
Aware	Conscious of sound event
Unaware	Not conscious of sound event
Spatial cues	**Position**
x axis	panning/left-right
y axis	depth/reverb/close-far
Type	**Category**
Speech	Spoken language
Music	Performed composition
Sound effect	Audible events and actions
Material	**Matter**
Gas	Airborne
Liquid	Fluids
Solid	Objects
Interaction	**Action**
Impulsive	Explosion/drip/impact
Intermittent	Whooshing/splashing/scraping
Continuous	Blowing/flowing/rolling
Temporal	**Duration**
Short	Brief
Medium	Neither long nor short
Long	Extended
Spectral	**Pitch**
High	High pitch/frequency *Treble*
Mid	Medium pitch/frequency *Alto*
Low	Low pitch/frequency *Bass*
Dynamics	**Volume/Loudness**
Loud	High volume *Forte*
Medium	Medium volume *Mezzo*
Soft	Quiet *Piano*
Content	**Relevance**
Informative	Relevant information
Neutral	Neither relevant nor irrelevant
Uninformative	Irrelevant information
Aesthetics	**Beauty**
Pleasing	Beautiful
Neutral	Neither beautiful nor ugly
Displeasing	Ugly
Clarity	**Quality**
Clear	Easy to hear and comprehend
Neutral	Neither easy nor difficult to hear
Unclear	Difficult to hear and comprehend
Emotions	**Feelings**
Positive	Joy, Love, Surprise
Neutral	No emotional content
Negative	Anger, Fear, Sadness

Appendix K: Designers' informed consent form

INFORMED CONSENT FORM

Soundscape mapping tool

Edinburgh Napier School of Computing requires that all persons who participate in research studies give their written consent to do so. Please read the following and sign it if you agree with what it says.

I freely and voluntarily consent to be a participant in the research project on the topic of soundscape mapping to be conducted by Iain McGregor, who is a PhD student in the Edinburgh Napier School of Computing. The broad goal of this research study is to explore listeners' experiences of the soundscape so that they can be compared with designers' expectations. Specifically, I have been asked to provide an existing sound design or designs and classify all of the sound events within it or them, which should take no longer than one hour to complete. The sound design or designs will be listened to by 10 participants and a map will be created from their responses as well as my own. These maps will be returned to me with a series questions, which should take no longer than forty five minutes to complete.

I have been told that my responses will be **anonymised**. I also understand that my participation in this study is completely voluntary, and I may withdraw from it at any time without negative consequences. ***In addition, should I not wish to answer any particular question or questions, I am free to decline***. My name will not be linked with the research materials, and I will not be identified or identifiable in any report subsequently produced by the researcher.

I have been given the opportunity to ask questions regarding the procedure, and my questions have been answered to my satisfaction.

I have read and understand the above and consent to participate in this study. My signature is not a waiver of any legal rights. Furthermore, I understand that I will be able to keep a copy of the informed consent form for my records.

________________________ ____________________________________

Participant's Signature Date

I have explained and defined in detail the research procedure in which the respondent has consented to participate. Furthermore, I will retain one copy of the informed consent form for my records.

________________________ __________________

Researcher's Signature Date

Appendix L: Listeners' informed consent form

INFORMED CONSENT FORM

Soundscape mapping tool

Edinburgh Napier School of Computing requires that all persons who participate in research studies give their written consent to do so. Please read the following and sign it if you agree with what it says.

I freely and voluntarily consent to be a participant in the research project on the topic of soundscape mapping to be conducted by Iain McGregor, who is a PhD student in the Edinburgh Napier School of Computing. The broad goal of this research study is to explore listeners' experiences of the soundscape so that they can be compared with designers' expectations. Specifically, I have been asked to listen to an audio recording and answer questions about what I heard, which should take no longer than 45 minutes to complete.

I have been told that my responses will be **anonymised**. I also understand that if at any time during the session I feel unable or unwilling to continue, I am free to leave. That is, my participation in this study is completely voluntary, and I may withdraw from it at any time without negative consequences. ***In addition, should I not wish to answer any particular question or questions, I am free to decline***. My name will not be linked with the research materials, and I will not be identified or identifiable in any report subsequently produced by the researcher.

I have been given the opportunity to ask questions regarding the procedure, and my questions have been answered to my satisfaction.

I have read and understand the above and consent to participate in this study. My signature is not a waiver of any legal rights. Furthermore, I understand that I will be able to keep a copy of the informed consent form for my records.

________________________ ____________________________________

Participant's Signature Date

I have explained and defined in detail the research procedure in which the respondent has consented to participate. Furthermore, I will retain one copy of the informed consent form for my records.

________________________ __________________

Researcher's Signature Date

Appendix M: Soundscape mapping tool designers' questionnaire

Introduction

Thank you for agreeing to take part in this survey.

The 29 questions are split up into three parts: classification, visualisation and general comments. It should take approximately 45 minutes to complete the questionnaire.

Please first refer to the two soundscape maps: the first represents the designer's responses (page 1) and the second the combined listeners' responses (page 2). The visualisation key is on page 3, and the explanation of the attributes can be found on page 4.

Classification

In order to compare sound designs with listeners' experiences, please rate the importance of the audio attributes used in the classification, using the drop down boxes.

Very Important

Important

Moderately Important

Of Little Importance

Unimportant

	Attribute	**Response**
1	**Awareness**: Aware/Unaware	
2	**Spatial cues**: x, y coordinates	
3	**Type**: Speech/Music/Sound effect	
4	**Material**: Gas/Liquid/Solid	
5	**Interaction**: Impulsive/Intermittent/Continuous	
6	**Temporal**: Short/Medium/Long	
7	**Spectral**: High/Mid/Low	
8	**Dynamics**: Loud/Medium/Soft	
9	**Content**: Informative/Neutral/Uninformative	
10	**Aesthetics**: Pleasing/Neutral/Displeasing	
11	**Clarity**: Clear/Neutral/Unclear	
12	**Emotions**: Positive/Neutral/Negative	

Visualisation

In order to compare sound designs with listeners' experiences, please choose the most appropriate way to display the audio attributes used in the classification.
Please choose only one form of visualisation per attribute, i.e. only check one box per question.

No.	Attribute	Inclusion of object	Position on grid	Symbols (abc)	Border colour	Border dashes	Fill gradient	Fill colour	Dimension	Shape	Border weight	Opacity	Emo-ticon
13	**Awareness**: Aware/Unaware	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
14	**Spatial cues**: x, y coordinates	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
15	**Type**: Speech/Music/ Sound effect	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
16	**Material**: Gas/Liquid/Solid	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
17	**Interaction**: Impulsive/ Intermittent/Continuous	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
18	**Temporal**: Short/Medium/ Long	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
19	**Spectral**: High/Mid/Low	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
20	**Dynamics**: Loud/Medium/Soft	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
21	**Content**: Informative/ Neutral/ Uninformative	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
22	**Aesthetics**: Pleasing/ Neutral/Displeasing	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
23	**Clarity**: Clear/Neutral/Unclear	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
24	**Emotions**: Positive/Neutral/Negative	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

General comments

When comparing the sound design with listeners' experiences, please state your level of agreement with the statement, and then responses to the subsequent questions.

	Statement	Response
25	The soundscape mapping tool allowed me to compare a sound design with the experience of listeners.	
	Questions	
26	How do you currently evaluate your sound designs?	
27	How could you use the tool?	
28	What would you like to change about the tool?	
29	Do you have any comments that you would like to add?	

Thank you for completing this questionnaire.

Appendix N: Soundscape mapping tool pilot study questionnaire

Thank you for agreeing to take part in this pilot study for the soundscape mapping tool survey.

This involves watching a short film (02:40) referring to the associated maps, answering a 30 question questionnaire, and then answering three questions which are at the bottom of this page.

The soundscape mapping tool is way of visually comparing sound designs with listeners' experiences. In order to evaluate their sound design, a designer first lists all of the sounds contained within the design using short natural descriptions. These are then classified according to a set of attributes and subsequently visualised. This process is repeated with listeners and then the two maps can be compared so that similarities and differences can be highlighted.

The example that has been included here is for a short film called xxxxxxxxxx, please watch this QuickTime file and then refer to the soundscape maps document.

Pages one and two show the designer's and listeners' maps for you to compare the two different listening perspectives. Page three is a key so that you can decipher the shapes used in the maps. Page four describes the attributes used to classify the sound events. Pages five and six list all of the sound events with the designer's and listeners' choices of attributes.

After watching the QuickTime file and looking at the maps please could you make a note of the current time, and then complete the soundscape mapping tool questionnaire. Once you have finished this please again take note of the current time.

Please could you now answer the following three questions.

1. Did you find any questions ambiguous or difficult to answer?

2. If you found any of the questions ambiguous or difficult to answer, please specify the questions and explain what made them ambiguous or difficult to answer?

3. How long did the soundscape mapping tool questionnaire take to complete?

Please could you return both this file and the completed questionnaire to Iain McGregor i.mcgregor@napier.ac.uk. Thank you very much for taking part.

Appendix O: Visualisation key