Strategies for Environmental Sound Measurement, Modelling, and Evaluation

Francis Kit Murfin Stevens

PhD

University of York

Electronic Engineering

August 2018

Abstract

This thesis is a portfolio of research into three aspects of environmental sound: its measurement, modelling, and evaluation. In each of these areas, this body of work aims to make use of soundscape methodologies in order to develop an understanding of different aspects of our relationship with our sonic environments. This approach is representative of the nature of soundscape research, which makes use of elements of many other research areas, including acoustics, psychology, sociology, and musicology.

The majority of prior acoustic measurement research has considered indoor recording, often of music, and measurement of acoustic parameters of indoor spaces such as concert halls and other performance spaces. One strand of this research has investigated how best to apply such techniques to the recording of environmental sound, and to the measurement of the acoustic impulse responses of outdoor spaces.

Similarly, the majority of prior work in the field of acoustic modelling has also focussed mainly on indoor spaces. Presented here is the Waveguide Web, a novel method for the acoustic modelling of sparsely reflecting outdoor spaces.

In the field of sound evaluation of sound, recent years have seen the development of soundscape techniques for the subjective rating of environmental sound, allowing for a better understanding of our relationship with our sonic surroundings. Research presented in this thesis has focussed on how best to improve these approaches in a suitably robust and intuitive manner, including the integration of visual stimuli in order to investigate the multi-modal perception of our surroundings.

The aim of this thesis in making contributions to these three fields of environmental sound research is, in part, to highlight the importance of developing a comprehensive understanding of our sonic environments. Such an understanding could ultimately lead to the alleviation of noise problems, encourage greater engagement with environmental sound in the wider population, and allow for the design of more positive, restorative, soundscapes.

Contents

List of Figures

List of Tables

Music is the cup which holds the wine of silence.

Sound is that cup, but empty.

Noise is that cup, but broken.

- Robert Fripp, 1980

Acknowledgements

It is unlikely that the thesis would have taken the form that it has without the contributions of several people.

Firstly, for his constant enthusiasm, guidance, and (of course) creativity, I would like to express my sincere gratitude to Prof. Damian Murphy. Damian has managed to stay positive throughout (including when I was being rather negative), and without him none of this would have happened.

Thanks must also go to Prof. Stephen L Smith and Dr Gavin Kearney for their respective contributions as my second supervisor and thesis advisor. Prof. Vesa Välimäki and Prof. Lauri Savioja also deserve many thanks for their advice in the design of the Waveguide Web. This work was supported by an EPSRC doctoral studentship.

I would also like to thank many of the people of the AudioLab for their friendship (and the interminable crosswords). Many thanks to Joe in particular for going through the mill with me. We will get that globe bar one day...

My family also have my love and gratitude for their support, in particular I would like to thank my parents for their constant encouragement, and their continuing attempts to understand what I've been doing for the last four years.

Working on this research over the last few years would definitely have been rather more difficult without my dearest Yasmin, to whom I offer all my love. Oh, and thanks for pouring all that beer and making all that tea.

Other people who must be acknowledged by name for their moral support include Dr Schneider, Jem, John, Caz, and Eden. Thank you!

Finally thanks to Kerry, whose simple love of gravy bones, naps, and walkies has to be the greatest inspiration of all.

Declaration of Authorship

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as references. I also declare that parts of this research have been presented in previous conference and journal publications, which are listed as follows:

- F. Stevens, D. T. Murphy, and S. L. Smith, "Emotion and soundscape preference rating: Using semantic differential pairs and the self-assessment manikin," in *Sound and Music Computing conference, Hamburg, 2016*, Hamburg, Germany, 2016
- F. Stevens, D. T. Murphy, and S. L. Smith, "The Self-Assessment Manikin and heart rate: Responses to auralised soundscapes," in *Interactive Audio Systems Symposium 2016*, York, UK, 2016
- F. Stevens, D. T. Murphy, and S. L. Smith, "Ecological validity of stereo UHJ soundscape reproduction," in *Proceedings of the 142nd Audio Engineering Society (AES) Convention*, Berlin, Germany, 2017
- A. Southern, F. Stevens, and D. T. Murphy, "Sounding out smart cities: Auralization and soundscape monitoring for environmental sound design," *The Journal of the Acoustical Society of America*, vol. 141, no. 5, pp. 3880–3880, 2017
- F. Stevens, D. T. Murphy, and S. L. Smith, "Soundscape auralisation and perception for environmental sound modelling," in *Sound + Environment 2017*, Hull, UK, 2017
- F. Stevens, D. T. Murphy, L. Savioja, et al., "Modeling sparsely reflecting outdoor acoustic scenes using the waveguide web," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 8, pp. 1566–1578, Aug. 2017, issn: 2329-9290. doi: 10.1109/TASLP.2017.2699424
- F. Stevens, D. T. Murphy, and S. L. Smith, "Soundscape categorisation and the Self-Assessment Manikin," in *Proceedings of the 20th International Conference on Digital Audio Effects (DAFx-17)*, Edinburgh, UK, 2017
- F. Stevens, D. T. Murphy, and S. L. Smith, "Soundscape auralisation and visualisation: A cross-modal approach to soundscape evaluation," in *Proceedings of the 21st International Conference on Digital Audio Effects (DAFx-18)*, Aveiro, Portugal, 2018

Chapter 1

Introduction

Sound is inescapable. It envelops us at all times in all places, and is able to bring us joy, pain, unhappiness, and discomfort. It can deeply move us, providing succour, catharsis, and comfort. Our sense of hearing is one of the most powerful ways in which we understand the environments we inhabit, and affects the development of memories and emotional connections to places and people. Indeed, whilst one can picture experiencing the absence of light (for example at the bottom of mineshaft with all the lights turned off), the absence of sound, silence, is more elusive. Even in the quietest, most peaceful, places, there remains the experience of auditory sensation. For example in an anechoic chamber, an environment designed completely to absorb sound [1], after a certain time one starts to hear 'new' sounds from within one's own head (thought of by the composer John Cage as the sound of blood circulation and activity of the central nervous system [2], although tinnitus is a more likely source [3]). In point of fact, there is not really such a thing as silence, at least not anywhere a human is present. In a way, this can be seen as an inversion of the old thought experiment about a tree falling in a forest with no-one around to hear it[1], where here the question might be *'if there is silence and someone hears it, is it silent?'*.

The experience of sound is unrelenting even to the point where sleep offers no reprieve, and it can also have significant long term effects on health and wellbeing [5], [6]. As such,

[1]Of course this question isn't really concerned with the acoustics of a falling tree, rather the nature of human knowledge and perception, and the dissimilarity between sensation and reality [4]. The gulf between objective truth and subjective experience exemplified by this thought experiment remains pertinent to this thesis, especially when considering the evaluation of sound.

it seems somewhat bizarre that the our relationships with our sonic environments are not yet fully understood.

An example where this lack of understanding has had (and continues to have) severe ramifications is noise, in this context the presence of meaningless, unwanted, sound resulting in a stressful and unpleasant experience of a place. The problem of noise is by no means new (the movement of heavy carts through the nighttime street of ancient Rome is an early example [7]), but in recent decades has become a bigger problem due to the proliferation of cars, aeroplanes, and the increases in global population [8]. Indeed the problem of noise is so severe that it is acknowledged as a form of pollution [9], and is even referred to as a modern plague [10], one which can do significant damage in terms of both physical and mental health [11], [12].

Despite this, however, existing legislature is insufficient and shallow in its approach, often making use of summary noise level measurements alone with the goal being simply to reduce those levels. Such an approach ignores the subjective experience of sound (i.e. what a place actually sounds like), and therefore is insufficient. Accordingly, there is a need for the combination of an objective approach with an understanding of the subjective experience of sound. This thesis posits that an approach based on soundscape research, its theories and methodologies, offers an opportunity for achieving this aim.

'Soundscape', in the context of this thesis, is a term used to refer to all of the sounds that are heard in a particular location, considered as a whole [13]. Whilst explicitly referring to the sound sources that comprise a particular location's acoustic scene, the term implicitly contains an understanding of the dynamic, personal, experience of sound, and that this experience has subjective and emotional components dependent on an individual's history, emotional state, and identity [14], [15].

'Soundscape methodologies' is a collective term for the sound categorisation, recording, playback, and evaluation tools that can be used to develop this listener-centric understanding of sound. Soundscape research is a field that provides an opportunity to develop an holistic approach to the understanding of environmental sound, bringing together strategies from multiple disciplines. This thesis is representative of the interdisciplinary nature of soundscape research, in that it makes use of state-of-the-art sound capture and measurement techniques, soundscape methodologies, and novel methods of acoustic modelling to present a portfolio of research making novel contributions to the understanding of environmental sound, and our relationship with it.

This measurement work includes the capture of Ambisonic soundscape data (alongside the shooting of panoramic video footage), and the measurement of acoustic impulse responses at Creswell Crags, a limestone gorge housing a series of caves. The soundscape and visual data collected is then examined in a series of listening tests. The ultimate purpose of these tests is to investigate how the visual features that accompany a soundscape can affect the experience of it. Most existing soundscape research has focussed on aural stimuli alone, an approach which can be informative, although without an examination of the visual features of an environment it seems unlikely that such research can fully describe the experience of a space. In order to develop this understanding, incremental steps must be taken (building on existing research) in order to validate such an approach.

As such, these tests progressively present the collected data in a variety of formats, building towards a full audiovisual test. These listening tests make use of both established and novel (in the context of soundscape research) subjective evaluation techniques in order to establish how the subjective experience of soundscapes can best be measured. This includes the evaluation of emotional state, description of the soundscape contents, and measurement of how the presence of particular visual features can alter these responses. The ultimate goal is to find the aural and visual features that result in a peaceful, restorative location: spaces which give people the chance to 'hear themselves think', and to self-reflect. The use of soundscape methodologies to find these features and spaces is intended as an example of how to take a positive approach to environmental sound, that being an approach which does not focus solely on the negative aspects of noise, but one that also looks to find soundscape that can have a positive impact on health and well-being.

The environmental sound modelling work presented here is in the form of the Waveguide Web, a novel method for the simulation of sparsely reflecting outdoor acoustic scenes (such as a forest, or, indeed, the limestone gorge at Creswell Crags). The work makes a significant contribution to the field of acoustic modelling, which has historically focussed primarily on indoor spaces.

As this summary implies, this thesis contains multiple branches of research, all related by central concepts. As introduced above this is a reflection of the multitude of factors involved in the study of environmental sound, and the multi-faceted approach that is absolutely required in order to approach any sort of complete understanding of the relationship between a soundscape and a listener. This aims and motivation of this research will now be formalised in a hypothesis.

1.1 Statement of Hypothesis

The portfolio of work presented in this thesis covers a variety of topics and disciplines, all within the purview of environmental sound. As such the following hypothesis can be seen as a guiding principle, summarising the aims and motivation for the thesis work:

> *The measurement, modelling, and evaluation of environmental sound are effective soundscape methodologies for the identification and assessment of positive sonic environments.*

The key aspects of this hypothesis will now each be explained as they relate to this thesis:

- **Soundscape methodologies.** These include the use of sound recording, auralisation, and subjective evaluation techniques in order to develop a listener-centric understanding of sound.
- **Environmental sound.** In this context environmental sound refers to the soundscapes of outdoor locations. This includes natural, human, and animal sounds, as well as the behaviour and propagation of sound in outdoor spaces.
- **Positive sonic environments.** These are environments that are considered by listeners to be pleasant and relaxing, that may also afford an opportunity for self-reflection. Environments such as these are known to be beneficial in terms of health and well-being [16], and do not confer the negative health effects associated with noisy environments.
- **Identification and assessment.** In terms of this thesis, the 'identification' of positive sonic environments will be where the subjective assessment results indicate where participants experience soundscapes they find to be pleasant and relaxing. Where these environments are identified (as well as where negative, unpleasant,

soundscapes are identified) this will establish the suitability of the tools used for the evaluation of environmental sound.

As this hypothesis implies, soundscape research is a field that brings together techniques from many different disciplines, and makes use of subjective and objective test methodologies in order to develop a complete understanding of our relationship with our aural environments. This thesis presents an example of how this multi-faceted approach can be beneficial, where different perspectives on the same issue (i.e. environmental sound), using a variety of strategies, can inform one another and provide context beyond that afforded by taking a single approach only. The aims and objectives of this body of research will now be specified.

1.2 Aims and Objectives

This section will now state the aims and objectives of this body of research. The aims are enumerated below, and each is supported by the relevant objectives.

1. Investigate the phenomenon of cross-modal perception in the context of soundscape evaluation and analysis.

    - Make use of spatial audio and spherical video recording equipment to gather audiovisual soundscape stimuli for use in listening tests.
    - Run a series of audio-only tests first (in line with extant soundscape evaluation research) in order to establish and develop suitable soundscape evaluation tools, and to establish the ecological validity of a variety of formats (i.e. the extent to which the soundscape playback is representative of reality).
    - Introduce the visual content to the listening tests in order to investigate cross-modal perception.
    - Analyse the test results to identify the aural and visual features most that most impact the experience of a soundscape. This analysis should include a focus on identifying the features most likely to indicate positive, restful environments, and where the presence of visual features may impact the evaluation of aural ones.

2. Make use of impulse response measurement techniques in an outdoor environment in order to evaluate the success of the application of those techniques to that sort of environment, and to develop some understanding of the acoustic properties of that environment.
    - Record a series of spatial acoustic impulse responses at Creswell Crags, a limestone gorge housing a network of caves.
    - Conduct a analysis of these impulse responses in order to determine their acoustic and spatial properties.
3. Develop a novel method for modelling sparsely reflecting outdoor acoustic spaces. This should in part be informed by real life measurement work and experience, and should also draw on extant modelling methodologies designed for the simulation of other environments.
    - Research and reproduce existing simulation methods that are, or could be, used for the modelling of outdoor acoustic spaces.
    - Develop a novel method for the modelling of these spaces informed by extant methods and the real-world impulse responses.
    - Evaluate the performance of the novel modelling method by conducting a series of case studies comparing newly simulated impulse responses with those generated using other modelling techniques and real-world recordings.

These aims and objective clearly state the intended achievements of this thesis work, and the steps to be completed in order to complete them.

1.3 Thesis Scope

Given the aforementioned portfolio nature of this thesis, it is worth explicating the scope of the work presented, and the relationships between the research areas contributed to in this work, as well as other, related, ones. Figure 1.1 is included for this purpose.

It shows a conceptual framework of environmental sound research, which can be considered as broadly synonymous with soundscape research in the context of this thesis. It shows the relationship between the following elements:

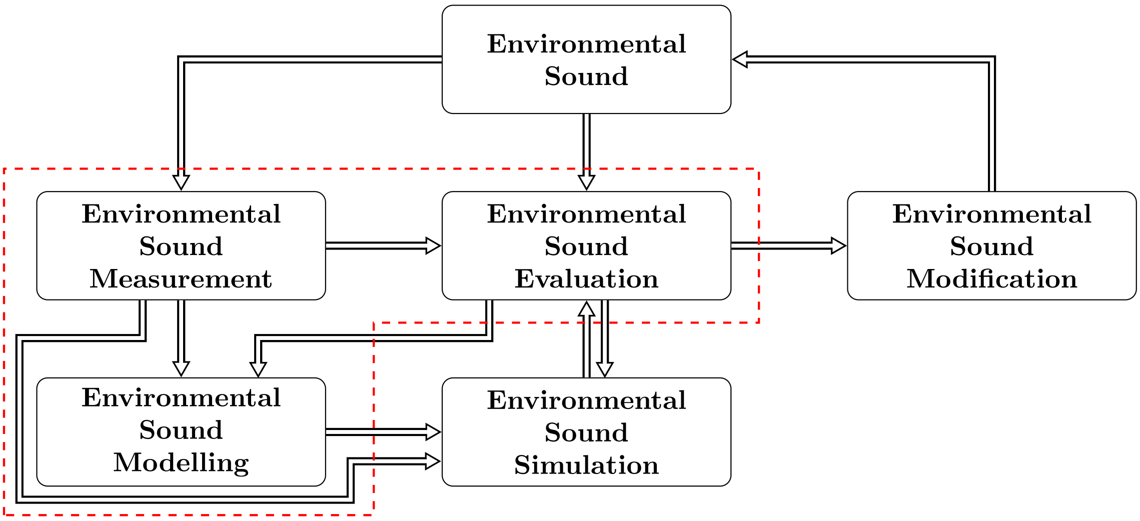


FIGURE 1.1: A conceptual framework of environmental sound (i.e. soundscape) research, indicating the relationship between its elements, and the scope of the work presented in this thesis (dashed box).

- Environmental sound: This refers to the soundscape of an environment under consideration (e.g. an urban street scene). This will typically be an environment that is in need of treatment to improve the experience of its soundscape. The relationship between this soundscape and the research tools represented by the other fields in this diagram determines how soundscape methodologies can be applied in order to effect such change.

- Environmental sound measurement: The processes by which the soundscape is sampled, quantified, or recorded. This can include the use of acoustic and psychoacoustic metrics, the measurement of impulse responses, and the recording of soundscapes (often in a spatial audio format). The data collected from this can then be used in the evaluation of environmental sound, including (as in this thesis) the auralisation of soundscape recordings in order to recreate the experience of the soundscape in a controlled environment.

- Environmental sound modelling: This includes the use of numerical techniques in order to develop acoustical models of environmental spaces (often informed by, or benchmarked with reference to, the measurement of extant environmental sound).

- Environmental sound simulation: This makes use of acoustical models of environmental sound in order to develop virtual sound scenes (that is to say, soundscape that are not solely based on recorded soundscape material).

- Environmental sound evaluation: This is the use of soundscape methodologies in order to evaluate sound scenes. These scenes can be: the real soundscape of the environment experience in-situ; the auralisation of a measured soundscape; and the auralisation of simulated sound scenes making use of acoustical models and/or recorded soundscape material. Results from this evaluation process can then, in turn, inform these acoustical models for the development of further simulated sound scenes that can then be evaluated.

- Environmental sound modification: This evaluation process then may determines real-world modifications that can be made to the environmental space in order to alter the soundscape of that space. Of course this 'new', modified, soundscape can then iteratively undergo the same analysis and evaluation processes as the original soundscape.

As detailed in the list above, Figure 1.1 clearly details what could be referred to as the 'complete cycle' of soundscape research, where an environment's sound scene is improved by the application of a soundscape methodologies. Having established this cycle then, the extent (and limitation) of the work presented in this thesis can now be stated. The red dashed box in Figure 1.1 indicates the boundaries of the thesis work.

This makes clear that certain research areas are not covered in this theses. For example, this thesis work does not evidence any research into the simulation of 'new' or altered soundscape stimuli, nor does it considered the physical alteration of any of the environments considered in order to make real-world changed to the sound scenes. Instead this thesis work has explored the use of measurement, modelling, and evaluation techniques (alone and in tandem) in order to develop the understanding of the acoustical behaviour of sparsely-reflecting outdoor spaces, and to establish the validity of a variety of soundscape measurement, auralisation, and evaluation methods. The structure of the following chapters of this thesis will now be described.

1.4 Thesis Structure

In order for the relevance of the completed thesis work to be understood Chapters 2, 3, and 4 have been written to provide an overview of existing research in environmental sound and related acoustic research in order to give an idea the context within which the complete thesis work exists.

Chapter 2 begins with a study of the fundamental principles of acoustics, providing a knowledge base upon which Chapter 3 and Chapter 4 build, where they provide a comprehensive review of current research in the fields of environmental noise and soundscape theory respectively. This approach deliberately starts with these, arguably 'basic', fundamentals to provide a knowledge base to support the introduction of further theories and ideas. This is also a worthwhile exercise given the aforementioned interdisciplinary nature of soundscape research. The inclusion of fundamental background information makes the topic available to the layman, which is crucial to ensure the ongoing inclusion of people with a variety of skills in the field of soundscape research (artists, for example, to whom the fundamental properties of sound might not already by known).

The findings from this review of the extant literature then inform the environmental sound measurement work presented in Chapter 5. This chapter includes details regarding the soundscape recordings made for use in listening tests, as well as the measurement of the acoustic properties of the caves and gorge at Creswell Crags.

Chapter 6 then presents four listening tests conducted making use of the collected soundscape data. These listening tests make use of multiple playback formats and evaluation tools in order to develop an understanding of the emotional responses evoked by certain soundscape elements, how these subjective reactions can be measured, and how the presence of visual features can affect these reactions.

Chapter 7 begins with a summary of existing sound modelling methods before presenting the Waveguide Web (WGW), a novel method for the simulation of sparsely reflecting outdoor acoustic scenes.

This thesis then concludes in Chapter 8 where the work presented is summarised, and the hypothesis is restated and considered in light of the findings from the research. The contributions to the field made by the thesis work are also summarised. Areas of future research based on this work are then considered and some final comments from the author are also given.

1.5 Summary

An overview of the aims and structure of this thesis has been presented. This has included a brief summary of the core tenets of soundscape research, and the soundscape methodologies that will be utilised in this thesis. The problem of environmental noise has also been identified, and it has been argued that the application of soundscape methodologies to the evaluation of soundscape recordings will allow for the development of a listener-centric understanding of environmental sound, and therefore an understanding of environmental noise beyond objective noise level measurement metrics.

The interdisciplinary nature of soundscape research has also been described, as has the manner in which this thesis reflects that fact by its portfolio structure. An hypothesis has been stated to formalise the aims of this research, and the program of research designed to support this hypothesis has been outlined.

Chapter 2

Fundamentals of Acoustics

This chapter will cover the fundamentals of acoustic theory, including: the basic properties of sound waves and their propagation, the mechanisms of the human auditory system and the perception of sound, and the behaviour of sound in space (including its measurement in the form of acoustic impulse responses). The purpose of this chapter is therefore to establish a suitable knowledge base and context within which the rest of this thesis can be understood. These fundamentals are, of course, of direct relevance to any soundscape based research, as any listener-centric understanding of sound must be underpinned by an understanding of the physical properties of sound.

2.1 Basic Properties of Sound

This section covers the fundamental properties of sound. Firstly, the nature of sound wave propagation in air, followed by an explanation of the relationship between frequency and wavelength. This section concludes with an examination of methods for quantifying and measuring sound levels, and the inverse square law.

2.1.1 Wave Propagation in Air

A sound wave is a pattern of vibration in a medium caused by the movement of energy, and is an example of what is known as a longitudinal wave [17], where the vibrations associated with the wave are parallel with the direction of wave propagation[1]. Figure 2.1 shows this type of wave movement as represented by the mass spring model (shown here in simplified one dimensional form), where the masses represent air molecules, and the springs represent the intermolecular forces [18].

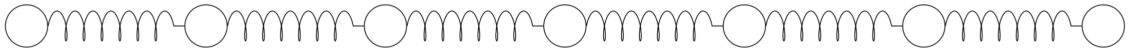


FIGURE 2.1: The mass-spring model of sound propagation, after [18].

The introduction of a sound wave to this set of particles is represented by a force applied to one of the masses. For example, if the first particle is pushed towards the next one, the spring between those two particles will compress. This compression will then later be released as the spring moves the first molecule back to its original position. This causes the spring to stretch out (to rarefy), and forces the next molecule onwards, allowing the pattern of compression and rarefaction to propagate through the medium.

Figure 2.2 shows a representation of a sound pulse propagating in the material presented in Figure 2.1. It clearly shows the progression of compression and rarefaction from one end of the masses to the next. It also shows a key concept regarding movement of sound waves: it is not movement of the molecules themselves that makes a sound waves, but rather the movement of pressure differential between them due to the changing intermolecular forces.

From this model of sound propagation two factors that determine the speed of sound in a given material can be inferred:

- **Stiffness.** The stiffer the material, the faster sound travels. This is represented by the stiffness of the springs.
- **Density.** The denser the material, the slower sound travels. The is represented by the density of the masses.

These two factors are represented in the equation for the speed of sound in a given material [19]:

$$c = \sqrt{\frac{B}{\rho}} \tag{2.1}$$

[1]The alternative case, a transverse wave, is where the vibrations are perpendicular to the direction of wave propagation e.g. electro-magnetic waves.

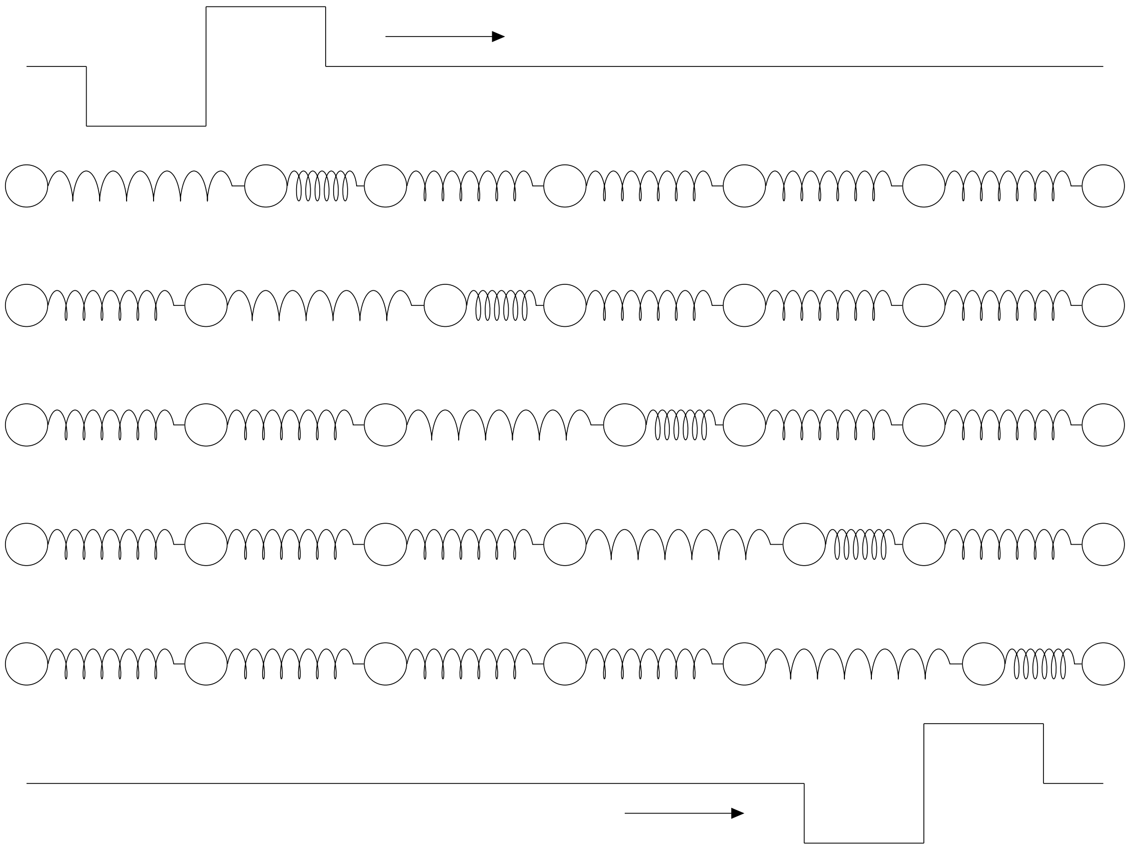


FIGURE 2.2: The mass-spring model showing the propagation of a sound pulse through a material, after [18].

where B is the bulk modulus (Nm^{-2}), a measure of the strength of the intermolecular forces in the material, and ρ is density ($\mathrm{kg\,m}^{-3}$) of that material. For a gas:

$$B_{\text{gas}} = \gamma P \tag{2.2}$$

$$\rho_{\text{gas}} = \frac{PM}{RT} \tag{2.3}$$

where γ is the Adiabatic gas coefficient of the particlar gas (1.4 for air), R is the gas constant ($8.31\mathrm{JK}^{-1}\mathrm{mole}^{-1}$), T is the absolute temperature (in K), and M is the molecular mass of the gas ($\mathrm{kg\,mole}^{-1}$). The speed of sound in a gas therefore given by [20]:

$$c_{\text{gas}} = \sqrt{\frac{\gamma RT}{M}} \tag{2.4}$$

This expression shows that beyond the medium specific quantities R and M, the only factor that will alter the speed of sound is temperature. Using this equation the speed of speed of sound in air at 20°C can be calculated:

$$c = \sqrt{\frac{\gamma RT}{M}} \quad (2.5)$$

$$c = \sqrt{\frac{1.4 \cdot 8.31 \cdot T}{2.89 \times 10^{-2}}} \quad (2.6)$$

$$c = 20.1\sqrt{T} \quad (2.7)$$

$$c = 20.1\sqrt{273 + 20} \sim 343.4\,\mathrm{ms}^{-1} \quad (2.8)$$

2.1.2 Frequency and Wavelength

Whilst Figure 2.2 considers the propagation of a single pulse, the vast majority of sounds we hear contain some sort of periodic component. The simplest example of a periodic signal is a sine wave, as it represents vibration at a single frequency.

Figure 2.3 shows an example of sinusoidal behaviour. Here the mass on the far left is being moved backward and forward at a particular frequency, resulting in a repeating pattern of compression and rarefactions propagating through the medium. These variations in pressure can be plotted, indicating the presence of a sine wave.

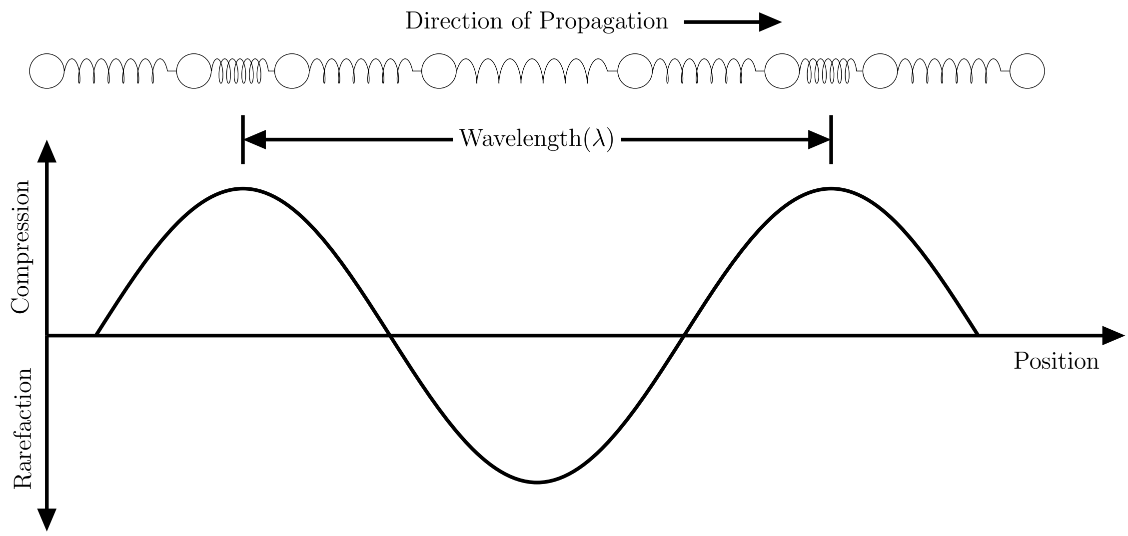


FIGURE 2.3: The propagation of sine wave through a material represented by the mass-spring model, after [18]

The distance between the repeating sections of the wave in Figure 2.3 is know as its wavelength, denoted by λ. Combining this fact with the previously derived speed of sound in the material, and the period T of a sine wave, the equation for speed as a relationship between distance and time (speed = $\frac{\text{distance}}{\text{time}}$) becomes:

$$c = \frac{\lambda}{T} \tag{2.9}$$

this can be combined with knowledge of the reciprocity between T and F ($f = \frac{1}{T}$) to give:

$$c = f\lambda \tag{2.10}$$

which describes the similarly reciprocal relationship between f and λ. Rearranging (2.10) to make λ the subject allows for the calculation of the wavelength of different frequency components. For example, the wavelength of a 1 kHz sine wave in air at 20°C:

$$\lambda = \frac{343.4}{1000} \tag{2.11}$$

$$\lambda = 0.3434\,\text{m} \tag{2.12}$$

2.1.3 Sound Pressure Level and the Inverse Square Law

Sound levels are measured on a decibel (dB) scale in the form of sound pressure level (SPL) values. SPL measurements indicate loudness relative to the weakest audible sound (a pressure value of 2×10^{-5} Pa [21]). The dB SPL level of a recorded RMS sound pressure value is given by [22]:

$$\text{dB SPL} = 20\log_{10}\left[\frac{\text{RMS Sound Pressure}}{2 \times 10^{-5}\ \text{Pa}}\right] \tag{2.13}$$

Some example noise sources and their associated SPL levels are shown in Table 2.1.

As the decibel is a unit of relative measurement, the difference between two recorded sources can be expressed in relative SPL (L_p):

$$L_p = 10\log_{10}\left(\frac{p_1}{p_0}\right)^2 = 20\log_{10}\left(\frac{p_1}{p_0}\right)\text{dB SPL} \tag{2.14}$$

where p_1 indicates the level currently being measured, and p_0 indicates the reference level [24].

Sound Pressure Levels Levels (dB SPL)			
Weakest sound heard	0 dB	Hand drill	98 dB
Whisper quiet library at 6'	30 dB	Power motor at 3'	107 dB
Normal conversation at 3'	60-65 dB	Sandblasting, live concert	115 dB
Telephone dial tone	80 dB	Onset of pain	125 dB
City traffic (inside car)	85 dB	Pneumatic riveter at 4'	125 dB
Train whistle at 500'	90 dB	Level at which short term exposure can cause permanent damage	140 dB
Jackhammer at 50'	95 dB	Jet engine at 100'	140 dB
Subway train at 200'	95 dB	12 Gauge shotgun blast	165 dB
Level at which sustained exposure may result in hearing loss	90-95 dB	Death of hearing tissue	180 dB
		Loudest sound possible	194 dB

Table 2.1: Some common environmental sounds and their associated SPLs, after [23].

Whilst Figure 2.1 and Figure 2.2 show the propagation of sound in a single dimension, in reality sound propagates in three dimensions. As a sound spreads out spherically from its source it becomes weaker as its energy is spread out increasingly thinly. As such the sound intensity (power per unit area) is given as function of distance by:

$$I = \frac{W_{\text{source}}}{A_{\text{sphere}}} = \frac{W_{\text{source}}}{4\pi r^2} \tag{2.15}$$

where I is the sound intensity in W m^{-2}, W_{source} is the power of the source in W, and r is the distance from the source in m. This indicates the inverse square relationship between sound intensity and distance from the sound source. It is important to note however that (2.15) indicates a theoretically infinite sound intensity level at the source, which ignores the fact that all real sources have a finite area. It also assumes free propagation from the source in all direction ('free field radiation'), which is an assumption that does not hold for real environments. As such, 2.15 can be rewritten taking into account factor Q which describes the directivity of the source relative to a sphere:

$$I = \frac{QW_{\text{source}}}{4\pi r^2} \tag{2.16}$$

The value of Q is defined by the presence of orthogonal boundaries near the source. The presence of one boundary effectively gives hemispherical spreading, two boundaries makes a half hemisphere, and three boundaries a quarter hemisphere. These states correspond to Q values of 2, 4, and 8 respectively, as shown in Figure 2.4.

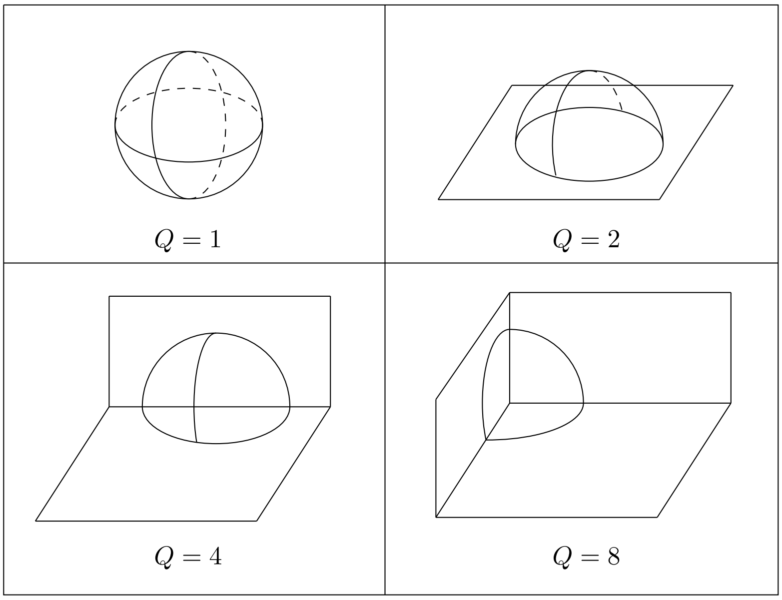


FIGURE 2.4: Representations of the approximate directivity factor Q for different orthogonal boundary cases, after [24].

2.2 Hearing and Perception

This section contains a brief introduction to human perception of sound, the study of which is known as psychoacoustics. This includes a description of the human hearing system, and how it relates to frequency perception and spatial hearing.

2.2.1 The Ear

The anatomy of the ear is formed of three sections: the outer, middle, and inner ear [18]. Figure 2.5 shows these three sections and their constituent parts. The outer ear consists of the pinna and the external auditory canal. The many ridges of the pinna enhance particular frequencies and aid with sound source localisation. The auditory canal directs incoming sound waves to the tympanic membrane, or eardrum. The auditory canal introduces further frequency modifying effects, with a main resonance around 4 kHz.

The tympanic membrane is the interface between the outer ear and the middle ear, and transforms incoming vibrations in air into bone vibrations transmitted to the ossicles. The ossicles are three small bones: the malleus, incus, and stapes. The role of the ossicles is to act as an impedance converter, receiving vibration from the tympanic membrane which

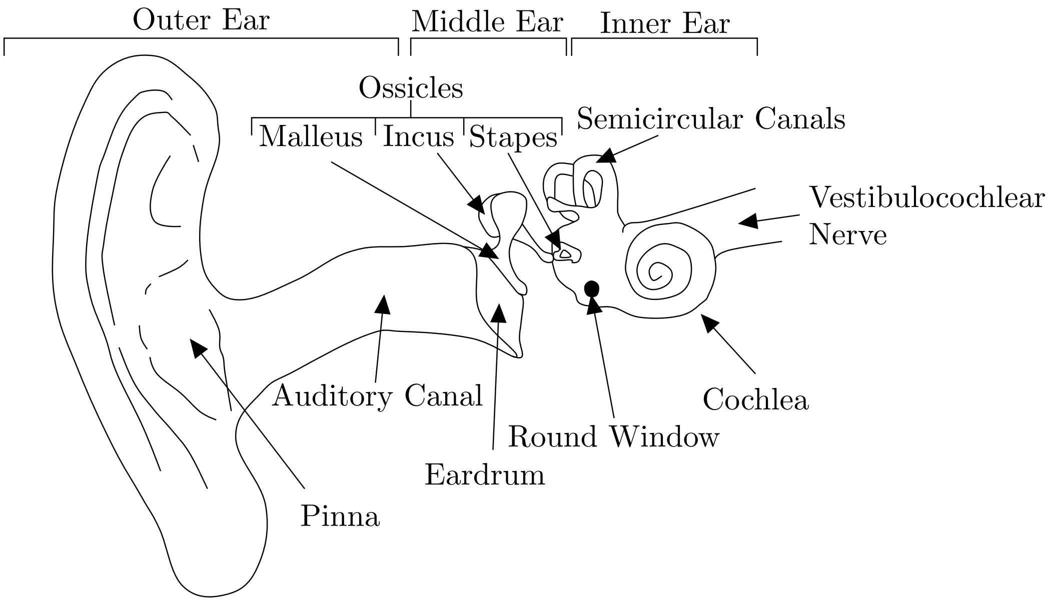


FIGURE 2.5: The anatomy of the ear indicating the outer ear, middle ear, and inner ear, after [25].

are then delivered to the base of the cochlea with increased effective pressure, which is required due to the higher resistance of cochlear fluid relative to air.

The cochlea is a coiled structure that receives mechanical vibrations from the stapes at its base and converts them into nerve firings identifying the frequency components in the incoming sound. Running along the length of the cochlea is the basilar membrane. As shown in Figure 2.6, the basilar membrane resonates along its length at frequencies ranging from 20 Hz (at the apex) to 20 kHz at its base (a phenomenon first identified by Von Békésy [26]). In this way the basilar membrane can be thought of as a *tonotopic* map [27], where resonance with different frequencies (or tones) is spread across the surface topography of the membrane. These resonances then stimulate the Organ of Corti, a set of hair cells (cilia) distributed along the length of the basilar membrane. As the cilia are stimulated they trigger a nerve bundle known as the vestibulocochlear nerve, which send this frequency and timing information to the brain.

Where the resonances of the basilar membrane define the frequency range of the hearing system, the physiology of the outer and middle ear dictates the hearing system's sensitivity to different frequencies. This changing sensitivity with frequency can be seen in the equal loudness contours (Figure 2.7) first measured by Fletcher and Munson [28]. Clearly

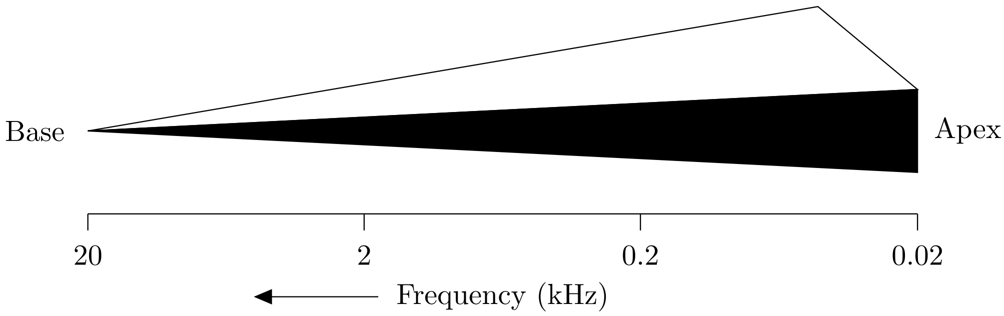


FIGURE 2.6: A simplified diagram of the basilar membrane indicating its resonance with different frequencies across the hearing range along its length, after [18].

indicated is the particular sensitivity of the hearing system between 2 kHz and 4 kHz (frequencies within the human vocal range crucial for intelligibility [29]), and decreasing sensitivity at lower frequencies.

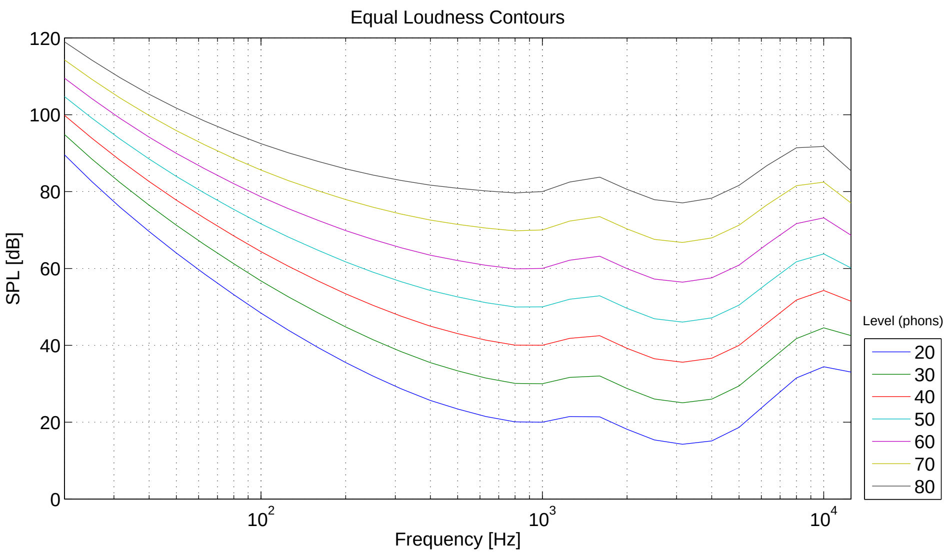


FIGURE 2.7: Equal loudness contours. The phon is a measure of loudness relative to the dB SPL level at 1 kHz.

In this figure, the plotted contours indicate the relative SPLs required for tones at different frequencies to be perceived equally as loud as a 1kHz tone at a particular reference level. This is known as the *Phon* scale.

2.2.2 Spatial Hearing

Having now considered the human auditory system in terms of the way a single ear perceives pitch and loudness, the way in which humans use two ears (binaural hearing) can now be considered.

Figure 2.8 shows a diagram of the human head with the positions of various features of relevance to an understanding of spatial hearing marked. In this figure it can be seen that a sound arriving from a particular direction has associated azimuth and elevation angles, expressed relative to the horizontal and median planes and their respective axes.

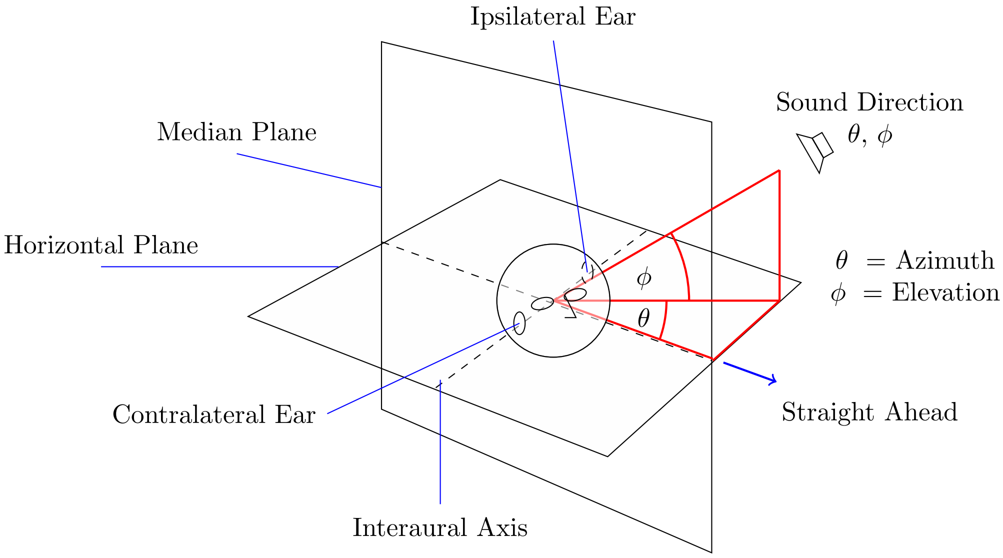


FIGURE 2.8: A diagram indicating aspects pertinent to the understanding of spatial hearing.

The following terms are displayed in Figure 2.8:

- The horizontal and median planes. These are effectively the references axes within which the direction of incoming sound can be understood. The horizontal plane is defined by the interaural axis, which bisects both ears, and the 'straight ahead' direction. The median plane is then vertically orthogonal to the horizontal plane, aligned with the 'straight ahead' direction.

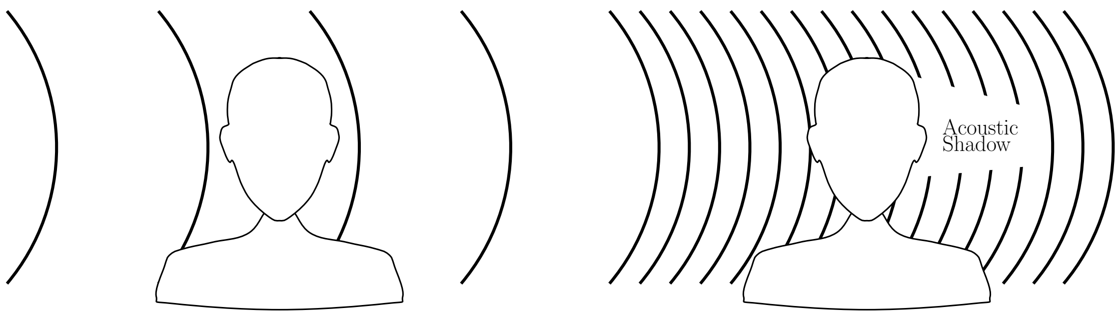
- Sound Direction, as defined by two angles: an azimuth angle (θ) expressing the direction in the horizontal plane, and an elevation angle (ϕ) expressing the relationship between its vertical position and the horizontal plane.

- The direction of the sound source then determines which ear is referred to as the ipsilateral ear, which is the ear on the same side of the head as the arriving sound. The contralateral ear is therefore the ear on the other side of the head to the arriving sound source [30]. These terms come from the Latin *ipse* meaning itself, and *contra* meaning against.

With these defined, the various cues, both directional and distance related, that comprise the ability to perceive where a sound is coming from can now be considered:

- The difference in the time of arrival of a sound at each ear, known as Interaural Time Difference (ITD)
- The difference in the amplitude of a sound at each ear, known as Interaural Level Difference (ILD).
- Spectral cues determined by the Pinnae.
- The balance of low and high frequencies in a sound.
- The ratio of direct-to-reverberant sound.

Interaural Level Differences (ILDs) are present when the head has a shadowing effect on arriving sound waves, leading to an attenuation in level at one ear relative to the other. This is due to the arriving sound having wavelengths with dimensions comparable to the head and features of the head. Figure 2.9 shows how the presence of the head in the path of a sound wave produces an acoustic shadow at high frequencies (typically 1600 Hz and above).



(a) (b)

FIGURE 2.9: (a) shows where low frequency sounds interact with the head there is little or no acoustic shadowing, whereas (b) shows that high frequency sounds with much smaller wavelengths result in acoustic shadowing, after [24].

Figure 2.9 also shows that where the wavelengths associated with the incoming sound are large when compared with the head, there is little acoustic shadowing. For lower frequencies then, binaural localisation relies on Interaural Time Differences (ITDs). Figure 2.10 shows a simplified model of the path difference introduced by the head that lead to these timing differences, where the path difference is dependent on the size of the head, and the angle of incidence of the sound source.

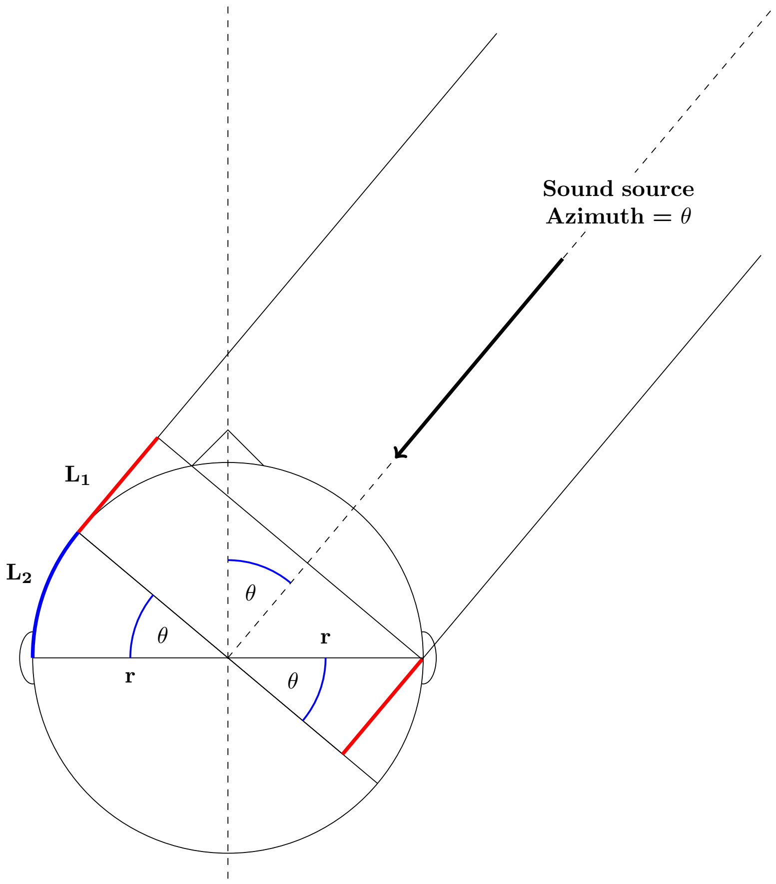


FIGURE 2.10: A simplified model of Interaural Time Difference (ITD) showing how the angle of incidence of the sound source and the size of the head introduces a path difference between sound arriving at each ear, after [18].

The diagram in Figure 2.10 can be used to derive Woodworth's formula [31] which can be used to estimate ITD for a head of a given size:

$$L_1 = r\sin(\theta) \tag{2.17}$$

$$L_2 = r\theta \tag{2.18}$$

$$\text{Path difference, } P = L_1 + L_2 \tag{2.19}$$

$$P = r\sin(\theta) + r\theta \tag{2.20}$$

$$\text{ITD} = \frac{P}{c} = \frac{r}{c}(\sin(\theta) + \theta) \tag{2.21}$$

where c is the speed of sound. Using this formula a maximum theoretical value for ITD, $\mathrm{ITD}_{\mathrm{max}}$ can be approximated (where r = 875mm, $c = 340\mathrm{ms}^{-1}$, and $\theta = 90°$):

$$\mathrm{ITD}_{\mathrm{max}} = 660\mu\mathrm{s} \tag{2.22}$$

Given this maximum value for ITD it is clear that, for frequencies with time period above this value, localisation ambiguity will occur due to the reduced phase difference between the signals reaching the two ears. This lowest frequency of ambiguity is therefore:

$$\text{Lowest frequency of ambiguity} = \frac{1}{\mathrm{ITD}_{\mathrm{max}}} \approx 1.5\mathrm{kHz} \tag{2.23}$$

Above this frequency ILD cues therefore become more important for sound localisation.

The Duplex theory [32] states that only ITD and ILD cues are required for localisation. However, this neither accounts for localisation in the median plane, nor for how pinnae cues are used to reduce the impact of the 'cone of confusion' which can be seen in Figure 2.11.

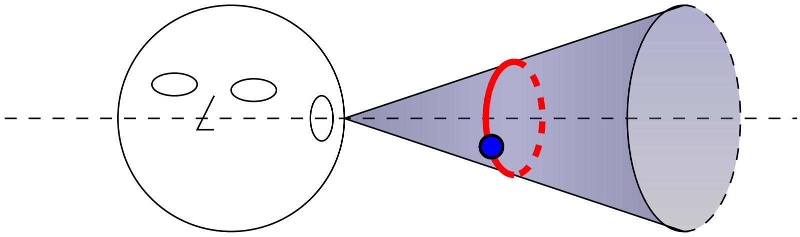


FIGURE 2.11: A demonstration of the cone of confusion and the ring of confusion as determined by ITD and ILD cues. Indicated by the blue circle is an example sound source position on the cone and ring that would require additional cues in order to determine its location.

The 'cone of confusion' as shown here is an illustration of sound source locations that will result in the same ITD. A sound source on a circle of the cone also has the same ILD. Cues from the pinnae are then used to localise the source at a single point. It is also important to consider that when listening in order to localise a sound an individual will turn and tilt their heads, in order to modify and minimise the ITD and ILD cues to identify the source of the sound more effectively.

2.3 Room Acoustics

Having now covered the basic properties of sound waves and their movement in air, as well as the fundamentals of sound perception, the behaviour of sound in physical environments will now be examined. This behaviour has typically been studied in the context of room acoustics, due to the importance of the acoustic design of concert venues and lecture theatres. This section will examine three main areas of room acoustics literature: acoustics impulse responses, room modes, and reverberation time calculation.

2.3.1 Acoustic Impulse Responses

An acoustic Impulse Response (IR) can be considered as the 'acoustic fingerprint' of a space, and describes the response of a space to sound within it. IRs are often used to add artificial reverberation to a sound through the use of convolution [33]. This technique is often employed for either the realistic auralisation of an acoustic scene [24], [34], or for artistic/musical purposes [35], [36]. The impulse response of any linear, time-invariant, system is defined by its output following an impulse at its input. An impulse is defined by the δ-function:

$$\delta(t) = \begin{cases} \infty, & \mathrm{t} = 0 \\ 0, & \mathrm{t} \neq 0 \end{cases} \tag{2.24}$$

The Fourier transform of the δ-function is given by:

$$\mathcal{F}[\delta(t)] = \int_{-\infty}^{\infty} \delta(t) e^{-j\omega t} \mathrm{d}t \tag{2.25}$$

The frequency spectrum of the δ-function is flat across all frequencies. Due to the rapid temporal response required, faithful reproduction of an impulse can be very difficult. Impulse-like sound sources, such as gunshots and balloon pops, can be used as the input signal for IR recording [37], but these sources are not repeatable, do not offer true impulses, and their own spatial and acoustic properties will affect the recorded results. This means that when recording IRs using such methods the environment will not be excited equally at all frequencies or in all directions, which are requirements for impulse response recording (as defined in ISO 3382 [38]). This being said, chapter 5 will show the use of a starter pistol for IR measurement to be suitable in certain situations, depending on the application of the recorded IRs. Figure 2.12 shows the time and frequency domain plots of the δ-function.

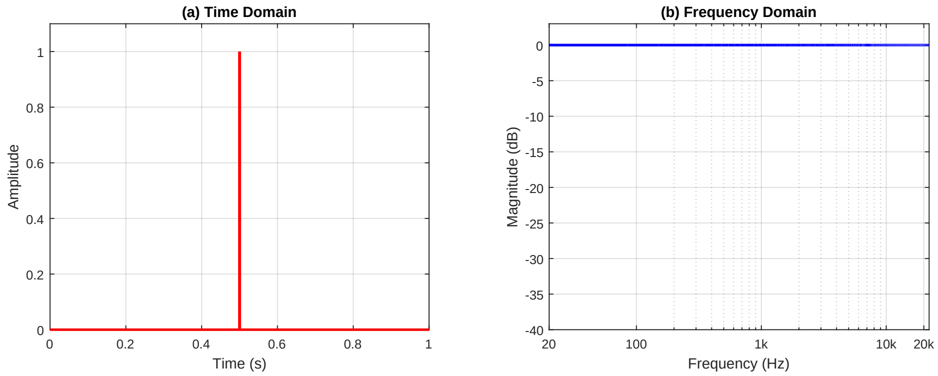


FIGURE 2.12: (a) Time and (b) frequency domain representation of the δ-function.

An alternative sound source for impulse response measurement is an Exponential Sine Sweep (ESS) [39]. An exponential sine sweep *s(t)* can be reproduced by a loudspeaker in a recording space, exciting it at many frequencies sequentially over an extended time period. This sweep will typically cover the range of human hearing (c. 20 Hz - 20 kHz). The input sine sweep can then be deconvolved from the recorded result of introducing it into the space, giving the IR. *s(t)* is given by:

$$s(t) = sin\left[\frac{\omega_1.T}{\ln\left(\frac{\omega_2}{\omega_1}\right)}.\left(e^{\frac{t}{T}.\ln\left(\frac{\omega_2}{\omega_1}\right)} - 1\right)\right] \tag{2.26}$$

where T is the length of the sweep, ω_1 is the start frequency, and ω_2 is the end frequency. Whilst the ESS method can suffer from certain issues, including pre-ringing in the resultant impulse at low frequency before the arrival of the direct sound pulse, and sensitivity to abrupt impulsive noises during the measurement process (which result the presence of reverse sweeps in the IR) and non-linearities in the signal chain, it can offer greater signal-to-noise ratio than the use of impulsive sound sources or linear sweeps, and is free of the time-ambiguity aliasing errors associated with the maximum length sequence (MLS) method [40].

It is often desirable to use one very long sine sweep for IR measurement, ideally with an omni-directional sound source. Use of a longer sine sweep results in an increased signal-to-noise ratio [41] (although this does increase the risk of interference), and the use of a single sweep avoids possible averaging errors associated with the use of multiple sweeps

[42]. Use of an omni-directional source is desirable in order to excite the space acoustically equally in all directions.

However, typical omnidirectional sources are expensive, can suffer from attenuated bass response [43], and are usually designed for band limited noise excitation [44]. A directional source can therefore be used to approximate an omnidirectional one by taking multiple recordings at the same position, rotating the sound source to a different angle for each measurement. These results can then be summed and normalised to emulate the use of a single omnidirectional source [41].

Whilst there has been considerable work regarding impulse response recording and sound propagation indoors, there has been relatively little has concerning the acoustic impulse responses of outdoor spaces [45]. Work conducted in the city of York considered the acoustic properties of a street historically used for dramatic performances, and showed that the use of a sine sweep for impulse response recording is a suitably robust method for use in outdoor measurement work [46]. Other work in this area has investigated sound propagation in urban environments [47] and forests [48], including recording impulse responses in Koli National Park, Finland [41].

Figure 2.13 indicates the three primary sections of an idealised[2] IR. The direct sound is the initial acoustic wave reaching the receiver from the source, following the straight line path between source and receiver. Early reflections are distinct echoes of the direct sound, formed by reflections from the surrounding geometry that result in longer propagation paths. These are followed by the reverberant tail section, where the acoustic energy has diffused into the space and there are no longer distinct reflections reaching the receiver, resulting in a statistically determined decay.

Acoustic Absorption Coefficient

The dimensions of the environment define the time of arrival of early reflections, and the later reverberation, but the tonal character of these reflections is also affected by the acoustic absorption properties of the surface materials. This is defined by a surface's acoustic absorption coefficient α at various frequencies, which is influenced by various

[2] An impulse response recorded in a bounded space, with a single point source, a single point receiver, and line-of-sight between the two.

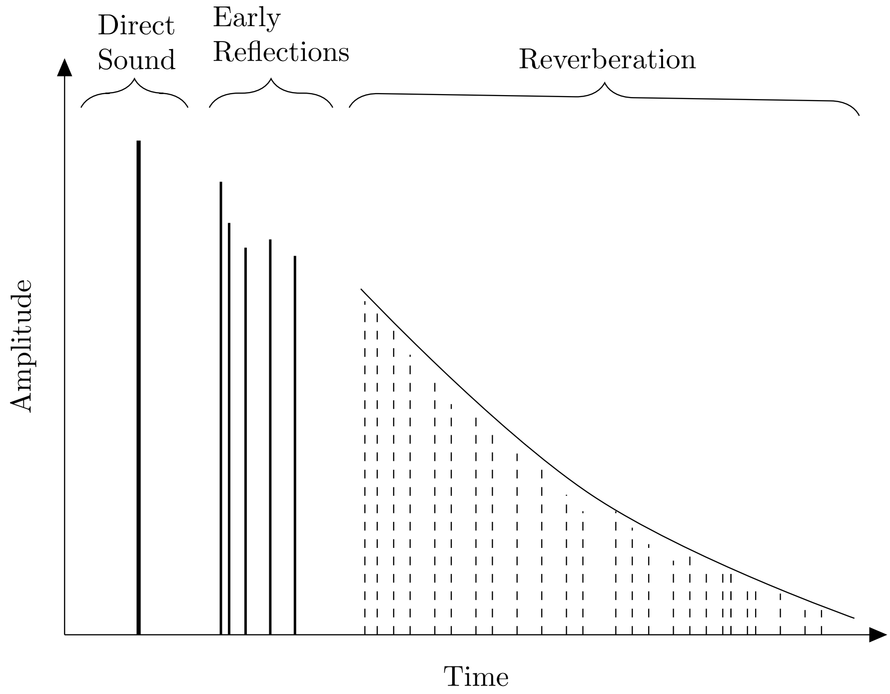


FIGURE 2.13: A simplified IR, indicating the three main sections: the direct sound, the initial distinct echoes of this sound in the form of early reflections, and the late tail of echoes forming the reverberation section. After [18].

factors, including the thickness of the material and its porousness [26]. Table 2.2 gives some example α values. These values are bounded between 0 and 1, where 0 represents no absorption, and 1 represents total absorption.

Materials	**Acoustic Absorption Coefficient (α)**					
	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4kHz
Brick - Unglazed	0.03	0.03	0.03	0.04	0.05	0.07
Ordinary Window Glass	0.35	0.25	0.18	0.12	0.07	0.04
Plaster on Lath (rough finish)	0.14	0.10	0.06	0.05	0.04	0.03
Plaster on Lath (smooth finish)	0.14	0.10	0.06	0.04	0.04	0.03
Carpet on Concrete	0.02	0.06	0.14	0.37	0.60	0.65
Open Doors and Windows	1	1	1	1	1	1

TABLE 2.2: The acoustic absorption coefficients (α) of some example materials at different frequencies, after [49].

2.3.2 Room Modes

Where an environment is excited by an acoustic source, the physical dimensions of that environment will result in resonances at particular frequencies which have wavelengths corresponding to those dimensions. These resonances, at least when considering indoor sound, are known as room modes. The simplest environment to consider in order to understand the frequencies at which resonances occur is a cuboid 'shoebox' room. Figure 2.14 shows the three types of modal paths present in such a space, where resonances are set up with soundwaves reflecting off of multiple surfaces. Resonances are found at frequencies that have corresponding half wavelengths equal to the modal path length.

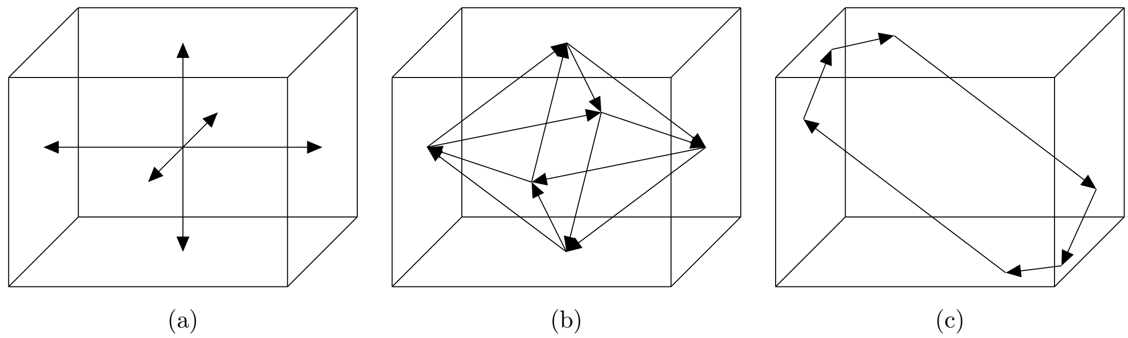


FIGURE 2.14: (a) Axial, (b) Tangential, and (c) Oblique modal paths in a cuboid room.

The frequencies of the room modes (f_{xyz}) illustrated in Figure 2.14 can therefore be calculated using the following equation:

$$f_{xyz} = \frac{c}{2}\sqrt{\left(\frac{x}{L}\right)^2 + \left(\frac{y}{W}\right)^2 + \left(\frac{z}{H}\right)^2} \tag{2.27}$$

where c is the speed of sound, L, W, H are the length, width, and height of the room, and x, y, z are the number of half wavelengths between the surfaces $(0, 1, 2, \ldots)$ [50]. Whilst 2.27 only describes the modes for cuboid rooms, it represents concepts that can be applied to irregularly shaped rooms in order to estimate the room modes present in such a space.

2.3.3 Room Acoustic Parameters

Acoustic impulse responses, as defined in section 2.3.1, have a variety of metrics associated with them (e.g. those outlined in ISO-3382 [38]). This section will consider the following parameters: reverberation time, early decay time, definition, and clarity.

Reverberation Time

Reverberation time (RT_{60}) is defined as the time taken for the impulse response to reach one millionth of its initial intensity, equivalent to -60dB [51]. The first method for calculating the reverberation time of a room was developed by Wallace Clement Sabine following his work on improving the Fogg Lecture Hall at Harvard in 1895 [52]. Through this work he established 'appropriate' acoustic properties. For example, a reverberation time of 2-2.25 seconds is ideal for a concert hall[3], whereas for a lecture a reverberation time of under 1 second is desired. As a result of this work Sabine was hired as acoustic consultant for Boston's Symphony Hall in 1900.

As well as developing these guidelines for reverberation time, Sabine developed an equation for the calculation of a room's reverberation time [53]:

$$T_r = 0.161\frac{V}{A} \tag{2.28}$$

where the reverberation time T_r is determined by room volume V and effective surface area A. This effective surface area is sum of products of each area covered by particular material:

$$A = \sum_{i+1}^{n} \alpha_i A_i = \alpha_1 A_1 + \alpha_2 A_2 + \ldots + \alpha_n A_n \tag{2.29}$$

where α_i is the acoustic absorption coefficient of surface A_i, and the unit of measurement for $\alpha_i A_i$ is the Sabin.

Whilst Sabine's findings were ground-breaking (and can still prove useful), following the advent of recording technology new, more reliable, methods have been developed to measure reverberation time. In 1964 Manfred Schroeder of Bell Labs developed a new method for RT_{60} calculation. Schroeder introduced the Energy Decay Curve (EDC, sometimes known as the Schroeder curve) which is the reverse time integral of the squared impulse response.

This gives a smooth curve describing the total signal energy in the impulse response over time, which can then be used to estimate reverberation time [51]. RT_{60} values are calculated by extrapolation from the EDC, for example between the -5dB and -35dB points (giving what is known as a T30 RT_{60} value) or between the -5dB and -25dB points

[3]Faster and more intricate pieces of music, by composers such as Frank Zappa and Phillip Glass will be better served by a shorter reverberation time relative to the sweeping romantic works of, for example, Arvo Pärt and Sunn 0))).

(known as T20) [54]. Figure 2.15. shows gives an example of this process finding the T30 RT_{60} value of an IR recorded at York minster.

Schroeder's method of reverberation time calculation remains in use today, alongside a host of other tools for measuring characteristics of impulse responses (such as those outlined in ISO-3382 [38], or Farina's set of *Aurora* plug-ins for Adobe Audition [56]). It is also worth noting that the formulation of the EDC can vary slightly depending on the signal-to-noise ratio (SNR) of the IR under analysis, but the process remains similar to that presented in Figure 2.15. Reverberation time measurements are typically made for multiple frequency bands [54]. This section will now examine some of the other parameters that can be calculated for impulse responses.

Early Decay Time

The EDT parameter (Early Decay Time) is calculated by taking an IRs EDC and making a linear extrapolation of the from the first 10 dB of decay to the -60 dB point. This parameter gives a measure of the decay of the initial section of the IR, and has been shown to be well aligned with perception of reverberation time (due to the perceptual importance of the initial characteristics of an IR [57]).

Definition

Definition (D_{50}) is an expression of the early to total sound energy ratio, where the early time is defined as the first 50ms of the IR (hence the subscript 50). Definition values are calculated by:

$$D_{50} = \frac{\int_0^{0.05} p^2(t)\mathrm{d}t}{\int_0^{\infty} p^2(t)\mathrm{d}t} \tag{2.30}$$

where p is the sound pressure level, and D_{50} is expressed as a percentage. The top limit of the numerator integral can be modified to change the definition of the duration of the early time parameter (e.g. 0.03 for a D_{30} parameter and an early time of 30ms.

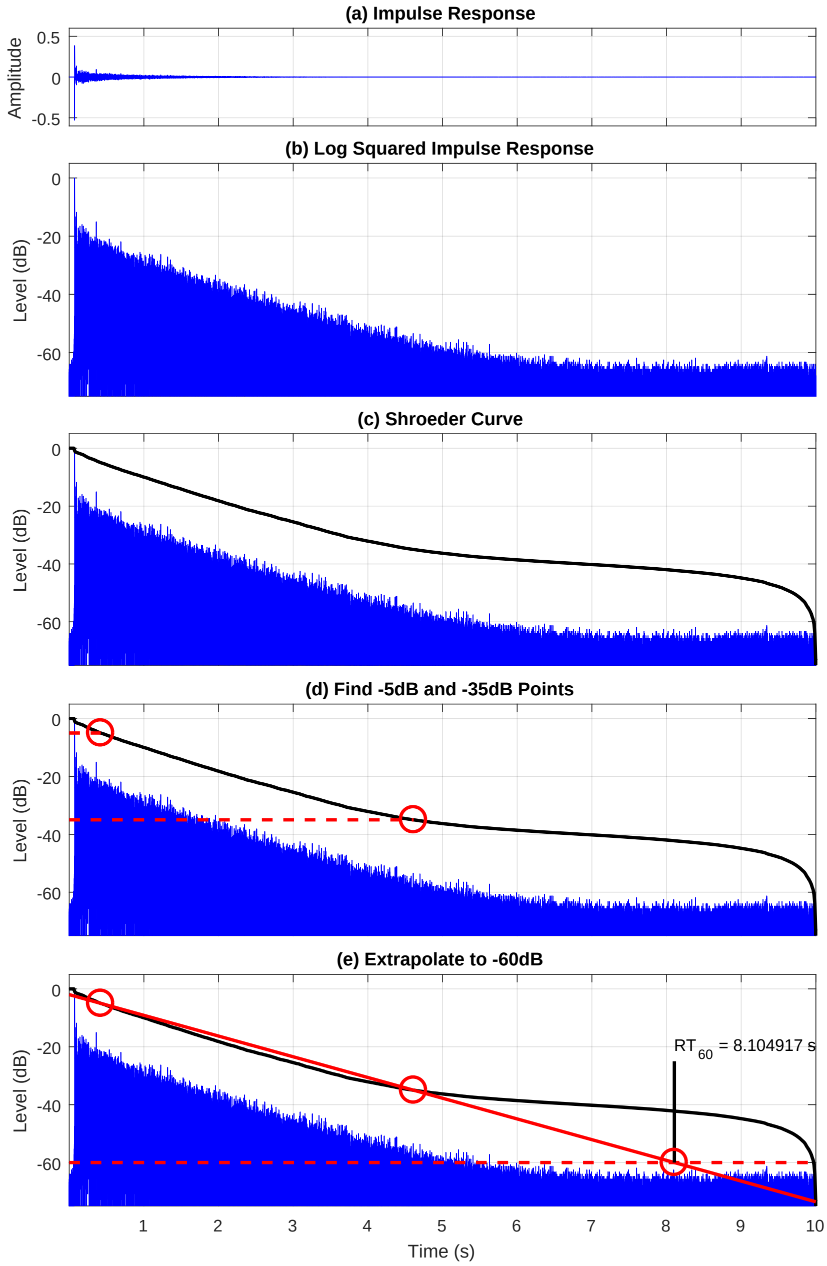


FIGURE 2.15: Plots showing the stages in finding the T30 RT_{60} value of an IR. The IR used in this example was recorded in the York Minster and was taken from OpenAIR [55]. (a) Shows the time domain waveform of the IR. (b) Shows the logarithmic squared version. (c) Adds the Schroeder curve, which is generated by reverse integration of the log squared IR (b). (d) Indicates the -5dB and -35dB points on the Schroeder curve used in T30 RT_{60} measurement. (e) Shows hows these points are used to extrapolate to the -60 dB to find the RT_{60} reverberation time. Note this example makes use of the broadband impulse response, whereas, typically, RT_{60} values are calculated separately for multiple octave bands.

Clarity

Clarity (C_{50}) is a measure of intelligibility of sound in a space (e.g. speech). A high value of clarity indicates a good level of intelligibility. Clarity is defined as the logarithmic ratio of early-to-late energy in an IR (i.e. it is a logarithmic expression of Definition) and is calculated by:

$$C_{50} = 10\log\left(\frac{\frac{\int_0^{0.05} p^2(t)\mathrm{d}t}{\int_0^{\infty} p^2(t)\mathrm{d}t}}{1-\frac{\int_0^{0.05} p^2(t)\mathrm{d}t}{\int_0^{\infty} p^2(t)\mathrm{d}t}}\right) \tag{2.31}$$

giving a value expressed in decibels (dB). As with Definition the subscript value indicates the early time parameter used.

2.3.4 Convolution

IRs can be used to make a recording sound as if it was made in the IR measurement space. This is achievable through the a process known as convolution, making use of the convolution integral [58], which for two functions $x(\tau)$ and $h(\tau)$ is expressed as:

$$y(t) = \int_{-\infty}^{\infty} x(\tau)h(t-\tau)\mathrm{d}\tau \tag{2.32}$$

where t represents a time shift. The four steps identified by this integral are:

1. Reflect: $h(\tau) \rightarrow h(-\tau)$
2. Shift: $h(-\tau) \rightarrow h(t-\tau)$
3. Multiply: $h(t-\tau)\,\mathrm{by}\,x(\tau)$
4. Integrate: $y(t)$ is area under product of $h(t-\tau)$ and $x(\tau)$

Performing this integral over all values of t required to fully 'slide' $h(\tau)$ across $x(\tau)$ results in the convolution of the two signals. The convolution integral can be expressed in its discrete time domain form as :

$$y[t] = \sum_{\tau=-\infty}^{\infty} x[\tau] \cdot h[t-\tau] \tag{2.33}$$

Figure 2.16 shows an example convolution of an audio signal (some synthesised drum beats) with an impulse response.

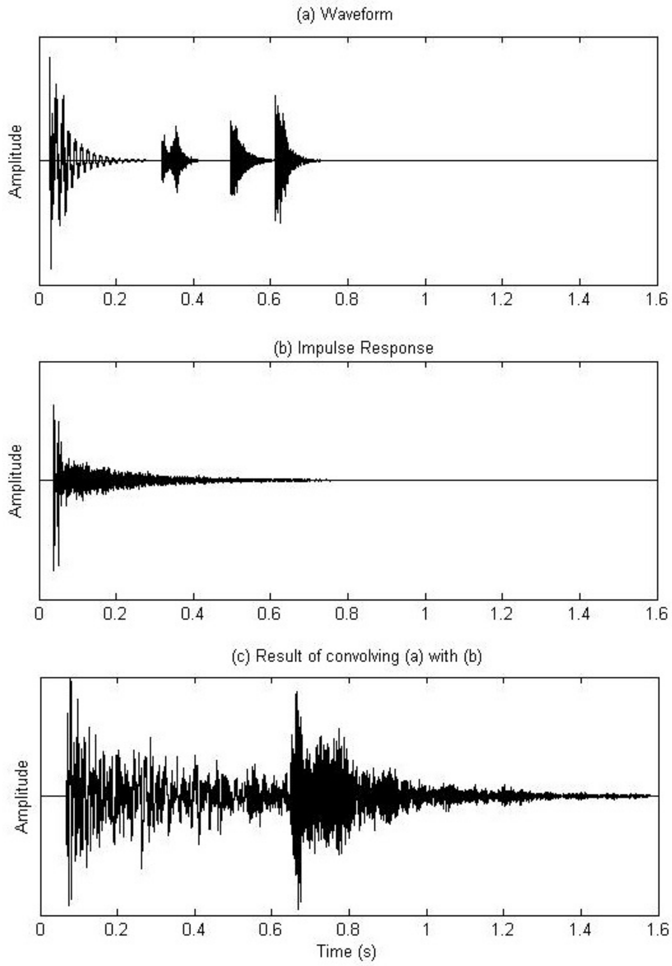


FIGURE 2.16: An example convolution of waveform (a) with impulse response (b), resulting in waveform (c). Sound files (a) and (b) from OpenAIR [55].

It can be seen that the convolution 'adds' the acoustic properties of the impulse response to each drum beat, resulting in a waveform that makes the drums sound as if they had been recorded in the impulse response recording space.

2.4 Summary

This chapter has presented a summary of the fundamentals of acoustics, providing a context for the further chapters in this thesis. This included the nature and behaviour of sound waves, their perception by via the human auditory system, and the behaviour of sound in indoor spaces. It was shown that the behaviour of sound in a space can be quantified in the form of an IR, and these can be analysed in order to extract various parameters quantifying the perceptual qualities of the space. An understanding of all of these areas is crucial for successful soundscape based research. The next chapter will build on some of this chapters content and will cover the phenomenon of environmental noise, its characterisation and measurement, and methods for its abatement and control.

Chapter 3

Environmental Noise

Before talking about noise it is worth trying to understand what the word actually means. It is possible to define noise as *'unwanted sound'*, or sound *'out of place'* [59]. These characterisations of noise are in accord with a strict signal processing definition of noise: interference of a desired signal. Such definitions certainly align with the derivation of the word itself, which has roots in *nausea* and *noxia* [13]. Incidentally, in 2007, Trevor Cox found the sound of vomiting to be amongst the least popular noises in a study of unpleasant sounds [60].

Acoustic noise can therefore exist in a variety of contexts, including communications and music [7]. This thesis is focussed on understanding various aspects of environmental sound, and as such the type of noise most relevant here is environmental noise, where noise is a form of pollution resulting from transport, industry, and recreational activities [12].

As this chapter will show, exposure to environmental noise can have effects that are physical, physiological, and emotional, and that the extent of these effects has deepened as noise pollution has become an increasingly big problem [61]. It is important to note that the effect noise can have is dependent on the life experiences and outlook of an individual. Indeed it has been argued that a certain amount of environmental noise is required to give meaning to a location, and to make it inhabitable in its representation of the presence of human society [8]. A totally 'silent' environment is more unsettling than a noisy one, representing a dangerous environment not fit for habitation [62].

The subjective nature of an individual's experience of noise therefore makes the creation of a comprehensive definition of noise exceptionally difficult. Indeed, without a listener present to experience it, noise cannot take on any real meaning [8]. The introduction of listeners to a noisy environment introduces the fundamental problem of noise definition: what may be considered as noise by one individual is not necessarily noise to another [7]. This inherently subjective element of noise (i.e. the *experience* of noise) is essentially what has necessitated the birth of the field of soundscape research.

This chapter builds on the content covered in Chapter 2 where the fundamental aspects of sound, its behavior in space, and its perception, were covered. This chapter will now consider existing methods for the control and abatement of noise. The effects of noise are then examined, followed by methods for the measurement and quantification of noise. One method for controlling noise, the use of noise barriers, is then considered. This chapter will therefore set up and understanding of the problem of environmental noise which will provide a context for the soundscape theories and methodologies covered in Chapter 4.

3.1 Noise Control

Noise control legislation has been in existence since ancient times. Juvenal, a Roman poet, described the problem of traffic noise in his time [7]:

> *'What sleep is possible in a lodging in Rome? Only those with great wealth can sleep in the city. Here is the reason for the trouble. The movement of four-wheeled wagons through the narrow, winding streets, the clamorous outcries of the cattle drovers when brought to a standstill would be enough to deprive even General Drusus of his sleep.'*
>
> - Juvenal, Satire 3 [63]

As a remedy for this problem, Julius Caesar passed a law in 44 BC prohibiting the movement of wagons in the city during the daytime [8]. The restriction of daytime rather than night-time traffic indicates the balance of power towards the rich and powerful, who would be able to retire to their villas away from the bustle of the city at night-time, so would not want their daytime experience in the city interrupted by noise. Whilst this does not represent hugely successful noise control, it illustrates that noise in urban environments is by no means a modern problem, and that developing a solution depends on distribution of wealth and influence.

With the rapid growth of cities and town following industrialisation, noise levels from industry and city inhabitants have risen significantly over the past few centuries. Perhaps the earliest example of modern noise abatement legislation is a result of Julia Barnett Rice's work in New York. Her antipathy towards the social, and therefore (to her mind) unnecessary, use of tugboat steam whistles resulted in the passing of the Bennett act in 1905, prohibiting such behaviour [7]. She then went on to form the New York Society for the Suppression of Unnecessary noise, which resulted in several successful policies, including control of fireworks as part of independence day celebrations, and the creation of quiet zones around schools and hospitals [8].

New York was also home to a pioneering study as part of Noise Abatement Commission in 1930. This involved a truck equipped with recording equipment driving through the various boroughs of New York, taking around 10,000 observations at 138 different locations [64]. These recordings were used to generate a taxonomy of urban noise, as shown in Figure 3.1.

These pioneering works of noise legislation mark significant breakthroughs in the extension of anti-noise campaigning beyond personal consideration or societal bias, and represents a movement towards a framework for identifying urban noise sources, and distinguishing the difference between necessary and unnecessary noise. These are the principles upon which all modern noise legislature is based, with examples including ISO 1999, ISO 1996, ISO 9613, BS 4142 [24], and the European Noise Directive (END) [65]. The END was adopted in 2000 and identifies a common approach for EU countries to follow in order to assess and manage environmental noise, including sharing relevant information with the public.

Besides explicit legislation other work has been conducted to encourage people to engage with and understand the noise that they themselves create, for example the noise code developed by Keizer allowing people to identify the noise they produce, and actions they can take to reduce it [61]

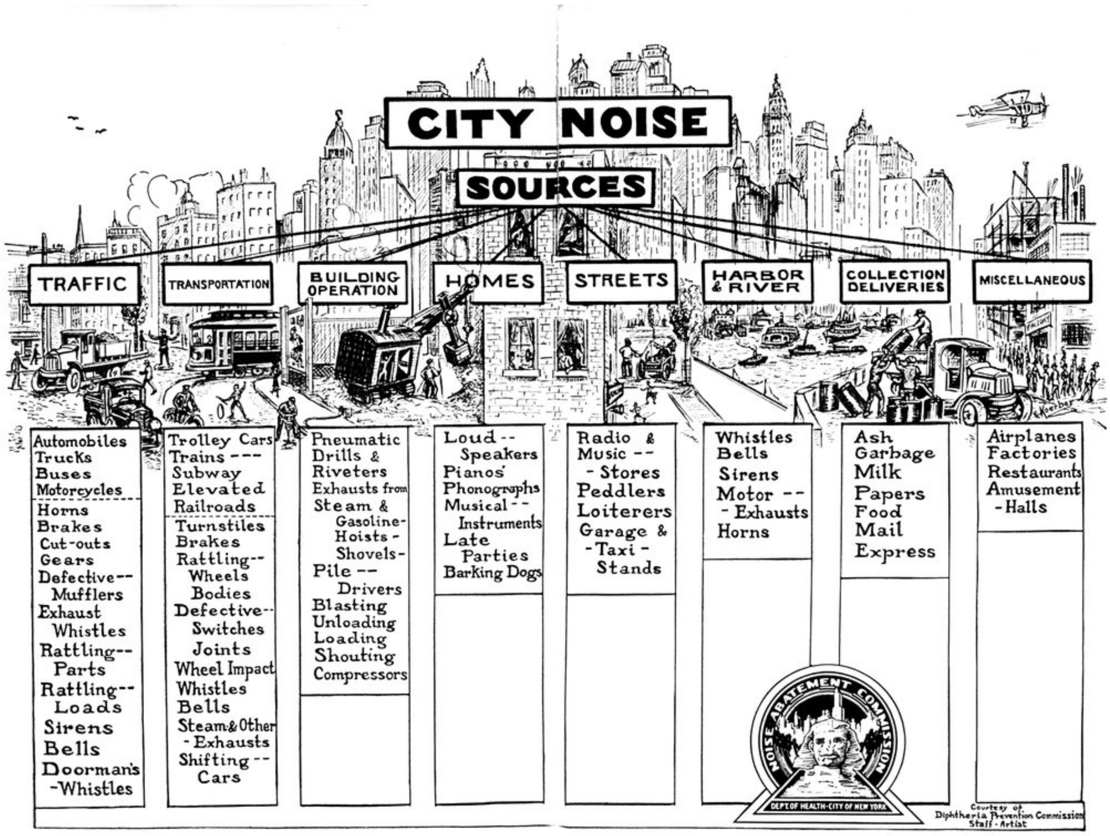


FIGURE 3.1: Urban noise taxonomy from the Noise Abatement Commission 1930 study, from [64].

Whilst there is variation between these directives, there are four typical approaches to environmental noise control: reduction at source, separation (increased source-receiver distance, often achieved by ‘zoning’, for example where industrial areas are located away from cities and residential areas [66]), administrative controls (specifying limits on operational hours, restricting certain activities), and screening [24].

3.2 Effects

Environmental noise can have significant negative effects on human health. These effects can be both auditory and non-auditory, and are a well known problem. For example, in 1909 a project in New York allowed children to make pledges regarding their behaviour, including mitigating noise making around hospitals [67]:

"I offer up this sacrifice so as to comfort the sick near hospitals and any place I know where sick persons are, and to prevent all sorts of noise that are not necessary"

Figure 3.2 indicates some of the effects of noise which will now be discussed in further detail.

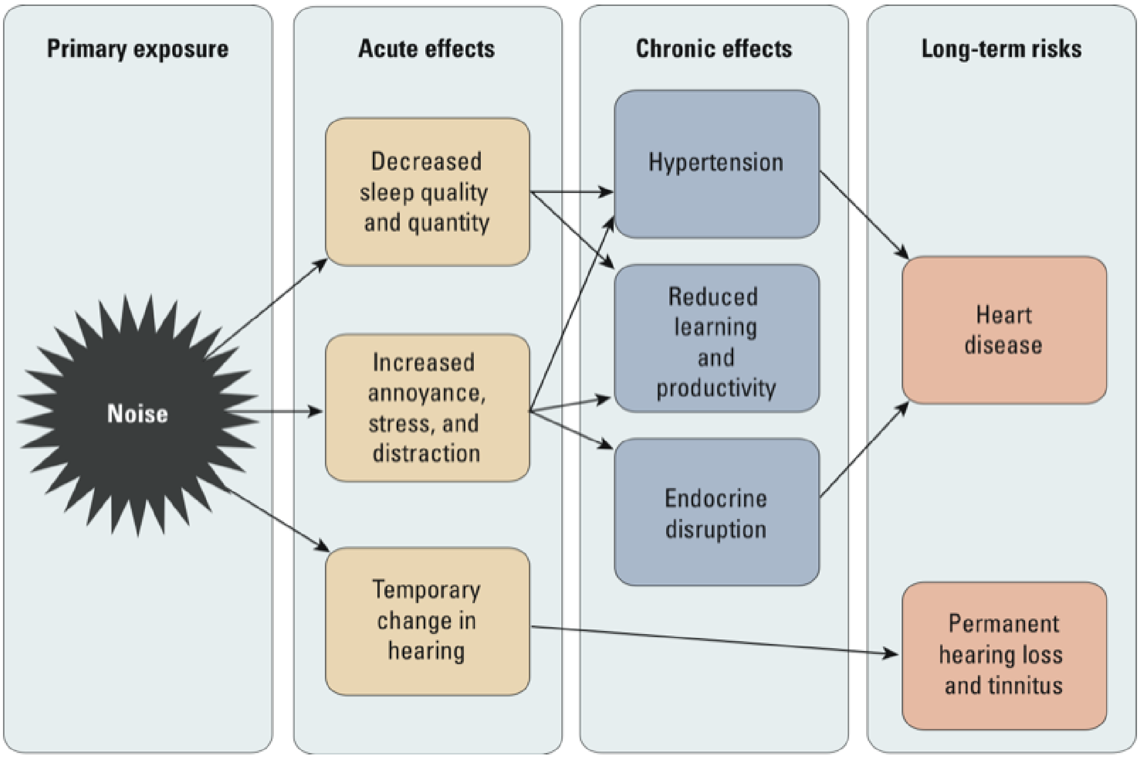


FIGURE 3.2: Select effects of noise, from [68].

.

3.2.1 Auditory Effects of Noise

Whilst environmental noise is less likely to cause hearing damage than noise due to occupational hazards (e.g. the use of heavy machinery) or leisure activities (e.g. attending a loud rock concert) [24], sustained exposure to noise levels can result in permanent impairment to the auditory system. Permanent hearing loss is referred to as permanent threshold shift (PTS), which results from continuous exposure to sounds sources greater than 80 dB(A)[1] [69].

[1]The precise meaning and formulation of decibel measurement such as this will be covered later in this chapter.

Temporary threshold shift (TTS) is the decreased sensitivity of the ear to certain frequencies for a short period of time, with the recovery dictated by the age of the individual, and the nature of the sound source [70]. The frequency range at which PTS and TTS are most likely to occur is around 1kHz - 4kHz (as shown by the ELC plot in Chapter2, the range at which the auditory system is most sensitive [18]. A common symptom experience as a result of hearing impairment is Tinnitus, where sufferers hear sounds with no external cause, typically experienced as a ringing or buzzing [71]. It is a result of symptoms such as these that noise has come to be recognised as a form of pollution which can have severe health effects [72].

3.2.2 Non-Auditory Effects of Noise

The main categories of non-auditory effects of noise, as listed in Stansfield and Matheson's paper in the British Medical Bulletin [72], are:

- Increased annoyance, and resultant social behaviour issues.
- Sleep disturbance.
- Impaired task performance.
- Increased risk of cardiovascular disease.
- Increased risk of psychiatric disorder.

Whilst the degree to which noise has an effect on these symptoms is in some cases debatable, perhaps of most interest from an experimental standpoint is what physical measures might be applied to assess the body's response to the non-auditory effects of noise. Following the stimulation of the auditory system, acoustic information is transmitted via the reticular formation to the Hypothalamus [73]. Figure 3.3 shows the anatomy of the brain and the positions of the Hypothalamus and the reticular formation within it.

Stimulation of the hypothalamus by noise results in certain effects being transmitted through the pituitary gland and onto various organs, as well as to the autonomic nervous system (ANS). The ANS is the 'vegetative' part of the human nervous system, only partially under concious control, that is regulated in part by the hypothalamus. Stimulation of the

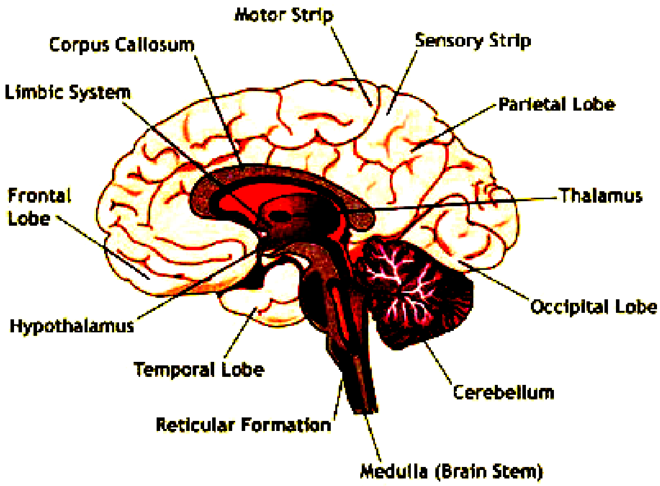


FIGURE 3.3: Diagram of the brain indicating various areas including the positions of the Thalamus and the Reticular Formation, from [74].

.

ANS can have a variety of side effects due to the release of adrenocorticotropic hormone (ACTH), a hormone produced in response to stress. ACTH stimulates the adrenal cortex, resulting in the release of adrenocorticoids which have the following effects (amongst others) [75]:

- Interference with metabolism.
- Decreased resistance to infection.
- Dissolution of lymphoid tissue.
- Reduction in fibroplast production, inhibiting wound healing.
- Elevation of cholesterol levels.
- Vasoconstriction and enlargement of the cardiovascular ventricles.

As these non-auditory, physical, effects of noise are primarily associated with stress, there has been several previous studies attempting to make use of physiological measurement to quantify the stress cause by exposure to noise. This has includes the measurement of galvanic skin response, heart and respiratory rate, saliva cortisol levels and EEG recording [76]–[78].

Whilst Hume et al. found no significant relationships between such physiological measures and subjective evaluations of the pleasantness and stressful nature of sounds [76], there have been multiple other studies indicating the opposite to be true. Bradley and Lang found a relationship between different emotional and arousal content of stimuli and changing heart rate and skin conductance [79]. Gomez et al. [80] found respiratory rate (RR) to become faster with increasingly arousing stimulus. Results from Magnetic Resonance Imaging (MRI) experiments have also indicated increased emotional response to a highly rated (whether positively or negatively) soundscape relative to a neutral one [81]. In Chapter 6 results are presented making use of heart rate measurements as part of a listening test.

Ultimately it is important to acknowledge that noise has a variety of potential negative effects on health, which can be auditory or non-auditory, and which depend on the level and duration of noise exposure. It is a health issue to the point where its effect can be quantified as Disability-Adjusted-Life-Years (DALYs), which quantify the impact of noise [82] in terms of its medical ramifications cost and reduction of life expectancy as well as financial cost. How to correctly calibrate such a measurement is, of course, open to debate, but the very existence of the DALYs metrics indicates the severity of the issue.

3.3 Measurement

Noise measurement quantifies noise sources in terms of SPLs, as introduced in Chapter 2. When evaluating such SPL measurements, it is important to consider if any weighting curve has being applied, and what other parameters are being used. Weightings are used to prioritise different frequency bands in the calculation of an overall SPL measurement. An example of perhaps the most commonly used weighting, known as A-weighting is shown in Figure 3.4.

These weighting curves are used to approximate the relative loudness perceived by the human ear at different frequencies [84], prioritising high frequencies. A-weighting is the most common weighting used, and is specified in many directives including the END [65]. Various other SPL metrics, as defined in the International Standards Organisation document of noise level measurement [85], are shown in Table 3.1.

One problem with such measurements is the fluctuation in definitions of the different periods of the day between countries and seasons, which indicates how the variety of SPL

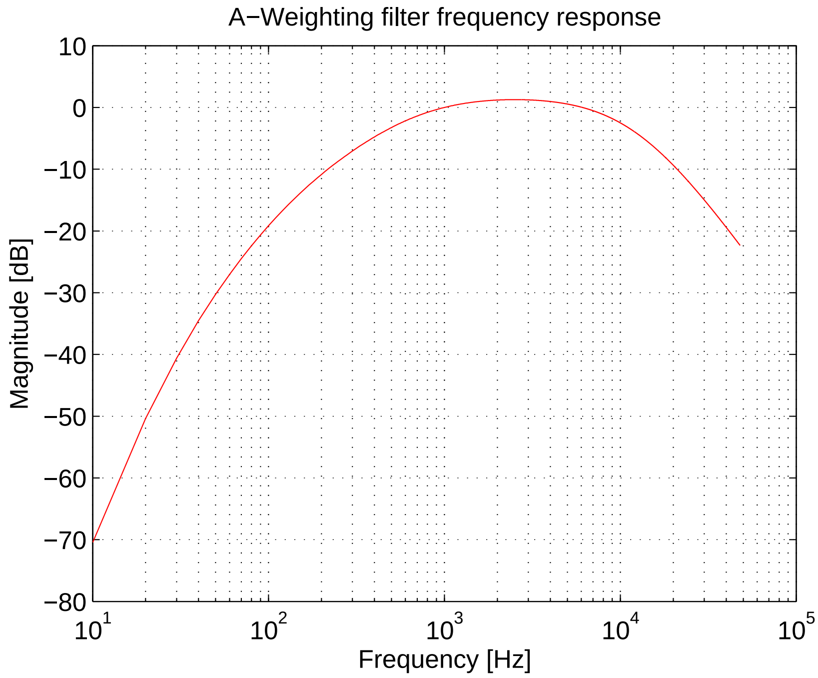


FIGURE 3.4: The A-weighting curve, as defined in [83].

Descriptor	Definition
$L_{Aeq,T}$	A-weighted equivalent continuous sound pressure level over time period T.
L_{Amax}	The maximum A-weighted sound pressure level in a given measurement period. This can be be expressed as either L_{AmaxS}, a 'slow' measurement averaged over 1 second; or L_{AmaxF}, a 'fast' measurement averaged over 0.125 seconds. L_{Amin} makes use of the same data and identifies the minimum A-weighted sound pressure level.
L_x	The noise level exceeded for x% of the measurement period. Common values used are: 90, which can be used as an indicator of background noise; 50, which gives a median value of fluctuating noise levels; and 10, which is sometimes used in traffic noise measurement.
L_{DEN}	This is known as the 'Day Evening Night' level and is used to express the noise levels over a 24 hour period calculated by weighting the noise levels recorded in the day, evening, and night periods. It is given by: $L_{DEN} = \frac{10\log}{24}\left((12 \times 10^{\frac{L_{Day}}{10}} + 4 \times 10^{\frac{L_{Evening}+5}{10}} + 8 \times 10^{\frac{L_{Night}+10}{10}}\right)$ [65].

TABLE 3.1: Some noise level descriptors as given in [85].

measurement methodologies can make the direct comparison between results of different studies difficult.

Other examples of extracting information from SPL measurements include measuring the number of 'emerging peaks' in a recording, where an emerging peak is identified as 5 dBA above the L_{90} value. The rate of these peaks R_{Em} can be calculated by comparing the time of these peaks T_{Em} with total time [86]:

$$R_{Em} = \frac{T_{Em}}{\text{Total Duration}} \tag{3.1}$$

A similar measure of more discrete noise events is the Noise and Number Index (NNI), designed by the Wilson committee to quantify the subjective noisiness of aircraft, and the recurrence of aircraft noise at different times of day [87]. It defines annoyance as proportional to the average peak recorded noise level L and the logarithm of the number of occurrences, N:

$$\text{NNI} = \text{L} + 15 \log \text{N} - 80 \tag{3.2}$$

For example, if the average peaks level is known to be 100 dB, and the total number of flights is recorded as 110, the NNI is given as:

$$\text{NNI} = 100\,\text{dB} + 15 \log(110) - 80 \tag{3.3}$$

$$= 100 + 15(2.04) - 80 \tag{3.4}$$

$$= 100 + 30.62 - 80 \tag{3.5}$$

$$\text{NNI} = 50.62 \tag{3.6}$$

The NNI scale ranges between 0 and 60, with a value above 50 indicating an unreasonable level. The subtracted value of 80 is derived from a social survey which indicated that an average peak noise level of 80 dB corresponds to a noise annoyance factor of zero [88]. The NNI is limited as it provides no information regarding the duration of noise, but was created as an empirical measure following a survey of noise annoyance due to overhead aircraft [89].

A limitation of SPL measurements is that the weighting applied can result in underestimation of the effect of low frequencies, or overestimation of the impact of higher frequencies.

As such, the spectrum gravity centre (centroid) G of an SPL recording can be calculated by:

$$G = \frac{\sum_i \left[10^{\frac{L_i}{10}} \times B_i\right]}{\sum_i \left[10^{\frac{L_i}{10}}\right]} \tag{3.7}$$

where i indicates the frequency bin number, L_i is the SPL measurement at that frequency, and B_i is the frequency value of that bin [24]. This can given an indication of frequency weighting present in the noise source itself, which may influence the choice of weighting curve used in further analysis.

Once a set of noise measurements have been collected, a method is required to present the information in a useful way. One common method for presenting recorded noise levels is sonic cartography, also known noise mapping.

3.3.1 Noise Mapping

Noise mapping is the process whereby a set of noise measurements taken at a number of sampling positions are evaluated, and extrapolated from, to allow noise level contours to be overlaid on a map of the space [90]. An example noise map of York is shown in Figure 3.5.

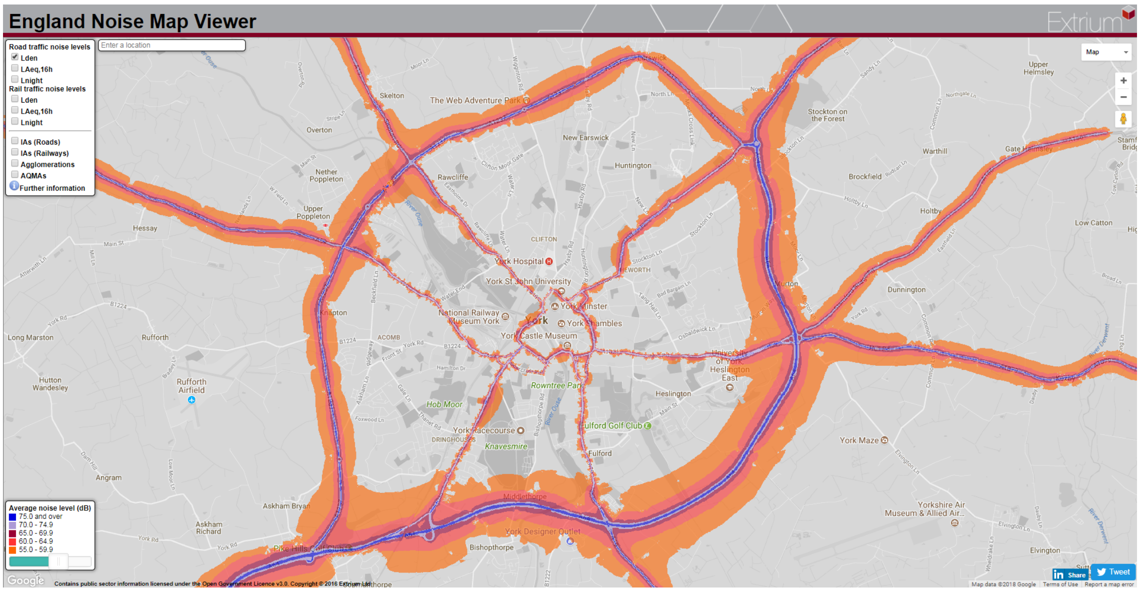


FIGURE 3.5: An example of a noise map of York showing L_{DEN} values, from [91].

Whilst noise mapping allows for the illustration of particularly noisy areas of a particular environment, its usefulness beyond that can be limited. This is partially due to the limitation of SPL measurements generally where indication of loudness does not necessarily give an idea as to what an area sounds like. Indeed such work can be limited in its usefulness in describing real-world scenarios due to the a reliance on computer modelling because of the lack of a sufficient number of *actual* noise measurements [61]. It is also important to consider that these maps often only consider traffic or rail noise and do not necessarily take into account noise made by human beings [92].

One way to mitigate this problem is the inclusion of perceptual and/or subjective information as part of the noise mapping process. WideNoise [93] is an app that allows users to make recordings at certain locations and upload them to a communal map. This noise level measurement is accompanied by subjective evaluation of the noise at that location, rating it in terms of the following categories: love/hate, calm/hectic, alone/social, and natural/man-made. This adds context to the level measurements that could offer greater insight into the the acoustic experience of the location than noise level measurement alone. For example, an area with a higher recorded noise level could have a more pleasant atmosphere, and therefore be a nicer environment than another area with a lower objective level measurement. Figure 3.6 shows an example.

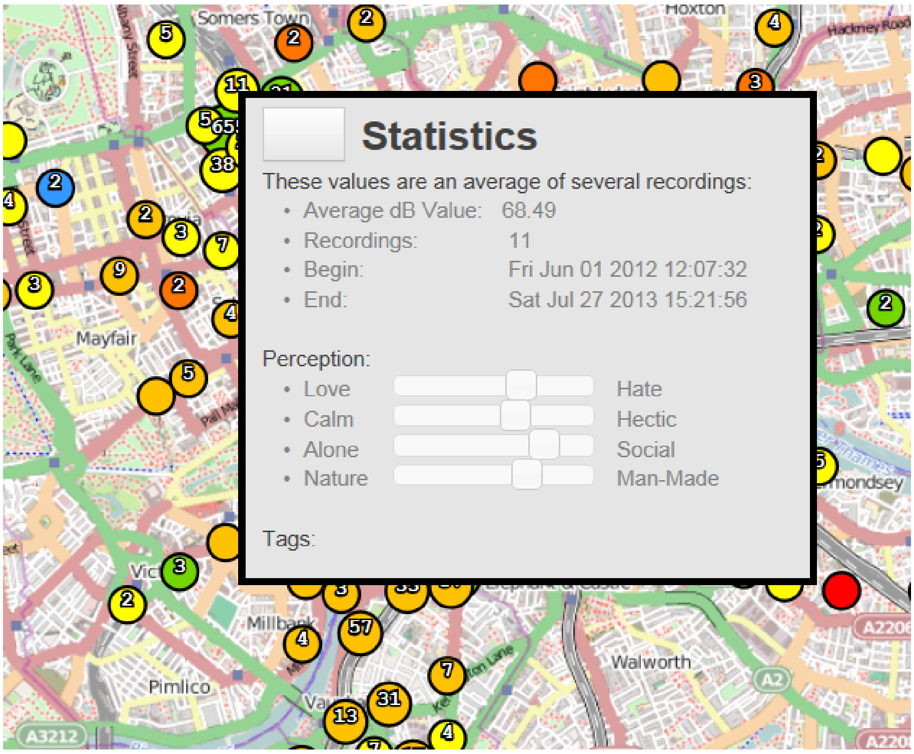


FIGURE 3.6: Screenshot of the WideNoise map, focused on Euston station, from [90].

One problem with WideNoise, and other similar crowd source noise surveys (e.g. the Noise Nuisance app [94]), is the potential for variation in the level recordings, due to differences between (and limitations associated with) the recording hardware used, and the lack of user training. The added subjective information, whilst potentially illuminating, can result in an overloaded and confusing noise map, especially when viewed by non-professionals [95]. Chapter 6 will show how subjective ratings similar to those shown in Figure 3.6 can be used for the evaluation of recorded soundscapes.

3.4 Noise Barriers

One method for reducing the impact of noise sources is the use of noise barriers. Noise barriers are structures that prevent sound transmission, through acoustic reflection and absorption [24]. In order to have any significant effect, a noise barrier is required to interrupt the line of sight between noise source and receiver [96]. In its simplest form, a noise barrier is a rigid wall that efficiently reflects sound. A rigid barrier reacts to various frequencies in different ways, changing the transmission loss (i.e. the reduction in noise level associated with the presence of the barrier) associated with it. A graph of this changing transmission loss with frequency is shown in Figure 3.7.

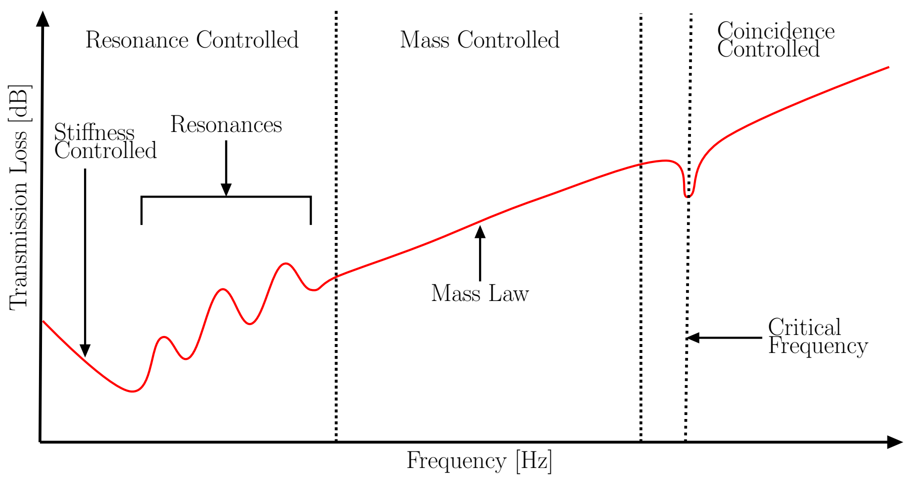


FIGURE 3.7: The frequency response of the transmission loss associated with a rigid barrier, after [97].

At particularly low frequency (in the 'stiffness region') the stiffness of the barrier material dictates the transmission loss [98]. At slightly higher frequencies the resonance of the barrier will dictate the transmission loss. The lowest resonance frequency f_r of such a barrier can be calculated by:

$$f_r = 0.454 h v_b \left[\frac{1}{x^2} + \frac{1}{y^2} \right] \tag{3.8}$$

where the resonant frequency f_r is determined by barrier thickness h, wave velocity in barrier material v_b, and the barrier dimensions x and y [98]. Above the barrier's resonant frequencies, in the 'mass controlled region' there is an increase in transmission loss with frequency at a rate of about 6 dB per decade, followed by a dip in transmission loss due to the coincidence effect where incoming sound waves coincide with diffracted sound waves and interfere constructively. The frequency at which this effect is most prominent is the critical frequency f_c:

$$f_c = \frac{c^2}{1.8 h c_b} \tag{3.9}$$

where c is the speed of sound in air, h is the barrier thickness, and c_b is the speed of sound in the barrier material [99]. Straight edge noise barriers are limited in their usefulness due to the sheer size required, and lack of effectiveness at particularly high and low frequencies [96]. Noise barriers can be designed in various shapes, as shown in Figure 3.8.

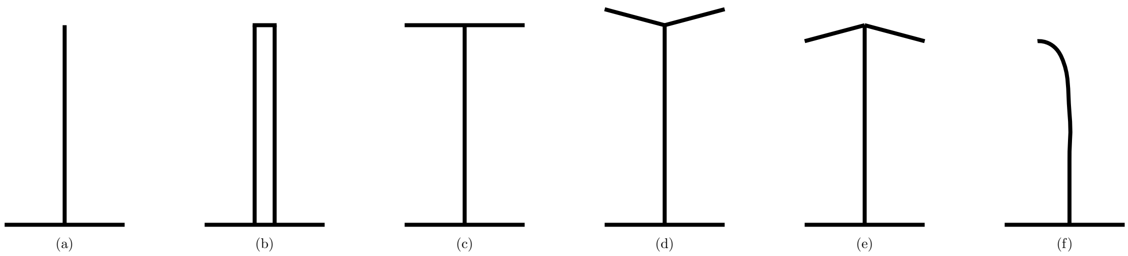


FIGURE 3.8: Examples of noise barrier profiles: (a) Conventional (b) Thick profile (c) T-Profile (d) Y-Profile (e) Arrow-Profile (f) Curved. After [24] and [100].

An alternative to traditional noise barrier design is the Sonic Crystal. The Sonic Crystal was 'discovered' by Martinez-Sala after noticing acoustic effects as a result of a sculpture by Eusebio Sempere [101], which can be seen seen in Figure 3.9.

A Sonic Crystal is an acoustic 'meta-material' that consists of solid elements periodically spaced in air. A wave propagating within the crystalline structure created by this arrangement will be scattered as it interacts with the elements, resulting in phase interference [103]. This behaviour results in a pattern of standing waves in the sonic crystal, giving rise

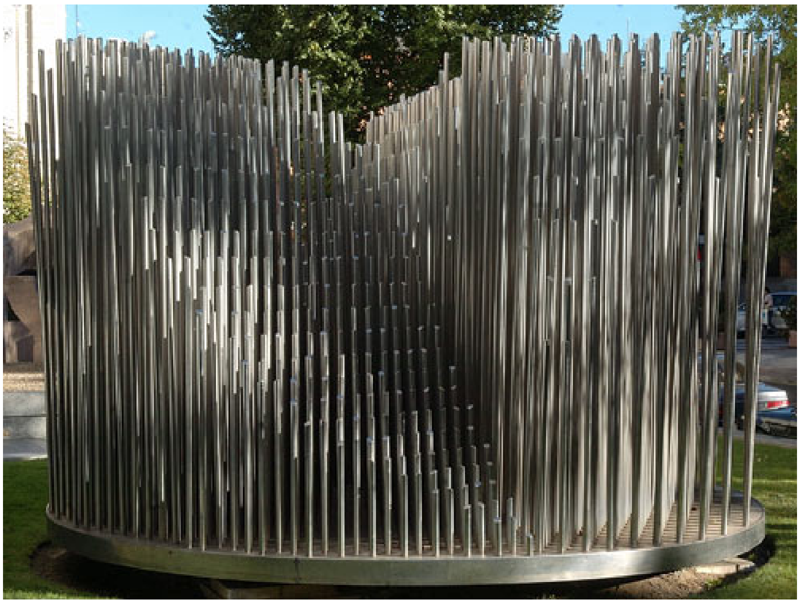


FIGURE 3.9: Eusebio Sempere's *Organo*, an example of a sonic crystal. From [102].

to a transmission band gap [104]. The frequencies at which peaks in scattered intensity occur are described by Bragg's Law:

$$2d \sin\theta = n\lambda \tag{3.10}$$

where n is the reflection order, λ is the wavelength of the incident wave, θ is the scattering angle, and d is the lattice spacing (as shown in Figure 3.10) [105]. This behaviour allows sonic crystals to be used to ameliorate the acoustic properties of their environment, with the added benefit of visual intrigue.

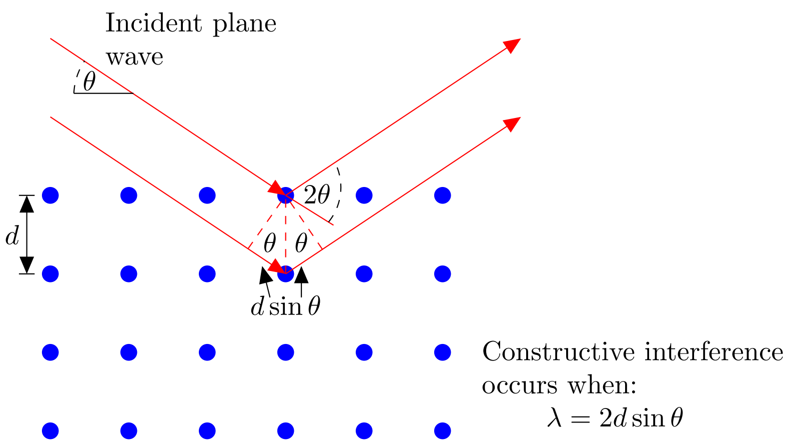


FIGURE 3.10: Bragg's Law in a 2D lattice, after [106].

Plants and vegetation are an alternative material to use for the creation of noise barriers, as they have been found to be effective in dispersing sound energy [107]. This could include the use of foliage for the absorption of high frequencies [108], a band of vegetation acting as a noise barrier [109], or even a set of trees planted in such a way to act similarly to a sonic crystal [110]. A patch of woodland will have noise screening properties defined primarily by three factors: the ground effect, scattering from trunks and branches, and absorption of high frequency energy by foliage (where the scattering can be seen as random, allowing phase information to be ignored) [111], [112]. Chapter 7 goes on to present a novel method for modelling this behaviour as an example of a sparsely reflecting outdoor acoustic scene.

3.5 Summary

This chapter has presented information regarding environmental noise, its definition, effects, and evaluation. As established in this chapter, the current techniques for noise evaluation can be rather limited, giving little indication of the quality of sound. An alternative approach is one based in soundscape research where the focus is on the experience of sound [113]. Such an approach can therefore avoid the dismissal of noise related symptoms extant in individuals where the SPL level is 'too low' to officially warrant changes to be made [114]. As such the next chapter will examine soundscape research, which integrates a user-centric understanding of sound into the consideration of environmental noise.

Chapter 4

Soundscape

Having established in Chapter 3 the definition of noise, its health effects, and methods of measuring it, this chapter will now consider a subjective approach to environmental sound in the form of soundscape research and methodologies. This includes: the fundamentals of soundscape theory and a listener-centric approach to environmental sound, the capture and categorisation of soundscapes, the use of soundscape in art and wider public engagement work, and the evaluation and analysis of soundscape recordings.

4.1 Soundscape Theory

In his seminal text 'The Soundscape: Our Sonic Environment and the tuning of the World', R. Murray Schafer defines a soundscape as [13]:

> *'The sonic environment. Technically, any portion of the sonic environment regarded as a field for study. The term may refer to actual environments, or to abstract constructions such as musical compositions and tape montages, particularly when considered as an environment.'*

Soundscape analysis looks at the holistic experience of all sound in a given location, and aims to explore the perception of, and interaction with, that environment [16]. It is a field that aims to avoid the negativity of a solely noise-based approach, as Schafer has said: *"Let's not concentrate on noises. Let's concentrate on the entire acoustic atmosphere"* [61].

In this way, soundscape analysis describes both the physical and perceptual properties of an environment [115]. This approach exposes soundscape research's position as a convergence of multiple disciplines, including acoustic ecology, musicology, sociology, psychology, architecture, and acoustics [13].

Figure 4.1 gives a representation of the conceptual framework of soundscape analysis. It shows the relationship between objective measures of acoustic and psychoacoustic features, perceptual identification of the sound sources present in the soundscape, and an appreciation of how the sound sources interact with one another to form an information rich, cohesive whole. This conceptual framework was specified as part of the Positive soundscape Project (PSP), an initiative designed to promote the active design of more pleasant sonic environments [81].

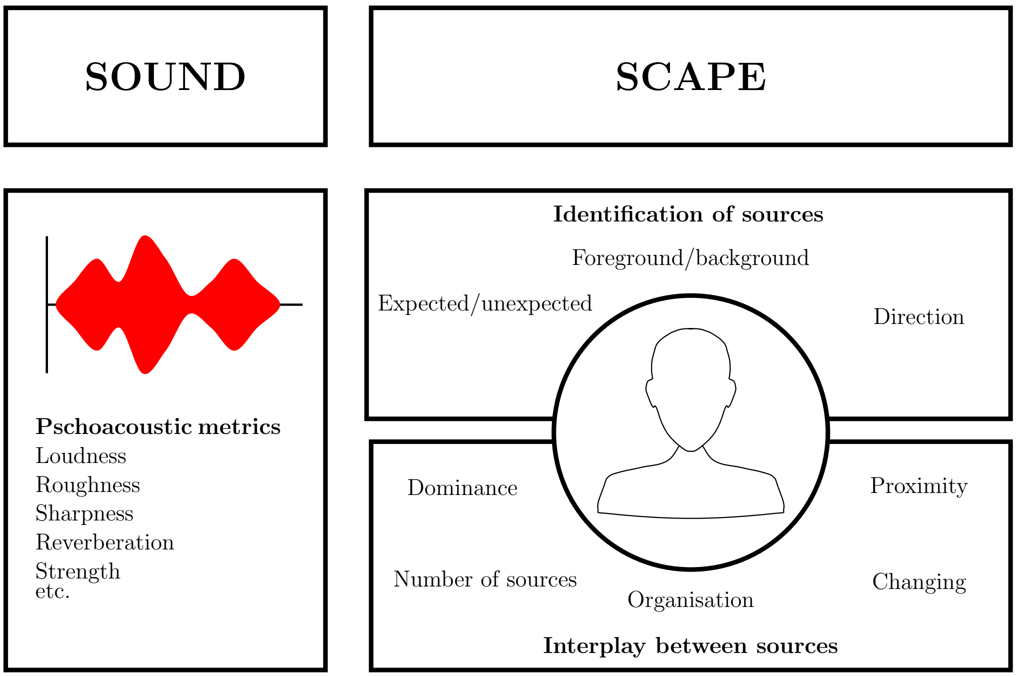


FIGURE 4.1: The conceptual framework of a soundscape, after [81].

The motivation of the PSP is to avoid the typical mindset where the acoustic qualities of an environment are ignored until they become a problem [1]. This reduces soundscape work to 'damage limitation' in the form of noise abatement and control legislation, rather than working to avoid such a situation in the first instance. One way in which the PSP aims to promote this approach is 'ear-cleaning', a concept introduced by Schafer [13] to promote 'clairaudience' where people actively listen to their environment in order to understand what is pleasant about it, what could be improved, and why. This contrasts with the typical response to negative soundscapes, whether the offending noise is either shut out (e.g. by closing windows) or controlled with ‘acoustic perfume’, where a new virtual soundscape acting as an ‘audioanalgesic’ defensive barrier is created with music or other sounds [116]; an approach that ultimately results in increased noise levels.

Given the emphasis on the interaction between sound sources and the information contained by that interaction, soundscapes will vary between locations, and between time periods at the same location. This, twinned with perceptual and emotional aspects, makes the soundscape experience an inherently personal one, defined by individual as well as societal interpretation of the sounds present in a space. A way in which the sounds present in a soundscape can be analysed is by decomposition into three categories of auditory event, as identified by Schafer[1]:

- **Keynotes.** The fundamental underlying sounds that provide situational context, for example the sound of waves at a beach.
- **Signals.** Sounds in the foreground of a soundscape that demand attention (e.g. alarms, sirens, horns) and may allow for the transfer of messages/information,
- **Soundmarks.** Analogous to landmarks, these are sounds that are of particular importance to the inhabitants of a given environment, and perhaps have cultural or historical significance. An example of this would be the clocks chimes of Big Ben in London, or the bells ringing at the York Minster.

These categories can be though of in terms of the visual perception terminology of ‘figure’ and ‘ground’. Broadly speaking, the ‘ground’ is the backdrop that gives context to the ‘figure’; the subject under consideration [117]. Without the ‘ground’, the ‘figure’ loses its outline, and its significance becomes obscured. Keynote sounds can therefore be seen as the ‘ground’. They do not require conscious listening to be understood, and provide a framework in which the ‘signal’ can operate. Accordingly, through active listening keynote sounds can become the ‘figure’, which is analogous to the typical birth of noise problems where previously background noise becomes the focal point of a soundscape [24].

[1] It is important to note that there may be some overlap between these categories, and that the same sound may change category depending on context.

The findings of the PSP, and Lercher and Shulte-Fortkamp's investigation of noise annoyance [81], [116] provide the following approaches for decomposing and understanding soundscapes:

- The balance of level and frequencies in a noise source is important, not just the overall level. This was seen in the consideration of available noise level measurement techniques in Chapter 3.
- The daily rhythm and the meaning of sound present in the soundscape must be taken into account. It must be understood that simple reduction in noise level is not the purpose of soundscape analysis and design, the soundmarks of the location should be preserved and the aim should be to enhance the pleasant characteristics already existing as part of the space.
- Sonological competence must be learned and incorporated into environmental design. For the successful design of acoustic environments, architects and other designers must consider the acoustic features of their designs from the first instance.
- If people perceive their environment consciously they will have a chance to change it. Active listening is required as part of general participation, ideally taking place prior to the existence of a noise level problem.

4.1.1 Environmental Psychology

In order to evaluate preferences for particular soundscapes, an understanding of the psychological basis of a human being's relationship with their environment provides a useful framework. A central question to environmental psychology is whether landscape preference, the degree to which a particular environment is liked, is innate or represents learned behaviour [62]. Whilst landscape preferences are likely to be influenced by both factors, a popular theory in support of innate behaviour is Wilson's Biophilia hypothesis. This states that humans have an *'innate affinity for life and lifelike processes'* [118], and explains in part why people seek out nature, wildlife, and countryside for stress relief, self-reflection, and convalescence.

It has also been hypothesised that in searching for an inhabitable environment our ancestors often sought out savannah-like locations (large, open spaces with considerable presence of vegetation and good visibility in multiple directions) [119]. The presence of vegetation is also supposed to have an auditory advantage. Thayer formulated a theory that, since for our pre-historic ancestors a lapse in concentration could be fatal, they would have sought environments that would allow for respite from the constant state of arousal that still provided sensory information regarding potential danger. Wide open grassland areas then give both visual information (ability to see a considerable distance, movement of other prey as a bio-indicator of danger), and aural clues from the soundscape environment (e.g. the sound of a predator disturbing the surrounding grass) [120].

A similar interpretation of this affinity with open, natural landscapes is given by Appleton's prospect-refuge theory, where an environment is require to allow the inhabitants to scout out new opportunities (prospect) without being detected in return (refuge) [121]. Landscape preferences can be explained further by humanity's desire to explore and understand. Kaplan and Kaplan's preference matrix (shown in figure 4.2) identifies four landscape characteristics as defined by the matrix dimensions of informational needs (Understanding-Exploration), and level of interpretation (Immediate-Inferred).

	Informational needs	
Level of interpretation	Understanding	Exploration
Immediate	Coherence	Complexity
Inferred	Legibility	Mystery

FIGURE 4.2: Kaplan and Kaplan's preference matrix, from [122].

An environment exhibiting coherence is one where the relationship between the elements contained within it are immediately understood. Complexity indicates an environment rich with immediately comprehensible visual information. Legibility represent an environment that gives an indication of what lies ahead. An environment exhibiting mystery is one that promises new opportunities and experiences. In their research, Kaplan and Kaplan found 'mystery' to be the strongest indicator of environmental preference, indicating an ancestral desire for stimulation and new experiences beyond survival [122]. This idea of mystery has also been expressed as *genius loci* or the 'spirit of the place', where preference is greatly influenced by visually striking features unique to particular environments [123]. The equivalent of *genius loci* in soundscape analysis is, then, the presence of soundmarks: sounds that are particular to a certain location and of importance to inhabitants and visitors.

The sound of running water is another natural sound that can influence environmental preference. Schafer theorised that the sound of water is reminiscent of time spent in the womb, and so is innately comforting and represents the first sound we 'hear' [13][2]. Perhaps more convincing is the absolute dependence on water for survival, the discovery of which would likely be dependent on following aural clues to find the source of water [124]. The above theories can be summed together and considered as an 'ecological aesthetic' where a knowledge of the ecological function of a particular landscape will lead landscape preferences. To our ancestors this knowledge would be the difference between life and death, resulting in an innate sensibility in modern times as a result of millennia of evolutionary bias.

4.1.2 Soundscape Categorisation

In order to classify the content and nature of different soundscapes, it can first be useful to define categories of environmental sound to which different sound sources belong. In much of the literature [24], [125]–[129] three main groups of sounds are identified:

- **Natural:** These include animal sounds (such as bird song), and other environmental sounds such as wind, rustling leaves, and flowing water.
- **Human:** Any sounds that are representative of human presence/activity that do not also represent mechanical activity. Such sounds include footsteps, speech, coughing, and laughter.
- **Mechanical:** Sounds such as traffic noise, industrial and construction sounds, and aeroplane noise.

These categories are congruent with the taxonomies of environmental sound as defined by both Schafer [13] and Krause [130]. Figure 4.3 shows these categories of environmental sound, extended here to make a distinction between different types of human made sound (i.e. distinguishing between industrial sounds and 'natural' human sounds). These categories can then be used as a guide when planning to record a set of soundscapes, particularly when the aim is to capture as wide a range of environmental sound sources as

[2] At this point, it is worth noting that Schafer's observations are occasionally muddled by a rather new-age mysticism, or a technophobic, agrarian, mentality.

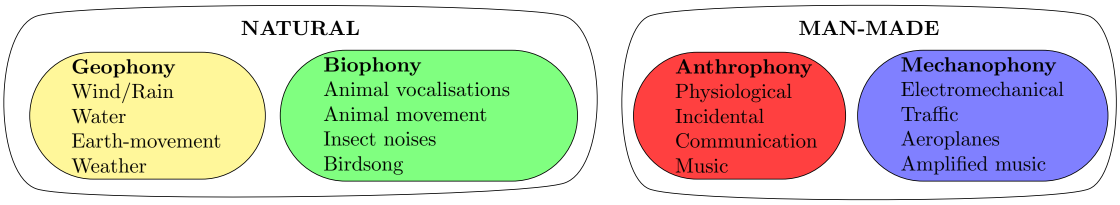


FIGURE 4.3: A framework for soundscape categorisation, combining elements of Krause [130] and Schafer's [13] work.

possible (as will be seen in Chapter 5). The use of category rating scales as a soundscape evaluation tool, and how this relates to emotional state, will also be examined in Chapter 6.

4.2 Soundscape Evaluation and Perception

The perception of auditory information is an exceedingly complex process. On a fundamental level it involves the interaction of two neural systems: primal system processing, assessing the immediate content of a sound for meaning; and pattern analysis, where the incoming auditory information is referenced against previous experiences to give it a contextual meaning [131].

This explanation of auditory perception has a clear evolutionary basis in allowing humans to detect danger, invoking a fight or flight response [62]. When considering the perception of the soundscape of an environment, while the basic processes involved ostensibly remain the same, the relationship becomes much more complicated. This relationship can be expressed by the Filter Model, which is shown in Figure 4.4.

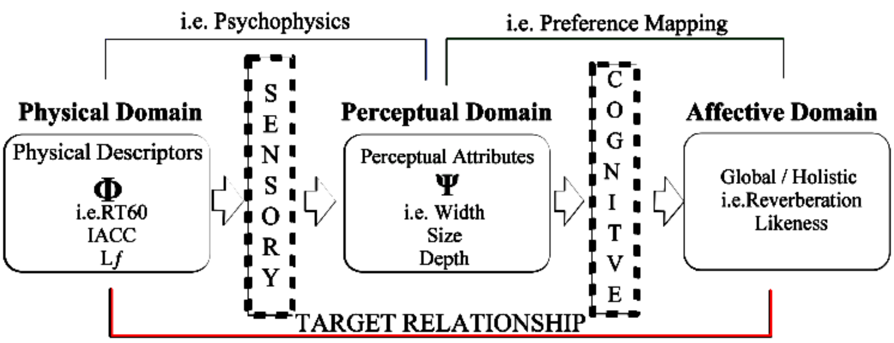


FIGURE 4.4: A diagram of the Filter Model, indicating the relationships between the physical, perceptual, and affective domains defined by sensory and cognitive filters, from [132].

The filter model indicates that the physical stimulus from an acoustic event is filtered by our senses, becoming a perceptual event. This percept is then passed through the cognitive filter (non-sensory interpretation of the event where pattern analysis relates it to expectation, memory etc.) resulting in an affective response [133]. The filter model identifies two methods required to map physical acoustic properties to affective response. These are: the creation of psychoacoustic and perceptual models, to link physical and perceptual attributes; and preference mapping, allowing affective responses to be linked to particular percepts. Whilst the relationship expressed in Figure 4.4 was developed within the context of room acoustics, it reflects the mechanism by which acoustic waves are translated into psychoacoustic features by the hearing system, and ultimately these features form an holistic subjective experience of that stimulus.

This chapter will move on to look at soundscape analysis methods: how they relate to environmental psychology principles, and their relation to the filter model. This is followed by an examination of some auditory percepts and their associated objective parameters, accompanied by investigation of available methods for the subjective analysis of these percepts (i.e. methods allowing for the preference mapping of percepts to affective responses). Finally cross-modal perception is explained, considering how concurrent visual stimulus can change auditory perception and vice versa.

4.3 Soundscape Analysis

4.4 Percepts

The sum psychoacoustic effect of a location is formed of a complex relationship between various physical acoustic attributes. Figure 4.5 shows an example diagram indicating how some of these attributes relate to one another. In this diagram the large circles represent primary attributes, small circles represent possible subcategories, and the connecting lines indicate the links between the attributes [132]. Table 4.1 contains a similar set of parameters, here paired with their equivalent objective measurements.

Figure 4.5 and Table 4.1 are useful in describing the relationship between percepts associated with spatial acoustics, which has relevance to how differently designed environments will

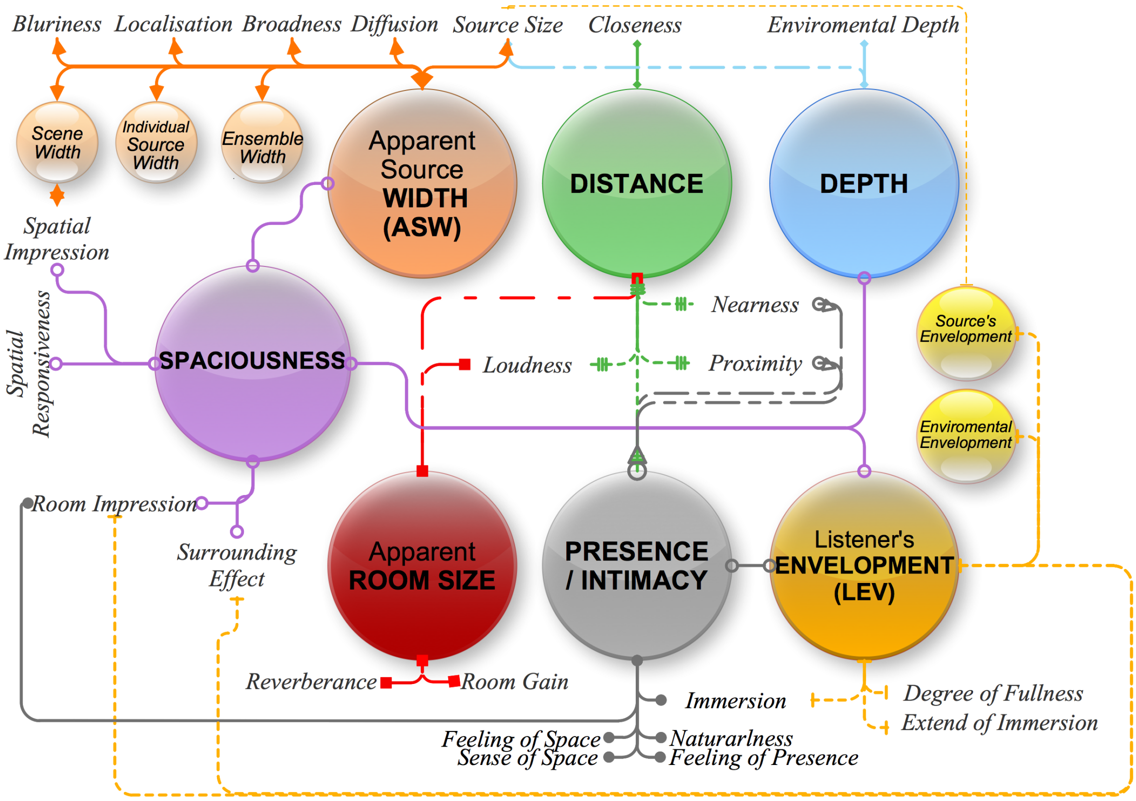


FIGURE 4.5: Diagram indicating the relationship between perceptual attributes associated with reverberation in the context of room acoustics. From [132].

Subjective Parameter	Objective Parameter
Liveness	Reverberation Time (RT60) Early Decay Time (EDT)
Warmth	Bass Ratio (BR)
Brilliance	Treble Ratio (TR)
Intimacy	Initial Time Delay Gap (ITDG)
Clarity	Clarity (C50, C80) Definition (D50, D80)
Strength	Sound Strength (G)
Spatial Impression	Inter Aural Cross Correlation (IACC)

TABLE 4.1: Table of some subjective acoustic parameters and their associated objective measurements, after [24] and [134]. Some of these were covered in Chapter 2

change perception of noise by altering the propagation of sound. It is important to note that these parameters were explicitly developed for use when discussing indoor acoustic spaces, although they can still be of use when considering environmental sound and indicate how objective measures and parameters can be considered alongside a subjective evaluation in a unified soundscape approach. Other non-spatial factors of importance are the volume and duration of the noise source (as measured by the various dB SPL measurements outlined in Chapter 3, as well as the spectral content of the noise source). Three further parameters to be used in evaluating a noise source are loudness, sharpness, and roughness [135].

Loudness and sharpness are easily understood as relating to the level of a noise source and its relative high frequency content respectively. Roughness is a slightly more complex parameter that measures the amplitude modulation of a sound source, where the modulation occurs at a rapid rate (between 15 and 300 Hz) [136]. The unit of measurement for roughness is the Asper, where one Asper is defined as the roughness associated with a 1 kHz tone at 60 dB which is 100% modulated at 70 Hz [137]. This measurement has been used in the analysis of the annoyance of wind-farm noise [138], and is representative of how spatial characteristics are perhaps less crucial to understanding the affective properties of acoustic noise than the spectral and temporal properties are.

4.5 Subjective Analysis

The earliest modern investigation into the emotional aspects of auditory experience came in 1921 when Edison aimed to explore the emotional effects of his famous tone tests. A psychologist from Carnegie Institute of Technology was employed to develop a Mood Change Chart for use at so called 'mood change parties'. The aim of these events was billed as the *'Analysis of Mental Reactions to Music, as Re-Created by the New Edison, the Phonograph with a Soul'*. The participants' reactions were gauged by having them evaluate their emotions against a set of dichotomous descriptors, for example *'serious or gay'*, or *'depressed or exhilarated'*. [131].

Elements of this work remain in use today in the subjective analysis of auditory scenes, where semantic differential pairs are commonly used to evaluate various dimensions of a soundscape [24], [81], [139]. Table 4.2 contains an example set of semantic differential pairs, as used by Kang and Zhang in their study of the soundscape of public spaces [140].

Semantic Differential Pairs			
Agitating	Calming	Comfort	Discomfort
Directional	Everywhere	Echoed	Deadly
Far	Close	Fast	Slow
Gentle	Harsh	Hard	Soft
Interesting	Boring	Like	Dislike
Meaningful	Meaningless	Natural	Artificial
Pleasant	Unpleasant	Quiet	Noisy
Rough	Smooth	Sharp	Flat
Social	Unsocial	Varied	Simple
Beautiful	Ugly	Bright	Dark
Friendly	Unfriendly	Happy	Sad
High	Low	Impure	Pure
Light	Heavy	Safe	Unsafe
Steady	Unsteady	Strong	Week

Table 4.2: Kang's semantic descriptor pairs, from [140].

Semantic descriptors can be used in the measurement of responses in the perceptual domain for the creation of a cognitive filters, linking the perceptual domain to affective domain 4.4. Table 4.3 shows how semantic descriptors can be related to perceptual acoustic features.

Perceptual feature	Attributes of Scales
Strength	Quiet - Loud
Spatial Occupancy	Little Attending - Very Attending
Spatial Arrangement	Organised - Disorganised
Spatial Localisation	Nearby - Far
Temporal Balance	Steady - Unsteady
Time Evaluation	Established - Evolutive
Clarity	Hubbub - Distinct
Activity	Monotonous - Varied
Assessment	Pleasant - Unpleasant

Table 4.3: Some perceptual acoustic features and their associated differential pair descriptors, from [86].

One issue with semantic descriptors (as exhibited by Figure 4.5) is the complexity of the relationship between different terms, which makes them inherently open to interpretation and potentially interchangeable. This becomes especially true when considering experimentation involving non-expert participants of multiple nationalities with different native languages. To remedy this problem in terms of the psychoacoustic perceptual descriptors used, the Spatial Audio Quality Inventory (SAQI) was recently developed [141] in an attempt to create a definitive set of descriptors that would be relevant for multiple languages.

In an attempt to streamline the relationship between various semantic differential pairs, Bradley and Lang were able to extract three major factors: Valence, Arousal, and Dominance [142]. These are the analogous to the three factors identified by Wundt to organise affective responses to any given stimuli: *lust* (pleasure), *spannung* (tension), and *beruhigung* (inhibiton) [143]. These findings allowed them to develop the Self-Assessment Manikin (SAM), a pictoral representation of the emotional response to stimuli. An example of the SAM indicating these three factors can be seen in Figure 4.6.

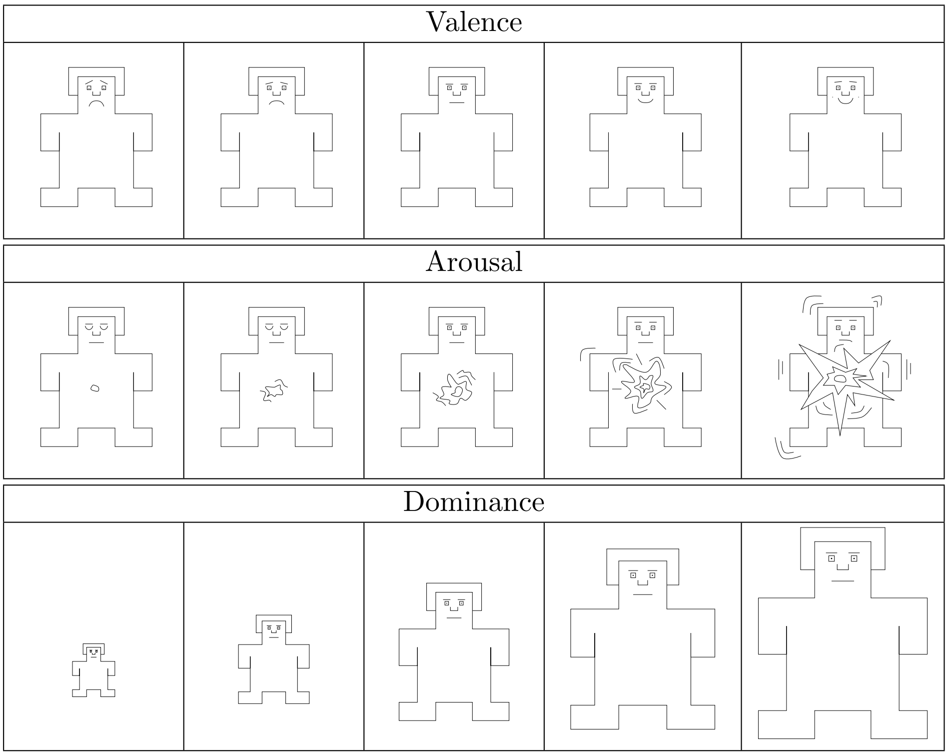


FIGURE 4.6: The Self-Assessment Manikin (SAM), after [142].

The SAM was initially developed using pictures taken from the International Affective Picture System (IAPS, a set of pictures developed especially for research into emotional responses [144]) as the stimuli, and Bradley and Lang went on to consider the SAM for acoustic stimuli (using the aural equivalent of the IAPS, the International Affective Digitised Sounds library, or IADS [145]). In this study they found the same three factors (valence, arousal, and dominance) to define emotional reaction, but found dominance to be a less important factor than valence or arousal [79]. The benefits of the SAM are that it is intuitive to use and provides relatively clear preference mapping between response to stimuli and resultant emotional state. It is important to note that Bradley and Lang's work used recordings of particular sound sources rather than soundscape recordings.

Similar to Bradley and Lang's conclusions, typical studies of soundscape perception focus on rating either 'annoyance' [126], or 'tranquillity' [124]. Annoyance in this context is somewhat analogous to valence rating, as tranquillity is to arousal (with inverted measurement scales in both cases).

In a similar study Cain, Jennings, and Poxon [146] identified two independent factors for soundscape analysis: 'calmness', and 'vibrancy'. Another pair of scaled descriptors found by Viollon and Lavandier echoing Bradley and Lang's SAM related findings were 'unpleasant/pleasant' and 'stressful/relaxing' [139]. The ratings can be combined with more general questions (e.g. pairing the question 'how tranquil was this environment?' with 'which soundscape did you prefer for what reason?') to explicitly relate semantic ratings of affective responses to the overall emotional effect of a soundscape [147].

The Valence and Arousal dimensions of the SAM can be rendered as a two-dimensional plot in what is known as the Circumplex Model of Affect (CMoA), a model introduced by Russell in 1980 to relate valence and arousal scores to distinct emotional states [148]. Figure 4.7 contains an example plot of the CMoA, showing valence as the x-axis and arousal as the y-axis. Also indicated are example emotional states on this model. The top-left quadrant (low valence, high arousal) is where one might expect SAM results for very noisy, unpleasant, soundscapes to lie. For calming, natural soundscapes anticipated SAM results would inhabit the bottom-right quadrant (high valence, low arousal) [149].

A slightly different interpretation of these dimensions has also been used to determine four types of soundscape, as shown in Figure 4.8. Here the arousal and valence scales are used to define four types of environments, each evoking a different emotional response.

Another method for allowing participants to express their feeling regarding a soundscape is Lynch's idea of urban cognitive mind maps [153], where participants are invited to draw a representation of their experience. Whilst the use of such a technique might not necessarily produce entirely, immediately, useful information (due to the lack of a strict framework in which a subject's responses can be evaluated), it does allow for intuitive

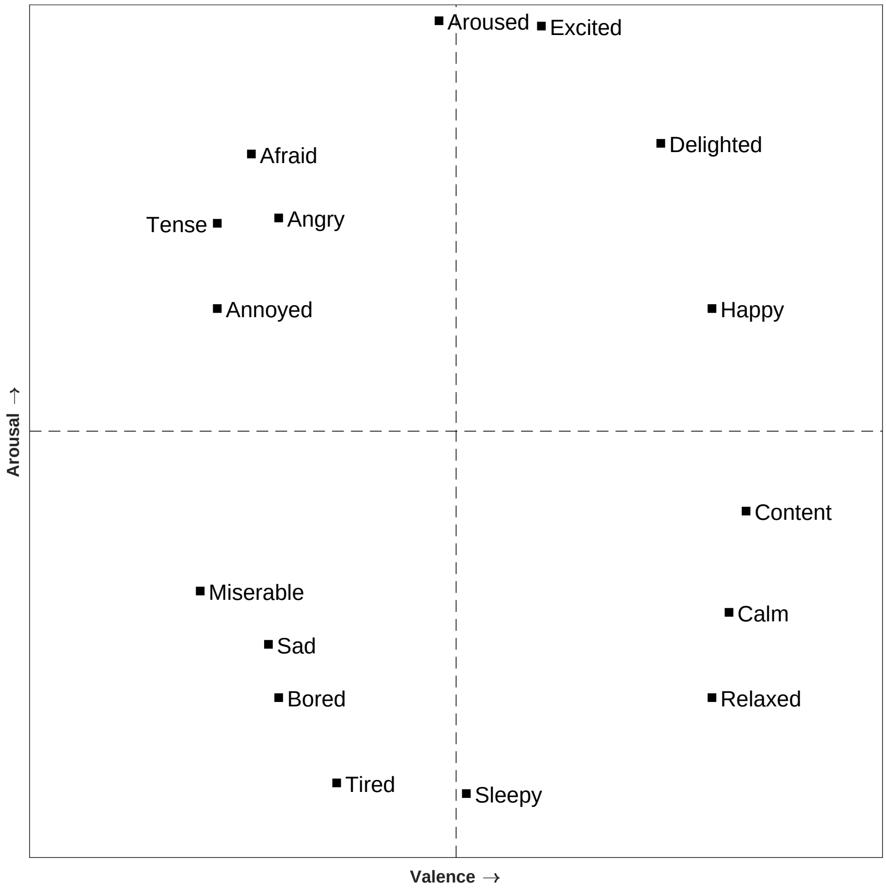


FIGURE 4.7: The circumplex model of affect, after [150].

spatial conceptualisation of their experience, which can be difficult to describe in words. An example cognitive mind map is shown in Figure 4.9.

The cognitive mindmap shows a central arrow representing the subject's perspective and movement through the scene. Shown to the left are stationary and moving vehicles, with more traffic and a construction site (complete with excavator and drill) to the right. Of note are the fumes emitting from the tailpipes of the cars, and the lines emanating from them that represents loud noise. Similar lines have been drawn next to the drill on the right, as well as two exclamation marks. these features indicate how this method can be used to identify the perceptually dominant sound sources in an acoustic scene, despite the potential difficulty in extracting data for comparison with other cognitive mindmaps.

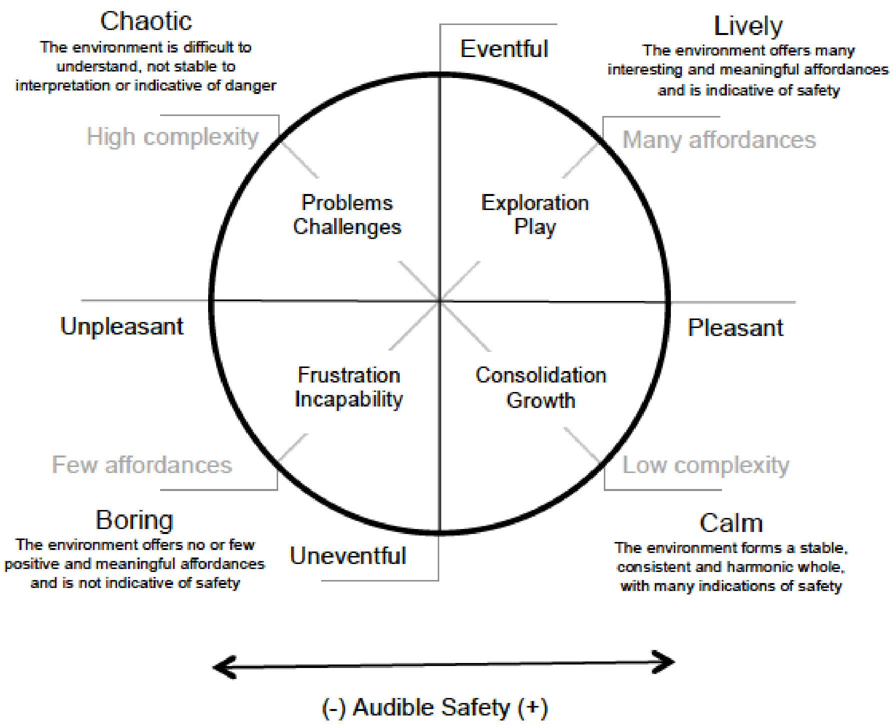


FIGURE 4.8: The CMoA and four types of soundscape from [151], adapted from [152].

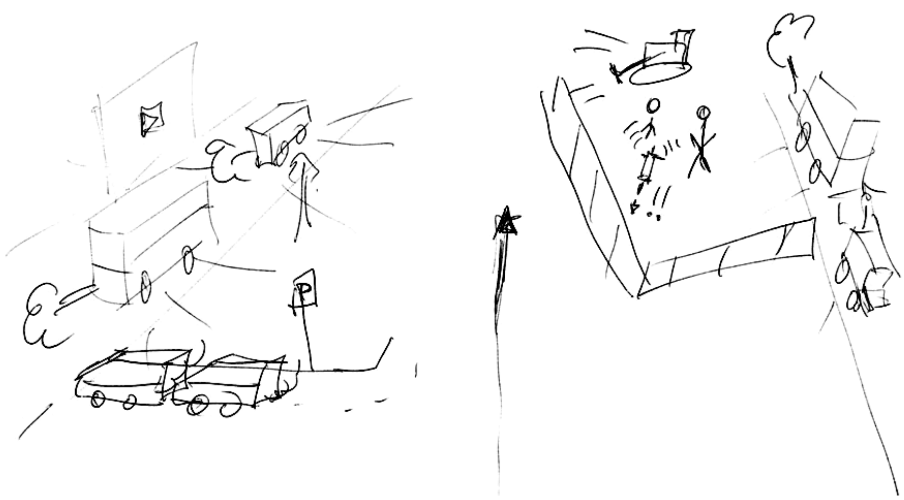


FIGURE 4.9: An example of an urban cognitive mindmap, from [153].

4.6 Cross-Modal Perception

As part of his tone test concerts, Edison used an on-stage performer alongside a recording to convince the audience that the recording system could be used to perfectly replicate real life sound [131]. This is an example of how visual elements can be used to affect auditory perception, in this case achieved by setting up certain expectations that will bias the coming experience. More generally though, there is a complex interaction between sensing modalities (such as vision and hearing) that will change how the qualities of a given environment are perceived.

An example of how this cross-modal perceptual influence relates to environmental noise exploits the spectral similarities between a busy road and crashing waves, using the presence of different visual stimuli to change the perception of the same sound [154]. The frequency spectra of the audio samples and example of the visual stimuli used in this study can be seen in Figure 4.10.

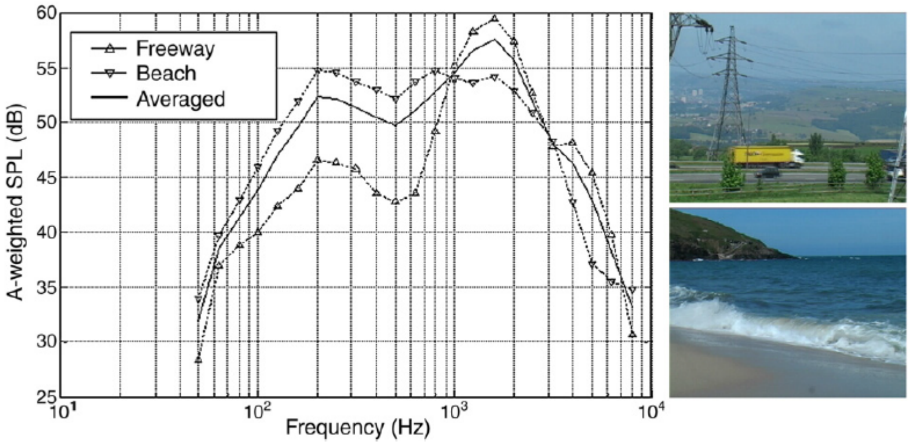


FIGURE 4.10: Frequency plots for traffic recording, beach recording, and averaged spectrum. Also indicated are the visual stimuli presented alongside the auditory information. From [154].

Similar studies have shown various other intriguing results, including finding that the presence of any image at all typically reduces perceived loudness [155] relative to auditory stimulus alone, and that red trains are perceived as louder than green or blue ones [156] (indicated by Cox as most likely due to a greater difference between visual elements comprising 'figure' and 'ground' [60]).

Study of how audio-visual interaction affects perception of environmental noise has included investigating feelings towards wind-turbines [22] (leading to the development of the metric 'vertical visual angle'). Banjun, Zhang, and Gouqing examined how the presence of a barrier changed reactions to traffic noise [157], and found that visual obstruction reduced the perceived annoyance independent of any significant change in the auditory information received. Living on a 'pretty' street has also been shown to reduce noise annoyance [116], similar to where a patient's recovery following an operation is typically faster if they have access to a window with a pleasant view [158].

Alongside theses studies, typically investigating specific audio-visual elements, there has also been investigation of how the more general visual aesthetic of an environment affects auditory perception. Viollon, Lavandier, and Drake compared user responses to 8 different auditory stimuli given the presence of 5 visual environments representing differing degrees of urbanisation [126]. Increased urbanisation was shown to negatively impact the subjective evaluation of sound natural sounds, and to increase the aggravating effect of mechanical noise, but to have very little effect on human sounds.

This has been speculated as being due to the degree of matching between audio and visual stimuli, or due to the orienting nature of the human sounds [159]. These results suggest that an understanding that integrates visual design into soundscape analysis is crucial to the design of better environments [24], indicating the value of investigating cross-modal perception following the synchronous presentation of auditory and visual stimuli.

4.7 Demography

Whilst generally speaking soundscape studies have found various demographic factors to have no significant effect on analysis results [24], [139], there are some factors that are still worth bearing in mind. For example Pedersen and Larsman found the noise annoyance experience due to wind turbines to differ between areas where the landscape is typically flat and more mountainous areas [22]. This was reflected in their measurement of the vertical visual angle (VVA) which describes the angle of inclination between a viewer's position and the top of the wind turbine, which is typically higher in a flatter environment.

This is one example of how environmental factors can have an effect on soundscape experience. Another example is the difference in response between climates, where the tolerance for outdoor noise is generally lower in colder countries than in more temperate climates where a greater amount of time is spent outside [13]. Also, whilst age difference has not been shown to be of significant importance in a great number of studies, Yang and Kang found that older test participants were less tolerant of higher noise levels than younger ones [160].

It is worth mentioning that other demographic factors may not be seen to have such a great effect due to the nature of the studies being conducted. Experiments investigating soundscape evaluation might typically be seen to draw from a relatively limited demographic group of researchers and academics in the same, or similar fields [24]. Combine this with the sheer difficulty of conducting larger scale studies (and the possibility of receiving incomplete demographic data as part of such a study [161]), and it can be seen that the lack of comprehensive data could potentially hide the effect of some demographic factors.

For example, various studies have shown that certain personality traits can have a significant effect on noise annoyance evaluation which has implications for a change in soundscape experience. Some attempts have been made to define noise sensitivity as a personality trait in itself, evaluated through the use of profiling methods including the noise sensitivity questionnaire [162] and Weinstein's noise sensitivity scale [75] which can be seen in Table 4.4.

Other studies have shown links between personality traits and noise annoyance. For example in the development of a noise sensitivity scale Bregman found a relationship between the introversion-extroversion scale and noise annoyance [163] where individuals showing introverted/neurotic tendencies were shown to be increasingly susceptible to noise annoyance. Similarly Verona et al. found the presence of psychopathic traits in criminals to result in deviant emotional response to affective aural stimuli [164].

Moreira and Bryan further explored the relation between personality traits and noise sensitivity through the use of the Rorschach test. Various metrics extracted from conducted Rorschach tests, such as determinant (M), Human content (H), total response, original response, and succession were compared to the same participant's noise sensitivity results. High H scores (a possible indicator of empathic skills) in particular were shown to link

Items on the Noise Sensitivity Scale	
1.	I wouldn't mind living on a noisy street if the apartment I had was nice.
2.	I am more aware of noise than I used to be.
3.	No one should mind much if someone turns up their stereo full blast once in a while.
4.	At movies, whispering and crinkling candy wrappers disturb me.
5.	I am easily awakened by noise.
6.	If it's noisy where I'm studying, I try to close the door or window or move someplace else.
7.	I get annoyed when my neighbours are noisy.
8.	I get used to most noises without much difficulty.
9.	How much would it matter to you if an apartment you were interested in renting was located across from a fire station?
10.	Sometimes noises get on my nerves and get me irritated.
11.	Even music I normally like will bother me if I'm trying to concentrate.
12.	It wouldn't bother me to hear the sounds of everyday living from neighbours.
13.	When I want to be alone, it disturbs me to hear outside noises.
14.	I'm good at concentrating no matter what is going on round me.
15.	In a library, I don't mind if people carry on a conversation if they do it quietly.
16.	There are often times when I want complete silence.
17.	Motorcycles ought to be required to have bigger mufflers.
18.	I find it hard to relax in a place that's noisy.
19.	I get mad at people who make noise that keeps me from falling asleep or getting work done.
20.	I wouldn't mind living in an apartment with thin walls.
21.	I am sensitive to noise.

TABLE 4.4: Weinstein's 21 point noise sensitivity scale, from [75].

with increased noise sensitivity. This lead to the conclusion that individuals particularly susceptible to noise may have personality types where they are sympathetic, empathetic, intelligent, and creative. Conversely, a low number of total responses was associated with high noise tolerance, an indicator of general lethargy and potential depressive traits [165]. These results reflect the impact that lack of empathy and general ambivalence has had on rising noise levels and lack of consideration of noise in the design of urban public environments [166].

4.8 Soundscape in Art and Engagement

Elements of soundscape research have also been utilised in the practice of many artists, often in order to encourage people to engage with their acoustic surroundings, and to explore the nature of certain places and sounds. An example of where this research has contributed to some soundscape based art is covered in Chapter 5, and as such this section will now examine extant work in the area applying soundscape methodologies to artistic and public engagement projects.

This can include the use of soundscape recordings to examine the nature of a particular environment and use those sounds to create a narrative. One recent example of this is Chris Watson's *Trent Falls to Spurn Point*, an aural journey along the Humber estuary [167] presented as a fixed installation. Other works can make use of directed sound walks, and even hydrophones, to take a group of participants and allow them to explore an environment and its ecological soundscape: for example, Leah Barclay's work [168]. Physically guiding people around a location clearly has limitations in terms of time, budget, and equipment. As such an alternative way to allow people to explore an environment's soundscape is by using audio streaming, where a portable recorder is used to transmit the sound it picks up in real time. An example of this is the 'The Overheard' project [169] where a set of audio monitors have been placed by six outdoor sound sculptures in various locations around Denmark to encourage active listening and soundscape engagement. Some other approaches do not make use of any recording techniques and instead focus on the introduction of structure into an environment that will alter the behaviour of sound in that space. The work of *Liminal*, a collaboration between architect Frances Crow and sound artist David Prior, includes examples of this, such as *Organ of Corti* (2010-11) and *Cochlea Unwound* (2010-) [170], [171]. These two works make use of the sonic crystal methodology outlined in Chapter 3, taking the form of a set of regularly arranged columns the result in patterns of acoustic interference changing the experience of sound in the space and encouraging interaction with sound.

All of the above techniques are essentially defined by the artist who dictates what the sounds of interest are, whether through recordings they have made or by deciding upon the listening location. An alternative approach can make use of citizen participation in order to determine the sounds of interest. An example of this is Peter Cusack's ongoing Favourite Sounds Project, a process of what he calls 'urban soundscaping' where the inhabitants of a particular city respond to a questionnaire and identify their favourite sounds in and around that location [172]. This information is then used by Cusack to determine which sounds to record. Questionnaire results from these projects have also been used to develop interactive online maps tagged with the locations of the favourite sounds along with embedded audio examples.

Auralisation and acoustic modelling techniques can also be used to create soundscape art for engaging with the public. Recent research at the University of York is making use of these techniques to *'recover the soundscape of debate as experienced by women listening through a ventilator in the old House of Commons c. 1800-34.'* [173]. The results of this project were incorporated into the 2018 "Voice and Vote' exhibition in Westminster Hall, London [174].

These are just some fairly recent examples of how environmental sound recording and soundscape methodologies can be used to cultivate a wider appreciation and engagement with our sonic environments, which is a first step towards effecting real change. There remains something of a gap, where an artistic/engagement focussed approach could be combined with a more rigorous scientific approach to sound recording and monitoring.

4.9 Summary

This chapter has covered soundscape research, including the underlying aspects of soundscape theory, techniques for the subjective analysis of soundscapes, and the use of soundscape methodologies in public engagement and sound art. The next chapter will consider the practical side of soundscape research in terms of how environmental sound can be measured and recorded in order to capture soundscape data. The soundscape data collected using these techniques was then auralised in a series of listening tests that are presented in Chapter 6, where the semantic differential pairs and the self-assessment manikin introduced in this chapter will be examined further.

Chapter 5

Environmental Sound Measurement

This chapter builds on the soundscape methodologies covered in the previous chapter, and presents the work completed as part of this research in the measurement and capture of environmental sound. This includes soundscape recording for auralisation and subjective evaluation (which is used in the listening tests presented in Chapter 6), and the measurement of acoustic impulse responses in an outdoor space (which have been used in a musical composition).

This chapter will first consider the extant soundscape recording literature in order to explicate the methods applied in recording soundscapes for this research, including the capture of spatial audio and visual information for virtual reality (VR) presentation. With this established, details of the conducted soundscape recording work are then given in order to set up the listening tests making use of these recordings presented in Chapter 6.

This chapter concludes with details of impulse response recording conducted at Creswell crags, a limestone gorge containing caves and some early examples of cave art. These impulse responses were recorded in B-format using both the ESS method (introduced in Chapter 2) and a starter pistol as the excitation source. These were recorded for use in *Refugium* [175], an artistic work exploring the archeological history of the site. This section also includes the use of Spatial Impulse Responses Rendering (SIRR) analysis to explore the acoustic nature of the scene and the direction of incoming sound in the IRs.

5.1 Spatial Audio Capture

This section will consider the available methods for the capture of environmental soundscapes. Chapter 4 covered several aspects of soundscape research and evaluation, all of which require recorded soundscape material. As such there is a need for a methodological approach to soundscape capture, in terms of the recording format and the audio content, alongside many other parameters. This is one of the contributions made in this chapter, which will go on to describe a novel methodology for the recording of soundscape data (and visual features) for VR use following a summary of other audio recording methods.

The simplest formats for audio recording are: mono recording, where only a single channel of audio is used; and stereo recording, where two channels (left and right) are produced. These formats have perhaps limited use in terms of allowing for the creation of an enveloping auralised soundfield. However, the relative simplicity of the hardware required, and the ease of playback on all systems, means mono and stereo recorders are still of use in a soundscape context; including, recently, for audio streaming. Audio streaming is where equipment can be placed on-site and then used to store or transmit the soundscape of that environment in real time. Examples of audio streaming include: the AudioMoth [176], [177], an acoustic logger designed for monitoring biodiversity; and 'The Overheard' project [169] (as mentioned in Chapter 4) where a set of audio monitors have been placed by six outdoor sound sculptures in various locations around Denmark to encourage active listening and soundscape engagement.

Spatial audio techniques can also be used to capture and recreate acoustic scenes more fully (i.e. beyond the capabilities offered by stereo and mono recording). These techniques include sound recording using microphone arrays, the synthesis of spatialised sound, and the reproduction of a soundfield using a technique known as auralisation (the aural equivalent of visualisation) [178], [179]. This section will consider such techniques for the capture and reproduction of spatial audio in an environmental sound context. This includes binaural audio recording and playback, and ambisonic recording techniques and sound field recreation.

An extension of stereo format designed to emulate the experience of presence in the recording space is binaural audio, where often a dummy head with microphones positioned in its ear canals is used to make a recording that captures the full 3D soundfield in two channels of audio designed for playback over headphones [180].

The theory that underpins binaural synthesis is that the pressure variation at each eardrum can be recorded, and then presenting a reproduction of this to a listener will cause them to perceive the sound as if, in theory, they were present at the recording location. Binaural recordings can be recorded using a dummy head, such as the Knowles Electronics Mannequin for Acoustic Research (KEMAR), with microphones placed in its ear canals [181]. This recording can either be field based recording where the actual soundfield is captured via the dummy head's microphones, or can be lab based where the head-related-transfer-function (HRTF) of the dummy head is measured for different sound source positions and then convolved with anechoic data [182].

Binaural audio produced in this way can be limited in its effectiveness due to the generality of the head-related-transfer-function (HRTF) of the dummy head. Binaual audio can be personalised by measuring an individual's HRTFs to take into account the unique nature of each persons morphology (particularly, but not exclusively, that of the pinnae). Such HRTFs are also known as Head Related Impulses (HRIRs).

Binaural recordings are then typically presented over (typically in-ear) headphones, but can be presented over loudspeakers. Headphone reproduction tends to result in better sound source localisation than loudspeakers, but can make the listener feel isolated, or (due to 'ill-fitting' HRTFs being used) result in sounds appearing to originate from within the head or from incorrect directions.

5.1.1 Ambisonics

Ambisonics is a set of audio techniques for the recording, modification, and recreation of three-dimensional sound [183]. It is a system that allows a spatial recording to be played back over a variety of loudspeaker configurations [184]. Ambisonic recordings are not encoded based on speaker information, and instead make use of spherical harmonic encoding to capture the soundfield [185]. A widely used form of Ambisonics is First-Order-Ambisonics (FOA), also know as B-format, which has been used in this research to make soundscape recordings.

B-format is a coincident multichannel measurement technique, producing recordings comprised of four channels, (labelled W, X, Y, and Z). These represent omnidirectional pressure information, and directional pressure gradient information for the front/back, left/right and top/bottom axes respectively. Figure 5.1 gives graphical representation of the response of the Soundfield microphone, a microphone specifically developed for B-format recording [186].

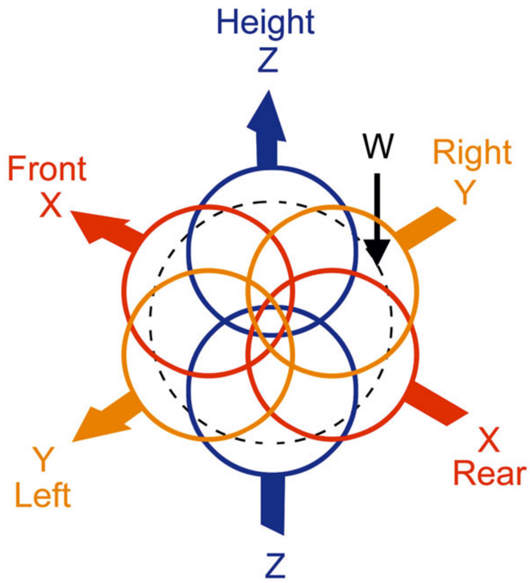


FIGURE 5.1: The directional response of a Soundfield B-format microphone, from [187].

The B-format signal was originally designed for a square speaker array, but can be decoded to any number of loudspeakers [188]. Once a B-format recording has been made it can be reproduced as a full soundfield over a multiple-loudspeaker array [189], or as a dynamic stereo rendering (for example via YouTube [190]). Higher orders of Ambisonic recording also exist, where a greater number of microphone capsules and audio channels can be used to capture the spherical soundfield to a greater degree of spatial accuracy. One example of this being used in soundscape research is the MH acoustics Eigenmike microphone array [191]. Low order Ambisonic playback has been criticised for its inherent instability and the small size of its 'sweet spot', outside which phase effects can negatively affect a listener's experience. It can also be difficult to correctly set up and calibrate an Ambisonic rig, especially in the context of consumer audio systems [192].

5.1.2 Soundscape Content

Following evaluation of semantic responses to soundscape assessment, Raimbault et al. found three primary factors described the range of sounds heard: sound strength, temporal dynamics, and spatial dimension [166]. The audio content used in this experiment must, therefore, cover a suitable range of environmental sounds to go some distance towards a comprehensive representation of a typical soundscape. In terms of specific stimuli this presents two requirements:

1. Noise levels. The recorded soundscapes must cover a range of noise sources with various sound pressure levels.
2. Sound sources. The sound sources recorded must include sources of various types; including natural, human, and industrial/mechanical sounds.

Sound Sources

In Chapter 4 three categories of sound sources were identified: natural, human, and mechanical. These categories were used as a guide for the content required to be captured as part of this recording work in order to cover a wide range of sound sources and evoke a wide range of emotional states.

Generally speaking, extant research indicates that natural sounds are most preferred, mechanical sounds are disliked, and the reaction to human sounds depends more on the context within which they are presented [128]. Viollon et al. found that while increased urbanisation of visual setting generally had a negative effect on the perception of auditory scene, this was shown to affect human sounds much less [126]. This was posited to be due to the orienting nature of human sounds, where familiarity causes people to gravitate to and identify with evidence of human activity (this reflects the topophilia (from the Greek *topos*, meaning ‘place’ and *-philia*, meaning ‘love of’) theory which states that landscape preferences are dictated at least in part by familiarity [62]).

Audio Format

When reproducing soundscape material in a lab environment, some studies have made use of stereo reproduction, either over headphones [129] or over a pair of loudspeakers [126]. Whilst this has been shown to be sufficient for certain stimuli, for increased background noise detail and believability (also referred to as ecological validity), a surround sound reproduction method can be used [193].

The surround sound method most commonly used for soundscape reproduction is Ambisonics, which has been used in multiple studies including Bruce and Davies's use of first order Ambisonic playback in a semi-anechoic chamber [194], or Harriet's use of third-order Ambisonics in a damped listening environment [24].

The recordings for this research have been made using a B-format Soundfield microphone, which is a widely used [41], [47], [195] single microphone surround sound recording solution that also allows for potential future conversion. Conversion options include Binaural [196] and conversion to stereo UHJ format, which will be covered in detail in Chapter 6.

5.1.3 Visual Content

This section will focus on the background theory on how the visual features of an environment can affect the perception of that environment's soundscape, and how extant research has informed the data collected in this recording work.

The typical focus of soundscape analysis work on urban areas means that, where the visual features of the soundscape environment are taken into account, a usual rating scale is the degree of urbanisation of that environment. One example of this is Viollon et al.'s study of the influence of various visual settings on the ratings of various sounds [126]. Figure 5.2 shows the four photographic visual settings used in this study, the fifth visual setting used in their experiment was a control case with no visual stimulus present [126].

The environments shown in Figure 5.2 indicate how the 'degree of urbanisation' increases at the expense of the number of natural features present. Indeed, the presence of trees and shrubs in a noisy environment has been shown to have a psychologically mediated

FIGURE 5.2: The four visuals settings used by Viollon et al. displaying varying degrees of urbanisation. (a) Woods with a bench (b) Small houses with trees (c) Apartment block with trees (d) Apartment blocks. From [126].

abatement effect on the noise level, even where the vegetation present has no appreciable physical noise abatement effect [197].

A similar approach to Viollon et al. was taken by Anderson et al. in their 1983 investigation into the effect of various sounds on environmental preferences [128]. Figure 5.3 shows a selection of settings used in their studies, showing a similar range of urbanisation to those in Figure 5.2. In their study Anderson et al. found while the presence of trees and vegetation did ultimately correlate with higher preference ratings, it could sometimes raise expectations about the pleasantness of that environment, resulting in greater disappointment following the addition of unpleasant/incongruous sounds [128].

A more formal method of describing the degree of urbanisation of a given visual setting is the Natural and Contextual Features (NCF) percentage measure developed by Watts and Pheasant, and used in assessment of tranquillity ratings for various locations in the Scottish Highlands and in Dartmoor National Park [129]. A selection of visual settings used in this study can be seen in Figure 5.4. In their study, Watts and Pheasant obtained a measure of NCF for each location by overlaying a 10×10 grid on a still image of each environment and counting the number of squares occupied by natural and contextual (N) features, and by man-made (M) features (not including the area of sky above the horizon).

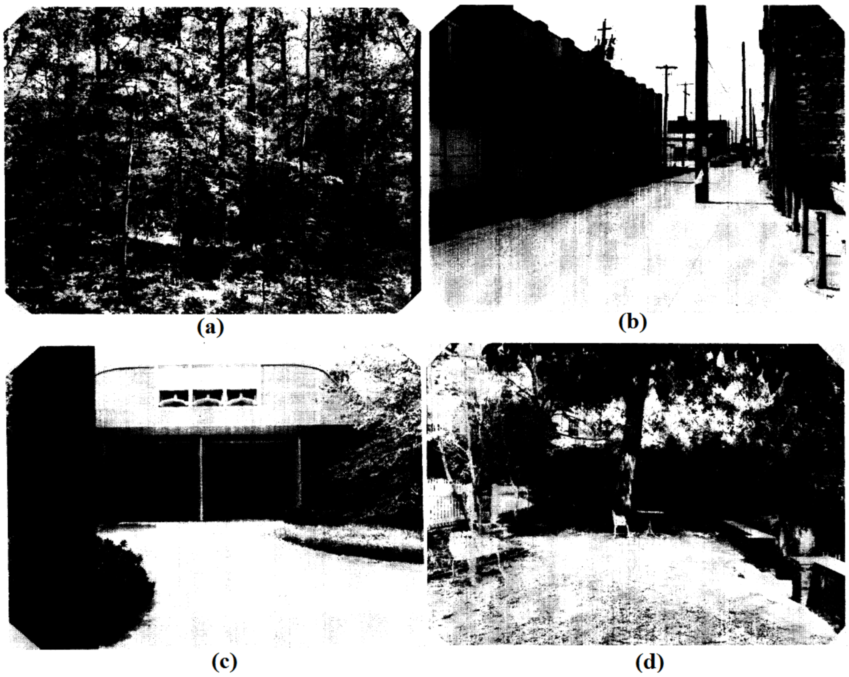


FIGURE 5.3: A selection of visual settings used by Anderson et al.: (a) Wooded area (b) Downtown street (c) Courtyard (d) Memorial garden. From [128].

They defined contextual features as elements contributing to the visual context of the environment in a positive way, for example listed buildings, landmarks/monuments, and religious or historical structures [1]. From these measure of N and M the NCF is then given by [129]:

$$NCF = 100 \cdot \frac{N}{N+M} \tag{5.1}$$

Watts and Pheasant have used the measure of NCF to create a method for predicting the level of tranquillity associated with a given environment, the Tranquillity Rating Prediction Tool (TRAPT) [198], which is given by:

$$TR = 9.68 + 0.041NCF - 0.146L_{day} + MF \tag{5.2}$$

where TR is tranquillity score between 0 and 10. NCF is the percentage of natural and contextual features visible within the landscape, and L_{day} is the A-weighted sound level equivalent for daytime (7am-7pm) [199]. MF is a moderating factor added to take account

[1]These contextual features could also be considered as elements that represent the *genius loci*.

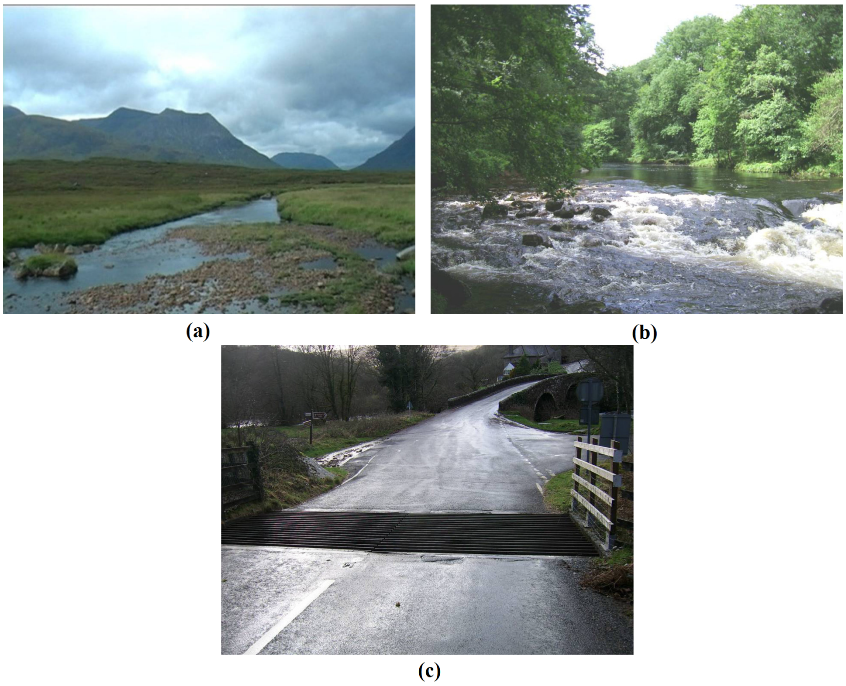


FIGURE 5.4: Examples of the locations used by Watts and Pheasant: (a) Glen Etive (b) Rapids on the river Dart (c) A cattle grid in Dartmoor National Park. From [129].

of other factors that would decrease/increase the rating (for example the presence of litter would likely reduce the perceived tranquility of a location [200]). Another example is Watts et al.'s 2013 study of tranquillity ratings in urban green spaces, where they used parks around Bradford to obtain subjective ratings of environments raging from natural environments with little or no man-made features through to environments with encroaching urbanisation and man-made features [199].

The urban environments, rural areas, and inner city parks shown in the above studies represent three of five generic landscape classes identified Pheasant et al.: mountains/wilderness, coastal areas, parks/gardens, rural areas, and urban environments [124], [201]. The focus on these three areas is suitable when considering soundscape assessment, as the problems associated with noise will naturally be most acute in urban environments and it is the interfacing between urban and rural areas and parks within cities that will perhaps gives the clearest indication of the cross-modal effect of visual setting on soundscape preference, and the impact of green infrastructure on the mechanism of perception.

When deciding on recording locations, one factor to consider is how particularly distinctive environments, or ones that are familiar to test participants, might skew the results. This could be due to the effect of prior experience on expectation of the soundscapes [194], or perhaps where well known land marks might move cognitive attention away from the soundscape (as acknowledged by Harriet where, in her study of responses to the soundscape of an urban park in Leeds, she found that the '*generic nature of this soundscape contributed to the success of its reproduction*' [24]).

Visual Format

Previous studies investigating the effect of visual factors on soundscape ratings have typically taken the form of soundwalks [198], [202], made use of either pictures or video clips presented in a lab environment [126], [128], [129], or presented a virtual reality environment [203].

The recent development of the virtual reality headsets (such as the Oculus rift) offers an opportunity for the novel presentation of recorded environmental visual information in the context of soundscape analysis: it allows for the presentation of immersive 3D visuals [204], [205]. In order to generate content suitable for playback over a typical VR headset, a 360° image or video of the space must be captured. One method for doing this is to use six GoPro cameras arranged in a suitable mounting system (such as the Freedom360 Mount, shown in Figure 5.5). A video stitching software program such as Kolor Autopano Video [206] can then be used to stitch together the footage from the six cameras into a single 360° feed.

FIGURE 5.5: The 'Freedom360 Mount', an example GoPro mounting rig for the capture of 360° video. From [207].

The use of this sort of set-up to record visual stimuli at the chosen locations is beneficial as it allows for the novel use of 360° video for presentation alongside soundscape recordings as well as affording further flexibility in presentation format.

Recording Duration

When collecting soundscape recordings it is important to consider the length of the recordings made, such that they are long enough to get a somewhat comprehensive sense of the soundscape of the location but short enough to be used in a listening test.

The duration of recording used for soundscape reproduction in previous studies varies considerably. Whilst Harriet made use of 7 minute long soundscape representations [24] constructed artificially from recorded material, other studies typically use shorter recordings (especially those presenting visual and aural stimuli simultaneously). For example both Anderson et al. [128] and Viollon et al. [126] used 20 second long recordings. Pheasant et al. have used 32 second long recordings [124], [201], and both Watts and Pheasant and Gifford and Ng make use of recordings lasting 1 minute [129], [202]. Axelsson's work as part of the Sound Cities project used binaural recording of 46 seconds in length, presented with a set of six still images of the recording site [208]. Rummukainen et al. used even shorter recording only 15 seconds in length [209]. One must bear in mind that these studies have all considered visual stimulus alongside aural information, and that most audio only studies have made use of soundwalks [24], [140], [194], [210], [211] which are naturally longer in duration. Work by Fröhlich et al [212] has previously confirmed the ecological validity[2] of short (c. 10 second long) video clips in quality of experience studies.

Ten minutes have been recorded at each location for this work, and from each of these recordings two shorter sections of 30 seconds in length have be extracted for experimental use. The A-weighted SPL level has also been measured at each location.

[2] Ecological validity, in this context, is where a reproduction of a soundscape is sufficiently realistic to evoke the same emotional response as the sound in its own real context [193]

5.2 Soundscape Recording Work

Following the findings in the previous sections regarding best practice for determining what soundscapes to capture, and the recording formats to use for both aural and visual elements, this section will describe how these findings have been applied to the capture of environmental soundscapes for use in this research. The requirements for the aural/visual stimuli for this research can be summarised as follows:

- Format: Audio recordings to be made in B-format using a single Soundfield microphone for FOA reproduction. Video recordings to be made using a 360° camera rig.
- Aural metric: The soundscapes recorded must cover a range of SPL levels (from a very quiet environment such as a forest to a very loud industrial/urban environment).
- Visual metric: The recording environments must cover varying degrees of urbanisation from a completely natural environment to a highly developed inner city location.
- Sound sources: The stimuli must include a variety of sources covering natural, human, and mechanical sounds.
- Environment types: Mountain/wilderness, coast, parks/gardens, rural areas, and urban environments have been identified as the main environment types.

These requirements were developed based on the findings presented in Chapter 3 and Chapter 4, and earlier in this chapter. The ultimate aim is to capture a range of soundscape stimuli covering the sound source and environmental categories identified in the literature. The equipment use to record this data has been chosen in order to provide ambisonic audio and panoramic video using compact equipment suitable for outdoor use.

5.2.1 Locations

This section contains details of the environments used to make recordings. Table 5.1 outlines the properties of the locations. The recording of stimuli at all of the locations shown in Table 5.1 was completed in two trips. Locations 1-5 are within 7 miles of each other near

Pickering toward the south east of the North York Moors National Park. Recordings at locations 6-8 were made as part of a second trip to Leeds.

FIGURE 5.6: Map of the North York Moors National Park with locations 1-5 indicated. Map from [213].

Figure 5.6 shows a map of locations 1-5, and Figure 5.7 show locations 6-8. Images of the recording locations can be seen in Appendix A.1, further details of the recordings made at each location can be found in Appendix A.2, and a photo of the recording equipment used can be seen in Appendix A.3. The soundscape recordings detailed in this chapter are those used as stimuli in the listening tests presented in Chapter 6, which also includes further detail on the recordings.

5.3 IR Recording At Creswell Crags

Another example of environmental sound measurement is the recording of IRs (which were introduced in Section 2.3.1. This chapter will now go on to present the use of this recording process to collect a series of IRs at Creswell Crags, a limestone gorge in England that comprises many caves and examples of early cave art. These IRs were captured for use in *'Refugium: Time, Stone, Voice and Sound'*, a musical composition created as part of a

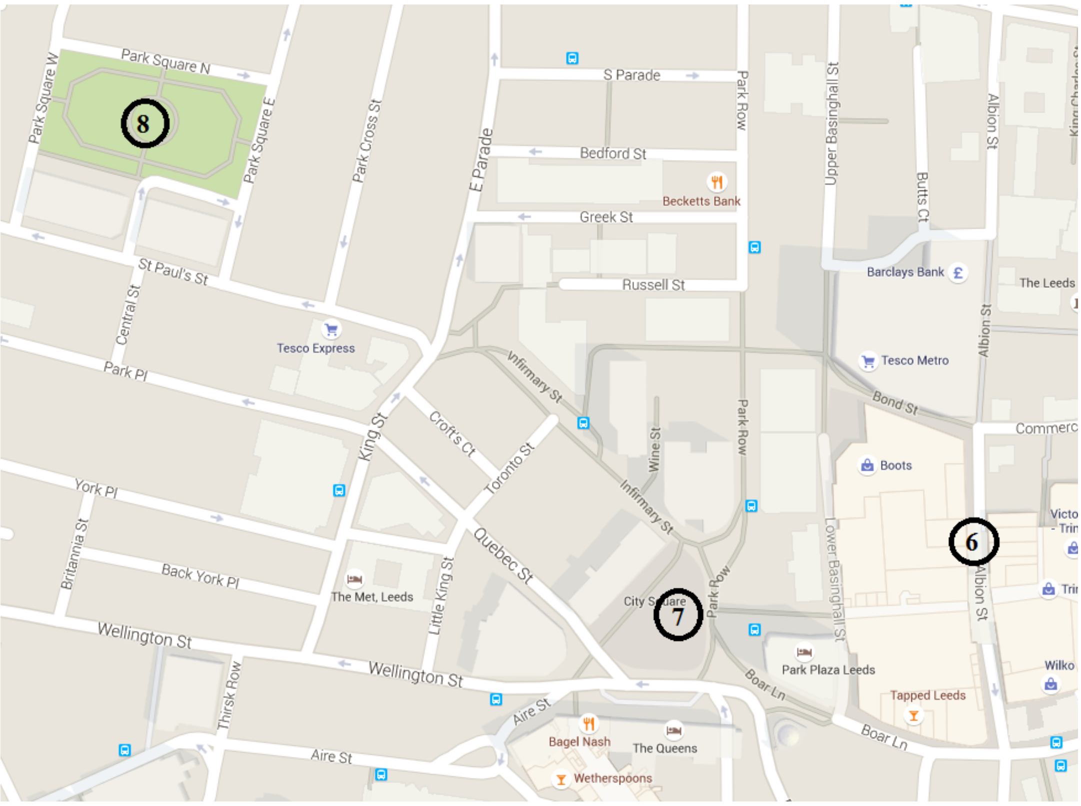


FIGURE 5.7: Map of Leeds with locations 6-8 indicated. Map from [213].

collaboration between the Electronic Engineering, Archeology, and Music departments at the University of York. This section will first consider the measurement process and then present an analysis of two of the recorded IRs.

5.3.1 IR Measurement Process

A total of twelve IRs were measured (in B-format). Three of these were made in Robin Hood Cave (the largest of the caves at Creswell) using the ESS method described in Section 2.3.1. The other nine were recorded using a starter pistol as the impulse sound source.

Figure 5.8 shows a map of Creswell Crags with the positions of the main caves and the other outside recording locations indicated. Figure 5.9 shows a plan view of Robin Hood Cave, in and around which a number of the impulse responses were measured. These two Figures shows the source and receiver positions used in recording the twelve IRs: Table 5.2 includes the details of the source and receiver positions used in each instance.

Location	Expected SPL Level (dBA)	Urbanisation	Sound Sources
1. Dalby Forest	40-45	Very Low	Natural sounds Cyclists and walkers
2. Dalby Forest Lake	45-50	Low	Natural sounds Cyclists and walkers Conversation Light traffic
3. A169 at the Hole of Horcum	50-55	Low	Natural sounds Cyclists and walkers Medium traffic Conversation
4. A169 at the Fox and Rabbit Inn	50-60	Low/Medium	Medium traffic Natural sounds Conversation
5. Pickering	50-55	Medium/High	Medium traffic Natural sounds Human activity
6. Albion Street, Leeds	55-60	High	Heavy traffic Pedestrians and cyclists Human activity Trains
7. Park Row, Leeds	65-70	High	Heavy traffic Pedestrians and cyclists Human activity Trains
8. Park Square, Leeds	50-55	Medium/High	Medium to heavy traffic Natural sounds Human activity

Table 5.1: Details of the recording locations, including the expected typical SPL level and degree of urbanisation, and sound sources expected to be present. Note that the sound sources in this table represent only what was expected to be recorded at each location. In Chapter 6 details are given of the sound sources captured in the actual recordings.

The starter pistol was used to record the impulse response at the mouths of the various caves, and in the gorge itself. Prior research by the author has shown the use of a starter pistol to be an adequate alternative to the ESS method (despite its lack of a flat frequency response), providing a sufficiently high signal-to-noise ratio when used to record the IR of an outdoor space [47], [215]. The use of a starter pistol also eliminates the need for a loudspeaker and power supply, and can be less sensitive to environmental noise than the ESS method. The ESS method was used to record the IRs in Robin Hood Cave due to concerns about the high SPL levels generated by the starter pistol disturbing bats residing in the cave.

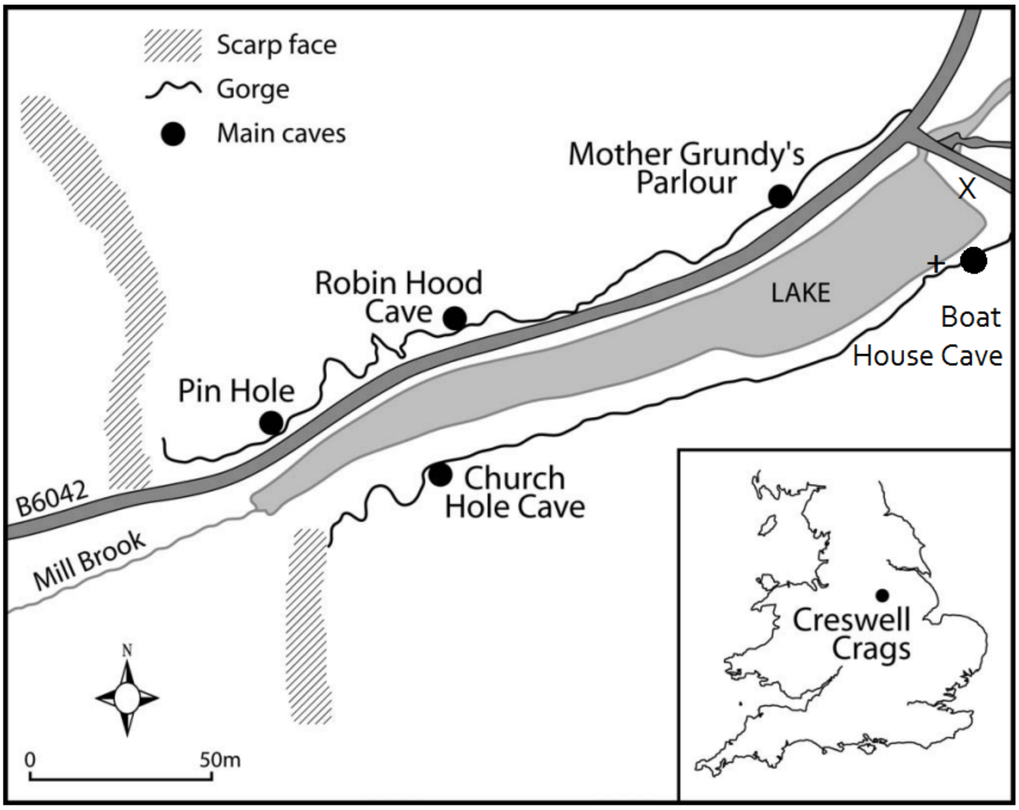


Figure 5.8: Map of Creswell Crags with the position of the main caves and other recording locations indicated by + and ×. Modified from [214].

Table 5.2: Summary of the recordings made, including filename, and source and receiver positions for each one. The source and receiver positions numbers and names make reference to the labels and numbers shown in Figure 5.8 and Figure 5.9.

Starter Pistol Recordings		
IR Number	Source Position	Receiver Position
1	7	5
2	3	5
3	6	5
4	Boat House Cave	x
5	+	x
6	5	3
7	Pin Hole Path	Pin Hole Mouth
8	Church Hole Path	Church Hole Mouth
9	Mother Grundy's Parlour Path	Mother Grundy's Parlour Mouth
Sine Sweep Recordings		
IR Number	Source Position	Receiver Position
10	1	2
11	1	3
12	1	4

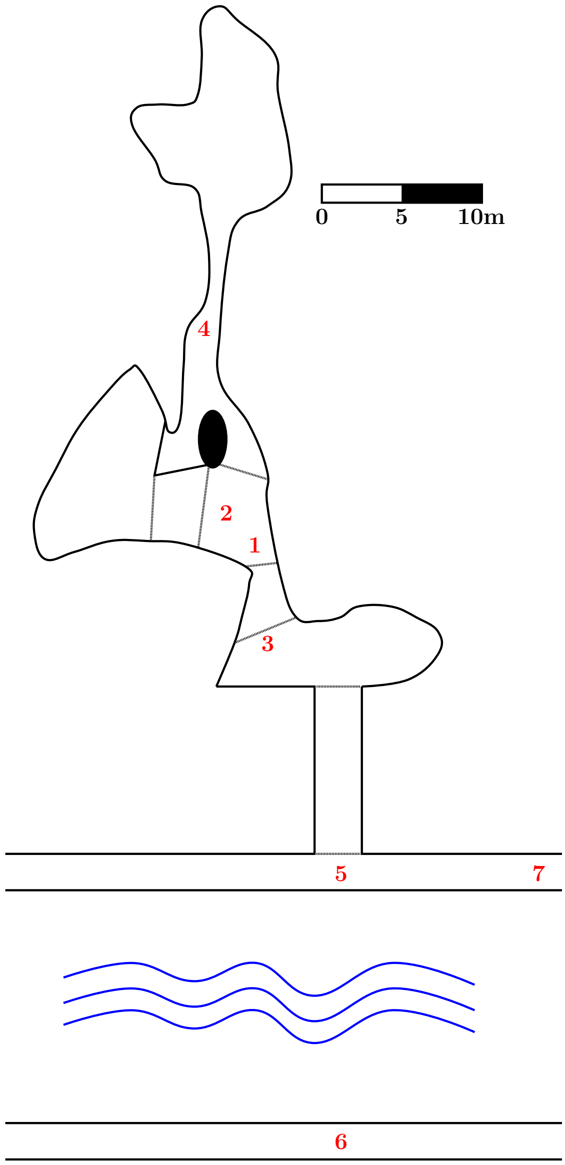


FIGURE 5.9: Plan view of the interior of Robin Hood Cave and adjacent paths in the gorger. Indicated by the numbers are the source and receiver position locations used in IR recording.

Figure 5.10(a) shows the B-format microphone (in its windshield) in its position on the path at the base of Robin Hood Cave. On the floor behind the microphone is the multi-track recorder. This setup was used to record IRs 1-9 (as listed in Table 5.2). Figure 5.10(b) shows the B-format microphone and Genelec loudspeaker set-up on the main level inside Robin Hood Cave. The loudspeaker was used as the sound source for IRs 10-12, generating an exponential sine sweep covering the frequency range of 22 Hz - 22 kHz for a duration of 15 seconds.

Having now given a summary of the IRs recorded at Creswell Crags the chapter will go on

(a) (b)

FIGURE 5.10: (a) B-format microphone positioned at the base of Robin Hood Cave. (b) Loudspeaker and B-format microphone positioned inside the main chamber of Robin Hood Cave.

to compare two of the recorded IRs: one recorded inside Robin Hood Cave, and the other recorded next to Mother Grundy's Parlour (one of the other, smaller, caves in the gorge). In order to demonstrate the differences between these two IRs, a process called Spatial Impulse Response Rendering (SIRR) analysis is used.

5.3.2 SIRR Analysis

The directional information encoded in B-format recordings can be taken advantage of to investigate the spatial characteristics of a recorded impulse response through the calculation of the instantaneous intensity vector, **I**. In order to do this, the B-format signal is divided into discrete time frames, each one of which is then windowed using a Hanning window, followed by performance of a short-time Fourier transform (STFT) on each channel. The resultant frequency domain signals can be used to estimate the intensity vector using the following equation [41]:

$$\mathbf{I}(\omega) = \frac{\sqrt{2}}{Z_0}\mathcal{R}\{W^*(\omega)\mathbf{U}(\omega)\} \tag{5.3}$$

where $\mathbf{U}(\omega)$ is comprised of three of the B-format channels in vector $[X(\omega), Y(\omega), Z(\omega)]$, Z_0 is the characteristic acoustic impedance of the air, and $*$ denotes the complex conjugate. **I** can be calculated at multiple time steps, and the time-frequency distribution of these vectors can be overlaid on a spectrogram of the recording's omni-directional (W-channel) response to create plots allowing concurrent analysis of the magnitude, and direction, of arriving acoustic energy. These plots will be referred to as SIRR analysis plots from this point. To generate these plots, $\mathbf{U}(\omega)$ is formed of $X(\omega)$ and $Y(\omega)$ only (ignoring the Z channel), resulting in a plot of the horizontal plane. Calculation of $\mathbf{I}$ is one of the steps involved in Spatial impulse response rendering (SIRR), a method of reproducing spatial acoustics over a multichannel loudspeaker system [216].

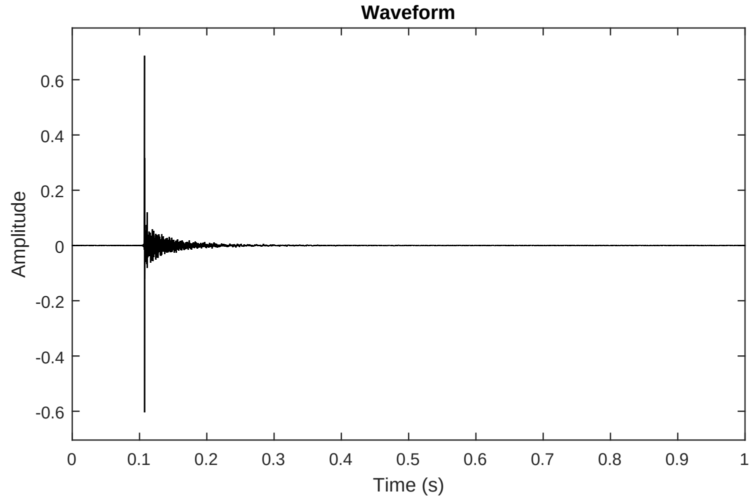
5.3.3 IR Comparison

Two of the IRs recorded at Creswell will now be analysed and compared. Figure 5.11(a) shows the W-channel of IR number 10, recorded in Robin Hood Cave. Figure 5.11(b) shows a SIRR analysis plot for the X and Y channels of this IR, showing the behaviour of sound in the horizontal plane over time at different frequencies. The source and receiver positions used in measuring this IR are detailed in Table 5.2 and Figure 5.9.

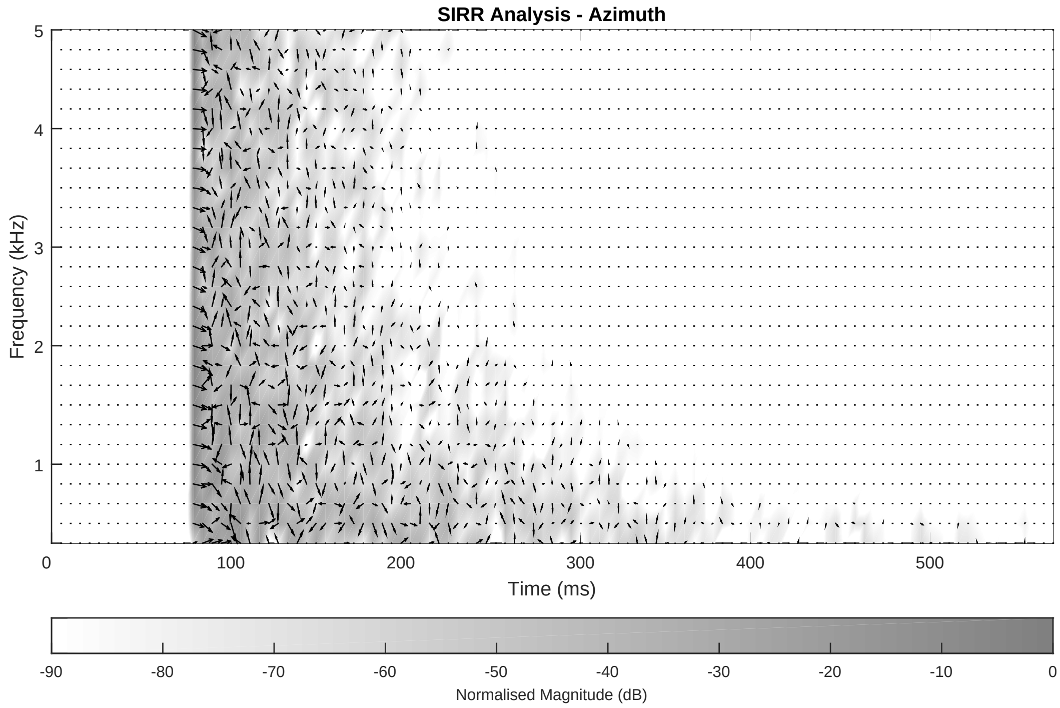
Figure 5.12(a) shows the waveform of the W channel of IR number 9. This IR was recorded using a starter pistol as the excitation source, positioned at the mouth of Mother Grundy's Parlour (Figure 5.8 shows the location of this cave within the gorge). The B-format microphone used to record this IR was positioned around 10 metres away from the starter pistol on the path next to the stairs leading up to Mother Grundy's Parlour.

The SIRR analysis plots in these two Figures each show a spectrogram of the IR under analysis overlaid with a series of arrows. These arrows represent the intensity and angle of arrival of incoming sound (in the horizontal plane) for each frequency bin, in each time frame (as determined by the spectrogram settings). An arrow pointing to the right indicates sound arriving at an azimuth angle of 0° (i.e. straight on to the microphone), an arrow point down indicates sound arriving at an azimuth angle of 90° and so on. The length of each arrow then indicates the magnitude of the arriving sound.

Figure 5.11(b) shows a distinct, highly directional, initial sound path followed by acoustic energy arriving incoherently from multiple directions with no clearly distinct later reflections. By contrast, the plot shown in Figure 5.12(b) again shows a highly directional initial impulse, followed c. 300ms later by a distinct directional reflection (somewhat spread out in time) caused by a reflection from the rock face of the opposite side of the gorge.



(a)



(b)

FIGURE 5.11: (a) A time domain plot of the W-channel of IR number 10, recorded at Creswell Crags in Robin Hood Cave. (b) A SIRR analysis plot of this IR. This IR was recorded using an exponential sine sweep as the sound source.

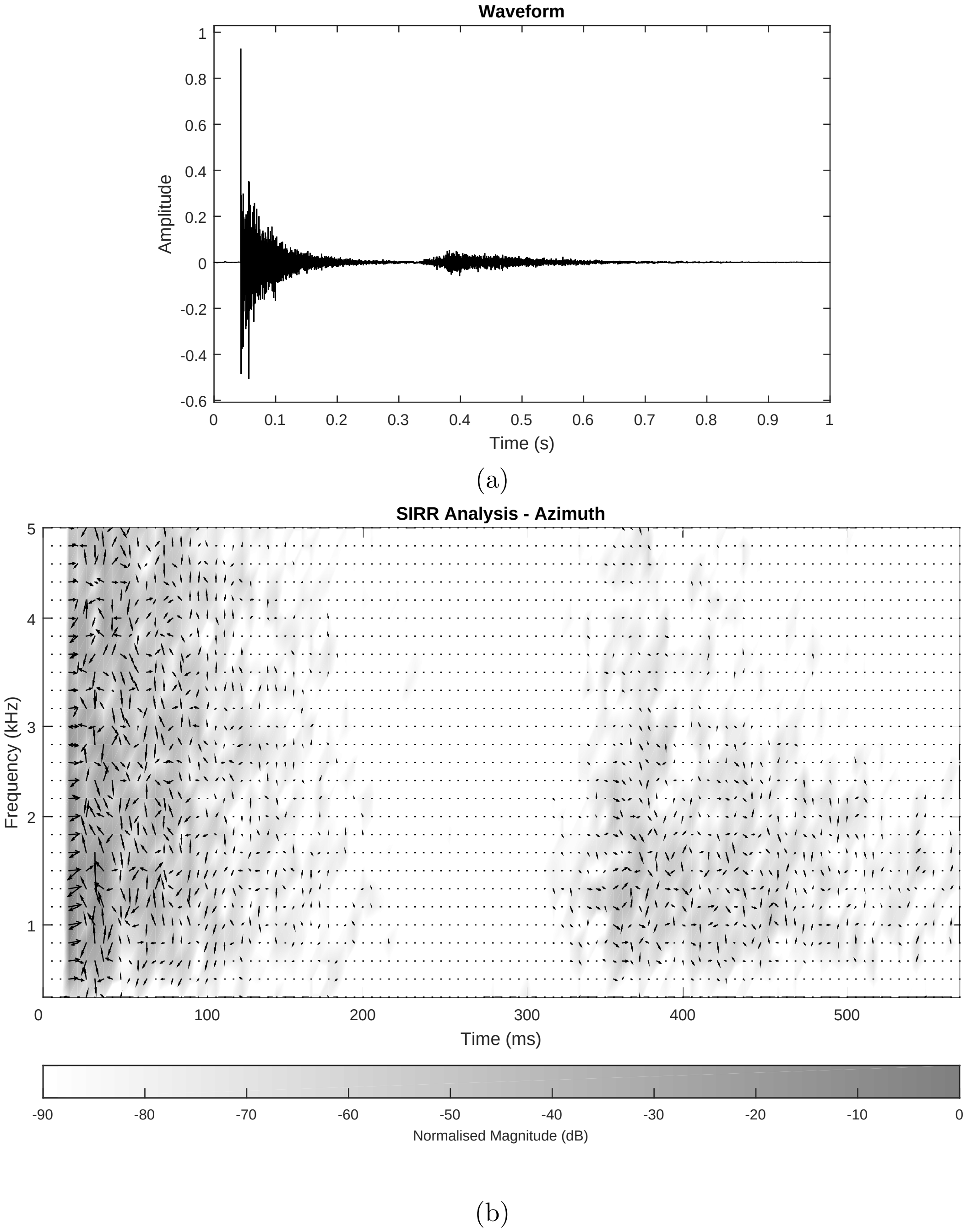


FIGURE 5.12: (a) A time domain plot of the W-channel of IR number 9, recorded at Creswell Crags next to Mother Grundy's Parlour. (b) A SIRR analysis plot of this IR. This IR was recorded using a starter pistol as the sound source.

These analyses give objective measures of the features to expect of IRs recorded in such locations. Figure 5.11 shows an IR recorded in a small cave formed of irregular curved surfaces. This results in diffuse reflected sound scattering throughout the space, and therefore no clear and distinct early reflections present in the IR (despite a clear initial impulse).

Figure 5.12 on the other hand shows an impulse response recorded in next to the mouth of Mother Grundy's parlour in the gorge itself. On the opposite side of the gorge there is a large, relatively flat, rock face. The specular reflection of sound from this surface results in the distinct reflection of the initial impulse seen in the SIRR analysis plot. This location is an example of a sparsely reflecting outdoor space, of the kind discussed further in Chapter 7. These plots show how SIRR analysis can be used to identify acoustic features of recorded IRs that cannot be identified from time or frequency domain plots alone.

All of the IRs recorded as part of this work are available online [217]. This link also contains a downloadable report on the recording work. These IRs have also been used in the artistic work *'Refugium: Time, Stone, Voice and Sound'*, a musical piece aiming to give the audience *'a feel of what Creswell Crags sounded like in the past'*. This piece was performed in the gorge itself and made use of composed material (some of it convolved with the recorded IRs) reproduced via loudspeaker alongside a live choir [175].

5.4 Summary

This chapter has presented two examples of environmental sound measurement: the capture of B-format soundscape recordings alongside panoramic visual data, and the measurement of IRs at Creswell Crags. These IRs show the validity of applying IR recording techniques to outdoor environments, and how SIRR analysis techniques can be used to demonstrate the effect that the geometry of an IR's recording location has on its acoustic properties. The use of these IRs in the creation of a musical composition is also an example of how IRs can be used to employ soundscape methodologies for artistic purposes. The limestone gorge at Creswell Crags is also an example of a sparsely reflecting outdoor scene (as shown in the SIRR analysis plot in Figure 5.12). A method for the modelling of acoustic scenes of this type is presented in Chapter 7.

The protocol for VR content capture presented in this chapter is a novel contribution to the field, especially given the time it was completed (the recording work took place in June of 2015). The next chapter will go on to make use of the captured soundscape and visual data in a series of listening test which establish the validity of the recording protocol. Following that, Chapter 7 will present a method for the acoustic modelling of sparsely reflecting outdoor acoustic scenes, building on the findings presented in this chapter regarding recording IRs in such a space.

Chapter 6

Environmental Sound Evaluation

This chapter presents the results of four listening tests conducted making use of the FOA soundscape recordings detailed in the previous chapter in Section 5.2. This listening tests have been developed to fulfill the relevant aim this thesis, as introduced in Chapter 1.

1. Investigate the phenomenon of cross-modal perception in the context of soundscape evaluation and analysis.

The objectives to fulfill this aim were also stated as follows:

- Make use of spatial audio and spherical video recording equipment to gather audiovisual soundscape stimuli for use in listening tests.
- Run a series of audio-only tests first (in line with extant soundscape evaluation research) in order to establish and develop suitable soundscape evaluation tools, and to establish the ecological validity of a variety of formats (i.e. the extent to which the soundscape playback is representative of reality).
- Introduce the visual content to the listening tests in order to investigate cross-modal perception.
- Analyse the test results to identify the aural and visual features most that most impact the experience of a soundscape. This analysis should include a focus on identifying the features most likely to indicate positive, restful environments, and where the presence of visual features may impact the evaluation of aural ones.

The first of these objectives was covered in Chapter 5, and the other three are covered in this Chapter. To achieve these objectives, the four tests presented here have been designed to build on established soundscape listening test results and techniques in order to validate each step of the research. To summarise, the four tests make use of this soundscape data in the following ways:

- The first listening test (LT1) makes use of the FOA soundscape recordings presented in full surround-sound using a 16-speaker listening rig.
- The second listening test (LT2) presents these recordings in UHJ stereo format, which makes use of the W, X, and Y channels of the B-format recordings only.
- The third listening test (LT3) uses the same UHJ stereo conversions as the second listening test presented alongside still panoramic images of the recording locations (created by stitching together images from the six GoPros cameras used to record visual information).
- The fourth listening test (LT4) makes use of YouTube as a platform to present full VR versions of the soundscape recordings, including dynamic binaural audio rendering, and spherical full motion panoramic video.

The tests make novel contributions to the field of environmental sound evaluation, including the development and validation of suitable evaluation tools, the comparison of FOA reproduction with stereo rendering, and the effect of the presence of visuals on the experience of soundscapes, including still panoramic images and full motion spherical video.

These listening tests were conducted following approval from the University of York Physical Sciences Ethics Committee (PSEC)[1].

6.1 LT1: FOA Soundscape Reproduction

This section covers the first listening test conducted in this research, which makes use of the FOA soundscape recordings detailed in Section 5.2 presented in a 16-speaker listening

[1] The first listening test, making use of FOA reproduction and biometric measurement, was approved in PSEC application **stevens150629**. The second listening test was approved in PSEC application **stevens160503** to allow for presentation of stereo soundscape renderings online. This application was then extended to include simultaneous presentation of images (still and moving) in order to allow the two further listening tests to be conducted. These two PSEC applications are included on the attached data CD.

rig. This includes a specification of the experimental methods used, a presentation of results, and a discussion of those results followed by a conclusion including consideration of further work. The hypotheses of this experiment can be stated as follows:

- Comparison of soundscape preference rating results between a set of SD pairs and the SAM will show the SAM to be a directly comparable and equally useful tool for the analysis of subjective soundscape experience.
- Biometric measurement of heart rate can be combined with preference rating results in order to make explicit connections between subjective and objective soundscape experiences.

This section is divided into two in order to address each hypothesis in turn.

6.1.1 Semantic Differential Pairs and the Self Assessment Manikin

In order to test the first of this experiment's hypothesis (as stated above) this test makes use of semantic differential pairs and the SAM as evaluation methods and compares the two sets of results.

The semantic differential is a method originally developed by Osgood to indirectly measure the interpretation of the meaning of certain words [218]. It is an indirect measure in the same way that an IQ test consists of multiple smaller tests, the results of which are combined to give an overall measurement of intelligence. The method involves the use of a set of bipolar (typically a 5- or 7-point [219]) descriptor scales, for example 'Weak - Strong', allowing a user to rate a given stimulus. Factor analysis can then be performed on the results to determine underlying patterns connecting the various descriptor pairs.

The use of Semantic Differential (SD) pairs for the assessment of soundscape quality is well established [24], [139], [140], [210], [219]–[222], and includes the use of both connotative and denotative scales. Denotative scales relate to the acoustic or psychoacoustic properties of the soundscape, whereas connotative scales measure the emotional meaning [221]. Table 6.1 shows a summary of SD pairs that have been used in prior soundscape research. This survey of the literature was conducted in order to determine which SD pairs to use in this listening test, making sure to select a set of SD pairs that are representative of extant research, and avoiding choosing too many (to keep the test duration reasonable).

#	Semantic Differential Pair	Harriet [24]	Kang [140]	Davies [220]	Viollon [139]
1	Quiet-Noisy	×	×	×	×
2	Comfort-Discomfort	×	×	×	×
3	Unique-Common (Interesting-Boring)	×	(×)	(×)	(×)
4	Monotonous-Varied (Varied-Simple) [Static-Changing]	×	(×)	(×)	[×]
5	Pleasant-Unpleasant	×	×	×	
6	Harmonious-Disharmonious (Gentle-Harsh)	×	(×)	(×)	
7	Soft-Rough (Soft-Hard)	×	(×)	(×)	
8	Natural-Artificial (Rural-Urban)		×	×	(×)
9	Social-Unsocial (Friendly-Unfriendly)		×	×	(×)
10	Calming-Agitating		×	×	×
11	Meaningful-Meaningless (Informative-Uninformative)		×	(×)	
12	Like-Dislike		×	×	
13	Fast-Slow		×	×	
14	Sharp-Flat		×	×	
15	Directional-Everywhere		×	×	
16	Echoed-Deadly (Reverberant-Anechoic)		×	(×)	
17	Far-Near		×	×	
18	Warm-Cold	×			
19	Communal-Private			×	
20	Reassuring-Unsettling				×

Table 6.1: Comparisons of different sets of SD pairs used in various studies.

SD pairs 1-11 as shown in Table 6.1 were therefore chosen for use in this listening test, as most of them have been utilised in three or more prior studies. The 'Meaningful-Meaningless' pair has also been included as it has been argued to be crucially important to the process of soundscape evaluation [127], and can be used as a measure of the presence of human activity. The 'Like-Dislike' scale will also be used as a separate factor to give an holistic measure of soundscape preference (after Anderson et al. [128]).

The Self-Assessment Manikin

The Self-Assessment Manikin (SAM) is a method for measuring emotional response developed by Bradley and Lang in 1994 [142]. It was developed from factor analysis of a set of connotative SD scales rating both aural [79], [145] and visual stimuli [223]. Table 6.2

contains the SD pairs used in generating the SAM, with the three identified underlying factors indicated. These factors are:

1. 'Valence'. A measure of how pleasant the stimulus is.
2. 'Arousal'. A measure of the extent of the emotional response evoked by the stimulus (i.e. the level of agitation and excitement).
3. 'Dominance'. The extent to which the participant feels they are in control of the scenario represented by the stimulus.

Semantic Differential Pair	
Dimension 1: 'Valence'	Annoyed-Pleased Unsatisfied-Satisfied Melancholic-Contented Despairing-Hopeful Bored-Relaxed
Dimension 2: 'Arousal'	Relaxed-Stimulated Calm-Excited Sluggish-Frenzied Dull-Jittery Sleep-Wide awake Unaroused-Aroused
Dimension 3: 'Dominance'	Controlled-Controlling Influenced-Influential Cared for-In control Awed-Important Submissive-Dominant Guided-Autonomous

Table 6.2: A list of the SD pairs used in generating the 3 dimensions of the SAM [142].

These three factors were then used by Bradley and Lang to create the SAM itself, a set of pictoral representations of the three identified factors. This version of the SAM used in this research can be seen in Figure 6.1.

The SAM has been used a select number of times for soundscape analysis, including recently by Watts and Pheasant [129] and combined with concurrent physiological measures by Hume and Ahtamad [224]. However, a direct comparison of SD pair ratings with SAM results has not been conducted. Similar studies, whilst not specifically soundscape related, have investigated the use of Russell's circumplex affect model to study urban environments. Hull and Harvey found a relationship between the physical characteristics of suburban

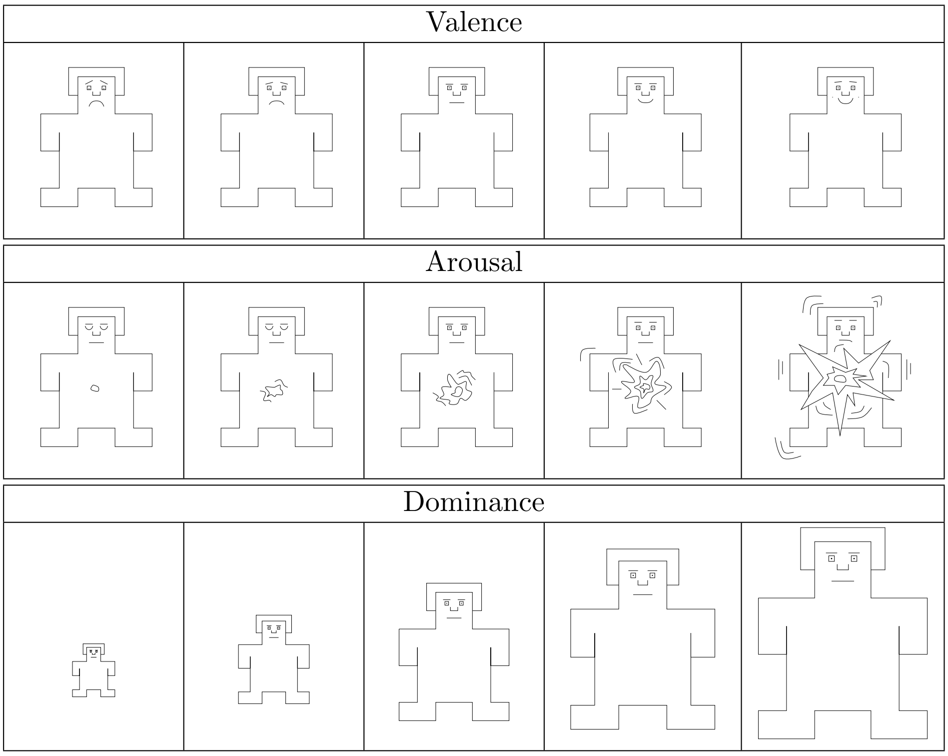


FIGURE 6.1: The Self-Assessment Manikin (SAM), after [142].

parks and affective states: tree density, and the presence of undergrowth and pathways [225], while Hanyu found that green, open, and well-kept spaces are related to positive valence, but that the presence of 'disorder elements' (vehicles, wires) is related to negative affective response [226]. Also of note is Viollon and Lavandier's identification of valence and arousal as the two main underlying factors in assessment of environmental quality [139].

Methodology

The aim of this experiment was to compare SAM and SD pair preference ratings for a set of B-format soundscapes, whilst simultaneously measuring biometric data: specifically, heart rate and galvanic skin response. Table 6.3 contains details of the sound sources present in the two, thirty second long, soundscape clips extracted from the recordings. The recording of these soundscapes was detailed in Chapter 5. From the ten minutes of recording at each soundscape location, two thirty second long sections were extracted. For each of the eight recording locations, then, there are two clips labelled A and B, as seen in Table 6.3.

Area	Location	Clip A Sound Sources	Clip B Sound Sources
Dalby Forest (Rural)	1. Low Dalby Path	Birdsong Owl Hoots Wind	Birdsong Insects Aeroplane
	2. Staindale Lake	Birdsong Wind Insects Single car	Insects Birdsong Water
North York Moors (Rural/Suburban)	3. Hole of Horcum	Birdsong Traffic Bleating	Birdsong Traffic Conversation
	4. Fox & Rabbit Inn	Traffic Car door closing Car starting	Traffic Footsteps Car starting
	5. Smiddy Hill, Pickering	Traffic Car door Car starting Conversation	Birdsong Distant traffic
Leeds City Centre (Urban)	6. Albion Street	Busking Footsteps, Conversation Distant traffic	Workmen Footsteps Conversation Distant traffic
	7. Park Row	Traffic Buses Wind Busking	Busking Footsteps Conversation Distant traffic
	8. Park Square	Birdsong Traffic Conversation Shouting	Workmen Traffic Conversation Birdsong

Table 6.3: A brief (partially complete) summary of the sound sources present in each of the two 30 second long clips (labelled A and B) recorded at each of the eight locations, as detailed in Chapter 5. A further discussion of the sound sources present (as identified by listening test participants) is presented later in this chapter.

The initial plan to play each of these clips twice to each test participant was changed following a pilot test. Doing this resulted in the test taking far too long (close to an hour and a half), with the pilot test participant remarking that they felt *'desensitised'* by the end. As such the decision was made to play the sixteen clips once each only. The clips were played back to the participants using *Spat~*[2] to decode the B-format recordings for presentation over a 16-speaker surround-sound rig.

[2] *Spat~* is a set of real-time spatial processing software features that run in MaxMSP and host a variety of spatial audio features [227], [228]. The MaxMSP patch used in this test is included on the attached data CD

Each participant was first presented with the pre-experiment statement and consent form, followed by a demographic questionnaire and a preview of the subjective assessment questionnaire they then used to rate each presented soundscape. This gave them the opportunity to raise any questions they may have had about the test, and to familiarise themselves with the test procedure.

FIGURE 6.2: Test participant in the listening space. Note the surface provided for the test participant to rest their hand (with Shimmer device attached - see Section 6.1.2 for details) when the clips are played.

They were then presented with the soundscape recordings in random order. After each recording has finished they were given time to fill out a subjective assessment form for each one. The duration of the entire procedure averaged at 37 minutes. For the duration of the test the subject's heart rate and skin conductance were be measured using a Shimmer device, and some of these results are analysed in Section 6.1.2.

All of the questionnaire forms were prepared for presentation online using Qualtrics, with only the pre-experiment statement and consent form, and a sheet of the term definitions presented as a hard copy[3].

A stopwatch was used to record the duration of the test, as well as to allow the start time of each clip being presented to be noted down on a time-log form. These time-log forms are used in the analysis of the biometric data.

Results

This results from this listening test will now be presented, including demographic data and a study of the rating scales responses for both the SAM and SD pairs. This presentation will be accompanied by a discussion of the how the results relate to the hypothesis for the experiment, and what conclusions can be drawn from them as a result. There was a total of 13 participants in this test.

Demographics

Figure 6.3 presents the demographic data collected in the experiment: nationality, gender, age group, profession, and experience with acoustics, and whether or not the individual considered themselves to be sensitive to noise.

It is worth noting that the only nationality present in the current data set it British. Whilst this is not necessarily a surprising fact, it implies that all of the test participants will have been raised in a similar cultural environment (to a certain extent), one that a group of people of a different nationality may not have experienced. Similarly, whilst there is some variation in terms of age and profession, there is a clear bias towards young PhD students who study some aspect of acoustics. This is again hardly a surprise given the circumstances of the conducted work.

One aim when finding test participants was to have an equal number of male and female participants. This is because, whilst no previous research has found gender to affect subjective responses to soundscapes, the physiological difference between men and women

[3]One test participant was presented with a pen-and-paper version of the survey instead, due to their lack of comfort in using the computerised version. This was considered to be suitable for inclusion after Bradley and Lang's findings of correlation values of 0.99, 0.94, and 0.79 for each dimension of the SAM (valence, arousal, and dominance respectively) compared between a pen-and-paper version and a computerised version [142].

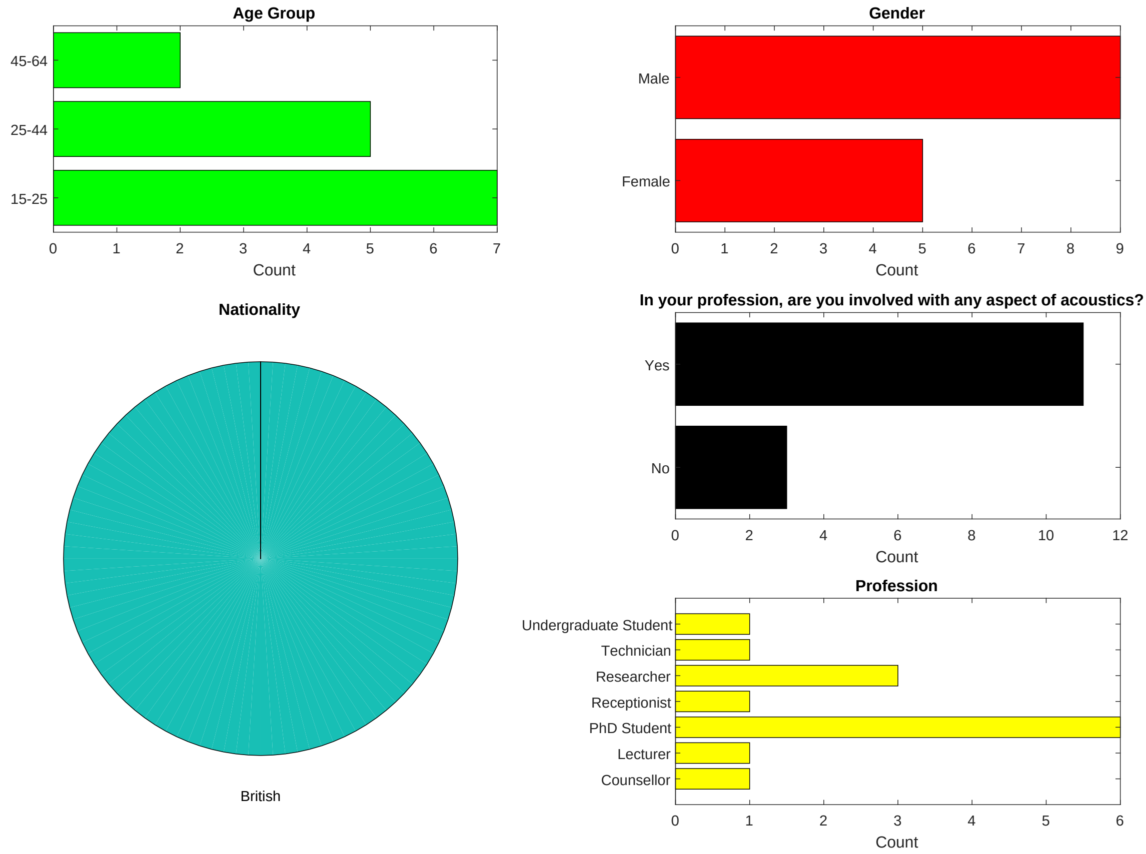


FIGURE 6.3: A summary of the demographic data gathered in the experiment.

can have an effect on biometric measurement [76], [79]. As such, the different number of male and female participants may affect future analysis of the biometric data due to differing levels of statistical significance between gender groups.

The only demographic question not included in Figure 6.3 is the question regarding the importance of a soundscape to experience of an environment. Every test participant said they thought that a soundscape was an important part of this experience.

Ratings

This section will consider some of the notable results from the SD pairs and the SAM. A full set of plots summarising these data are included in the attached data disc. See Appendix B for details.

Location 2: Dalby Forest Lake

Location 2 was the source of clip with the highest mean valence and tranquillity ratings: clip 2B. A plot of the results from the various ratings scales for this location's clips can be seen in Figure 6.4.

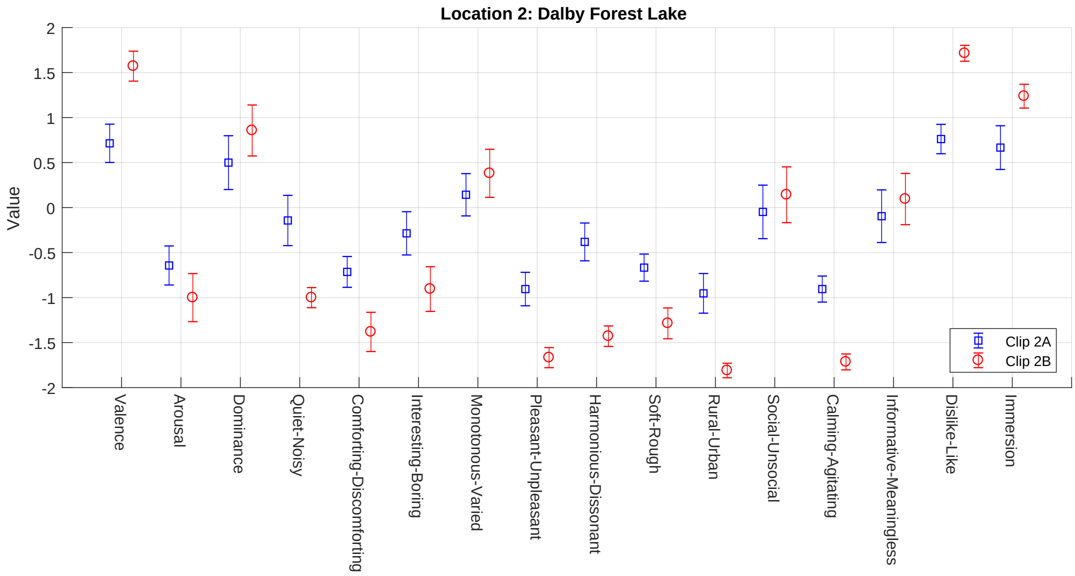


FIGURE 6.4: Mean values and associated standard error bars for each rating scale for the two clips recorded at location 2.

This is significant as the other clip recorded there (clip 2A) was rated very differently due to the presence of single car driving past. For example, one participant said *'The car driving by did affect the perception of a tranquil environment'* which is emblematic of the somewhat different ratings for the two clips recorded at this location.

It is interesting in this case the extent to which the presence of just a single car has had such an effect on an individual's attitude to an environment. This also presents the further question of whether or not the presence of the car in this environment will have such a negative affect on its preference ratings when the recordings are presented simultaneously with visuals.

Location 3: Hole of Horcum

The results for the two clips recorded at location 3, in contrast to those from location 2, show fairly neutral ratings. A plot of the results from the various ratings scales for this location's clips can be seen in Figure 6.5.

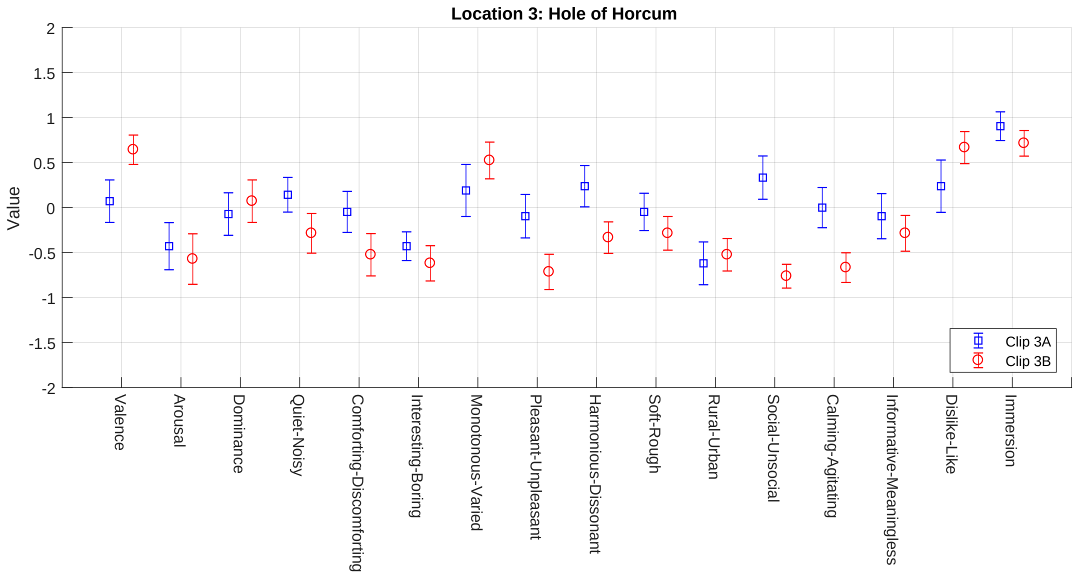


FIGURE 6.5: Mean values and associated standard error bars for each rating scale for the two clips recorded at location 3.

Whilst the two clips recorded here are broadly similar (featuring some fairly close traffic, birdsong, and bleating sheep), where they differ is that the second clip (Clip 3B) includes a man talking just audibly about the location. The 'positive' change in the rating scales as a result of this is possible evidence of the orienting nature of human sounds [159], and is also illustrative of how people will focus on human speech ahead of other soundscape factors. For example, one test participant said it was *'difficult not to try and listen to what the men were saying at the expense of everything else!'*.

The fairly neutral ratings associated with the clips recorded at this location make it suitable for use in later listening tests, as the fairly nondescript nature of the soundscape is at odds with the extraordinary visual setting of the location. It will therefore be interesting to see how results from the multi-modal presentation of the audiovisual recordings made at this location might differ from those resulting from audio-only presentation.

Location 7: Park Row, Leeds

As shown by the valence (and other) ratings the two clips recorded at location 7 were considered by test participants to be amongst the most unpleasant. A plot of the results from the various ratings scales for this location's clips can be seen in Figure 6.6.

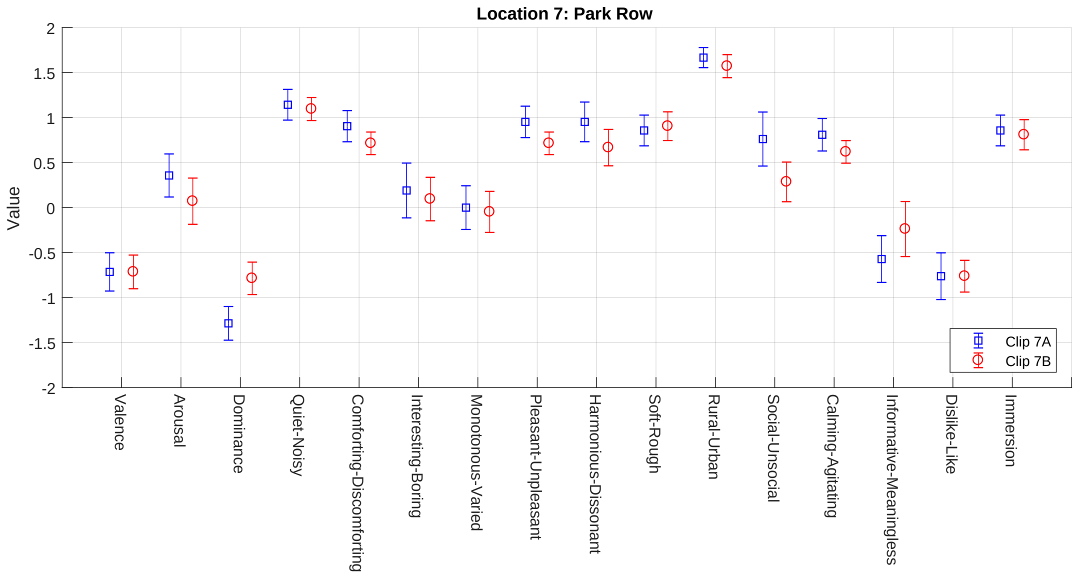


FIGURE 6.6: Mean values and associated standard error bars for each rating scale for the two clips recorded at location 7.

The two primary factors that make the clips recorded at this location rated as so unpleasant are the presence of large amount of traffic noise, and a distant flute player. The recordings made at this location exhibited the highest recorded SPL levels (70 dB L_{Aeq}), and feature constant background traffic and a succession of buses driving past and making loud noises when setting down.

The flute player seemed to be a cause for negative preference ratings from the test participants, as evidenced by comments pointing out *'the flute/whistling playing was very irritating'*, *'is it someone from the Fast Show mariachi band?'*, and identifying an *'annoying piping sound'* and *'tuneless whistling'*. The complete set of listening test data, for this and all of the listening tests in this chapter is included on the attached data disc. See Appendix B for details.

Location 8: Park Square, Leeds

Location 8 represents an interesting companion to location 3 as the listening test results are similarly neutral, as seen in Figure 6.7.

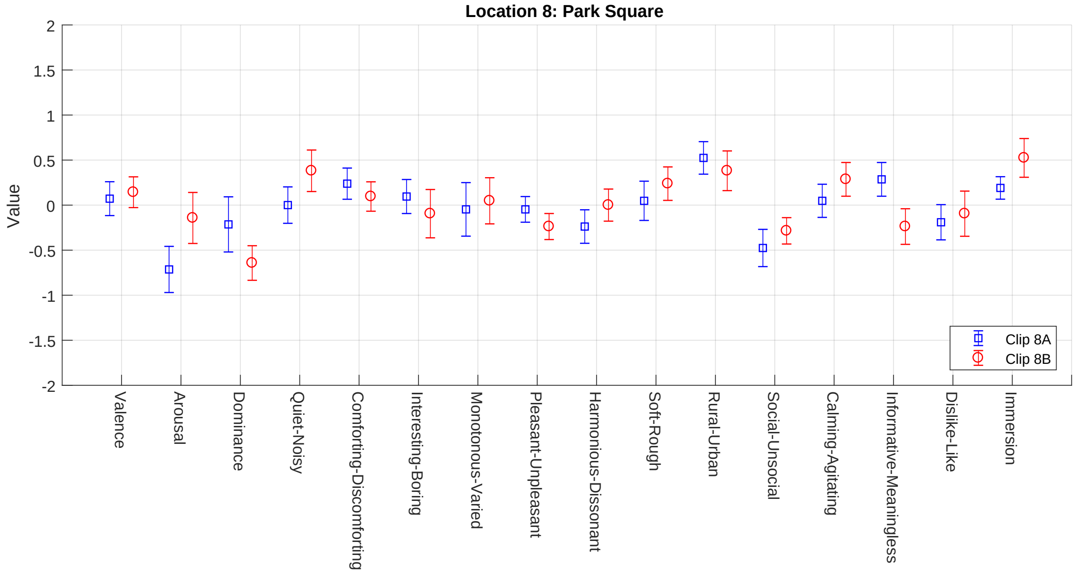


FIGURE 6.7: Mean values and associated standard error bars for each rating scale for the two clips recorded at location 8.

It is perhaps encouraging to see the more positive rating of the clips recorded at location 8 relative to those made at location 7, representing as it does an 'urban green space' (identified by Watts, Miah, and Pheasant as important to soundscape tranquillity rating [198]). Even when presented as a soundscape alone this fact was picked up by some participants (one participant said of one clip recorded here that it *'sounded like a rural setting within an urban one, like a park in a city'*). Also, even where this fact was not identified explicitly, other test participants could identify the combination of the two aural scenes (*'The rural sounds (birds etc...) made the more urban sounds less irritating.'*) as well as identifying how the increased physical distance between the listener and the inner city traffic changed how the presence of that traffic affected their experience (*'Background traffic is much less disturbing than being able to hear individual vehicles'*).

Analysis

Having now presented a summary of some of the more interesting elements of the data, this section will now present a statistical analysis of the result in order to determine how suitable the SAM is for soundscape evaluation.

Correlation Between Rating Items

The correlation between the normalised results was calculated between each pair of rating scales for each participant, using the Pearson's R correlation calculation method [229]. These correlation values were then averaged to give a mean correlation score r for each pair combination of rating scales across all test participants. These r values were then used to determine whether the correlation between the rating scales in each case was significant and either positive or negative. This was determined by comparison with the calculated 'r critical' value which describe the lowest value of r that represents a significant correlation for a given pair of variables [230], [231].

Figure 6.8 shows the results of this significance testing for each pair of rating scales. A white square represents no significant correlation (the white squares with black crosses indicate where the correlation value is for a rating scale's correlation with itself i.e. $r = 1$). The red squares indicate positive correlation, and the blue squares indicate negative correlation. For the SD pairs the direction of the correlation is given where the second descriptor is positive, and the first descriptor is negative. For example, the negative correlation between Valence and the Quiet-Noisy SD pair indicates a significant correlation between increased Valence rating and Quiet-Noisy ratings closer to the Quiet end of the scale.

There are several features of the collected data indicated by Figure 6.8 that merit discussion. One is that the Social-Unsocial, Informative-Meaningless, and Immersion scales are not correlated significantly with any other rating scales. In the case of the Immersion scale this is a positive result, as it indicates that all of the recordings presented in this test are similarly immersive.

The result for the Social-Unsocial and Informative-Meaningless scales is likely a reflection of feedback from test participants indicating a perceived ambiguity between these two scales. In the case of Social-Unsocial there is a certain paradox present where an ostensibly

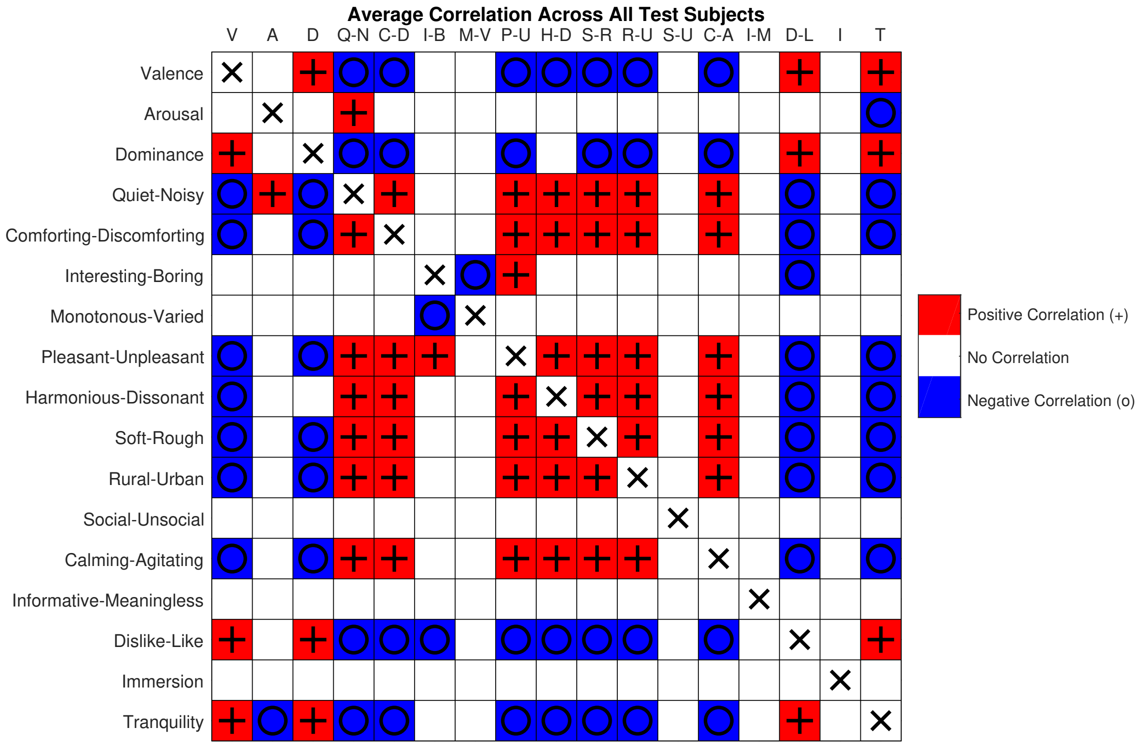


FIGURE 6.8: Plot showing the correlations between the rating scales used in the experiment.

sociable environment (e.g. Location 6 Albion St, Leeds) sounds overly busy and does not feel like an inviting place to take part in social activities: a fact evidenced by the different ratings for this scale for the two clips recorded at this location. On the other hand, a 'quiet' soundscape (e.g. Location 1 Dalby Forest) may sound like an encouraging place for an activity to take place, even if the soundscape itself does not contain any 'social' sounds.

A similar confusion is evident in the Informative-Meaningless scale, as there is no explication of what constitutes meaning and information. For example, traffic noise conveys information regarding the rate and volume of passing cars present in the soundscape but is in many other ways meaningless. This might indicate a blurred boundary between the 'keynote' and 'signal' elements of a soundscape identified by Schafer [13] (as described in Chapter 4). Without a wider context (such as might be provided by a presented visual setting), it is not facile to identify which sounds provide a backdrop, and those sound which comprise the foreground of the soundscape (if indeed there are any such sounds present).

The Interesting-Boring and Varied-Monotonous scales (beyond their correlation with one another) are not correlated with any other scales, apart from Dislike-Like which is negatively correlated with Interesting-Boring. This indicates that the two scales are in effect synonymous. The lack of significant correlation with any of the other scales is again evidence of rating scales that are relatively ambiguous. This is perhaps due to the dependence on an individual's perspective in the determination of how 'boring' a soundscape is. For example an ornithologist might find a recording rich in birdsong to be very interesting in a way that someone apathetic to birds might not. In this way these scales can be seen as 'overly subjective' where it is not just where a stimulus rates on a scale that is a subjective point, but where the meaning of the scale itself also varies between individuals.

Regarding the other SD pairs used in the experiment, Figure 6.8 indicates that the scales Quiet-Noisy, Comforting-Discomforting, Pleasant-Unpleasant, Harmonious-Dissonant, Soft-Rough, Rural-Urban, and Calming-Agitating are all similar correlated with one another and can therefore be considered as representing the same rating scale (with Dislike-Like representing the same scale again but with reversed polarity).

Almost all of these scales are also negatively correlated with Valence (apart from Dislike-Like with which, as one might expect, it is positively correlated). This fact evidences that the Valence dimension of the SAM is just as informative as several of the SD pairs, meriting its future use as a replacement subjective measure.

Another result to note is the lack of significant correlation between Valence and Arousal. The correlation of Arousal with the Quiet-Noisy and Tranquillity scales shows that the meaning of the Arousal scale has been correctly understood by the test participants, with the lack of significant correlation between Valence and Arousal indicating that the two scales are indeed measuring different elements of the subjective experience, even if they can be indirectly related to one another due to their significant correlations with other rating scales. This justifies the future use of the Arousal dimension of the SAM instead of the SD pairs that correlate significantly with it.

It is interesting to see in Figure 6.8 the significant correlation of Valence and Arousal with Dominance, as well as the correlation of Dominance with many of the SD pairs. This may be unexpected given Bradley and Lang's previous findings with the Dominance dimension of the SAM. They found that for certain stimuli the meaning of the Dominance scale could be confusing; for instance when rating the dominance of a photograph of a mutilated corpse the question arises as to whether the Dominance should be rated from the perspective of the viewer or the subject of the photograph [142].

In the case of this experiment the significant correlation of Dominance with other ratings scales indicates that whilst it is a scale that might not provide too much information beyond that given by the Valence and Arousal scales, it is at least explicit to participants what the Dominance dimension means. This is most likely due to the fact that the recordings place the listener in the environment making it clear that the rating is in terms of their own feelings of dominance, not those related to a subject presented in the stimulus.

Ultimately the results shown in Figure 6.8 support the first of this experiment's hypotheses: that the dimensions of the SAM are directly comparable to relevant SD pairs, and can therefore be used in the future for the measurement of subjective experience.

Circumplex Model of Affect

Another way of visualising the Valence and Arousal ratings for the soundscapes presented is to plot them as a Circumplex Model of Affect, a two-dimensional emotional space with arousal as one dimension and valence as the other [148]. This is shown in Figure 6.9 where mean valence and arousal values for each clip have been plotted.

The first thing apparent from Figure 6.9 is that the presented clips indicate a lack of aural scenes that are both highly valent and highly arousing. This leads to the question of whether there can even be such a thing as a valent and arousing soundscape. This would require an exciting soundscape with plenty of activity, but in a pleasant context (for the IAPS and IADS this has included examples such as erotic images and the sound of a roller coaster).

For the recorded locations presented in this experiment the pattern shown in Figure 6.9 is due to increased arousal being associated primarily with the increased presence of traffic (particularly evidenced by the difference between in results for clips 2A and 2B).

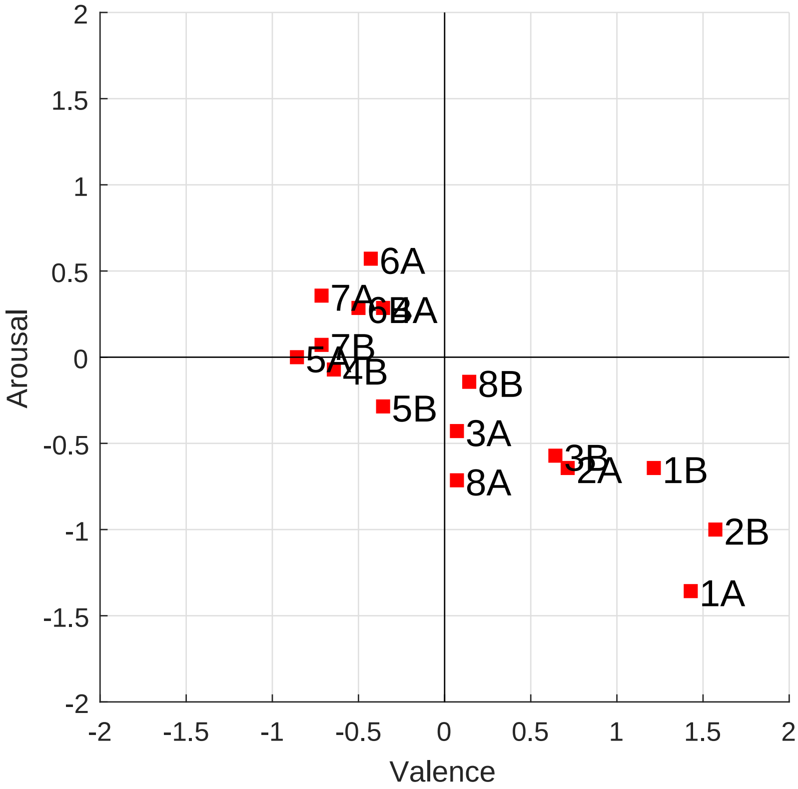


FIGURE 6.9: A plot of the mean arousal and valence values for each clip on the 'Circumplex Model of Affect', identifying their positions in 2D emotional space.

Discussion

The following was given as the first of the stated hypotheses for this experiment: Comparison of soundscape preference rating results between a set of SD pairs and the SAM will show the SAM to be a directly comparable and equally useful tool for the analysis of subjective soundscape experience.

The results from experiment have been shown to support the hypothesis, as shown in Figure 6.8 which indicated the relationships between the ratings pairs and showed the three dimensions of the SAM to be directly comparable and equally as meaningful as the SD pairs.

6.1.2 Biometric Data Analysis

This section will present an analysis of the heart rate (HR) biometric data collected as part of this experiment. The biometric data was collected in this listening test using a

Shimmer device (shown in Figure 6.10) to simultaneously measure heart rate and galvanic skin response. This section considers an analysis of the HR data only. The way in which previous studies have used HR data in a soundscape context has differed significantly. Making use of 30-second long clips, Gomez and Danuser compared the mean HR in the last 15-seconds of presented stimuli with the mean HR in the last 15 seconds of the preceding rest period [80], while Hume and Ahtamad made use of 8-second long soundscape clips and compared the mean HR across the entirety of the clip's duration with the mean HR during the last 8-seconds prior to the clip being played [224]. As such it was decided to examine the HR data recorded here in two ways: firstly comparing mean HR for the entirety of a clip's duration with the preceding 30-seconds; and secondly comparing the mean HR in the last 15-seconds of each clip with the first 15-seconds. These will be referred to as Clip-Rest and Within-Clip comparison respectively from here on.

FIGURE 6.10: The Shimmer GSR+ device, from [232].

Figure 6.11 shows an example HR measurement from the experiment. It shows 1 minute of recorded HR data, where the last 30-seconds correspond to the presentation of a soundscape recording, and the first 30-seconds were part of the preceding rest period. The sample rate for all of the HR measurement made was 51.2Hz.

In order to account for the variation in HR measurement between subjects, rather than using the raw beats-per-minute (BPM) values the change in BPM will be expressed as a percentage. The percentage change in BPM ($BPM_{\%}$) is calculated by using:

$$BPM_{\%} = 100 \cdot \frac{BPM_2 - BPM_1}{BPM_1} \tag{6.1}$$

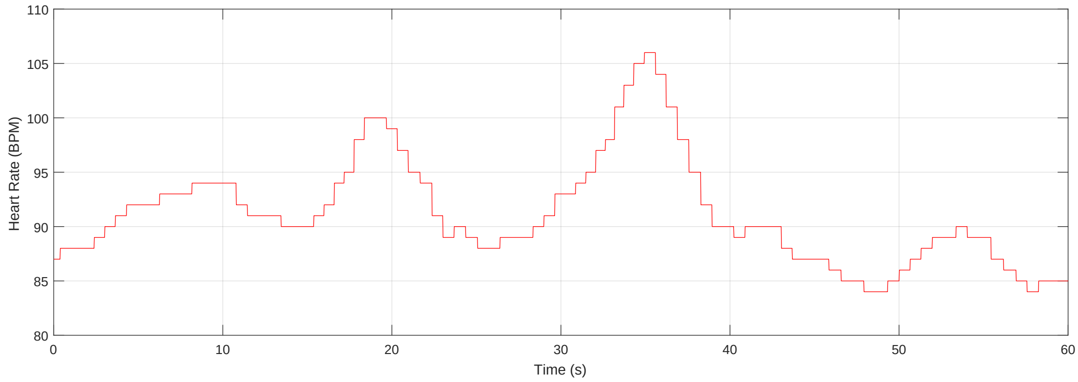


FIGURE 6.11: Example HR measurement taken during a particpant's exposure to a soundscape recording.

where BPM_1 is the mean BPM of baseline measurement that the mean BPM value BPM_2 is being compared with [233]. In the case of the Clip-Rest comparison BPM_1 is the mean BPM of the last 30-seconds before the clip is presented, and BPM_2 is the mean BPM for the whole 30-second duration of the soundscape recording. For the Within-Clip comparison BPM_1 is the mean BPM in the first 15-seconds of the recording, and BPM_2 is the mean BPM in the final 15-seconds.

Clip-Rest Comparison

Figure 6.12 shows the same HR measurement as Figure 6.11 with added lines to show the mean BPM during the rest period prior to the presentation of the soundscape recording, and during the presentation of the clip itself. The calculated percentage change in BPM is also indicated.

Within-Clip Comparison

Figure 6.13 shows only the section of HR measurement from Figure 6.11 during which the participant was being exposed to a soundscape recording. Again there are added lines to show the mean BPM during the first and final 15s of the clip, with the calculated percentage change in BPM also indicated.

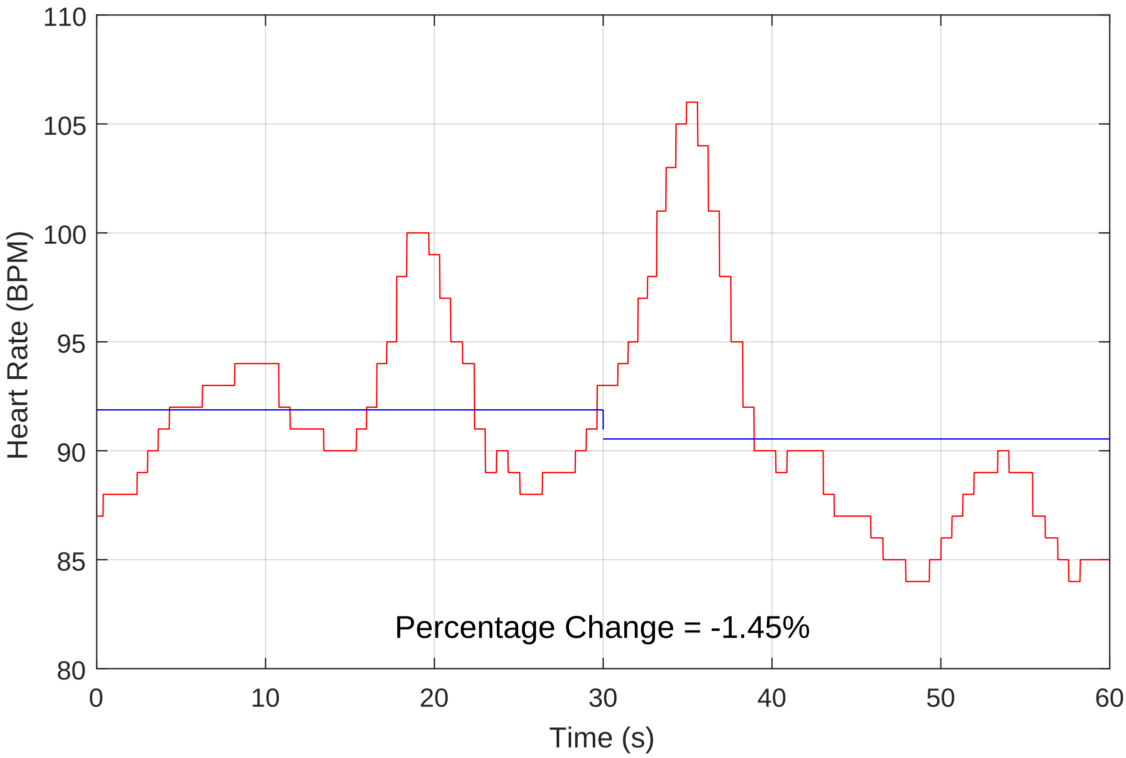


FIGURE 6.12: Plot showing the same BPM reading as presented in Figure 6.11. Also indicated are the mean BPM values in the first 30-seconds (before the participant was exposed to the clip) and the last 30-seconds (during which the participant was exposed to the clip).

Comparison Results

Figure 6.14 shows the mean BPM percentage change for all test participants for each clip (using both comparison methods). The Within-Clip results consistently show an average drop in HR, where the Clip-Rest results include some measurements indicating a rise in HR. The Within-Clip results are therefore in line with previous research [79], [224] that found HR to drop with the presentation of any stimuli. Where these results differ is in the level of change, where here the percentage change is between -8 and -10%; much bigger differences than those found in previous studies. This could be in part due to the use of the Shimmer device itself, as previous studies have made use of different HR measurement techniques.

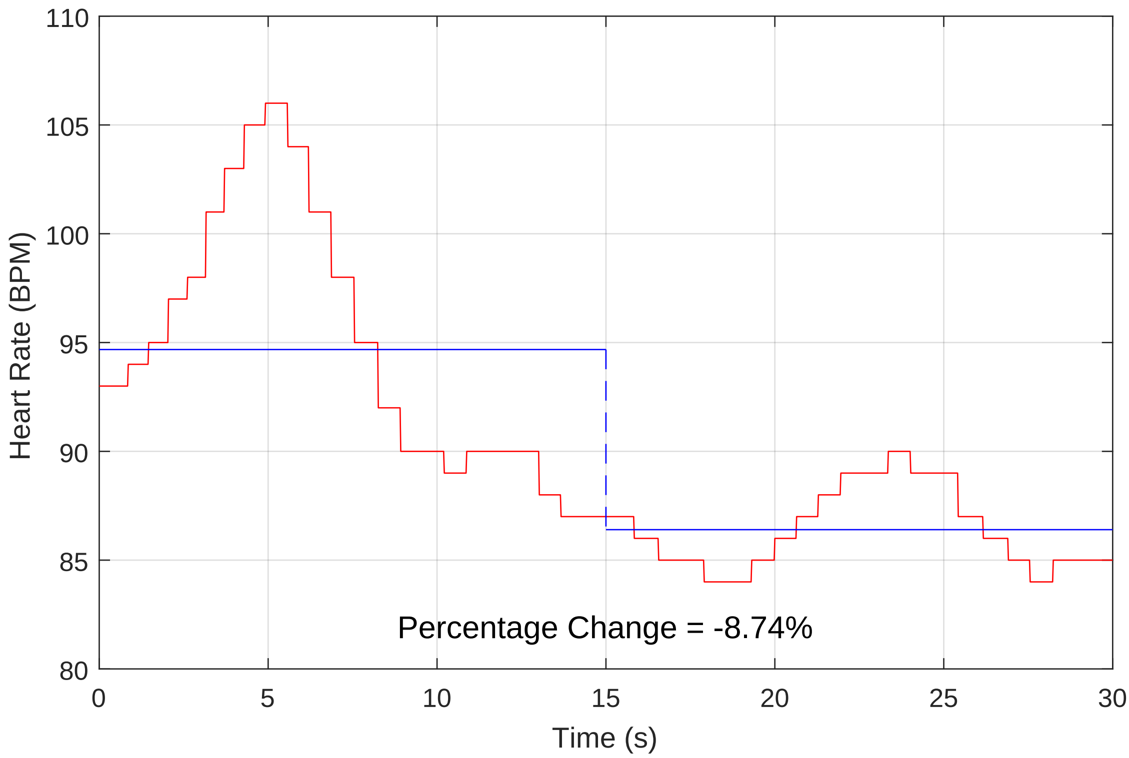


FIGURE 6.13: This is the same example BPM reading as shown in Figure 6.11. Added here are horizontal lines showing the mean BPM value for the first and last 15-seconds of the clip.

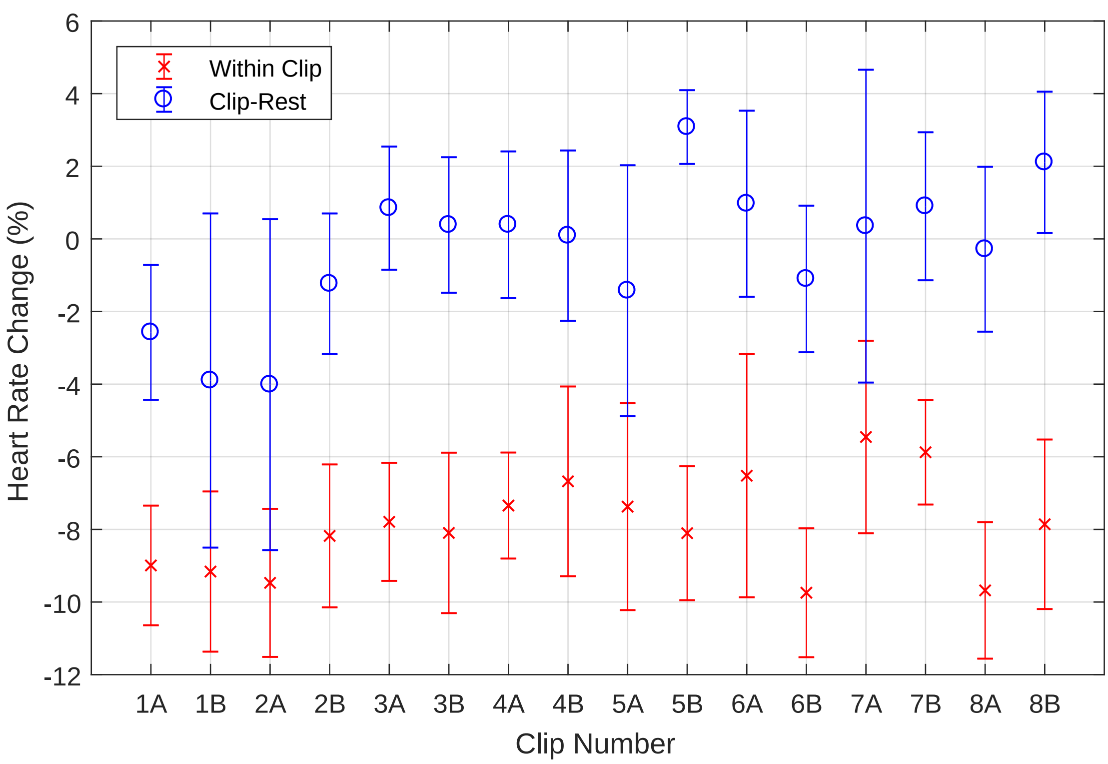


FIGURE 6.14: Mean BPM percentage change (and standard error) across all participants for each clip. Shown are the results for both the Clip-Rest and Within-Clip comparisons.

Analysis Of Subjective And Heart Rate Data

Figure 6.15 is a plot of the mean percentage change (for all subjects) in BPM for both comparison conditions against the mean valence rating for each clip.

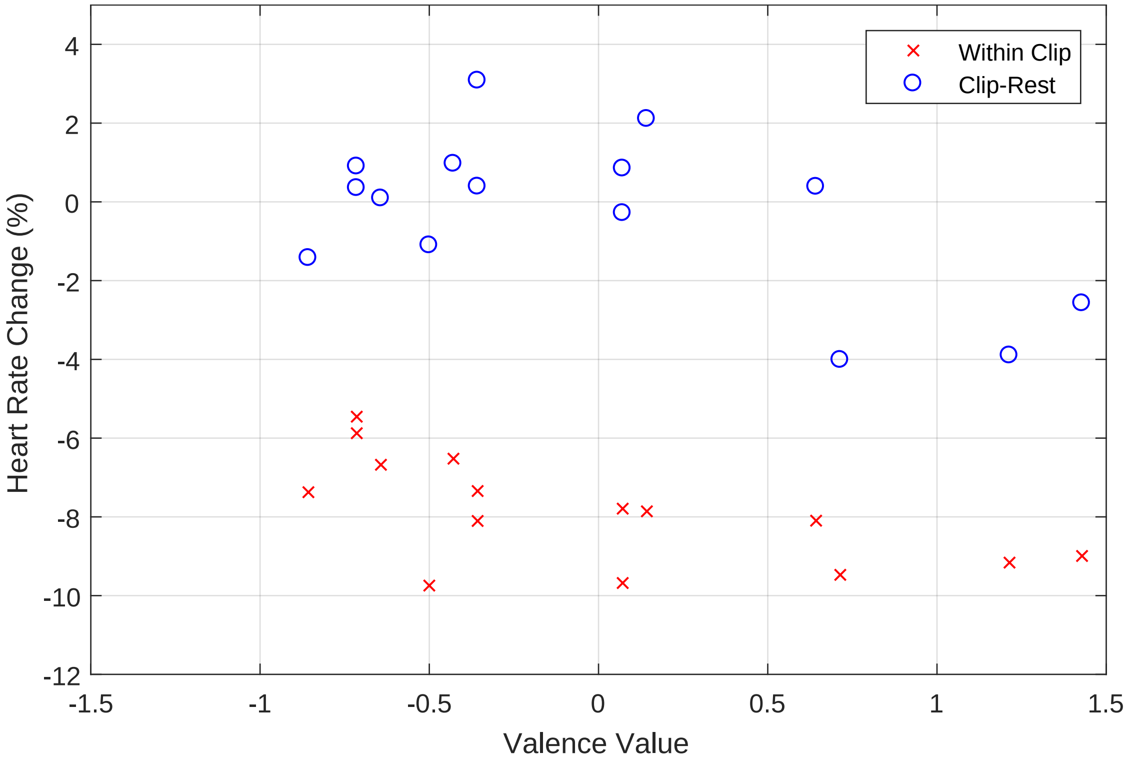


FIGURE 6.15: Plots of the percentage change in BPM (for both Clip-Rest and Within-Clip comparisons) against mean valence ratings.

For both the Within-Clip and Clip-Rest conditions one can see a slight negative correlation between the Valence and HR change. These correlations are statistically insignificant however ($r = -0.1673$, $p = 0.4608$ for the Clip-Rest condition, and $r = -0.0855$, $p = 0.6237$ for the Within-Clip results), indicating no significant relationship between HR and Valence. This is somewhat contrary to the findings of Hume and Bradley [79], [224], although one must bear in mind that their results, whilst significant, only indicate a slight change in HR.

Figure 6.15 is a plot of the mean percentage change in BPM for both comparison conditions against the mean arousal rating for each clip. Here there is a very slight positive correlation where increased arousal rating results in greater increase of smaller decrease in HR depending on the comparison method being considered. Again however these correlations are not statistically significant: $r = 0.0638$ ($p = 0.6128$) in the Clip-Rest case; and $r = 0.1368$ ($p = 0.5735$) in the Within-Clip case. All of the above correlation values have been calculated in MATLAB according to Pearson's R, which is a method suitable for use in the comparing the distribution of variable that makes no assumptions about the normality of the data [234].

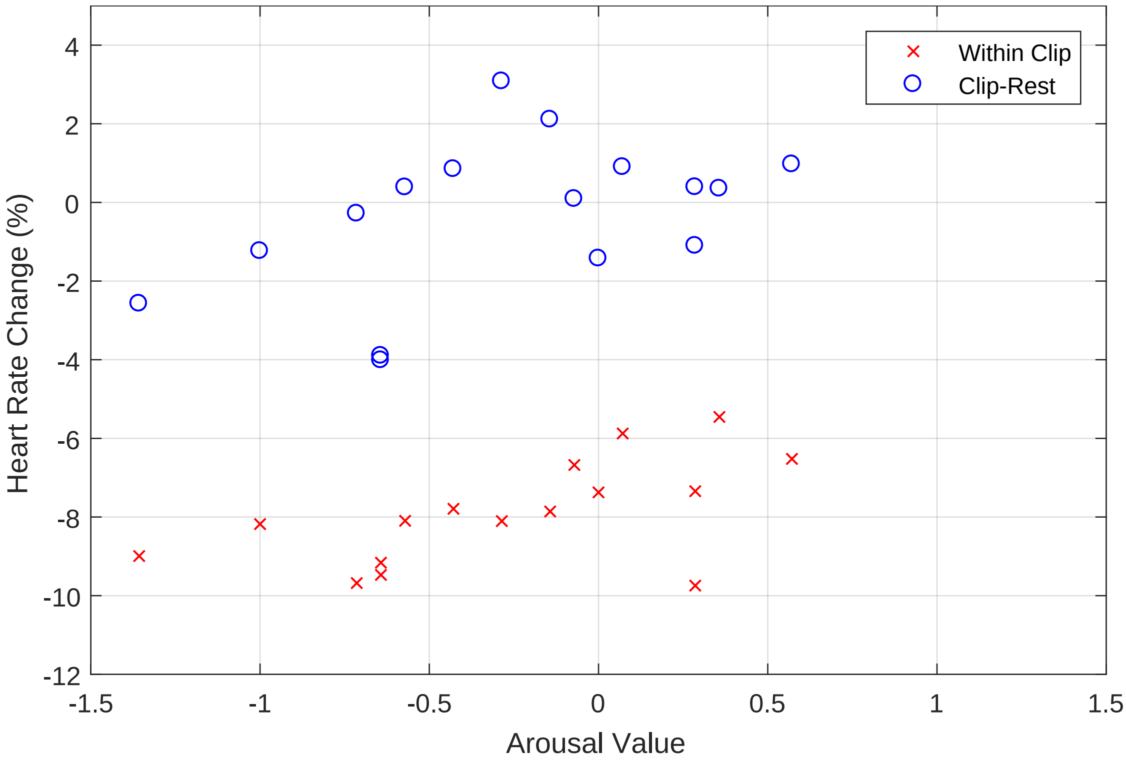


FIGURE 6.16: Plots of the percentage change in BPM (for both Clip-Rest and Within-Clip comparisons) against mean arousal ratings.

Discussion

This section has presented an analysis of HR data and Subjective assessment ratings obtained as part of a listening test presenting test participants with sixteen 30s-long soundscape recordings. The results show that the presentation of these soundscapes to the participants does have an effect on their heart rate, but correlational analysis has not shown this effect to relate to subjective scores of arousal and valence in a statistically significant way. These results therefore do not support the HR related hypothesis for this listening test, which stated: Biometric measurement of heart rate can be combined with preference rating results in order to make explicit connections between subjective and objective soundscape experiences. A similar analysis of the galvanic skin response data collected in this experiment shows that measure to be similarly insignificantly related to the SAM results. It is for this reason that these results have not been included in this thesis. Due to the inconclusive nature of the biometric data analysis as part of this listening test, it was therefore decided to abandon the use of biometric measurements in future experimentation as part of this thesis. It is also worth noting that even if further experimentation had been conducted, subjective evaluation would have been required in order to understand the HR or other biometry in the first instance.

6.2 LT2: Stereo Soundscape Rendering

The second listening test, as covered by this section, makes use of the same FOA soundscape recordings used previously rendered to stereo UHJ and presented in an online listening test. The results are presented in two section: firstly, a comparison of SAM results from this test with those from the first listening test (making use of multi-loudspeaker reproduction) in order to assess the validity of the stereo rendering method used; and secondly, an analysis of the use of soundscape categorisation tools and the SAM, and how soundscape categories and sound sources can be related to the evoked emotional state.

The hypothesis for this first section is that the stereo renderings will result in SAM ratings that are significantly statistically similar to those gathered in the previous listening test making use of FOA renderings of the soundscape clips.

For the second section it is hypothesised firstly that the soundscapes that are rated as being more mechanical will exhibit low valence and high arousal in the SAM results, and that highly natural soundscapes will exhibit high valence and low arousal. Soundscapes highly rated in the human category are expected to exhibit high arousal with valence determined by contextual cues from the other sound sources present.

In terms of the sound sources identified it is hypothesised that birdsong and traffic will be the most commonly identified sources, and that traffic noise and human activity (e.g. conversation or footsteps), when present, will have a significant effect on the category rating.

6.2.1 Stereo-FOA Comparison

In order to present the recorded soundscape material as part of an online listening test, the B-format signals have to be converted to a suitable two-channel format. It was decided to make use of the UHJ stereo format, where the W, X, and Y channels of a B-format recording are used to translate the horizontal plane of the soundfield into two-channel UHJ format (so-called *super-stereo* [235]).

This format was chosen for three reasons. Firstly, since it is a two channel format it can easily be shared online and reproduced over headphones. Secondly, unlike a binaural auralisation, no head tracking is required for accurate reproduction of the soundscape. Thirdly, the use of the stereo UHJ format allows the same B-format recordings presented in the previous listening test to be used here with the spatial content of the W, X, and Y channels preserved in the ultimate reproduction.

The following equations are used to convert from the W, X, and Y channels of the B-format signal to two stereo channels:

$$S = 0.9397W + 0.1856X \quad (6.2)$$

$$D = j(-0.342W + 0.5099X) + 0.6555Y \quad (6.3)$$

$$L = 0.5(S + D) \quad (6.4)$$

$$R = 0.5(S - D) \quad (6.5)$$

where j is a $+90°$ phase shift and L and R are the left and right channels respectively of the resultant stereo UHJ signal [236]. Note that the Cartesian reference for B-format signals is given by ISO standard 2631 [237].

As mentioned in [238], *'Listening directly at UHJ through a normal stereo system sounds usually too much reverberant, as You hear from front any sound which should come from behind You.'* [sic]. It is true that, when listened to over a standard two-channel setup (in this case over headphones), stereo UHJ format audio can sound overly spatial, but it is hypothesised in the context of this work that the inclusion of more spatial aural information could contribute to a more enveloping sense of the environment, compensating for the lack of surround-sound speakers or any form of head tracking (which would be needed for an accurate binaural reproduction).

6.2.2 The SAM

In the previous listening test a direct comparison between the SAM and a set of Semantic Differential (SD) pairs showed the SAM to be an equally informative method that was more intuitive and less time-consuming [189]. It is for this reason that the SAM was chosen for use here in order to make the online listening test as intuitive as possible, especially given the lack of control over each test instance.

The Dominance dimension was not used here as the results from that study indicated some confusion regarding its use in a soundscape context, as well as to keep the test duration as short as possible. The Valence and Arousal dimensions were chosen for use in this study because they are the two dimensions that comprise Russell's circumplex model of affect [148], where they can be used to map emotional states in two-dimensional space.

Methodology

This listening test was presented online using Qualtrics [239]. This was to investigate the validity of using the internet for conducting soundscape related listening tests. Using the traditional method of a lab-based listening test with the material presented over a surround-sound listening rig can be limiting in terms of the number of participants, and the diversity within those participants.

Indeed, Cox identified in [240] that *'the advantage of a web-based methodology is that it enables hundreds of thousands of judgements to be obtained over a diverse population'*. Whilst this study was not conducted on so large a scale, this point holds true as a positive aspect of the use of the internet as a listening test platform.

Obviously, it is also true that the use of the internet as a listening test platform results in a lack of control over playback conditions which may negatively affect the test's results. The ongoing question then is whether this lack of control can be compensated for by increasing the statistical power of the results with a larger number of participants, and through appropriate listening test design. Comparing the results of this test with those from the previous lab-based surround-sound study will therefore go some way to assessing the extent to which this lack of control is an issue.

If the SAM results from each listening condition are shown to be statistically similar to a significant degree then it can be said, for this listening test and these recordings at least, that lack of control over listening conditions in the case of the online version has not negatively affected the results.

The procedure of this listening test included the following steps:

1. An introductory demographics questionnaire including questions regarding age, gender, nationality, and profession. From the results of previous studies [189], [241] it is not expected that any of these demographic factors will meaningfully correlate with the results.

2. An introduction to the questions accompanying each soundscape, covering the SAM, soundscape categorisation, and the sound source identification entry field. Two example questions (using soundscape clips 1A and 7B) are presented in order to get baseline results for the participant. After listening to the first clips the participant was then instructed not to change the playback volume.

3. The 16 soundscape recordings were then presented in a random order. After listening to each one they were asked to rate their experience of the soundscape using the SAM, to categorise it, and to identify the sound sources present. They were also given the opportunity to add any comments they had in an optional text entry field.

Results

This section begins with a brief summary of the demographics of the test participants. This is followed first by a presentation of the SAM results for the online listening test, which are then compared with the same results from the previous surround-sound format test. This comparison is made both graphically and statistically.

Demographics

In total 31 people took part in this study, 19 of whom were male. 16 of these people were aged between 20 and 25, with other participants' ages going up 45-64 age group. People of multiple nationalities took part; whilst the majority were British there were also Swedish, French, Costa Rican, Bulgarian and other nationality participants. The overwhelming majority of participants were either academics or university students; those who were not identified as artists. Of the 31 participants 18 said they had studied, or had worked in, some aspect of acoustics. The average completion time was 27 minutes.

SAM Results

Figure 6.17 shows mean valence and arousal results for the online listening test. Here the clips are numbered from 1A to 8B, corresponding with the detail in Table 6.3.

As shown in Figure 6.1, the SAM was presented to the test participants with 5 images making up each dimension. These images therefore comprise a scale for each dimension ranging from -2 to $+2$, where a negative value is a low rating in that scale (and vice versa), and a rating of 0 sits in the middle of the scale (i.e. the central image).

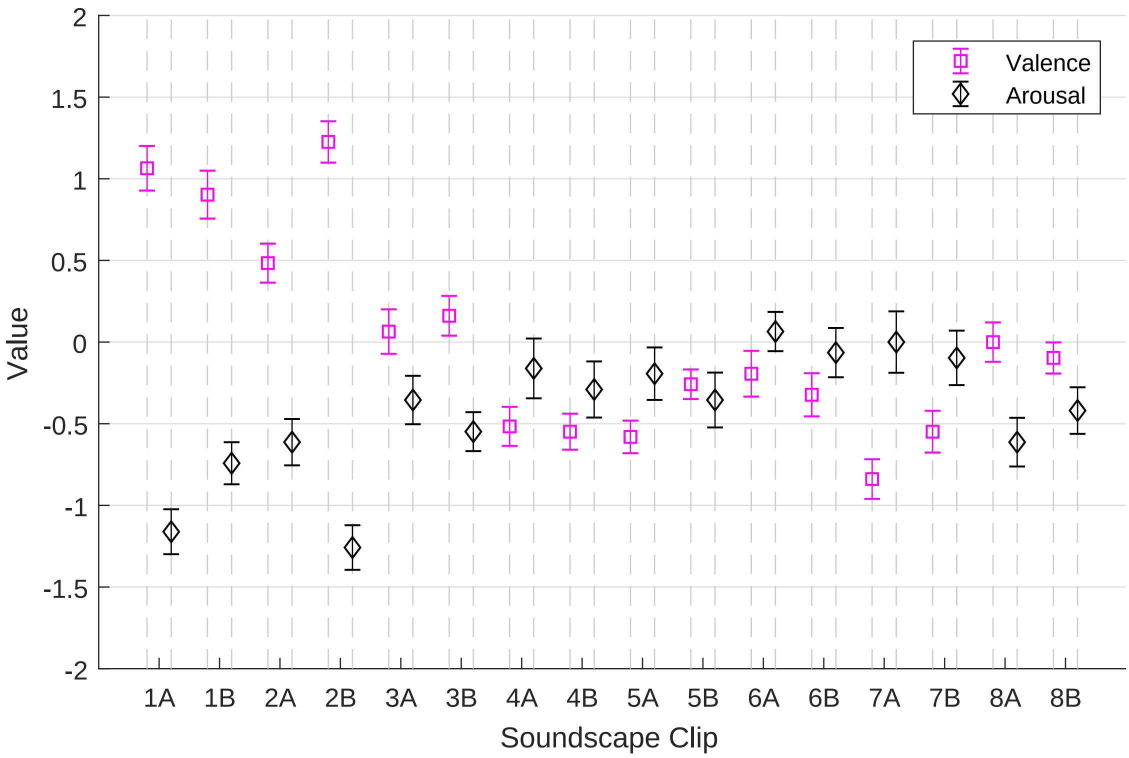


FIGURE 6.17: Mean valence and arousal scores for each clip presented to the test participants in this, the second listening test. The bars indicate the standard error associated with each mean value.

The results shown in Figure 6.17 show that the ratings attributed to each of the clips cover a considerable range. In order to evaluate their significance they will now be compared with the SAM results from the previous listening test presenting the same clip over a 16-loudspeaker surround sound setup (as seen earlier in this chapter). Figure 6.18 shows the SAM results from Figure 6.17 alongside the SAM results from this previous study.

Looking at Figure 6.18 the similarities between the two listening conditions' valence and arousal results become clear. These similarities can be formalised by using a Pearson's R correlation calculation [229]. The results of performing this calculation are shown in Table 6.4.

In this table the correlation coefficients and *p*-values for the comparison of valence and arousal results between the two listening conditions are shown. This coefficient is a value between 0 and 1, where a value of 1 indicates total correlation and a value of 0 indicates no relationship between the two data sets. The *p*-value relates to the statistical significance of the correlation results, where a value of less than 0.05 indicates that the result is statistically significant.

TABLE 6.4: Correlation coefficient and *p*-values comparing the valence and arousal results between the stereo and surround-sound conditions.

	Correlation Coefficient	*p*-value
Valence	0.980	3.0×10^{-11}
Arousal	0.933	1.2×10^{-7}

The results in Table 6.4 show that the both the valence and arousal results exhibit a high level of statistically significant correlation between the surround-sound and stereo results. It is worth noting that, despite the results in Table 6.4 suggesting an overall statistically significant correlation between the two set of results, Figure 6.18 shows some pairs of results that suggest differences in the emotional states evoked by the two listening conditions.

For example, in Figure 6.18(a) clip 3B shows a higher mean valence rating in the surround listening condition. This soundscape clip is one identified as fairly pleasant by most test participants, containing mainly birdsong, wind and water sounds, and a single car driving past. The surround presentation was rated as more valent than the stereo presentation. This is likely due to the car dominating the stereo presentation, where the surround presentation creates a larger aural space, reducing the impact of the car on the scene.

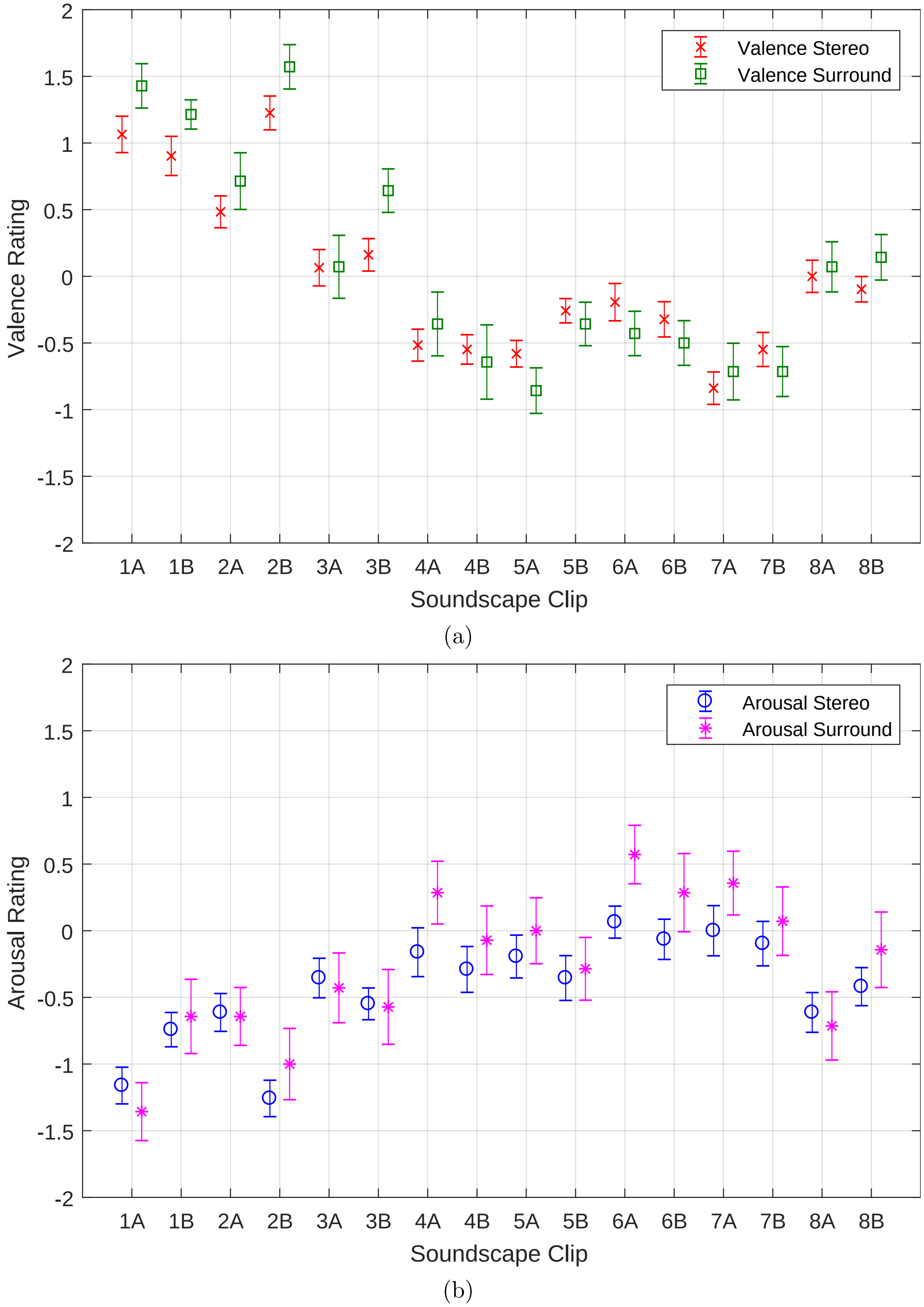


FIGURE 6.18: SAM results comparison between the stereo and surround-sound listening conditions for (a) the Valence dimension, and (b) the Arousal dimension.

Therefore, although there are clearly differences in the valence and arousal results between the two listening conditions, the overall trends of the two sets of data are statistically similar to a significant degree.

Discussion - Stereo and FOA Comparison

The results shown in Section 6.2.2 indicate statistically significant correlations between the SAM results for stereo and surround-sound listening conditions. These results support the initial hypothesis that the stereo UHJ format is robust enough to reproduce soundscape at a level of ecological validity comparable with surround-sound auralisation.

The results are encouraging for future online listening test work. The two-channel UHJ conversion process essentially adds in spatial information that 'shouldn't' be present when played back in standard stereo. Whilst this might create technically 'unreal' soundscapes, the results here indicate that this did not significantly alter the test participants' experiences relative to the presentation of the soundscape recording in the surround-sound case. This is a significant result as it enables future listening test work to be conducted using an online questionnaire tool (such as Qualtrics), allowing for larger numbers of people to be reached than in lab-based experiments.

6.2.3 Soundscape Categorisation and the SAM

The soundscape recordings used in this test were selected in order to cover as wide a range of sound sources as possible. In order to determine what such a set of soundscape recordings would contain, a review of soundscape research indicated that in a significant quantity of the literature [24], [125]–[130] three main groups of sounds are identified:

- **Natural:** These include animal sounds (such as bird song), and other environmental sounds such as wind, rustling leaves, and flowing water.
- **Human:** Any sounds that are representative of human presence/activity that do not also represent mechanical activity. Such sounds include footsteps, speech, coughing, and laughter.

- **Mechanical:** Sounds such as traffic noise, industrial and construction sounds, and aeroplane noise.

The aim of this test is to see how subjective assessments of soundscapes made using the SAM relate to the categorisation of those soundscapes into these three groups. Figure 6.19 shows the category ratings question as presented to the test participants. They were asked to give a rating in each category for each of the sixteen soundscapes. They were also given a free text entry field alongside each one where they were asked to identify the sound sources present. Participants were required to enter at least one sound source into this field in order to progress, but no further requirements were in place. This was done in order to identify the sound sources that were most noticeable to test participants and so potentially dominated their perception of the aural scene.

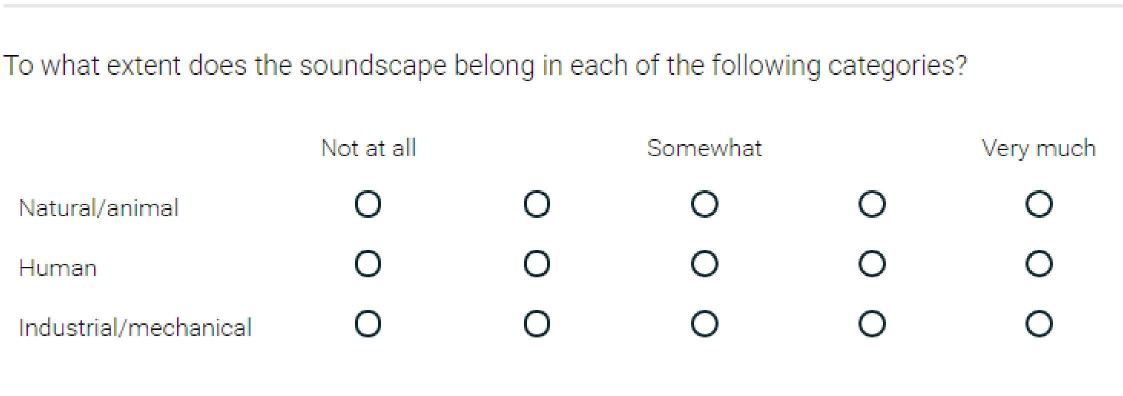


FIGURE 6.19: The soundscape categorisation question as presented to test participants.

Categorisation Results

This section presents the categorisation results from this listening test. A summary of the category ratings is examined first, then an overview of the identified sound sources is shown, after which the sound sources identified for each soundscape clip are categorised themselves and the percentage of these sources that belong to each of the three categories is presented. Correlational analysis is then used to examine the relationships between these metrics.

Category Ratings

Figure 6.20 shows a summary of the category ratings results. As with the SAM results shown in Figure 6.17 it is clear from this plot that the selected soundscapes cover a wide range of ratings in each category.

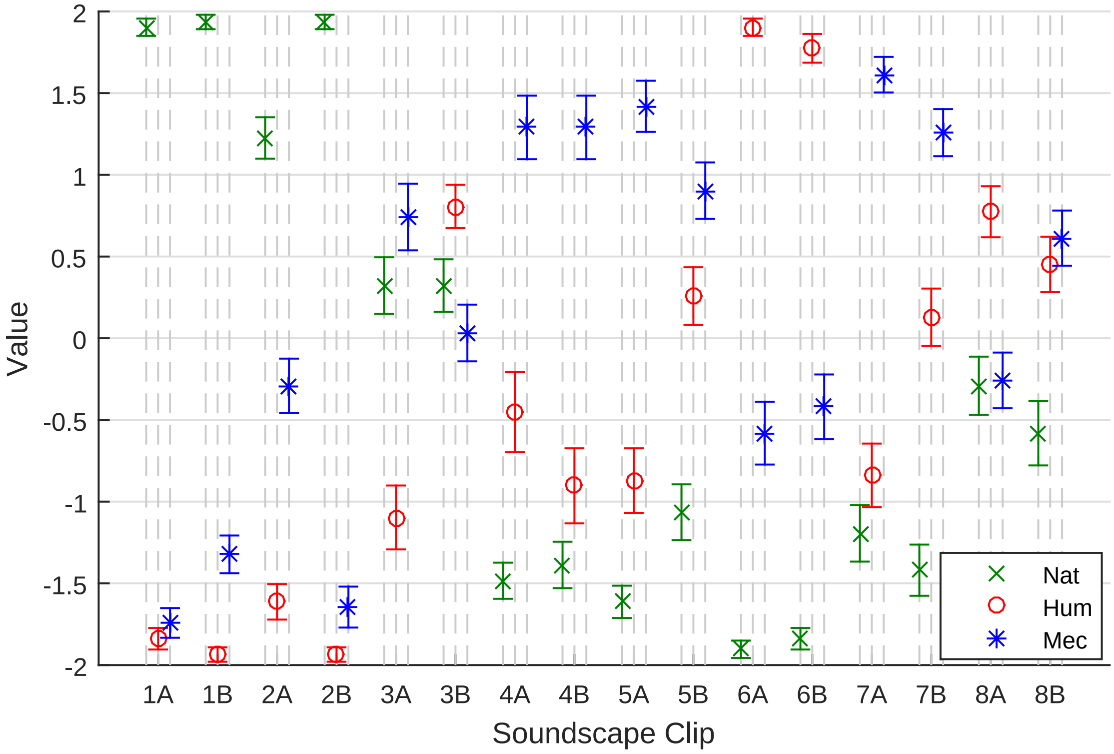


FIGURE 6.20: Summary of category ratings results from this listening test, showing the mean ratings for each category, and the standard error associated with each mean, for each of the 16 clips.

Sound Source Identification

Figure 6.21 shows a pie chart of all of the sound sources identified by test participants. In total there were 1369 sound source instances identified which contained 24 unique sound sources (as listed in Figure 6.21). The overall breakdown of these instances between the three categories is as follows: 38.9% natural, 26.9% human, and 34.2% mechanical.

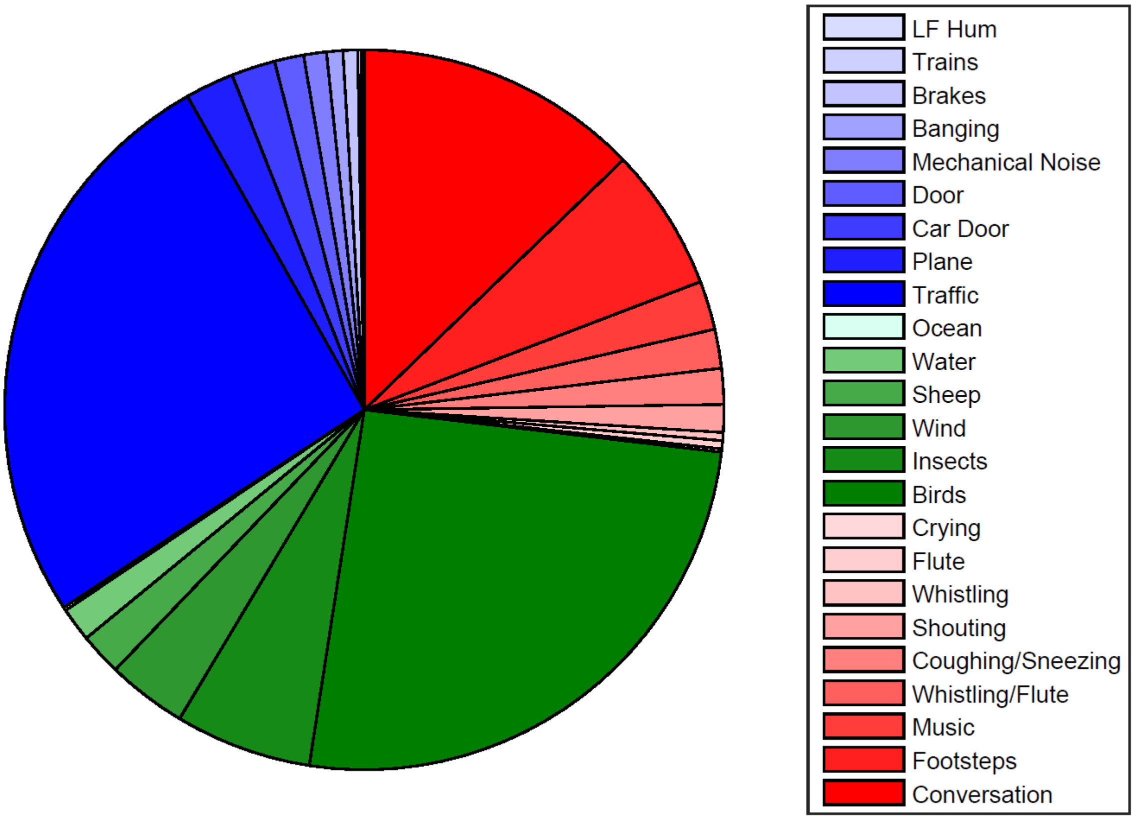


FIGURE 6.21: Pie chart showing the occurrences of the 24 unique sound sources identified in this listening test. For these 24 sources there were 1369 total instances across all clips and all participants. These sources have been colour-coded into categories where blue is mechanical, green is natural, and red is human.

Percentages

In order to make a comparison between the sound sources identified and the categorisation of each clip, it was decided that the sound sources identified for each clip should be grouped by category and then the number of sound source instances in each of these categories should be expressed as a percentage of the total number of sound source instances in each case. Figure 6.22 shows the sound source category percentage break down for each soundscape recordings.

Correlational Analysis

Having now presented a brief summary of the SAM and categorisation results, and having made use of the sound sources identified to create a data set suitable for comparison with

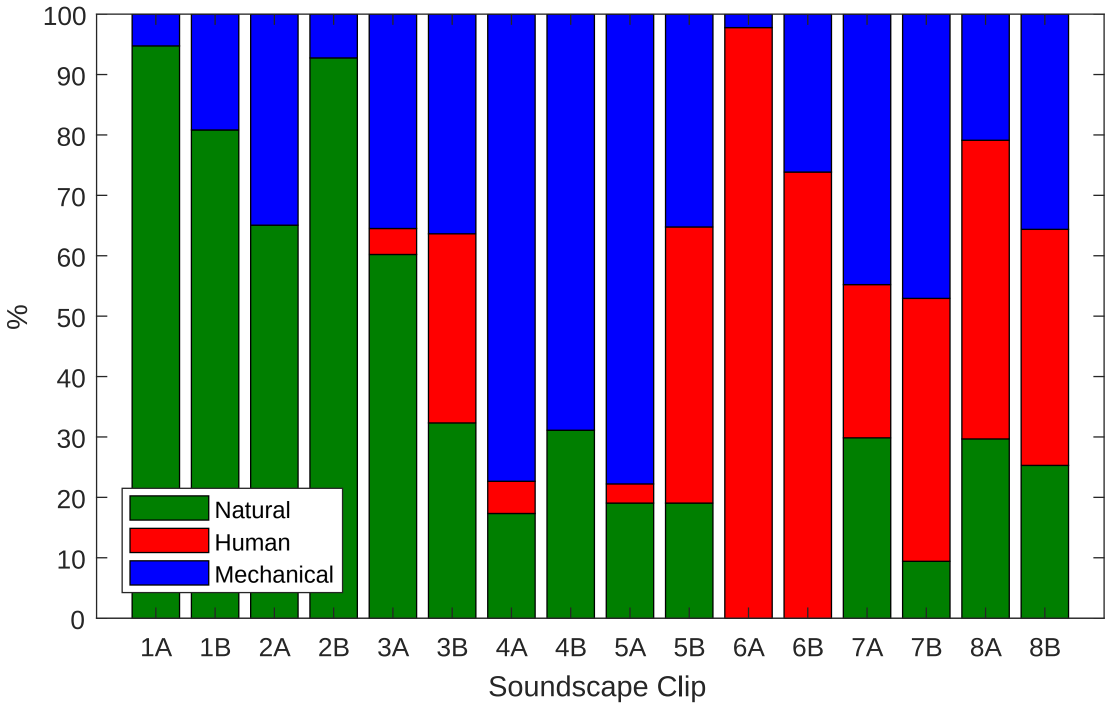


FIGURE 6.22: Sound source identification instances for each clip split into categories and expressed as a percentage.

those result, correlational analysis can be used to compare these three sets of variables to identify relationships between them.

SAM and category ratings

Figure 6.23 shows scatter plots of the mean category ratings against the mean valence and arousal ratings. These valence and arousal results were presented earlier in this chapter in Figure 6.17. The result of a Pearson's correlation analysis [242] of the data presented in these plots can be seen in Table 6.5.

TABLE 6.5: Correlation coefficient and p-values comparing the valence results with category ratings.

	Natural	Human	Mechanical
	R-values		
Valence	0.93	-0.51	-0.90
Arousal	-0.91	0.65	0.70
	p-values		
Valence	2×10^{-7}	0.04	2×10^{-7}
Arousal	9.4×10^{-7}	0.006	0.002

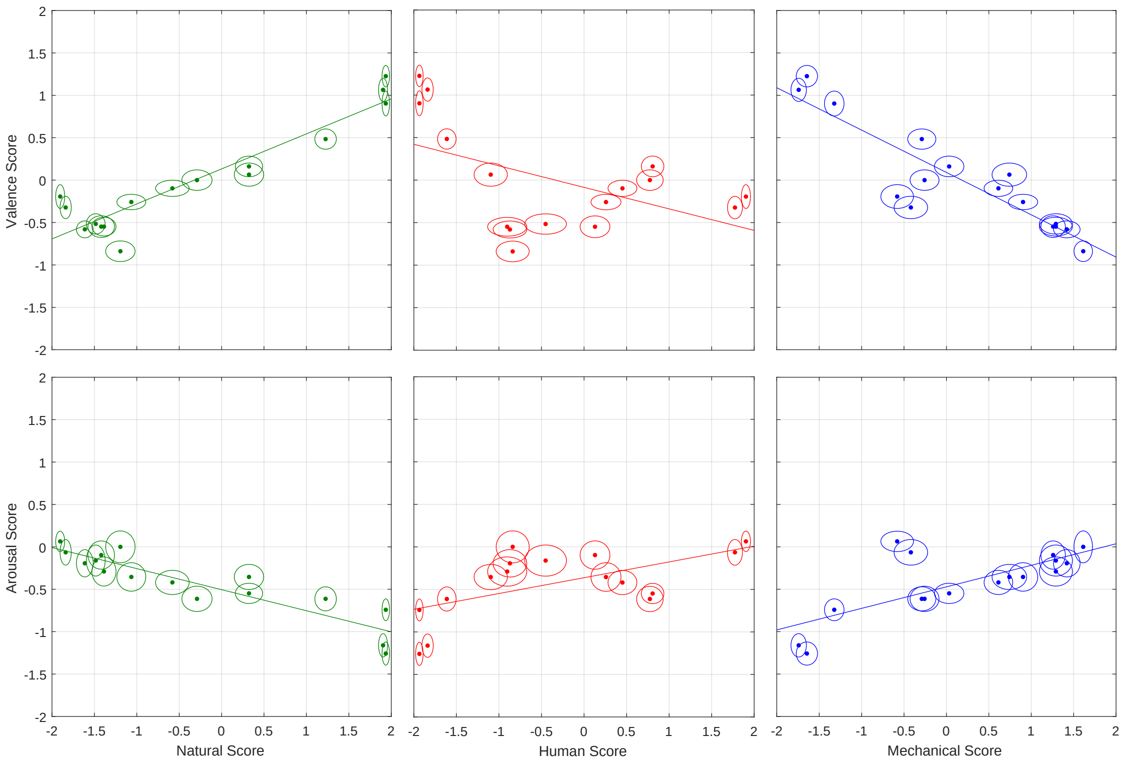


FIGURE 6.23: Scatter plots showing the mean category and SAM ratings for each clip. The ellipses surrounding each data point indicate the standard error associated with each mean value. Trend lines have been included to reflect the correlation results shown in Table 6.5.

The R-values indicate the degree of correlation between two variables, and the p-values indicate the statistical significance of this relationship where, $p \leq 0.05$ indicates a statistically significant correlation [242].

The natural and mechanical category ratings show a very strong set of correlations with the SAM results, where the Natural category is shown to have a highly significant positive correlation with the valence dimension, and a highly significant negative one with the arousal dimension. The mechanical category exhibits the opposite relationship with each dimension. From this it can be said that the locations rated as highly natural are typically rated as pleasant and relaxing, and locations that are rated as highly mechanical are unpleasant and invoke feelings of stress and anxiety. These are the emotions indicated by these SAM results when considering the circumplex model of affect [148].

The human category rating results also show a (just-significant) negative correlation with valence, and a positive correlation with arousal. This is perhaps due to the fact that most of the soundscapes that include human sounds also contain mechanical sounds, particularly traffic noise. Examining the top middle plot in Figure 6.23 gives an indication of why the relationship between the human category ratings and valence is only just significant. Whilst the lowest human category ratings correspond to high valence, as the human category rating increases there is a 'dip' where the valence values for human ratings around 1 are in fact lower than those for higher human ratings. This is due to a disproportionate number of the soundscapes containing very few or no human sounds (i.e. clips 1A-2B, as seen in Figure 6.22).

Table 6.6: Correlation results for the category ratings and sound source percentage results. Indicated here are the R values. The numbers in boldface indicate correlation results where $p \leq 0.05$ (a significant result); and the presence of an asterisk indicates $p \leq 0.01$ (a highly significant result).

	NatCat	HumCat	MechCat	Nat%	Hum%	Mech%
NatCat	-	-	-	-	-	-
HumCat	**−0.69****	-	-	-	-	-
MechCat	**−0.73***	0.19	-	-	-	-
Nat%	**0.95***	**−0.83***	**−0.64***	-	-	-
Hum%	**−0.58**	**0.92***	0.01	**−0.71***	-	-
Mech%	**−0.52**	−0.08	**0.84***	−0.41	−0.35	-

Category ratings and categorised sounds source percentages

Table 6.6 shows the correlation results indicating the relationships between the category ratings and the sound source percentage results.

Here significant negative correlation results between natural and human, and natural and mechanical category ratings can be seen, indicating that soundscapes rated as more natural were typically also rated as less human and less mechanical. Perhaps surprisingly there is not a significant relationship between the human and mechanical category ratings.

Sound source percentage results

The correlation results comparing the sound source category percentage results in Table 6.6 show that the only significant relationship is the negative correlation between the percentage of sound sources identified that are in the natural category, and those belonging to the human category. This is indicated by the sound source categorisation results for clips 6A and 6B as shown in Figure 6.22.

For both of these clips the majority of sound sources identified (> 70%) were in the human category, with the remaining sound sources belonging to the mechanical category and no identified sounds in the natural category. Likewise the results for clips 1A-2B contain very high percentages of natural sound sources, some mechanical sources, and no human ones.

Categories and percentages

There is a significant correlation between the category ratings and percentage of identified sound sources for each of the three categories. This gives some indication that the percentage breakdown of the identified sound sources into each category is a valuable metric that in some way reflects the overall categorisation of a soundscape.

The natural category rating shows significant negative correlation with the human and mechanical sound source percentage metric, as does the natural sound sources percentage metric with the human and mechanical category ratings. The two pairs of variables that show no significant correlation in this group are human category rating and mechanical percentage metric, and the mechanical category rating and the human percentage metric.

Key Sound Sources

As identified in Figure 6.21, the three most commonly identified sound sources were traffic noise, birdsong, and conversation; sound sources belonging to the mechanical, natural, and human categories respectively. Each one will now be considered to see how the presence of these sound sources in the soundscapes has impacted on the SAM and category ratings of those soundscapes.

Birdsong

Table 6.3 indicates that the soundscape clips recorded in Dalby Forest contain many instances of birdsong. As can be seen in Figure 6.20, these soundscapes (clips 1A-2B) were rated as most belonging to the natural category, and, as shown in Figure 6.18, received the highest valence and lowest arousal ratings of all the soundscape clips. This result is in accord with the correlation results presented in Table 6.5 that indicate the same relationship between higher natural category ratings and the SAM results.

It is also interesting to note the difference in category rating between clips 7B and 8A. These clips were recorded at two nearby locations in Leeds: clip 7B was recorded next to a busy road, and clip 8A was recorded in the middle of a small park. These two clips are fairly similar, but the presence of birdsong in the latter is partly responsible for a significant difference in the natural category rating, which contributes to the relative increase in valence rating (and decrease in arousal rating) as seen in Figure 6.18.

Traffic

The correlation between the mechanical category ratings and the SAM results (negative for valence, positive for arousal), as indicated by Table 6.5, is shown clearly in the difference between SAM results and category ratings for clips 2A and 2B. These two clips were recorded in the same location next to a lake in a forest: Clip 2A contains the recording of a single car driving on a nearby road where clip 2B does not. As shown in Figure 6.20 the presence of this car is likely responsible for the big increase in the mechanical category ratings, and a big decrease in the natural category rating, for clip 2A relative to clip 2B.

Figure 6.18 shows the effect of this car on the SAM ratings for this clip. It is interesting to note how big a difference the presence of a single car can make. This same pattern can be seen to a lesser extent in the results for clips 1A and 1B. These were both recorded at different positions in the forest next to a footpath away from any roads. In this case it is the presence of an aeroplane passing overhead in clip 1B that is not present in clip 1A that results in a similar pattern of differences between the two clips.

It is also worth noting that the clips recorded at roadside locations were rated as having the lowest valence levels (clips 4A-5A and 7A-7B), and that these clips were also rated as most belonging to the mechanical environment category.

Conversation

The most significant sound source identified as belonging to the human category was conversation. Its presence in clip 3B and absence in clip 3A, whilst not responsible for any appreciable difference in SAM results, produced a big change in category ratings resulting in a much higher human category rating for clip 3B.

Figure 6.22 also indicates the perceptual dominance of conversation when present in a soundscape recording - clips 6A and 6B are the only soundscapes for which no natural sound sources were identified by test participants, despite some birdsong and wind noise being present in those recordings.

6.2.4 Summary

This section has shown the results of an online listening test presenting stereo UHJ renderings of B-format soundscape recordings. Test participants were asked to describe the emotional state evoked by those soundscapes using the SAM, rate the extent to which each soundscape belonged in three environmental categories, and identify the key sound sources present in each one.

The purpose of this test was to identify any relationships between the subjective assessment of a soundscape and the extent to which that soundscape is perceived to be natural, human, and mechanical. A secondary aim was to use the sound sources identified by test participants to identify the sound sources that are the most important to soundscape perception, both in terms of subjective experience and category rating.

Correlation analysis indicated strong relationships between the natural and mechanical category ratings and the valence and arousal dimensions of the SAM. The natural category ratings were shown to be positively correlated with valence, and negatively correlated with arousal. The mechanical category ratings showed the opposite correlations in each case. The human category ratings showed similar but less strong correlation results to the mechanical category ratings with some ambiguity in the relationship between the human category ratings and the valence dimension of the SAM. These results support the study's hypotheses which predicted a negative emotional response to highly mechanical soundscapes, and a positive emotional response to highly natural soundscapes. The hypothesis regarding emotional response to soundscapes rated as highly human was also suggested with a clear correlation between human rating and arousal and a more ambiguous relationship with valence.

These results indicate that more natural soundscapes invoke a relaxation response (low arousal, high valence), and that more mechanical soundscapes invoke a stress response (high arousal, low valence). The human category rating results' relative lack of significance indicate a contextual dependence where human sounds are not necessarily disliked or stressful, but are mainly present in soundscapes where traffic noise (i.e. mechanical sound) is also present.

An analysis of the sound sources identified by test participants has indicated a key sound source for each category: conversation, birdsong, and traffic. The identification of these sources supports the hypothesis predicting which sound sources would be most commonly identified, and as such would be the most useful predictor of soundscape categorisation. The presence and absence of these sources in the soundscapes presented was examined by studying particular examples. This examination demonstrated where these sound sources in particular produced differences in the SAM and category ratings that exhibited with correlation relationships previously identified.

The results presented in this section certainly confirm the SAM as being a powerful tool for soundscape analysis, with soundscape categorisation and sound source identification offering suitable methods to pair with the SAM to offer further insight into which aural features can most dramatically affect emotional responses to environmental sound.

6.3 LT3 and LT4: Audiovisual Tests

This chapter will now present the results from two listening tests making use of audiovisual stimuli to investigate the phenomenon of cross-modal perception. The first, preliminary, listening test (LT3) presents participants with stereo UHJ renderings of B-format soundscape recordings and still panoramic images. The results from this preliminary test then inform a second, main, listening test (LT4). This presents some of these same environments in virtual reality (VR), making use of YouTube as a platform for binaural renderings of the original first-order-Ambisonic (FOA) audio and full motion spherical video.

These two tests have been designed in order to answer the following questions arising from the previous two listening tests presented in this Chapter:

- Will the presence of different visual features change the SAM results, indicating a change in the evoked emotional states?
- Will the key visual features identified be the same as the aural features?

- Will the presence of visual features change which aural features are identified by test participants?
- Will the presence of visual elements change the category ratings for the presented environments?

The necessary knowledge base, required in order to understand the research motivation and development, will now be presented. The methods applied are then examined, including: the collection of the soundscape and visual data; the soundscape evaluation tools used; and the content creation methods and test procedures used.

6.3.1 Background

Cross-modal Perception

Cross-modal perception is where the stimulation of one sensing modality (for example vision) can influence the experience of another (e.g. hearing). A famous example of this phenomenon is the McGurk effect [243] where a change in the appearance of mouth movement can alter the phoneme heard in recorded speech.

In a soundscape context, cross-modal perception has been considered as a way of understanding how the visual setting of an environment can change the perception of that environment's soundscape. For example, Lercher and Schulte-Fortkamp showed living on a 'pretty' street could reduce noise annoyance [116] and Viollon et al. found that exposure to still images of natural environments incorporating natural features reduced the perceived 'noisiness' of a soundscape [126]. Research into this area is of great importance to human health and well-being, both in terms of reduced stress due to lower levels of noise annoyance and other health effects (for example, a patient's recovery following an operation has been shown to be faster if the patient has access to a window with a pleasant view [158]).

Green Infrastructure

Broadly speaking, when considering noisy soundscapes, the kind of visual features that may be present to improve one's experience of noise can be collected under the term Green Infrastructure. A definition of Green Infrastructure is given in [244]:

> *'It can be considered to comprise of all natural, semi-natural and artificial networks of multifunctional ecological systems within, around and between urban areas, at all spatial scales.'*

Whilst the acoustic impact (in terms of noise level reduction, acoustic absorption to reduce reverberation times etc.) of green infrastructure may be minimal, the impact on perception of sound may be much more pronounced [245], [246]. An underlying motivation for this research is to investigate to what extent the presence of green infrastructure and other natural, pleasant, visual features can negate (or at least curtail) the negative effects of acoustic noise in a soundscape. The Biophilia thesis originates from the field of environmental psychology, and posits that human beings have an innate appreciation for, and affinity with, natural environmental features: particularly water and vegetation [62].

6.3.2 LT3: Still Panoramic Images and Stereo UHJ Audio

The context, procedure, and results of the preliminary listening test (making use of still panoramic images) will now be presented.

Still Panoramic Images

Figure 6.24 shows two example panoramic images as used in the preliminary test. The photos used to create these images and the video footage used in the next listening test presented in this section were captured using the GoPro camera rig as described in Chapter 5. These panoramas were created using Kolor Autopano [206] to stitch together the still images from the GoPro cameras into single equirectangular images.

Preliminary Test Procedure

The listening test was presented using Qualtrics [239] to administer the questions for the test participants to respond to, and MATLAB to play the stereo UHJ audio and present the panoramic images using FSPViewer [247] (a freely downloadable viewer for spherical panoramic images). Presenting the images in this way allowed participants to

(a)

(b)

FIGURE 6.24: Two example panoramic images as used in the preliminary listening test: One in Location 1, Dalby Forest (a), and the other in Location 7, Park Row in Leeds (b).

click-and-drag the panoramic image to 'look' around the environment, which they were instructed to do.

All 16 soundscape clips (in stereo UHJ format as previous) were presented to the test participants in a random order (with corresponding still panoramic images) and were preceded by two orienting stimuli. After listening to and viewing each one they were asked to rate their experience in terms of valence and arousal, to rate it in terms of the three soundscape categories, and to identify the sound sources and visual results present. This test was completed by 11 test participants. The purpose of this preliminary test was to compare SAM, category, and sound source results with those gathered in the previous listening test (presented in Section 6.2) in order to determine the which of the recorded soundscape scenes are of sufficient interest for use in the main listening test.

Preliminary Test Results

In order to compare the SAM and categorisation results between this test and the previous audio-only test, the suitability of the statistical method applied must be determined. The application of a Shapiro-Wilk's test [248] to all of the rating scales for each test stimuli in the audio-only case showed only one of the sets of results to be normally distributed. The application of the same test to the preliminary visual test results showed 16 of the 80 sets of results to be normally distributed.

As such, in order to make comparisons between the results for the different stimuli, the Mann-Whitney test was used [249]. This test is suitable for comparing the values of two random variables where those variables are not normally distributed [24]. It is also suitable for comparing variables with small, arbitrary, sample sizes, including where the sample sizes of the two variables are different.

The purpose of applying the Mann-Whitney test was to indicate where the test results were significantly different for each of the five rating scales (Valence, Arousal, Natural, Human, and Mechanical) when comparing the results for the audiovisual stimuli with the audio alone. Figure 6.25 shows the Mann-Whitney test results for the preliminary listening test data, indicating these significant differences. Dark squares indicate a significant difference at 95% confidence ($p < 0.05$), and Light marked squares at 90% confidence ($p < 0.1$). White squares indicate no significant difference at either confidence level. The next section will discuss theses results.

Significant Differences

The three clips showing a significant difference in Arousal values are 6A, 6B, and 7B. For all three of these clips the arousal rating value was significantly larger when the clip was presented with the visual stimuli, as seen in Figure 6.26. Both of these recording locations

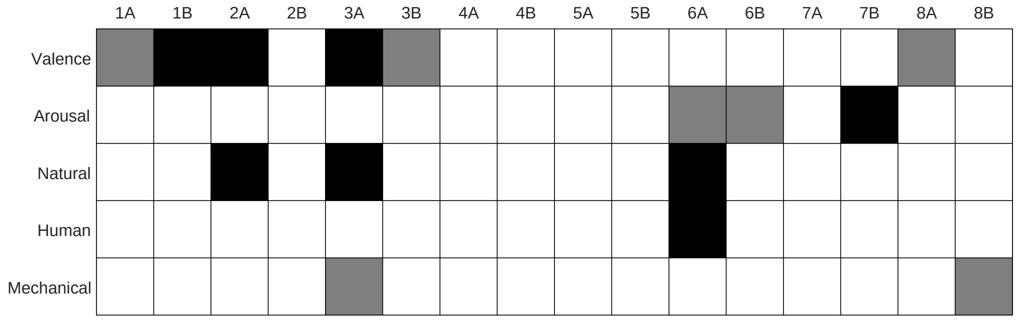


FIGURE 6.25: Mann-Whitney test results for the preliminary listening test, comparing results for each of the five rating scales for each of the 16 test stimuli when presented as the soundscape alone and with accompanying still panoramic images. Dark squares indicate a significant difference at 95% confidence ($p < 0.05$), and Light marked squares at 90% confidence ($p < 0.1$). White squares indicate no significant difference at either confidence level.

were in Leeds city centre: one next to main road (7); one on a pedestrianised street (6). This increase in arousal is therefore possibly due to the presences of cars and people in the images of the scenes that are not so pronounced in the soundscape recordings.

The 6 clips showing a significant difference in Valence values are 1A-2A, 3A-3B, and 8A. As with the arousal results, for all of these clips the presence of visual stimulus results in an increase in valence. The mean valence rating for each clip in each listening condition can be seen in Figure 6.27. For clips 1A and 1B this is unsurprising: the soundscape clips contain some birdsong and insect noise, but despite their hi-fidelity (where the sound sources present are clearly defined with little background noise [13]) there is little information given to indicate the features of the recording location. As such it is to be anticipated the presence of the visual features with the soundscape results in an increased valence rating.

For clip 2A a similar effect can be observed, due to the presence of a single car driving past. These results suggest that the visual setting (greenery and trees, peaceful lake, big sky) results in a significantly increased valence rating.

The significant increases in valence value for the audiovisual presentation of clips 3A and 3B also show the same effect: the aural information in these clips contains some natural sounds and traffic noise that indicate little about of the surrounding countryside of the North York Moors national park.

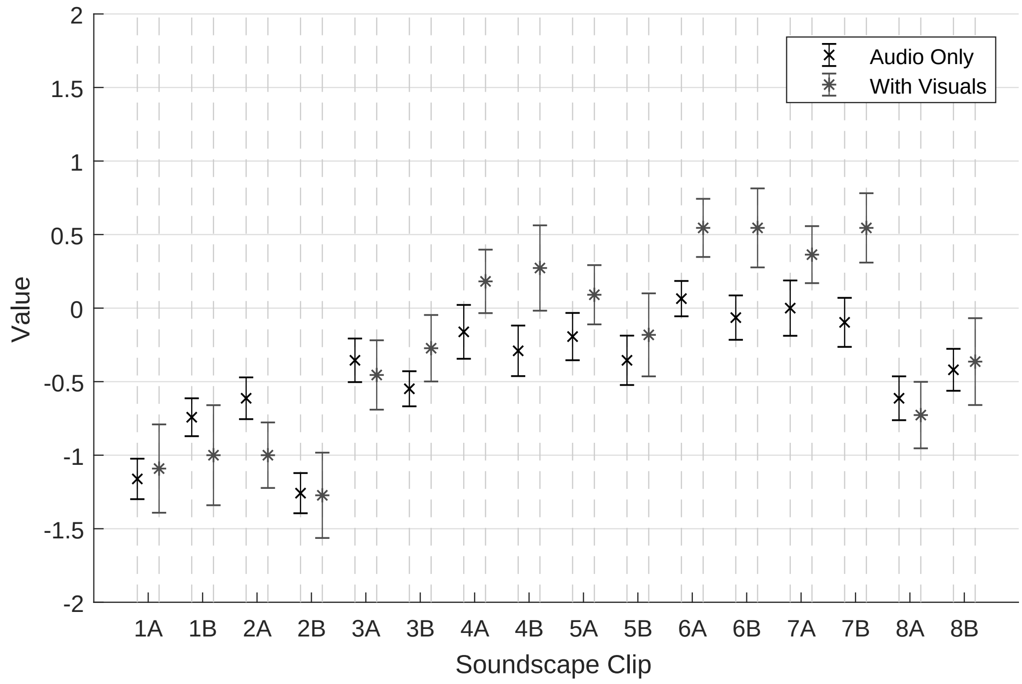


FIGURE 6.26: A comparison of the listening test results for Arousal between this listening test ('With Visuals') and the previous audio-only test. Indicated for each clip for each listening condition is the mean rating with error bars indicating the standard error.

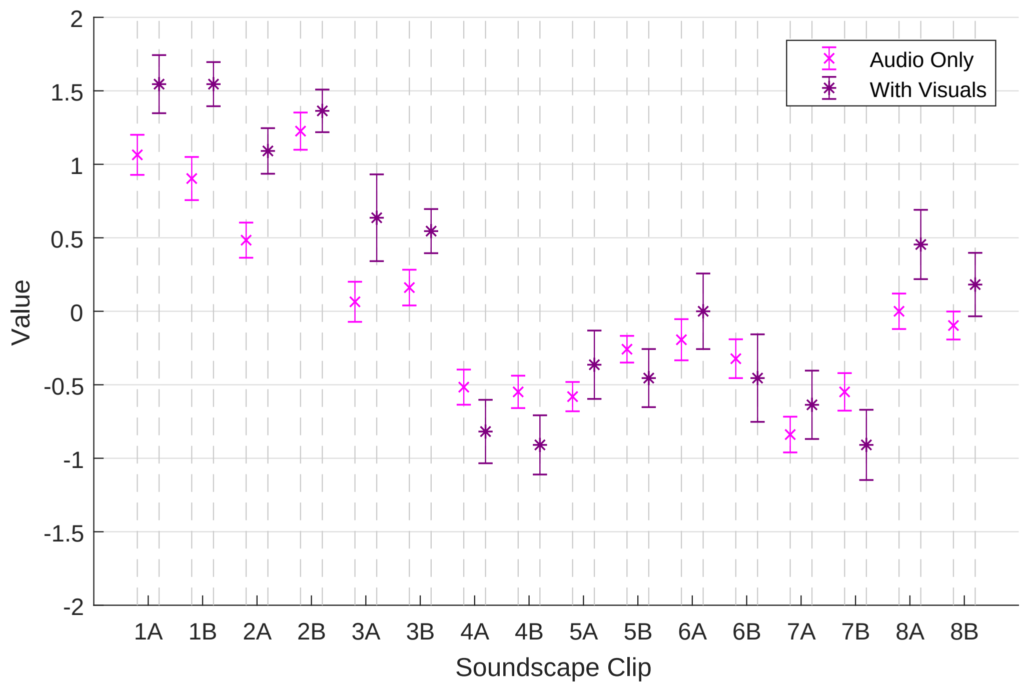


FIGURE 6.27: A comparison of the listening test results for Valence between this listening test ('With Visuals') and the previous audio-only test. Indicated for each clip for each listening condition is the mean rating with error bars indicating the standard error.

Likewise, the soundscape of clip 8A contains some birdsong alongside quiet traffic noise (and some sounds of human activity), but the visuals recorded at that location show an inner city park with a verdant foliage, blossoming flowers, and some trees. This green infrastructure is clear when viewing the scene, but not evident in any explicit way in the audio-only presentation, and is possibly responsible for evoking an alternative emotional state where reported valence levels (i.e. how pleasant the scene is) are higher.

The significant differences in the natural rating scale support this argument in part: clips 2A and 3A show a significant increase in the natural rating with the presence of visual stimuli, which includes a forest and countryside respectively. Clip 6A (recorded on a pedestrianised shopping centre street) also shows a significant increase in the natural rating with the presence of visual information. The mean natural rating for each clip is shown in Figure 6.28 This environment contains some very minor elements of green infrastructure in the form of a couple of trees in some small pots. Whilst this cannot directly be correlated with a change in the valence rating for the environment, it does perhaps indicate how even a very slight presence of green infrastructure can change an individual's experience and perception of a location.

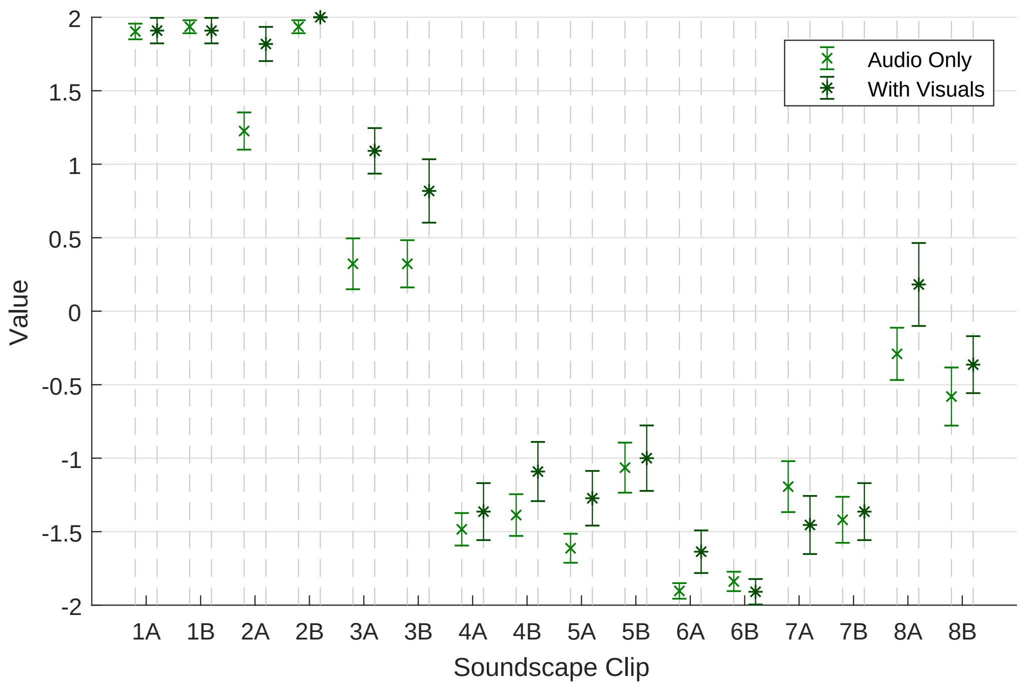


FIGURE 6.28: A comparison of the listening test results for the Natural category rating between this listening test ('With Visuals') and the previous audio-only test. Indicated for each clip for each listening condition is the mean rating with error bars indicating the standard error.

This location also sees a significant decrease in the human category rating for the audiovisual presentation of the clip relative to the soundscape alone. The mean human cateogry ratings for each clip can be seen in Figure 6.29. This is possibly due to the difference between reality and expectation of the visual setting: the dominant sound sources in this clip are human sounds (including very loud conversation, footsteps, and some shouting) with only some distant traffic noise. However the visual setting is dominated by concrete in the form a pavement, shop-fronts, and some larger inner city buildings, reducing the impact of the human activity.

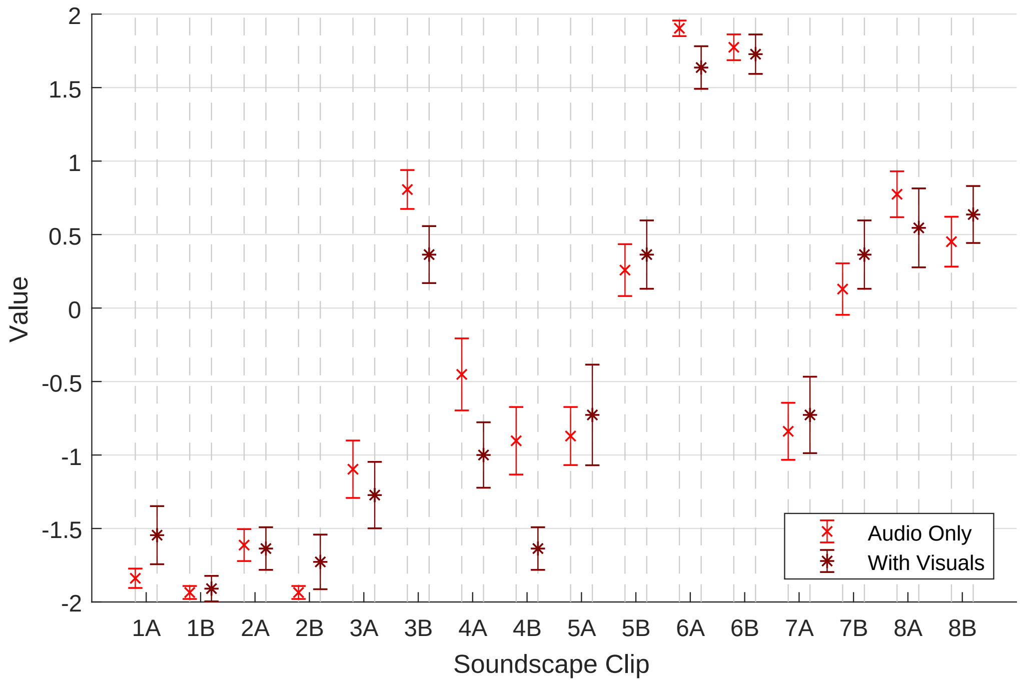


FIGURE 6.29: A comparison of the listening test results for the Human category rating between this listening test ('With Visuals') and the previous audio-only test. Indicated for each clip for each listening condition is the mean rating with error bars indicating the standard error.

The two soundscapes showing a significant difference in the mechanical category rating are 3A and 8B, both of which saw a decrease in mechanical rating with the introduction of visual stimuli. The mean mechanical category rating for each clip in each listening condition can be seen in Figure 6.30. In a way these two clips can be considered as the corollary of one another: clip 3A shows a natural environment 'interrupted' by the presence of a busy road; and clip 8B shows a green-infrastructure (a park) in the context of a large city. As such both of these soundscape clips indicate little about the features of the visual settings, resulting in a decreased mechanical rating for the audiovisual presentation.

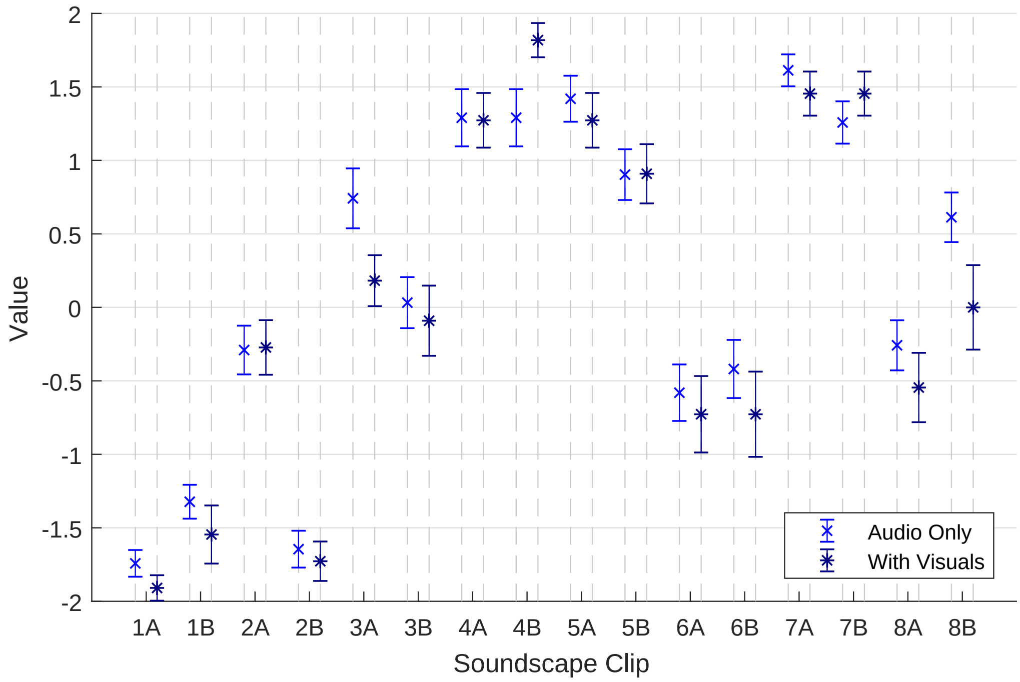


FIGURE 6.30: A comparison of the listening test results for the Mechanical category rating between this listening test ('With Visuals') and the previous audio-only test. Indicated for each clip for each listening condition is the mean rating with error bars indicating the standard error.

Perceptual Noise Impact Rating

In order to further investigate the effect of certain visual features on the emotional state evoked by a soundscape, the valence and arousal rating scales can be combined to form a single measure of the emotional state evoked by a noisy soundscape. This new measure is called the Perceptual Noise Impact Rating (PNIR) and was introduced as part of this body of research in [250]. It is formulated by:

$$\mathrm{PNIR} = 1 - 0.5(1 - \mathrm{A} + \mathrm{V}) \tag{6.6}$$

where A and V represent the Arousal and Valence scores respectively (where the scores are normalised between 0 and 1). Figure 6.31 contains a pair of CMoA axes overlaid with a heatmap of PNIR values. The intention of introducing the PNIR rating system is to reflect the different emotional states represented by the top-left and bottom-right quadrants of the CMoA, as covered in Chapter 4.

Figure 6.32 shows a summary of PNIR results from the preliminary listening test. Indicated

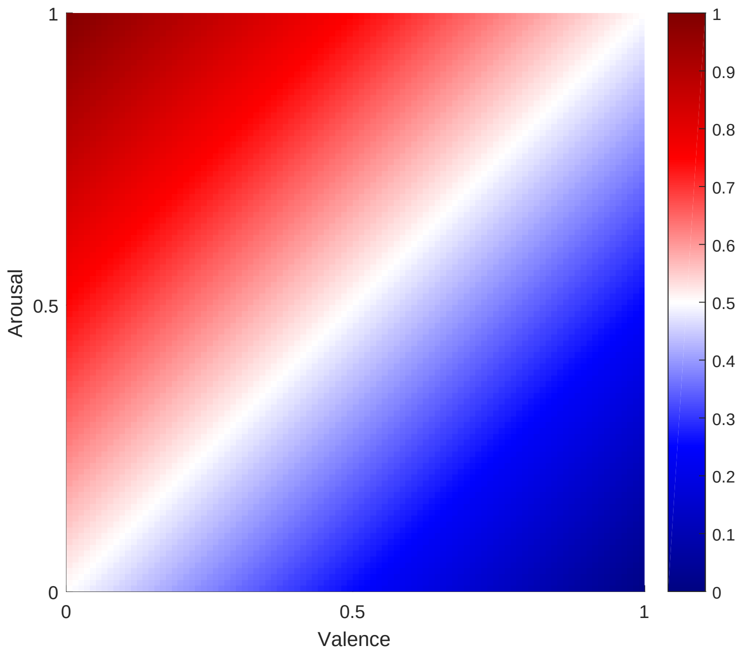


FIGURE 6.31: Heatmap of PNIR values on the circumplex model of affect axes.

in this plot are the mean PNIR values across all participants for each of the 16 stimuli for both the audio-only and audiovisual listening conditions. These results show a trend in the data towards three groups of PNIR values:

1. Clips 1A-2B: These soundscapes were recorded at two locations in Dalby forest, and are comprised of many natural sounds (birdsong, insects, wind) and visual features (trees, a lake, open sky).

2. Clips 4A-7B: These soundscapes were recorded in highly developed environments, including various location in the centre of the city of Leeds, and next to a road in the town of Pickering. The most commonly identified sound sources in these clips were traffic noise, other mechanical noise, and human sounds (footsteps and conversation).

3. Clips 3A-3B and 8A-8B: These soundscapes were recorded in environments that can be considered as being on the interface between the recording locations of the two above categories. Location 3 was next to a country road overlooking a wide expanse of countryside, and location 8 was in a park in Leeds city centre. Both of these environments contained a mixture of mechanical and natural sounds (i.e. relatively quiet traffic noise and birdsong) and visual features (i.e. flowers, trees and other greenery alongside the roads and buildings).

These three emotional groups, as indicated by the PNIR results, were used, alongside the Mann-Whitney test results, to identify which of the soundscape clips to use in the main listening test.

Clips 1B and 2A were chosen to represent group 1: clip 1B was recorded in Dalby forest and contains natural sounds and visual elements; clip 2A was recorded at a nearby lake and again presents many natural sounds and visual elements, as well as a single car drive by.

Clips 6A and 7B were chosen to represent group 2: clip 6A was recorded on a pedestrianised street lined with shops; clip 7B was recorded next to a busy road in Leeds city centre. Both of these clips contains mainly human and mechanical sounds, with little in the way of natural sounds or visual elements.

Clips 3A and 8A were chosen to represent group 3: clip 3A was recorded next to a road in the North York Moors national park; clip 8A was recorded in a small park in the centre of Leeds. As stated above, these locations both represent something of an interface between natural and developed habitats and contain both human and natural sounds and visual elements, including the presence of green infrastructure.

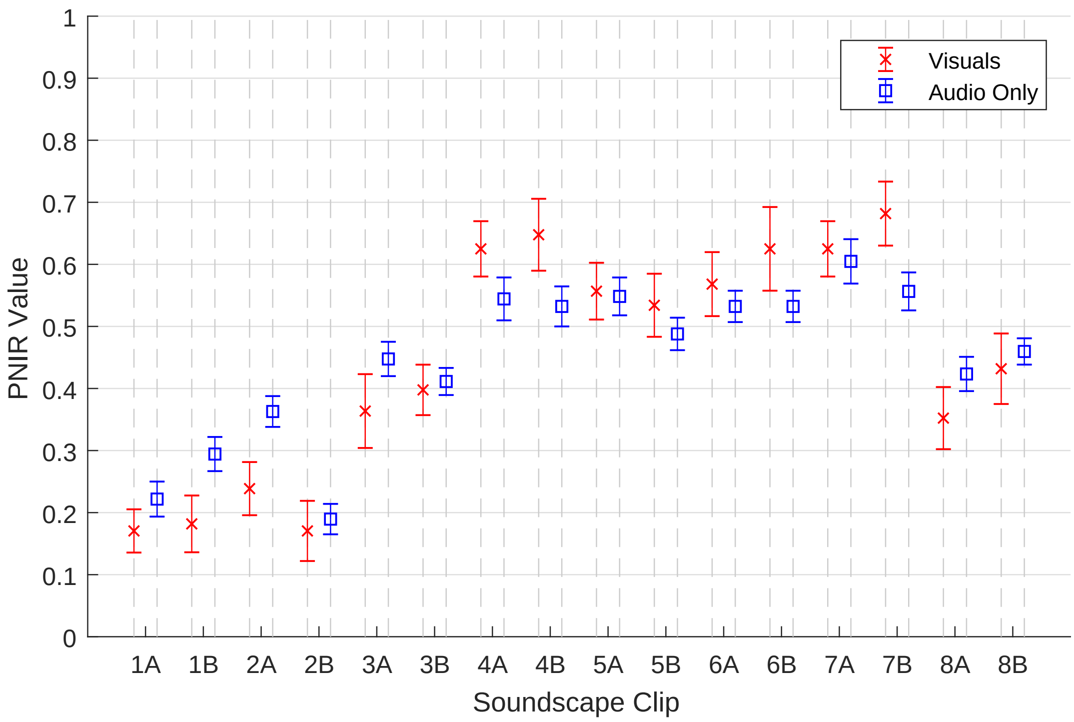


FIGURE 6.32: A summary of PNIR ratings from the preliminary listening test results (labelled 'Visuals' compared with PNIR results from the previous, audio-only, test).

6.3.3 LT4: YouTube VR Presentation

This main listening test will now be considered, starting with the VR content creation and test procedure used. After this the results of this listening tests will be presented, including some comparison with the preliminary test results.

Virtual Reality Content Creation

Figure 6.33 depicts a flow diagram for the creation of full motion spherical audio-visual content ready for VR playback on YouTube, possibly either via a VR headset or on a standard computer monitor. Firstly Kolor Autopano is used to stitch together the six feeds of GoPro footage into a single equirectangular panoramic video [206]. Details of this visual capture work were seen in Chapter 5. FFMPEG [251], a free software project designed for handling multimedia data, is then used to add the Ambisonic audio (with its channels in ACN, rather than Furse-Malham, order) to the panoramic footage [190]. In order for this file to then be uploadable to YouTube [252] the Spatial Media Metadata Injector [253] is used to indicate that the file contains a panoramic video. For the 'audio-only' stimuli a still image of equirectangular perspective lines was used as the visual component, in order to give the test participants some sense of orientation [254][4].

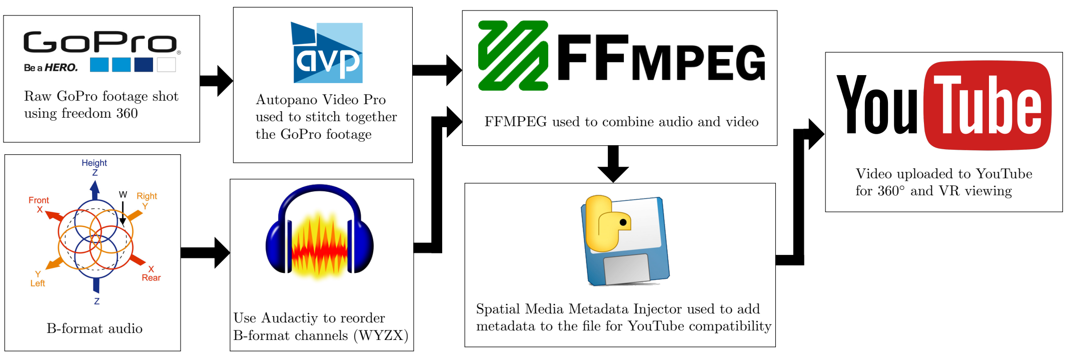


FIGURE 6.33: A flow diagram showing the method used in this study for VR content creation.

[4]Note that this workflow should make use of SN3D normalisation [255] of the ACN audio. This normalisation step was mistakenly not completed for this experiment, although this appears to have had little effect on the results. This is most likely due to the fact that the normalisation step is introduced to avoid clipping in the output signal, which is unlikely with the relatively low-level collected audio recordings. Also, it is worth saying that the second listening test presented in this chapter indicated that stereo UHJ playback was ecologically valid (despite its lack of true 'reality' in soundscape reproduction), so regardless of the lack of SN3D normalisation it is perhaps no surprise that test participants still found the YouTube VR presentations to be sufficiently immersive and representation of the experience of the recording locations.

The resultant content can be viewed in the following two YouTube playlists: the 'audio-only' playlist [256]; and the full audiovisual playlist [257].

Main Test Procedure

For the main listening test there were 20 participants, split into two groups of 10. Each group was exposed to the six chosen soundscape recordings: one group experienced the audio-only soundscapes first, and then experienced them with accompanying video footage; the other group of participants experienced the same stimuli but with the order reversed. As with the audiovisual stage of the preliminary listening test no demographic data were collected here. In each viewing condition participants were encourage to pan and 'look around' the environment, with YouTube updating the binaural rendering of the FOA audio according to the current perspective. It was decided to present the stimuli on a desktop PC screen, rather than on a VR headset, in order to present the visuals in the highest possible resolution, and to keep in line with previous listening tests. Future work should consider the use of a VR headset.

The soundscapes were presented as YouTube content embedded in Qualtrics. The presentation order within each set of stimuli was randomised. As with the preliminary test, each stimulus was rated in terms of valence and arousal, and in terms of the three established soundscape categories. Test participants were also asked to list the sound sources and visual elements in the scene.

Main Test Results

This section presents an evaluation and analysis of the results of the main listening test. As with the preliminary listening test a Shapiro-Wilks test for normality was used here. Similarly, only a very small number of variables were shown to demonstrate a non-normal distribution. The main listening test results were therefore suitable to be compared using the Mann-Whitney U-test.

Initially the results for all test participants are all compared with no consideration of the order in which the two sets of stimuli were presented. Further analysis is then presented in order to investigate how the order in which test participants were exposed to the aural and audiovisual stimuli has affected their experience of the soundscape. Before this analysis takes place, plots of the results from this test will now be presented.

Figure 6.34 shows the test results for the arousal rating scale, showing the mean rating value for each listening condition for every clip. Also shown are error bars indicating the standard error for each set of data. For this plot, as with the other plots showing the results of this test, the responses from all 20 test participants were used, with no consideration of the order in which participants were exposed to the two listening conditions.

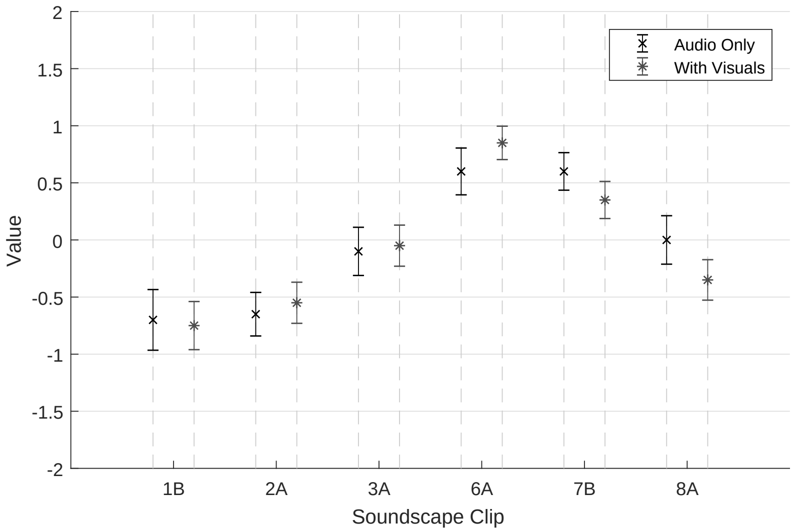


FIGURE 6.34: Comparison of arousal rating results from this test. For each clip error bars are plotted (indicating the mean ratings value and associated standard error) for both the audio-only and audiovisual listening conditions.

In Figure 6.34 it can be seen that for many of the soundscape clips there was very little change in arousal rating between the two listening condition. The only apparent exception is for clip 8A which suggests a lower arousal rating for the audiovisual presentation than for the audio-only case.

Figure 6.35 shows the results from this test for the arousal rating scale. This plot suggests greater difference in valence ratings between listening conditions than indicated by the arousal ratings shown in Figure 6.34. Clips 7B and 8A in particular show differences between the conditions, where the audio-only presentation ratings are lower than the audiovisual case.

Figure 6.36 plots the results from this test combining the valence and arousal scores into PNIR values. This measurement describes the perceptual noise impact of a soundscape, where a lower value indicates a peaceful, less-noisy, soundscape, and a higher value indicates

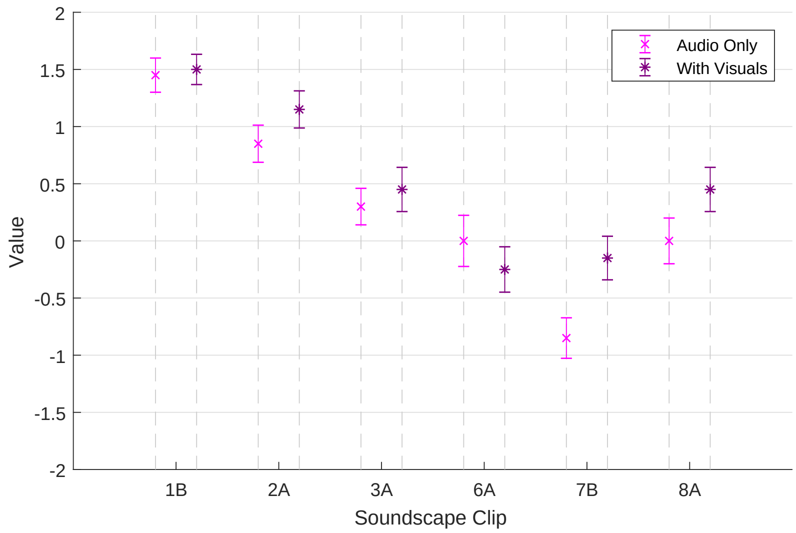


FIGURE 6.35: Comparison of valence rating results from this test. For each clip error bars are plotted (indicating the mean ratings value and associated standard error) for both the audio-only and audiovisual listening conditions.

a noisy, agitating, soundscape. As with the valence results shown in Figure 6.35, Figure 6.36 suggests significant differences in the ratings for clips 7B and 8A, where the audiovisual presentations have received lower PNIR ratings.

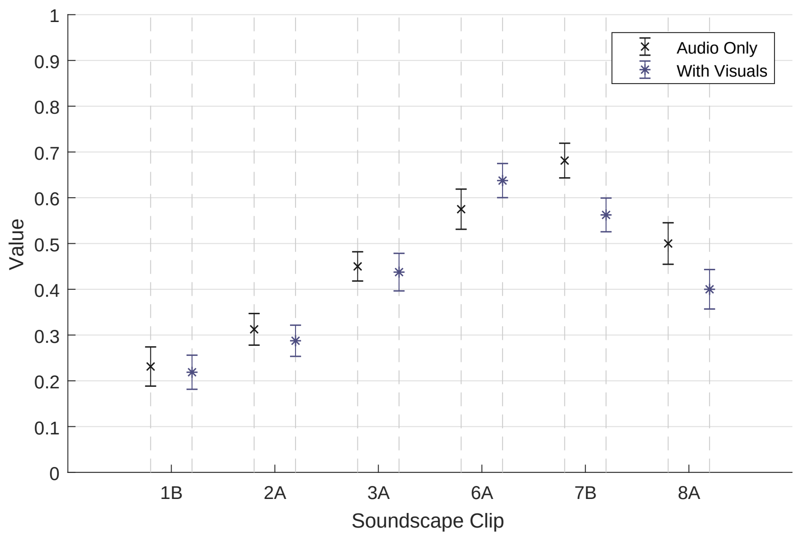


FIGURE 6.36: Comparison of PNIR results from this test. For each clip error bars are plotted (indicating the mean ratings value and associated standard error) for both the audio-only and audiovisual listening conditions.

Figure 6.37 shows the natural category ratings for this listening test. It shows a broad range of values across the 6 different clip, with most clips showing little difference between the results for both listening conditions. The exception is shown in the results for clip 3A, which shows a higher natural rating for the audiovisual presentation. This clip was recorded next to a busy road next to Hole of Horcum, a visually very impressive national park area.

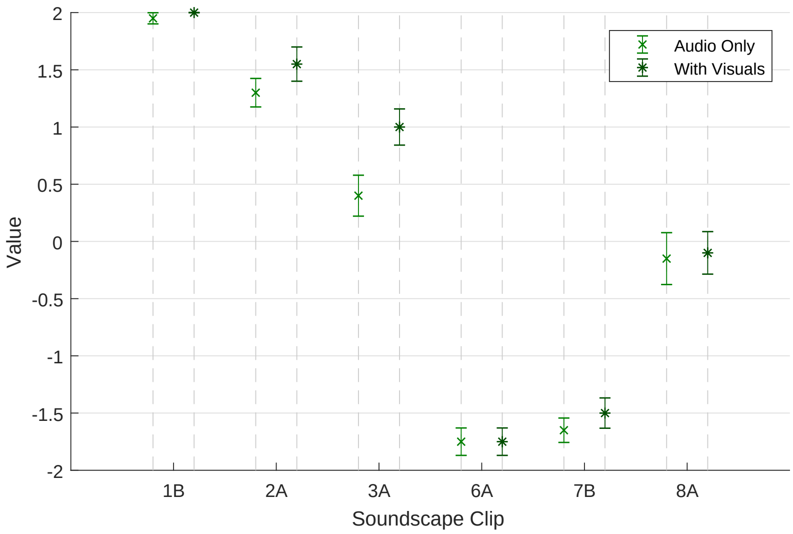


FIGURE 6.37: Comparison of natural category rating results from this test. For each clip error bars are plotted (indicating the mean ratings value and associated standard error) for both the audio-only and audiovisual listening conditions.

Figure 6.38 shows the human category ratings for this test, As with the natural category results shown in Figure 6.37 there is a great variation in the values across the six soundscape clips. This plot suggest a significant difference in human category rating for clip 7B.

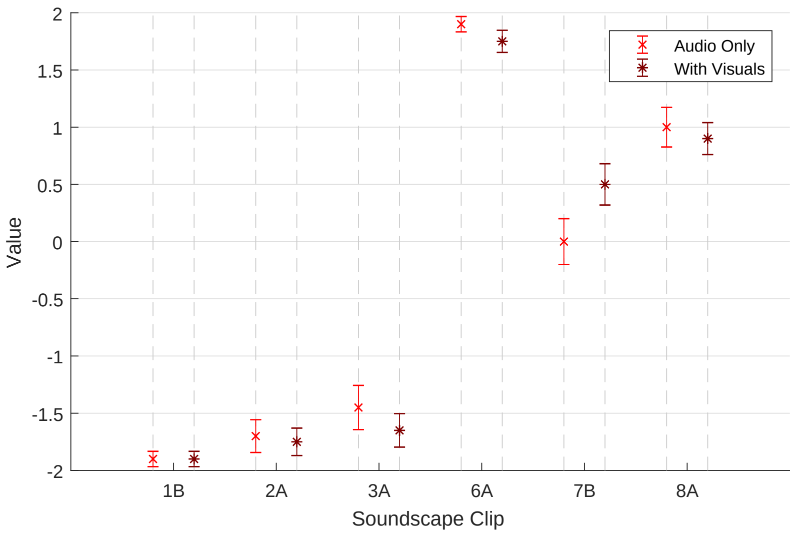


FIGURE 6.38: Comparison of human category rating results from this test. For each clip error bars are plotted (indicating the mean ratings value and associated standard error) for both the audio-only and audiovisual listening conditions.

Finally Figure 6.39 shows the mechanical category rating results for this test. These results again show a large variation across all of the soundscape clips, although there appears to be very little difference between the listening conditions for each clip.

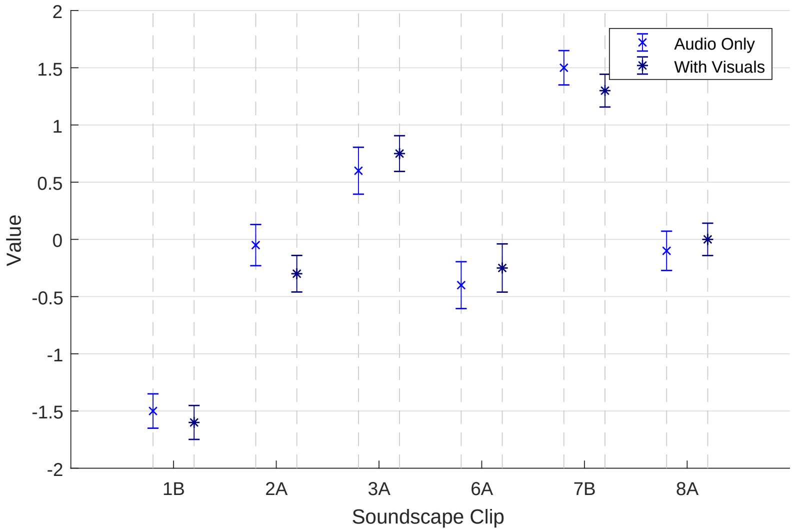


FIGURE 6.39: Comparison of mechanical category rating results from this test. For each clip error bars are plotted (indicating the mean ratings value and associated standard error) for both the audio-only and audiovisual listening conditions.

Having now presented Figures showing the results from this test, including plots of all the rating scales for each listening condition for each soundscape clip, the chapter will now make use of statistical analysis to determine where there are significant differences between the results for both listening conditions.

Overall Comparison

Figure 6.40 shows the results from Mann-Whitney U-test applied to the main listening test results, comparing the results for the audio-only soundscape presentations with the audiovisual ones. Dark squares indicate a significant difference at 95% confidence ($p < 0.05$), and Light marked squares at 90% confidence ($p < 0.1$). White squares indicate no significant difference at either confidence level.

As this figure indicates, there are relatively few significant differences in any of the rating scales when comparing the two listening conditions. The clip that shows the most significant differences is clip 7B, which was recorded next to a busy road in Leeds city centre. Compared to the audio only presentation of this soundscape clip, the ratings for the audiovisual presentation show significantly increased valence and human ratings, and a significantly reduced PNIR rating.

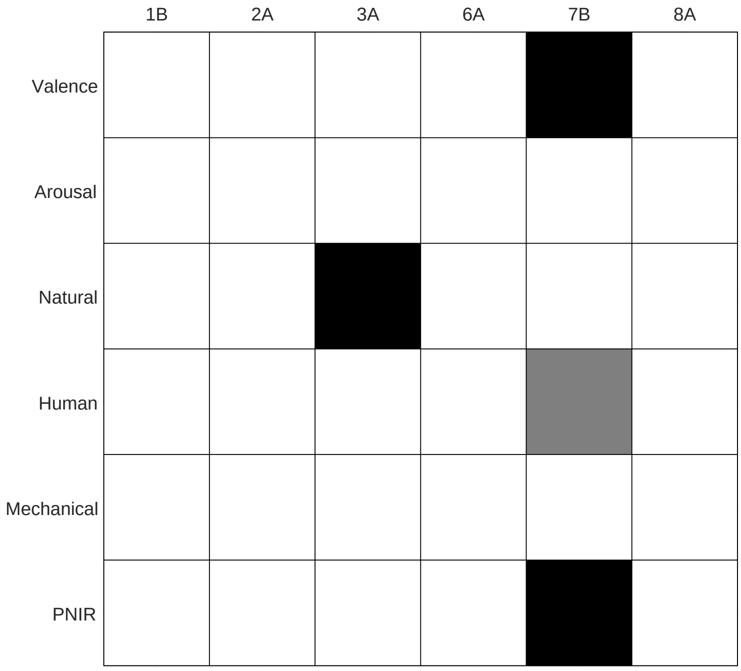


FIGURE 6.40: Mann-Whitney test results indicating significant differences between the two listening conditions for each soundscape clip for all test participants. Dark marked squares indicate a difference at 95% confidence ($p < 0.05$), and light marked squares indicate a difference at 90% confidence ($p < 0.1$).

There are two aspects of the visual setting of this clip that have likely contributed to these differences: firstly, it is difficult from listening to the soundscape alone to get a sense of how close to the road the listener is, as the traffic sounds are very loud, whilst the visual setting makes it clear that recording position is safely away from the road; secondly, the square that this recording was next to is lined with some trees which were identified by test participants as a major visual feature of the scene (as shown in Figure 6.44.

The only other significant difference shown in Figure 6.40 is for clip 3A, where the presence of visuals alongside the soundscape results in a significantly higher natural rating (as expected from the preliminary test results).

Order Dependence

Having now considered all of the results for both listening conditions for both groups of test participants, a breakdown of results by presentation order will now be considered.

Figure 6.41 shows the results of applying the Mann-Whitney U-test to just the first listening condition experienced by each group: i.e. the audio-only results for the group that experienced those clips first compared with the audiovisual results from the other group (who experienced the audiovisual versions first).

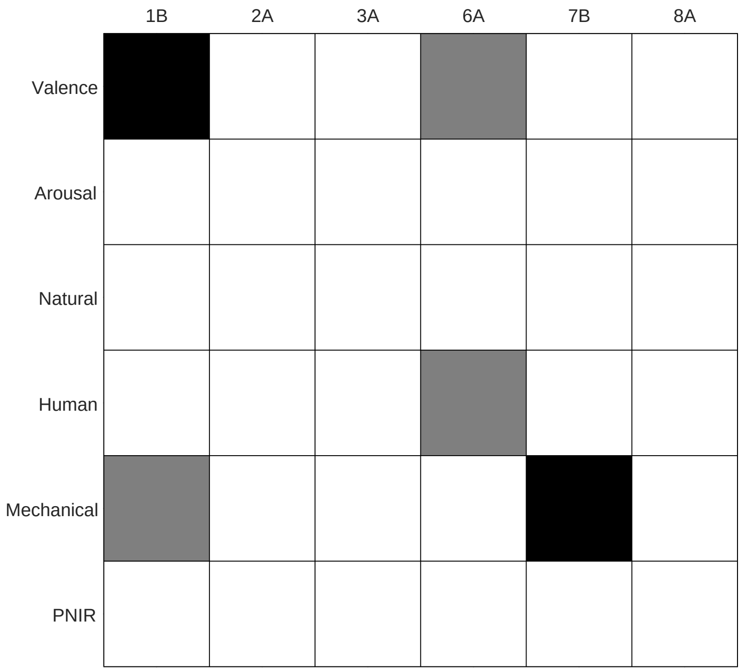


FIGURE 6.41: Mann-Whitney test results indicating significant differences between the two listening conditions for each soundscape clip for the first listening condition experience by each group. Dark marked squares indicate a difference at 95% confidence ($p < 0.05$), and light marked squares indicate a difference at 90% confidence ($p < 0.1$).

Firstly it is interesting to note that the significant difference shown in this figure are not the same as those shown in Figure 6.40. These results show that for clip 1B, recorded at Dalby forest, the version of the clip presented with the accompanying visuals received a significantly greater valence rating, and a significantly lower mechanical rating. As with the preliminary test results, the change in valence rating is most likely due to the pleasantness of the trees and open sky in the visual setting. The mechanical rating most likely is lower with the presence of visuals as the soundscape contains some ambiguous noise that may be distant traffic, wind, or aircraft flying overhead. When listening to the soundscape alone this presents the environment as having some mechanical features that are not readily apparent, or are overshadowed by the other natural features, in the environment.

A significant difference in mechanical rating can also be seen for clip 7B; this is most likely due to the human aspects (people walking past) and minor elements of green infrastructure (some trees lining the square) that reduce the impact of the mechanical noise on the audiovisual experience of the soundscape.

Also shown in Figure 6.41 are two significant differences in the ratings for clip 6A: the audiovisual presentation of this clip received significantly lower valence and human ratings than the audio-only version. This is most likely due to, again, elements of the visual environment that are not manifest in the soundscape itself: in this case the inner city shopping district buildings. In the audio-only presentation the dominant features are conversation and footsteps, whilst in the visual presentation the large buildings are the dominant feature. The presence of these buildings and paved streets also possibly gives some orientation for the background noise in the clip, grounding its otherwise ambiguous nature and indicating to participants that there is some distant traffic noise present.

Figure 6.42 shows the Mann-Whitney U-test results comparing the two listening conditions for the group who experienced the audio-only soundscapes first, followed by audiovisual presentation.

For clip 3A, recorded next to the Hole of Horcum in the North York Moors national park, there is a significant increase in the natural rating for the audiovisual presentation of the clip relative to the audio-only version due to the rolling countryside (something not obviously present in the soundscape itself, beyond the birdsong).

The category ratings for all other soundscapes show no significant differences between listening conditions, but for clips 7B and 8A there are some differences in the emotion ratings. For clip 7B this means a significantly higher valence rating, and a significantly lower PNIR, once again showing how the presence of a relatively small amount of green infrastructure can improve the experience of a location.

Also of note in Figure 6.42 is that for clip 8A, recorded at an inner city park in Leeds, there is an indication of a significant decrease in the PNIR for the clip presented with visuals relative to the audio alone. This is interesting as neither the valence nor arousal ratings on their own show significant differences, but when these ratings are combined a significant difference can be demonstrated.

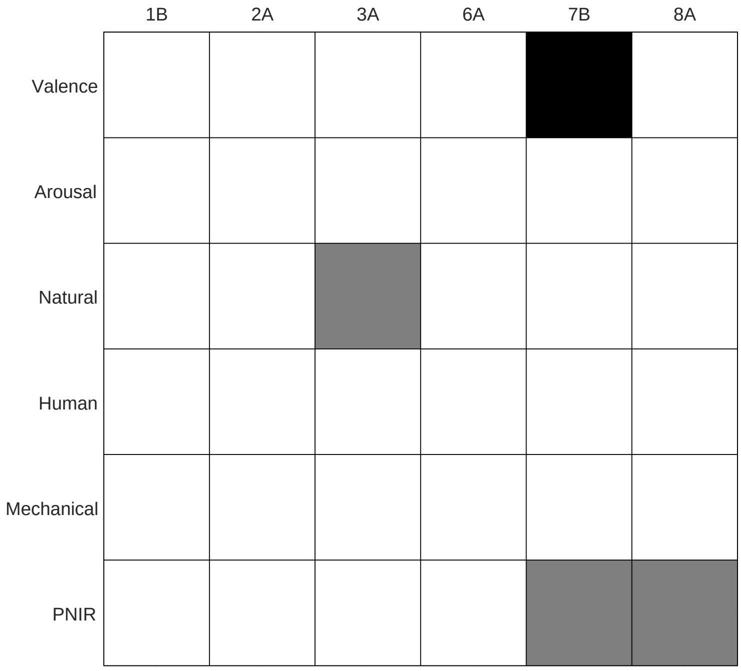


FIGURE 6.42: Mann-Whitney test results indicating significant differences between the two listening conditions for the group of participants that experience the soundscape as audio-only first and then audiovisually. Dark marked squares indicate a difference at 95% confidence ($p < 0.05$), and light marked squares indicate a difference at 90% confidence ($p < 0.1$).

Figure 6.43 shows a three-dimensional plot of the category rating results for this study, showing the results for both listening conditions for each clip. In this plot, the significant differences identified above when comparing the responses of all test participants between the two listening conditions can be seen: the difference in the natural category ratings for clip 3A between the two listening conditions, and the difference in human category ratings for clip 7B.

Aural and visual features

The results collected in this test where test participants were asked to identify the sound sources present in the soundscape clips (for both listening conditions), and to identify the the visual features present in the audiovisual condition, will now be presented

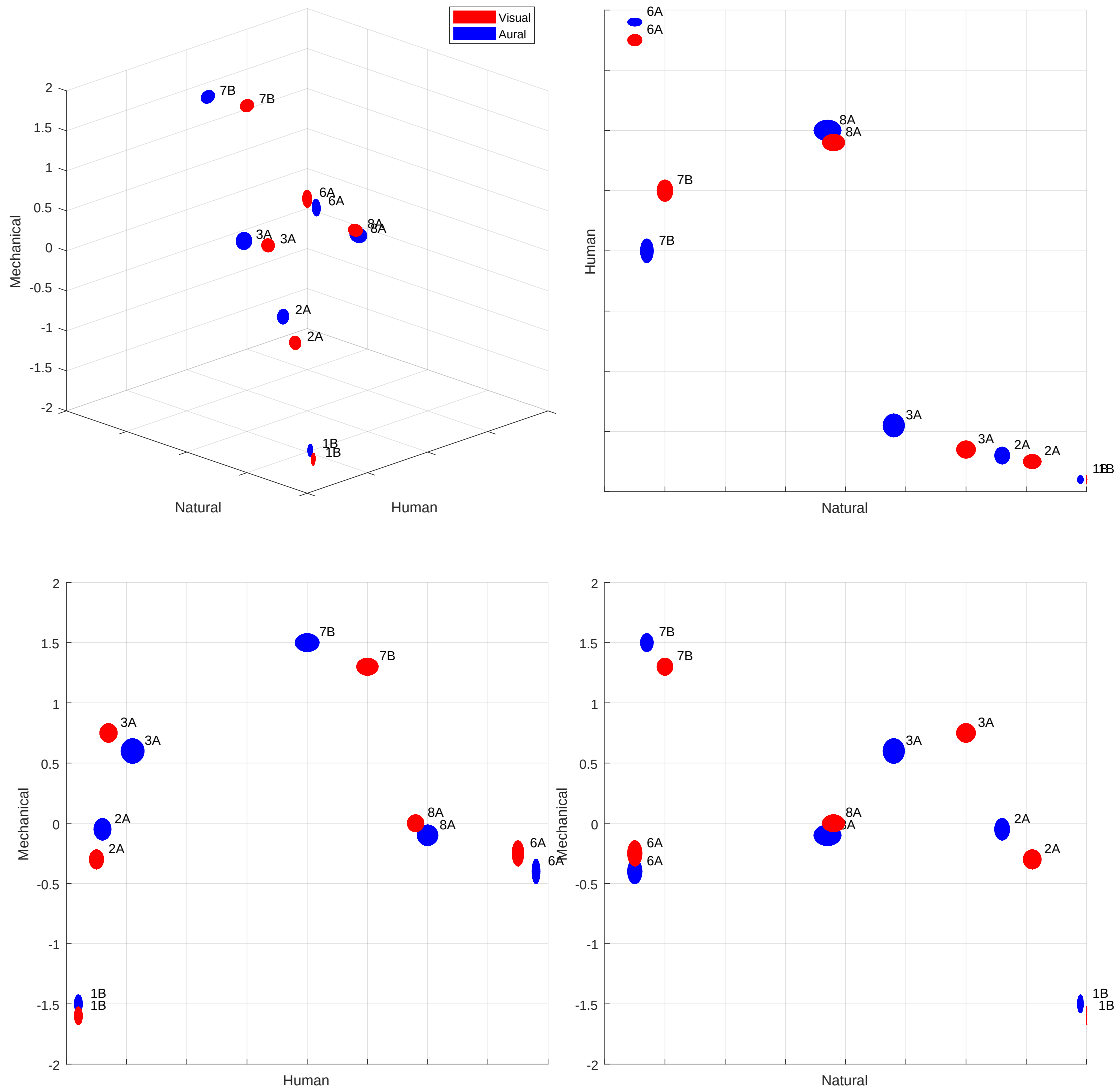


FIGURE 6.43: A plot summarising the category rating results from this test. Each of the three categories is represented as an axis in the plot, and the results for both listening conditions for each clip are presented. In each case the results are represented by ellipsiods plotted centred on the mean value for each category with width in each dimension determined by the standard error of the results for the rating scale corresponding to that axis.

Figure 6.44 contains histograms indicating the aural and visual sources identified by test participants. The top row shows the sound sources identified by participants in the audio-only listening condition. The middle row shows the sound sources identified in the audiovisual listening condition, and the third row contains the identified visual features.

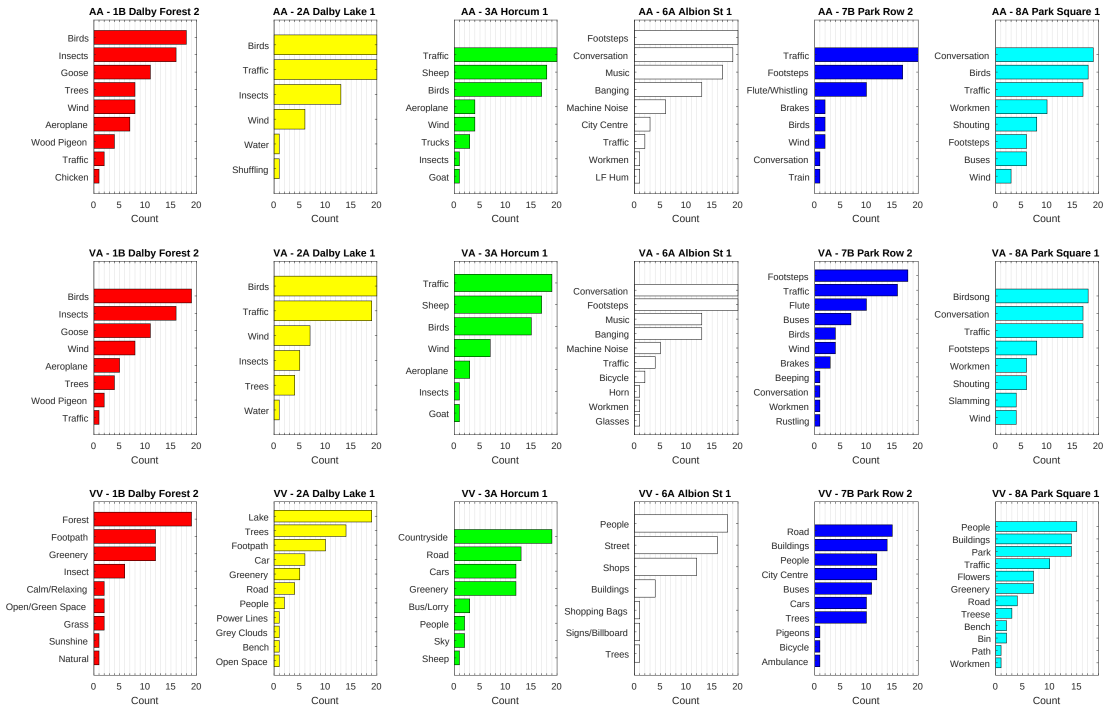
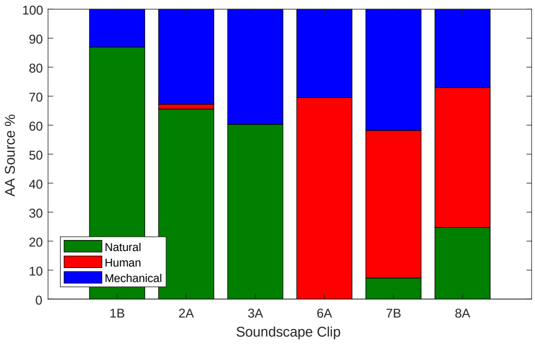


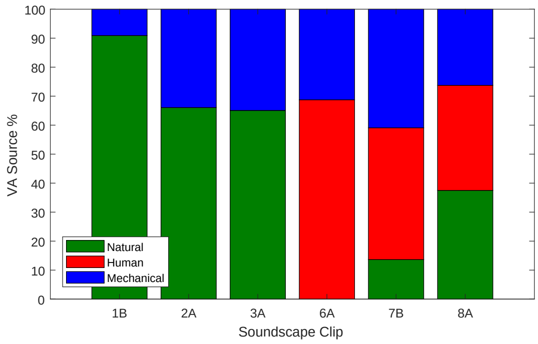
FIGURE 6.44: Histograms of the aural and visual features identified by test participants. The top row shows the sound sources identified in the audio-only condition, the middle row shows the sound sources identified in the audio-visual condition, and the bottom row shows the identified visual features.

The results shown in Figure 6.44 are presented in different form in Figure 6.45 where each identified source has been categorised and then the number of sources identified in each category are presented as a percentage of all of the identified sources. Figure 6.45a shows the sound sources identified for the audio-only listening condition, Figure 6.45b shows the sound sources identified for the audiovisual condition, and Figure 6.45c shows the identified visual features. In Figure 6.45c there is a fourth category (shown in grey) where a test participant identified 'calm/relaxing' as a visual feature. whilst of course this is not a visual feature it was considered worth keeping as it represented the strong connection between natural visual elements (this comment was made for clip 1B) and a relaxed emotional state.

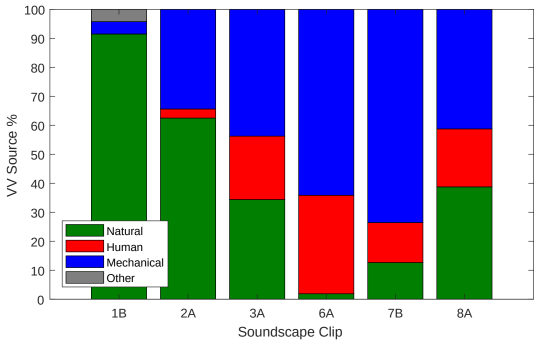
Finally these aural and visual feature results are presented as pie charts in Figure 6.46, Figure 6.47, and Figure 6.48. In each of these charts each the total number of mentions of



(A)



(B)



(C)

Figure 6.45: Percentage breakdown of the visual and aural features identified by test participants: (A) shows the sound sources identified in the audio-only condition, (B) shows the sound sources identified in the audio-visual condition, and (C) shows the identified visual features.

each unique identified aural or visual feature is shown, colour coded according to category. As with the previous listening tests the most commonly identified sound sources in each category were birdsong, footsteps, and traffic. These sound sources are reflected by their visual feature counterparts: greenery/trees, people, and roads.

Discussion

When taken together the above results can be summarised as three main findings:

- Many of the significant differences in emotional or categorical ratings for the different soundscape clips are (perhaps unsurprisingly) due to the visual features that are not manifest in the soundscape clips. This makes clear the need for a cross-modal approach to soundscape evaluation as any real-life soundscape evaluation procedure will have to consider the visual context of that soundscape.

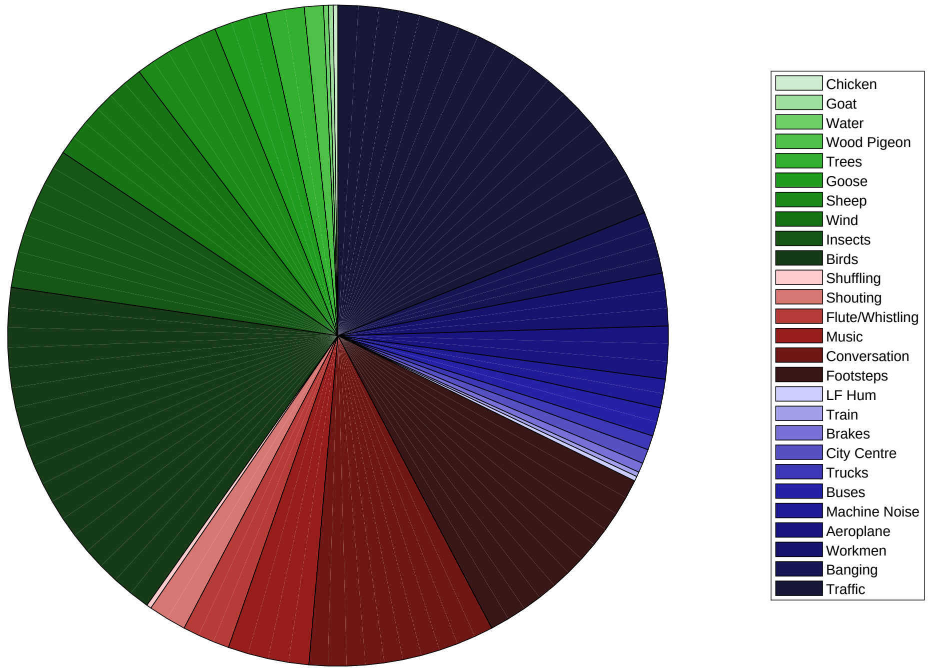


FIGURE 6.46: Pie chart of the sound sources identified in the audio-only listening condition.

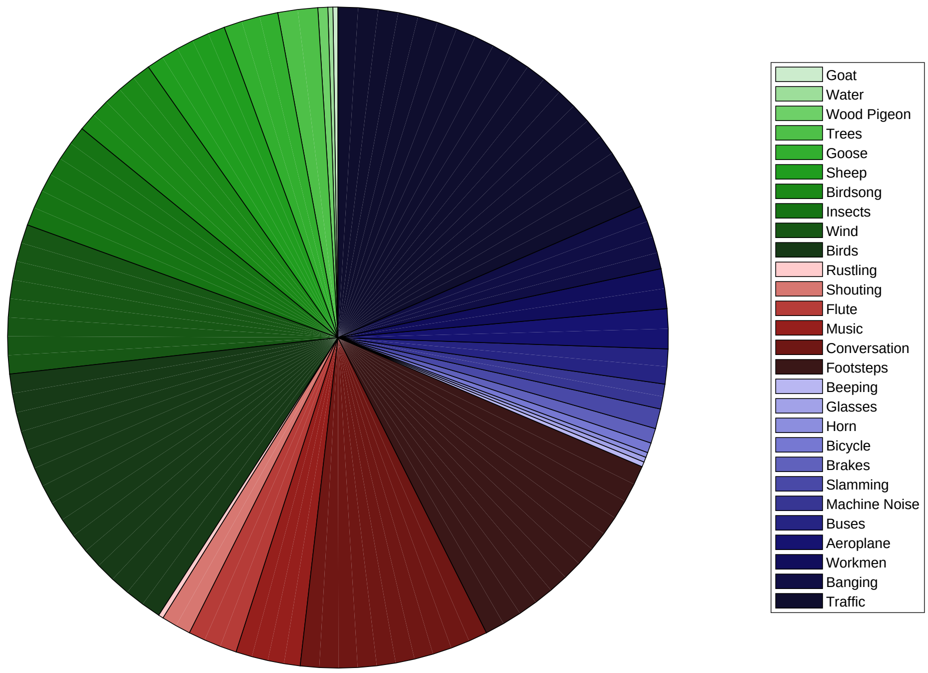


FIGURE 6.47: Pie chart of the sound sources identified in the audiovisual listening condition.

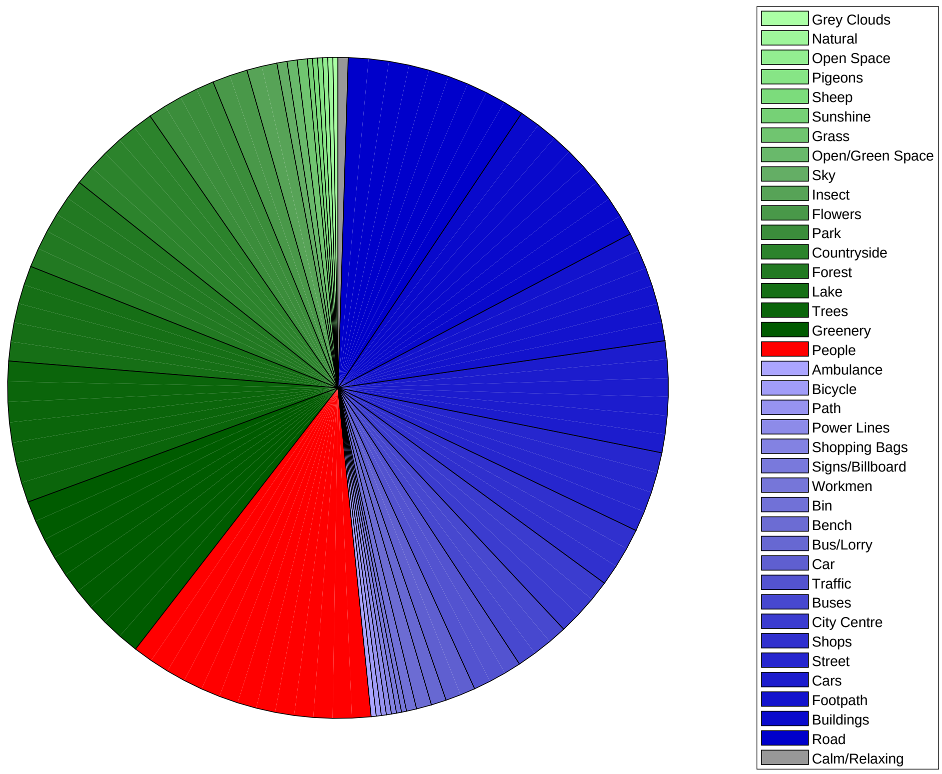


FIGURE 6.48: Pie chart of the visual features identified in this test.

- For many of the differences in perception of the soundscape clips, the presence of aspects of green infrastructure can be identified. This lends credence to the idea that green infrastructure, whilst not necessarily resulting in a significant change to an environment's acoustic properties, can improve the experience of that location.

- The SAM, which has been examined thoroughly throughout this research in terms of its usefulness for soundscape evaluation, has been shown to be very useful in examining differences between the emotional states evoked by different soundscapes. The PNIR, a combination of the valence and arousal dimensions of the SAM into a single perceptual rating that was introduced in a prior stage of this research, has also been shown to be useful in the study for discerning significant differences between emotional states evoked by soundscapes.

Whilst the results presented in this section show some significant differences in emotion and category rating between the audio only and audiovisual clip presentation, further work should be conducted comparing ratings for audiovisual soundscape presentation where the visual setting is changed, for example through the addition of more trees or other aspects of green infrastructure. Such research would build on the results presented here, which validate the methodology in terms of the rating scales used, and the VR content creation.

6.4 Summary

This chapter has presented the results of four listening tests making use of soundscape recordings collected in this recording work, this has included the evaluation of spatially rendered auralisation of soundscapes when compared with stereo UHJ format renderings and investigation of cross-modal perception resulting from concurrent presentation of visuals. The key findings from these listening tests are as follows:

- The SAM has been shown to be a suitable evaluation tool for use in soundscape analysis, where it is instructive as a means of measuring the emotional effect of soundscape exposure. It has been shown to be more intuitive and faster to use than SD pairs, an established analysis tool. These two points are supported by the test results: firstly the feedback from participants universally described the SAM as more intuitive than the SD pairs, and the nature of the results showed (for example, the correlation analysis of the SD pairs results showed confusion regarding some of the SD rating scales); secondly and, perhaps, obviously, the SAM only contains three scales (in this test) and is related solely to emotional experience, which means it takes less time to use than the SD pairs.
- The use of soudnscape category ratings has also been shown to be a useful tool in the evaluation of soundscape, with connections shown between categorisation results and the emotional state evoked. These connections include where more natural soundscapes typically evoke a relaxed, calm emotional state (high valence, low arousal), and where mechanical soundscape induce a more stressed response (high arousal, low valence). Soundscape rated highly in the human category typically results in greater levels of arousal. For each of these categories the results from listening tests three and four indicated a particular sound source that was most commonly identified. These sound sources were shown to be birdsong (natural), conversation (human), and traffic (mechanical). These categories were selected following a review of extant

soundscape categorisation work, and the results from this (and other) listening tests supports the use of this categorisation paradigm in future listening tests.

- The use of UHJ stereo format (where horizontal FOA soundscape data is condensed into a two channel format) was shown to be ecologically valid, evoked statistically similar emotional states to full surround-sound presentation of the FOA recording. This is interesting as UHJ stereo results in a representation of the soundscape that is not strictly 'correct', due to the loss of elevation information, and lack of head tracking. This result encourages the possible future use of the UHJ stereo format in order to present acoustic data beyond standard stereo recording in environment where a full surround-sound or binaural presentation is not possible (for example in an online listening test).

- YouTube has been shown to be a suitable platform for the presentation of VR soundscape stimuli, offering an accessible method for the presentation of FOA recordings and spherical video. This is supported by the results from LT4, which at least indicate a degree of ecological validity to stereo UHJ, or FOA, reproduction. This is also supported by informal feedback from test participants.

- The results of the third and fourth listening test showed some of the effects of cross-modal perception, where they demonstrated how the visual elements of an environment can affect its perception. The clear statistical differences in SAM results between the audio-only and audiovisual presentations, as well as the comments and feedback from test participants support this claim. This change in emotional state was shown to potentially the result of a mismatch between visual and aural features. Green infrastructure was also shown to have a potentially have positive emotional effect, and to increase natural rating. The results shown here are modest in this regard but represent a suitable basis from which future research could be developed.

The chapter has presented a methodical and rigorous approach to soundscape evaluation, building on and developing from extant research in the field to move from the subjective analysis of soundscape recording using SD pairs on to the investigation of cross-modal perception using the SAM and novel VR technologies. The next chapter will present novel work completed as part of this thesis in the field of environmental sound modelling, specifically a method for modelling sparsely reflecting outdoor acoustic scenes.

Chapter 7

Environmental Sound Modelling

The simulation of environmental sound is an area of significant interest due to the propagation of noise pollution over distances, and its related impact on well-being, particularly in urban spaces. This chapter introduces the waveguide web digital reverberator design for modelling the acoustics of sparsely reflecting outdoor environments; a design that is, in part, an extension of the scattering delay network reverberator. The design of the algorithm is based on a set of digital waveguides connected by scattering junctions at nodes that represent the reflection points of the environment under study. The structure of the proposed reverberator allows for accurate reproduction of reflections between discrete reflection points. Approximation errors are caused when the assumption of point-like nodes does not hold true. Three example cases are presented comparing waveguide web simulated impulse responses for a traditional shoebox room, a forest scenario and an urban courtyard, with impulse responses created using other simulation methods or from real world measurements. The waveguide web algorithm can better enable the acoustic simulation of outdoor spaces and so contribute towards sound design for virtual reality applications, gaming and auralisation, with a potential for use in the design of urban environments.

7.1 Introduction

Artificial reverberation research has, until recently, striven mainly for the realistic imitation of rooms, concert halls, and other indoor acoustic spaces [45], [258]. For this purpose, there are several specialised modelling techniques, such as the ray-tracing [259], image-source [260], [261], digital waveguide [262], [263], feedback delay networks (FDNs) [264]–[266], and finite-difference time-domain (FDTD) [267]–[270] methods. However, there has been relatively little research looking at modelling of sparsely reflecting outdoor acoustic scenes, although environmental sound is of significant importance, particularly because of the propagation of noise pollution over distances and its impact on human health and well-being [9]. This chapter proposes a new modelling technique for reverberant open acoustic environments, considering both forests and urban scenes. Such spaces are important for wider study in relation to how green infrastructure (e.g. trees and other similar natural interventions) and architectural design might have a positive impact on the soundscape of urban areas [160], [198], as well as in more creative applications such as sound design for film soundtracks and interactive computer games [271].

Early research contributions have considered the acoustic properties of outdoor environments, such as streets [272] and forests [273], both of which have a special character that contributes to the sonic experience a person has of these particular locations. For instance, in [111] it was shown that acoustic scattering from the tree trunk is a key factor affecting the behaviour of mid-frequency attenuation for sound propagation through trees. Surprisingly, a bamboo forest has good acoustic characteristics for certain types of music [274] and outdoor acoustics also have a connection to concert halls, as shown in a study by Lokki et al. who demonstrated that the finite difference time domain (FDTD) technique is suitable for modelling the acoustics of ancient amphitheaters, which are open but can have excellent acoustic properties [275]. A related study, that considered the acoustic characteristics of an historic street, combined impulse response measurement of the existing site and geometric acoustic modelling of the historic environment as part of an analysis of its suitability for dramatic performances [46].

Recent research has shown a growing interest in the modelling of urban environments. Kang modelled the acoustics of a town square using image-source and radiosity methods to predict the sound pressure level [276]. Collecchia et al. studied the acoustic characteristics of narrow alleyways and simulated their interesting behaviour using the image-source method [277]. Recently, Stienen and Vorländer demonstrated how to auralize the propagation of traffic noise in an urban environment [278]. An outdoor urban environment was also at the centre of a study that explored how spatial impulse response measurement, and reflection analysis can be used to help determine source localisation [47], [215].

Some researchers have also been interested in modelling natural environments, such as those containing forests and hills, which strongly reflect sound. Pieren and Wunderli [279] have studied how cliffs in an Alpine valley reflect noise and have proposed a model to account for this phenomenon in sound propagation calculations. Shelley et al. measured forest acoustics in a distant location in Finland, both in the summer and in the winter, to provide impulse responses for convolution-based reverberation [41]. In [280], recorded sound scenes were combined with modelled soundscape interventions, in this case, a FDTD simulation of a sonic crystal noise barrier, and used as part of a virtual soundwalk perceptual evaluation in order to elicit the effectiveness of applying such design strategies.

Spratt and Abel have proposed a general waveguide method called *Treeverb* for modelling the acoustics formed by trees in a forest [281]. Their model can be interpreted to be 2.5-dimensional, as it describes the geometry, including the locations of the source, trees, and the receiver, on a plane, and the structure is then extended in the third dimension. However, as more trees are added into the model, it quickly becomes too large for efficient computation, in terms of both number of operations and memory use, so Spratt and Abel chose to implement an image-source version of their method for faster simulation [281].

Another example digital reverberator representing an extension of the waveguide method is the scattering delay network (SDN) [271], [282], a method conceptually similar to the feedback delay network (FDN) [264] that consists of a set of discrete nodes representing the reflection points of a given environment connected by a set of waveguides.

The work in this chapter extends the Treeverb method by formulating a novel type of waveguide network [262], [263] called a Waveguide Web (WGW). Like the SDN, a WGW is a network of discrete nodes, including, source, receiver and a number of reflection points, which can be, for example, trees or other reflective surfaces. The WGW differs distinctly from the digital waveguide mesh, which is a regular grid structure of scattering nodes and interconnections, used for modelling multidimensional wave propagation, also usually in closed systems [283]–[286]. This work shows that the proposed WGW has similarities to, but is also different from, the recently introduced scattering delay network (SDN) approach to reverberation design as it extends the design of the SDN to incorporate directionally-dependent filtering at the node positions.

The rest of this chapter is organized as follows. Section 7.2 gives an overview of extant acoustic modelling methods, including Section 7.2.3 which gives a brief overview of the Treeverb and SDN methods, which are those closest to the proposed WGW method. Section 7.3 then introduces the theory of the WGW model. Section 7.4 presents the evaluation of the method by way of three case studies: the acoustic simulation of a shoebox room, a forest-like environment, and an urban courtyard. Section 7.5 concludes this chapter with a summary of the results from this work and recommendations for future development.

7.2 Acoustic Modelling Techniques

Existing acoustic modelling methods, broadly speaking, can be divided into three categories (although hybrid methods combining multiple approaches to this problem do also exist) [45]:

- Geometrical methods, where the transmission of sound is effectively treated analogously as light and modelled by 'rays' or 'beams'
- Wave-based approaches, where the transmission of sound waves in the transmission medium itself is modelled.
- Delay network methods, where a set of delay lines and digital filters are used to describe the acoustics of a space.

This chapter will first consider the geometrical and wave-based approaches individually before presenting the Waveguide Web, a novel extension of the Scattering Delay Network (SDN) method for use in modelling sparsely reflecting outdoor acoustic scenes. This work has been published as a journal paper [287]. This chapter will also cover background theory for delay network based acoustic modelling methods, as the Waveguide Web is an acoustic modelling method that makes use of these techniques.

7.2.1 Geometrical Modelling

Geometric modelling methods are based on the assumptions that the wavelength of the modelled sound is small in comparison to he geometry of the environment [288]. Example geometric modelling methods include, although are not limited to:

- The Image Source Method (ISM): The ISM computes, for a given auditory scene, the shortest reflection paths between source and receiver via all possible combinations of objects in the scene [289].
- The beam tracing method. This method also computes these paths but builds a *beam tree* which is a data structure storing reflection paths up to a certain order [290].
- Acoustic ray tracing is a method that builds the IR of an acoustic scene by sampling the space with 'sound rays' [291]. In the case of forward ray tracing, rays are cast from the sound source in many directions and traced through the scene, taking account of iterations with surface boundaries, ultimately to 'hit' the receiver where an impulse is registered in the output. Backward ray tracing is where the rays are cast using the same principle but are instead traced from the receiver to the source [292].

Geometric models can be advantageous as they are relatively fast when compared with wave-based modelling. However, since these methods only relate to the dispersion and magnitude of 'sound rays' no phase information or interaction effects are taken into account [293].

7.2.2 Wave-based Modelling

Wave-based modelling makes use of mathematical solutions to the wave equation in order to model the interaction of a sound wave with the transmission medium (air) and the sourrounding environment. Examples include the finite element method (FEM) and boundary element method (BEM) [45]. In order to best explain the fundamentals of wave-based techniques this section will consider the finite-difference time-domain (FDTD) method [294].

FDTD modelling is an acoustic modelling method where the vibrating medium (air) is modelled as a grid of particles. This gird is formed of a discretised set of interconnected nodes representing either boundary or free space elements [295]. Figure 7.1 shows a simple example of such an FDTD grid in 2D.

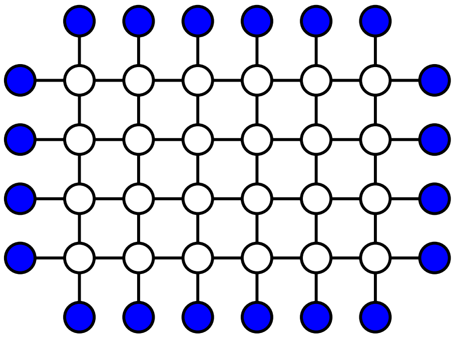


FIGURE 7.1: A particle mesh formed of a grid of sampling points. White denotes a freely vibrating node and blue denotes a boundary node, after [296].

Note how this model of a space is analogous to the mass spring model of sound propagation as shown in Figure 2.2. FDTD modelling is based on a solution to the wave equation discretised in time and space (i.e. by sampling time and grid spacing). The two dimensional form of the wave equation can be expressed as [296]:

$$\frac{\partial^2 p}{\partial t^2} = c^2 \nabla^2 p \tag{7.1}$$

$$\frac{\partial^2 p(t,x,y)}{\partial t^2} = c^2 \left(\frac{\partial^2 p(t,x,y)}{\partial x^2} + \frac{\partial^2 p(t,x,y)}{\partial y^2} \right) \tag{7.2}$$

Where p is sound pressure, c is speed of sound in the relevant medium (i.e. air), t is time, x and y indicate the x and y directions, and ∇ is the Laplace operator on p.

In order to perform FDTD modelling a finite difference form of 7.2 is required. Finite difference formulation represents an approximation of the mathematical derivative of a given function $f(x)$, defined as:

$$\frac{df(x)}{dx} = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} \tag{7.3}$$

Where h represents a small step along the x-axis (as shown in Figure 7.2). Given that in practice h cannot equal 0, we can consider the above using a small value of h to give an

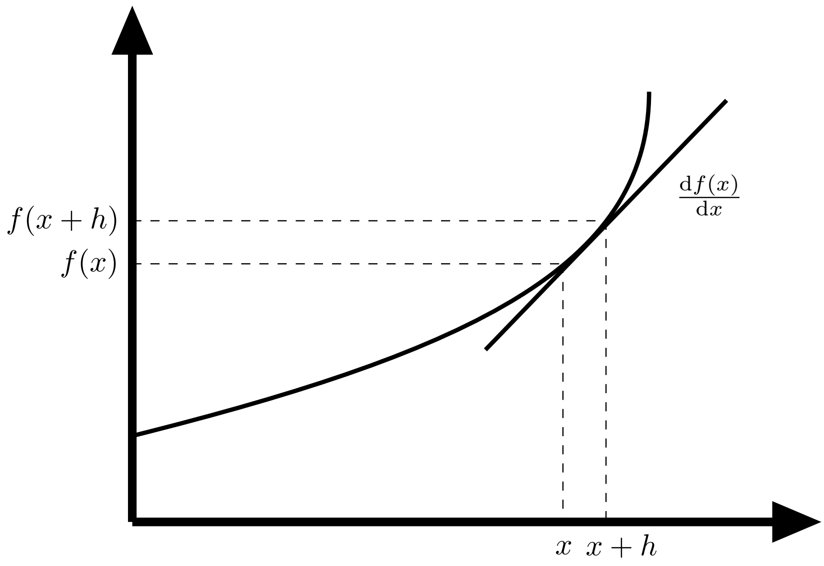


FIGURE 7.2: The derivative $\frac{df(x)}{dx}$ of the function $f(x)$, from [297].

approximation:

$$\frac{df(x)}{dx} = \frac{f(x+h) - f(x)}{h} \tag{7.4}$$

This is called the forward difference formula. When applied again to the result, the second derivative approximation is given by:

$$\frac{d^2 f(x)}{dx^2} = \frac{f(x+h) - 2f(x) + f(x-h)}{h^2} \tag{7.5}$$

These finite difference formulations can be used to substitute the following for $\dot{x}$ and $\ddot{x}$:

$$x \cong x(n) \tag{7.6}$$

$$\dot{x} \cong \frac{x(n) - x(n-1)}{\Delta t} \tag{7.7}$$

$$\ddot{x} \cong \frac{x(n+1) - 2x(n) + x(n-1)}{(\Delta t)^2} \tag{7.8}$$

where Δt is the sampling time step of the system ($\frac{1}{f_s}$). Given this formulation for a finite difference, it can be applied to the 2D form of the wave equation:

$$\frac{\partial^2 p}{\partial t^2} = c^2 \nabla^2 p \tag{7.9}$$

$$\frac{\partial^2 p(t,x,y)}{\partial t^2} = c^2 \left(\frac{\partial^2 p(t,x,y)}{\partial x^2} + \frac{\partial^2 p(t,x,y)}{\partial y^2} \right) \tag{7.10}$$

Applying the (7.8) finite difference formulation to (7.10) gives these second derivatives:

$$\frac{p_{l,m}^{n+1} - 2p_{l,m}^{n} + p_{l,m}^{n-1}}{T^2} = c^2 \left(\frac{p_{l+1,m}^{n} - 2p_{l,m}^{n} + p_{l-1,m}^{n}}{X^2} + \frac{p_{l,m+1}^{n} - 2p_{l,m}^{n} + p_{l,m-1}^{n}}{Y^2} \right) \quad (7.11)$$

$$\frac{p_{l,m}^{n+1} - 2p_{l,m}^{n} + p_{l,m}^{n-1}}{T^2} = c^2 \left(\frac{p_{l+1,m}^{n} + p_{l-1,m}^{n} - 4p_{l,m}^{n} + p_{l,m+1}^{n} + p_{l,m-1}^{n}}{D^2} \right) \quad (7.12)$$

where n is the time frame, l is the x grid position, m is the y grid position, T is the time step duration of the system (Δt), and D is the grid spacing (which is the same in both the x and y direction - i.e. $D = Y = X$). Then rearrange and let $\lambda = \frac{cT}{D}$:

$$p_{l,m}^{n+1} - 2p_{l,m}^{n} + p_{l,m}^{n-1} = \lambda^2 \left(p_{l+1,m}^{n} - 4p_{l,m}^{n} + p_{l-1,m}^{n} + p_{l,m+1}^{n} + p_{l,m-1}^{n} \right) \quad (7.13)$$

Rearrange again to make the pressure at the the current position at the next time step the subject:

$$p_{l,m}^{n+1} = \left(2 - 4\lambda^2\right) p_{l,m}^{n} + \lambda^2 \left(p_{l+1,m}^{n} + p_{l-1,m}^{n} + p_{l,m+1}^{n} + p_{l,m-1}^{n} \right) - p_{l,m}^{n-1} \quad (7.14)$$

Which can also be expressed as:

$$\underbrace{p_{l,m}^{n+1}}_{\text{next value}} = A \underbrace{p_{l,m}^{n}}_{\text{current value}} + B \underbrace{\left(p_{l+1,m}^{n} + p_{l-1,m}^{n} + p_{l,m+1}^{n} + p_{l,m-1}^{n} \right)}_{\text{surrounding values}} + C \underbrace{p_{l,m}^{n-1}}_{\text{previous value}} \quad (7.15)$$

Such that:

$$A = 2 - 4\lambda^2 \quad (7.16)$$

$$B = \lambda^2 \quad (7.17)$$

$$C = -1 \quad (7.18)$$

$$\lambda = \frac{cT}{D} \leq \frac{1}{\sqrt{2}} \quad (7.19)$$

Where the value of λ is limited by the Courant stability condition. Node spacing d_0:

$$d_0 = \frac{c\sqrt{n}}{f_s} \quad (7.20)$$

where c is the speed of sound in the medium, f_s is the sampling rate of the system, and n indicates the number of dimensions represented. To reproduce a certain bandwidth, the sampling rate of the system must be at least 4 times the desired top frequency [34].

7.2.3 Treeverb

Although the acoustics of forests have been studied previously (e.g. [273], [274]), to date, as cited in [45], only one study has attempted to simulate this reverberant effect. Spratt and Abel's *Treeverb* is a digital reverberator designed to model the scattering of acoustic waves between a number of trees as might be found in a forest environment [281].

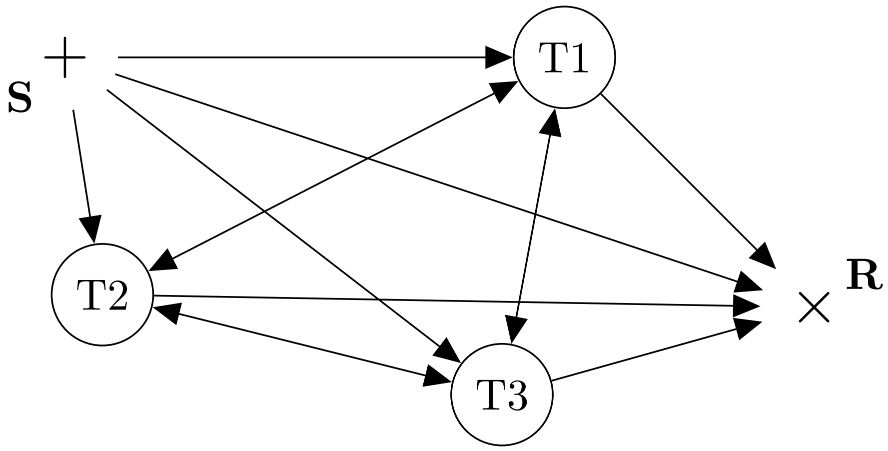


FIGURE 7.3: A simple *Treeverb*, or digital waveguide, network topology for modelling forest acoustics, consisting of three tree-nodes, T1, T2, T3, and a single source, S, and receiver R. The tree-nodes are connected via bidirectional delay lines, with the source and receiver connected to these tree-nodes via unidirectional delay lines. Each delay line has an attenuation factor associated with it and directional dependent filtering and scattering takes place at each tree-node connection. After [281].

In this work, the forest environment is considered as a two-dimensional geometry, with defined source and receiver locations, and a random arrangement of trees. This establishes a fixed network of connected paths between each node, defined as either source, receiver or tree. Each path is modelled using a time delay and attenuation factor, or spreading loss, in much the same way as in other reverberation algorithms. However, in this case boundary reflections are better considered as tree-node interactions, where a tree is modelled as a rigid cylinder. Signals incident on a tree-node are scattered in a frequency dependent manner, with appropriate proportions of the incident signal being transmitted to other connected nodes (either receiver or tree), or reflected and returned along the path of the incident signal. Hence Treeverb can be considered as a closed network of lossless *bidirectional waveguides* connected via tree-node *scattering junctions* with attenuation losses also lumped at these discrete points in the network, and so belongs to the *digital waveguide network* family of digital reverberators, first proposed in [262] and developed further in [263]. A simple example network is shown in Figure 7.3 consisting of three interconnected tree-nodes, T1, T2, T3, and a single source, S, and receiver, R. The tree-nodes are connected via bidirectional delay lines, with the source and receiver connected to these tree-nodes via unidirectional delay lines. Each delay line has associated with it a distance dependent attenuation factor.

Scattering at a tree-node interaction point takes place in a frequency dependent manner, based on Morse's solution to the acoustic scattering from a rigid cylinder [298]. In [281] an approximation of this solution is used to model the scattering occurring at each tree. In this, a plane wave incident on a rigid cylinder produces a result formed of two parts, defined by movement of the acoustic wave both clockwise and anti-clockwise around the cylinder.

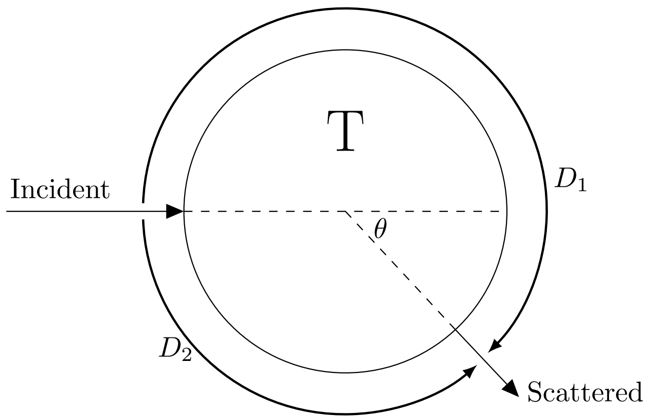


FIGURE 7.4: An acoustic wave incident on a tree trunk T, represented as a rigid cylinder, and the resultant scattered wave at angle θ as formed by the two paths, D_1 and D_2, around the tree. After [281].

Figure 7.4 demonstrates this effect where the scattered wave of interest is at an angle θ, with respect to the angle of incidence, due to interaction with a rigid cylindrical tree trunk, T, with radius, r. The two path lengths around T are defined as D_1 (clockwise) and D_2 (anticlockwise) and hence the path length difference $D_\theta = 2r\theta$, or, for a given sampling rate $1/f_s$, $\tau_\theta = 2r\theta f_s/c$ where c is the speed of sound and τ_θ is the angle dependent delay in samples.

The signal to arrive first has a high-pass characteristic for all scattering angles. The second signal has a high-pass characteristic for small scattering angles, where the wave essentially passes straight through the cylinder, and a low-pass characteristic for large (and hence back) scattering angles.

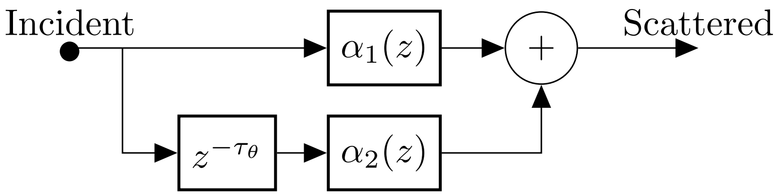


FIGURE 7.5: A block diagram of the tree-node scattering filter defined according to τ_θ, the angle dependent delay in samples. After [281].

Figure 7.5 expresses the required angle dependent scattering filter in block diagram form as used in [281]. $z^{-\tau_\theta}$ is the sample delay equivalent to the path distance between the two parts of the scattered signal, $\alpha_1(z)$ represents the filtering action associated with the shorter of the two scattering paths, and $\alpha_2(z)$ represents the filtering associated with the longer path.

As stated previously, the two-dimensional Treeverb network geometry definition is essentially 2.5-dimensional, and both spherical spreading losses, assuming a perfectly absorbing forest floor, and cylindrical spreading losses can be considered, resulting in different, if non-physical in the latter case, reverberant effects. Although Treeverb was conceptually derived as a digital waveguide network, computational limits in terms of both run-time costs and memory requirements resulted in an implementation based on the geometrical acoustics image-source method instead [261]. This implementation creates an offline impulse response as the system output for use as part of a convolution based reverberation algorithm.

7.2.4 Scattering Delay Networks

Another digital waveguide network based digital reverberator design is the scattering delay network (SDN) [271], [282]. An SDN reverberator is similar to the Treeverb system in that it decomposes the space to be modelled into a set of nodes representing the first-order reflection points, interconnected with bidirectional waveguides. A signal is introduced to the SDN from source node, S, and output at receiver, R, both connected via unidirectional delay lines. In the Treeverb reverberator the tree-nodes make up the whole of the physical surroundings, resulting in a relatively sparse set of possible reflection paths for multiple reflection orders. However, the wall-nodes in a SDN system are accurate for first-order reflections but only approximate acoustic behaviour at higher orders as the much larger surfaces involved and typically closed nature of the system result in many more possible connected paths.

An example SDN is shown in Figure 7.6 for the case of a two-dimensional rectangular model of a simple room, with four wall-nodes, W1, W2, W3, W4, interconnected with bidirectional delay lines. A signal is introduced to the SDN from source node, S, connected via unidirectional delay lines, and output at receiver, R, also connected via unidirectional delay lines.

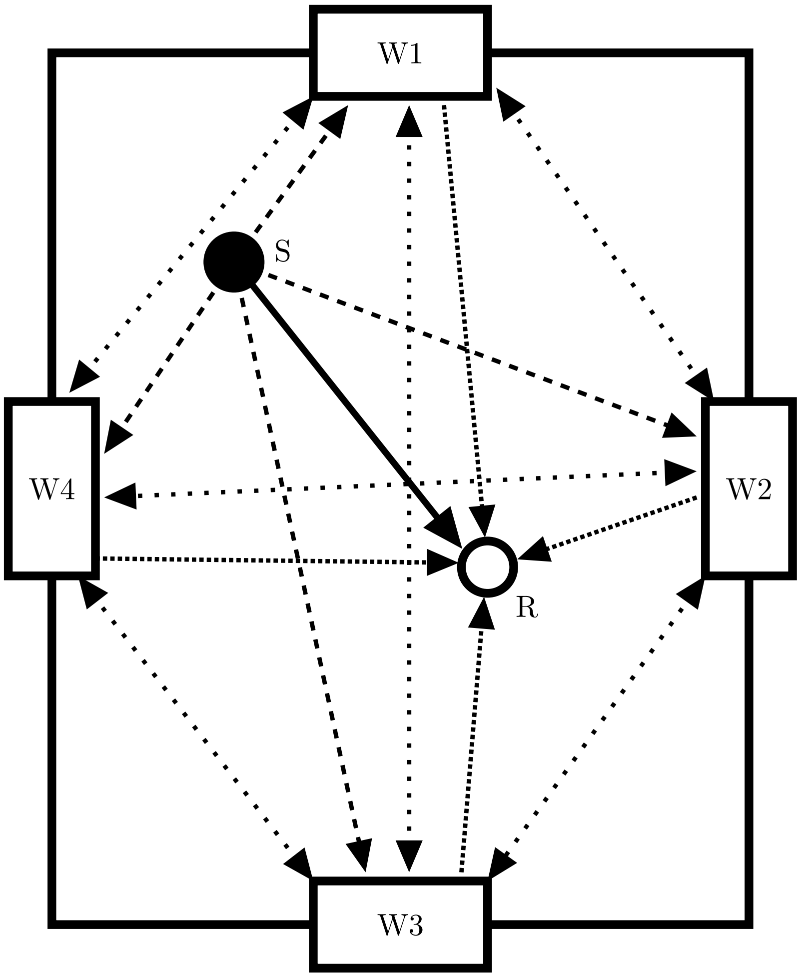


FIGURE 7.6: A representation of a scattering delay network (SDN) reverberator for a two-dimensional rectangular room, showing wall nodes, W1, W2, W3, W4, corresponding to first order reflection points, interconnected using bidirectional delay lines, for a given source and receiver node, both connected into the SDN using unidirectional delay lines.

A block diagram representing the operation of the SDN reverberator is shown in Figure 7.7 after [282] where a detailed description of each stage can be found and is presented here in overview. The input signal, $x(n)$, is applied at the source node, and through the application of input delay and attenuation matrix operators ($\mathbf{D}_s(z)$ and $\mathbf{G}_s$ respectively), this signal is transmitted to each wall-node. The scattering matrix $\bar{\mathbf{S}}$ is then applied to scatter this incoming signal between wall-nodes with $\mathbf{H}(z)$ applying frequency dependent absorption at each.

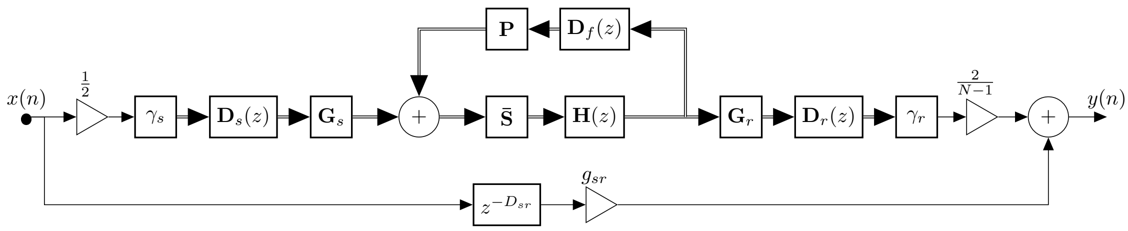


FIGURE 7.7: The scattering delay network overview block diagram, taking account of explicit direct sound and first reflection paths via wall-nodes for a given source and receiver position, with higher order reflections derived via the main feedback loop, after [282].

For higher than first-order reflections, a feedback loop comprising inter-node delays $\mathbf{D}_f(z)$ in series with permutation matrix $\mathbf{P}$ is applied to recursively model the higher order reflection behaviour of the system. Attenuation matrix $\mathbf{G}_r$ and delay factors $\mathbf{D}_r(z)$ associated with each wall-node to receiver connection are then applied for generating output signal $y(n)$. The matrices γ_s and γ_r represent the directivity patterns between the SDN nodes and the source and receiver respectively. Note that for all of the simulations included here the source and receiver are modelled as omnidirectional, so neither γ_s nor γ_r will be considered further. Finally, $z^{-D_{sr}}$ and g_{sr} represent the direct path delay and attenuation respectively.

Of particular interest are the scattering matrix $\bar{\mathbf{S}}$ and permutation matrix $\mathbf{P}$. The former represents the scattering associated with the SDN system as a whole, and is formed of identical smaller scattering matrices, $\mathbf{S}$, representing the scattering at each individual wall-node:

$$\mathbf{S} = \frac{2}{N-1}\mathbf{1}_{(N-1)(N-1)} - \mathbf{I} \tag{7.21}$$

where N is the number of wall-nodes in the system, $\mathbf{1}$ is a matrix of ones, and $\mathbf{I}$ is an identity matrix. For an outgoing wave pressure signal p_{ij}^{-}, and incoming wave pressure p_{ij}^{+}, from wall-node i to wall-node, j, $\mathbf{S}$ determines the outgoing pressure signal from one wall-node to the other wall-nodes in the system given knowledge of the incoming pressure signal:

$$p_{ij}^{-} = \mathbf{S} p_{ij}^{+} \tag{7.22}$$

where ij denotes the signal direction, $-$ indicates the outgoing wave pressure from node i to node j, and $+$ indicates the incoming wave pressure. This can be extended to characterise the whole system $\bar{\mathbf{S}}$:

$$\bar{\mathbf{S}} = \mathrm{diag}(\underbrace{\mathbf{S} \ldots \mathbf{S}}_{N}) \tag{7.23}$$

The scattering matrix $\bar{\mathbf{S}}$ is therefore a $N(N-1) \times N(N-1)$ matrix that determines the spread of acoustic energy among the bidirectional delay lines connecting the wall-nodes. For the system to work recursively and model higher order reflection paths it is required to re-arrange the result of this scattering in the feedback loop ready for input and a new scattering operation. Note that, after scattering, the outgoing pressure value p_{ij}^{-} is equivalent to the incoming pressure values p_{ij}^{+} at the next scattering instance, which, in order to be the correct input to the next scattering operation, must be rearranged to form the vector multiplied by the scattering matrix $\bar{\mathbf{S}}$. Hence the permutation matrix $\mathbf{P}$ is defined to rearrange the elements appropriately.

For an N node SDN, each node has $N-1$ permutation terms. Given wall-node m such that $1 \leq m \leq N$, connected to n other wall-nodes where $1 \leq n \leq N$ and $n \neq m$, we define:

$$P_{m,n} = \begin{cases} m - 1 + (n-1)(N-1), & n < m \\ mN + (n - m + 1)(N - 1), & n > m \end{cases} \tag{7.24}$$

Such that the required permutation, σ, is given by:

$$\sigma = \begin{pmatrix} 1, & 2, & \ldots, & N-1, & \ldots, & (N-1)(N-1)+1, & \ldots, & N(N-1) \\ P_{1,2}, & P_{1,3}, & \ldots, & P_{1,n}, & \ldots, & P_{m,1}, & \ldots, & P_{m,n-1} \end{pmatrix} \tag{7.25}$$

The input scaling factor of $\frac{1}{2}$ is included in order to provide the intended pressure at each node [282]. This input scaling is then compensated for by the output being scaled by a factor of $\frac{2}{N-1}$. The denominator value of $N-1$ in this case compensates for the $N-1$ 'copies' of the input signal being applied to each of the N nodes by the input delay and attenuation matrices, copies which are made to allow the output result of the application of $\mathbf{G}_s$ to be combined with the results of applying the permuation matrix.

The application of **P** to the incoming pressure values p_{ij}^{+} therefore results in the required input reordering. This alternating scattering-permutation matrix operation allows the SDN to successively model the interaction defined by the bidirectional delay elements connecting the wall-nodes. The SDN is, therefore, an efficient and effective method of reverberation design for room acoustic simulation, with accurate first-order reflections and good perceptual accuracy [282]. However, for correct scattering paths between nodes, to potentially incorporate direction dependent filtering as found when considering reflections from objects similar to a rigid cylinder, a modified approach is required.

7.3 The Waveguide Web

This section introduces the WGW, a novel type of waveguide network that has been designed to allow for the implementation of directionally dependent filtering at each node. As such it allows for the precise characterization of second-order reflection attenuation, following previous work indicating the importance of first- and second-order reflections in the characterization of the acoustics of sparse outdoor spaces [47].

7.3.1 Design Overview

The design of the Waveguide Web is similar to the SDN where the modelled space is represented by a set of scattering nodes connected to one another via bidirectional delay lines. Source and receiver nodes are also connected to these scattering nodes by unidirectional delay lines. Where the WGW differs from the SDN is in the scattering action at each node. Whereas the SDN implementation allows for one filtering action only at each node, the WGW design allows for directionally dependent filtering to be implemented. Like the SDN, the WGW method presents an abstracted representation of a space, based on an interconnected network of significant reflection points. These points can be at any 3D position, as required by the geometry of the system being modelled.

WGW Connections

Figure 7.8 shows all of the connections and filters associated with a node j in an N-node structure, including the source-to-node connections, inter-node connections, and node-to-receiver connections. In this diagram K is a vector, formed of $N-1$ elements, denoting the indices of all nodes present in the system apart from node j. For example, if $N=4$ and $j=3$, then $K=[1,2,4]$. If the case was that $j=2$ however, then $K=[1,3,4]$.

In the SDN, each node only has a single filter associated with it. As shown in Figure 7.8, in the WGW design each node has N^2 filters: one filter for the first-order reflection between source and receiver occurring at that node; $N-1$ filters corresponding to incoming signals from all other nodes and outgoing to the receiver; $N-1$ filters corresponding to the signal incoming from the source and outgoing to all other nodes; and $(N-1)(N-1)$ filters corresponding to recirculating signals incoming from all other nodes and then also outgoing to all other nodes.

The notation for these filters is as follows: H_{ijk} represents a filter at node j acting on a signal arriving from node i that will ultimately be sent to node k. In the case of the signal arriving from the source node, S is used in place of i, and where the signal is ultimately being sent to the receiver node, R is used in place of k.

The delay lines present in Figure 7.8 are absorptive, as represented by the lumped gain factors placed at the end of each one. In the case of second order reflections the gain values for the 'middle path' of each one are combined with the attenuation multiplier for the final node to receiver path. In Figure 7.8 it is shown that there are three steps in the WGW structure where this attenuation takes place: between the source and each node (g_{Sj} in Figure 7.8), between each node and the receiver (g_{jR}), and between each node and the receiver where the incoming signal is from each of the other nodes and not from the source ($g_{K_ijR} \cdots g_{K_{N-1}jR}$).

In order to maintain correct attenuation according to $\frac{1}{r}$, where r is the distance travelled, the delay line attenuation values present at each of the points identified in Figure 7.8 are

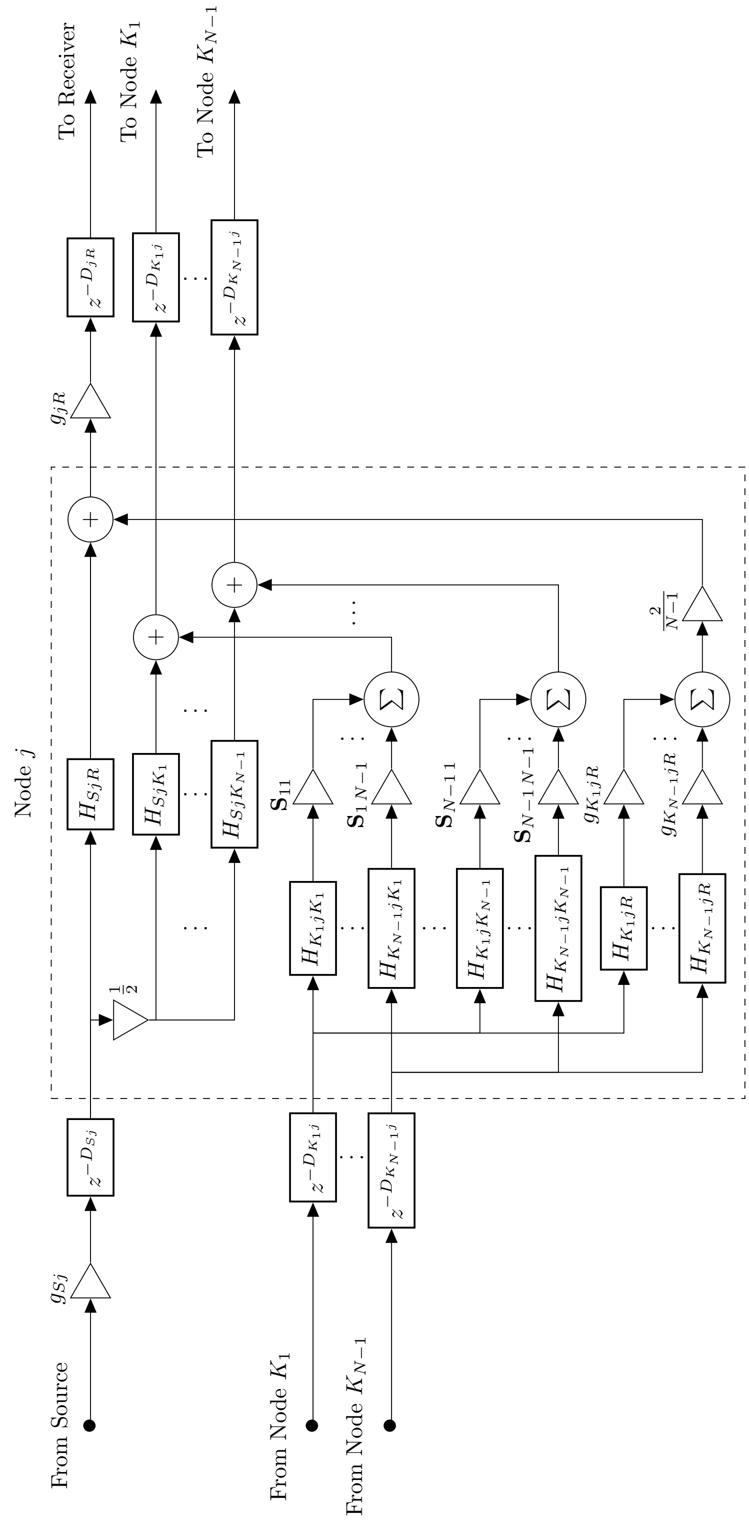


FIGURE 7.8: Structure of the source-to-node, inter-node, and node-to-receiver connections at a single WGW node. The implementation of the directionally dependent filtering at each node is shown, including how the the elements of the scattering operator **S** are applied to the incoming signal from each node.

calculated using the following equations:

$$g_{ST_M} = \frac{1}{\|x_S - x_{T_M}\|} \tag{7.26}$$

$$g_{T_M R} = \frac{1}{1 + \frac{\|x_{T_M} - x_R\|}{\|x_S - x_{T_M}\|}} \tag{7.27}$$

$$g_{T_M T_N R} = \frac{1}{1 + \frac{\|x_{T_N} - x_{T_M}\| + \|x_R - x_{T_N}\|}{\|x_S - x_{T_M}\|}} \tag{7.28}$$

where g_{ST_M} is the attenuation between the source and the node indicated by T_M, $g_{T_M R}$ is the attenuation between node T_M and the receiver, and $g_{T_M T_N R}$ is the attenuation associated with the total path from node T_M via node T_N to the receiver. Note that (7.26) and (7.27) are formulated identically in the SDN [271] to give correctly attenuated first-order reflections. The WGW extends the SDN algorithm in this regard with the addition of (7.28) providing correctly attenuated second-order reflections. Higher order reflections are reproduced less accurately, and do not follow the $1/r$ law, but still produce a rich reverberation tail. In Figure 7.8, the gain quantities defined by (7.26)-(7.28) are represented by the multiplication operators labeled g_{Sj}, g_{jR}, and $g_{K_i jR} \ldots g_{K_{N-1} jR}$ respectively.

Also shown in Figure 7.8 are the elements of the scattering matrix $\mathbf{S}$ and how they are applied to incoming signals at each node. In Figure 7.8 each multiplier marked $\mathbf{S}_{ij}$ indicates the element in row i and column j of the matrix $\mathbf{S}$, as in (7.21). There is an important distinction here between the WGW and the SDN. In the case of the SDN, the scattering operation is applied to incoming signals at each node regardless of their point of origin and further destination. Here the scattering operator allows for directionally dependent filtering to be applied by making $N - 1$ copies of each incoming signal and filtering as appropriate.

7.3.2 WGW Structure

Presented in Figure 7.9 is the overall structure of the WGW in block diagram form. A comparison with Figure 7.7 shows the similarity between the designs of the WGW and the SDN as well as their differences. In the case of the WGW the first-order reflections are calculated separately along with the direct path. This is because the filters associated with first-order reflections (i.e. from source-to-receiver via a single node) are not appropriate to

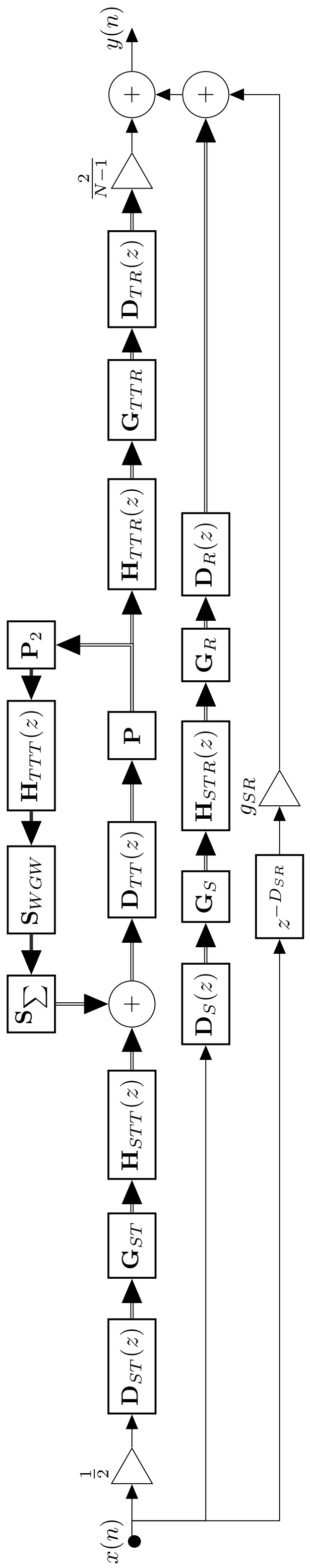


FIGURE 7.9: Block diagram of the WGW reverberator. Here there are paths for the explicit calculation of first-order reflections, alongside the direct sound path, separate from the main path for the calculation of reflections of second-order and above.

be fed back via scattering and permutation operations. Instead, the main section of the WGW is used to model the second-order reflections represented by the given node layout, with reflections of third-order and above calculated using the feedback loop.

Starting with the first-order reflection feed-forward path in the block diagram of Figure 7.9,

$$\mathbf{D}_S(z) = \mathrm{diag}(z^{-D_{S1}}, z^{-D_{S2}}, \ldots, z^{-D_{SN}}) \tag{7.29}$$

$$\mathbf{D}_R(z) = \mathrm{diag}(z^{-D_{1R}}, z^{-D_{2R}}, \ldots, z^{-D_{NR}}) \tag{7.30}$$

are the source and receiver delay matrices for the first-order section.

$$\mathbf{H}_{STR}(z) = \mathrm{diag}(H_{S1R}(z), H_{S2R}(z), \ldots, H_{SNR}(z)) \tag{7.31}$$

is the $N \times N$ matrix that contains the filters representing the first-order reflections, and $z^{-D_{SR}}$ and g_{SR} are the direct path delay and attenuation respectively. Considering the second-order reflection feed-forward path in the block diagram of Figure 7.9,

$$\mathbf{D}_{ST}(z) = \mathrm{diag}(\underbrace{z^{-D_{S1}} \ldots z^{-D_{S1}}}_{N-1}, z^{-D_{S2}} \ldots z^{-D_{SN}}) \tag{7.32}$$

$$\mathbf{D}_{TT}(z) = \mathrm{diag}(\underbrace{z^{-D_{12}} \ldots z^{-D_{12}}}_{N-1}, z^{-D_{1N}} \ldots z^{-D_{N(N-1)}}) \tag{7.33}$$

$$\mathbf{D}_{TR}(z) = \mathrm{diag}(z^{-D_{2R}}, \ldots, z^{-D_{NR}}, \ldots, z^{-D_{1R}}, \ldots, z^{-D_{(N-1)R}}) \tag{7.34}$$

are the source-node, inter-node, and node-receiver delay matrices[1], and

$$\mathbf{H}_{STT}(z) = \mathrm{diag}(H_{S12}(z) \ldots H_{S1N}(z), \ldots, H_{SN1}(z), \ldots, H_{SN(N-1)}(z)) \tag{7.35}$$

$$\mathbf{H}_{TTR}(z) = \mathrm{diag}(H_{12R}(z), \ldots, H_{1NR}(z), \ldots, H_{N1R}(z), \ldots, H_{N(N-1)R}(z)) \tag{7.36}$$

are the source-node-node, node-node-receiver filter matrices. Note that these filters also include the effect of absorption at a node due to the process of reflection.

$$\mathbf{G}_{ST} = \mathrm{diag}(\underbrace{g_{S1} \ldots g_{S1}}_{N-1}, g_{S2} \ldots g_{SN}) \tag{7.37}$$

[1] There are effectively $N-1$ 'copies' of each delay to allow for the implementation of the second-order filters.

$$\mathbf{G}_{TTR} = \mathrm{diag}(g_{12R}, \ldots, g_{1NR}, \ldots, g_{N1R}, \ldots, g_{N(N-1)R}) \tag{7.38}$$

are the second-order source-node and node-receiver attenuation matrices respectively, and

$$\mathbf{G}_S = \mathrm{diag}(g_{S1}, g_{S2}, \ldots, g_{SN}) \tag{7.39}$$

$$\mathbf{G}_R = \mathrm{diag}(g_{1R}, g_{2R}, \ldots, g_{NR}) \tag{7.40}$$

are the source and receiver attenuation matrices for the first-order section. The permutation matrix $\mathbf{P}$ is formulated according to the SDN design as described in Section 7.2.4.

7.3.3 Feedback Loop

Besides the feedback path, there are $N(N-1)$ channels used in the calculation of second-order reflection paths. In order to accommodate directional dependent filtering at each node $N-1$ copies of each channel must be made. This action is in effect performed by the matrix $\mathbf{P}_2$, creating $N-1$ copies of the output from $\mathbf{P}$.

In order to do this $\mathbf{P}_2$ must take the form of an $N(N-1)(N-1) \times N(N-1)$ matrix. $\mathbf{P}_2$ is formed of multiple copies of a sub-matrix $\mathbf{P}_{2s}$ given by:

$$\mathbf{P}_{2s} = [\underbrace{\mathbf{I}_{N-1} \ldots \mathbf{I}_{N-1}}_{N-1}]^T \tag{7.41}$$

$\mathbf{P}_{2s}$ is then an $(N-1)(N-1) \times (N-1)$ matrix. If we then apply the following tensor product operation (as denoted by $\otimes$:

$$\mathbf{P}_2 = \mathbf{I}_N \otimes \mathbf{P}_{2s} \tag{7.42}$$

we get the desired $N(N-1)(N-1) \times N(N-1)$ matrix $\mathbf{P}_2$. The result of applying $\mathbf{P}_2$ can then be input to $\mathbf{H}_{TTT}(z)$, which is formulated according to

$$\mathbf{H}_{TTT}(z) = \mathrm{diag}(H_{212}(z), \ldots, H_{N1N}(z), \ldots, H_{1N1}(z), \ldots, H_{(N-1)N(N-1)}(z)) \tag{7.43}$$

and represents the $N(N-1)(N-1) \times N(N-1)(N-1)$ node-node-node filters.

Following the application of the directionally dependent filtering, the scattering operation can take place. $\mathbf{S}_{WGW}$ is given by:

$$\mathbf{S}_{WGW} = \mathrm{diag}(\underbrace{\mathrm{vec}(\mathbf{S})^T \ldots \mathrm{vec}(\mathbf{S})^T}_{N}) \tag{7.44}$$

and allows for the correct element of $\mathbf{S}$ to be applied to each incoming filter signal. The result must then be summed appropriately to give the total scattering output at each node, using the operator $\mathbf{S}_{\sum}$:

$$\mathbf{S}_{\sum} = \mathbf{I}_{N(N-1)} \otimes [\underbrace{1 \ldots 1}_{N-1}] \tag{7.45}$$

In this way, the application of both $\mathbf{S}_{WGW}$ and $\mathbf{S}_{\sum}$ recombines the $N(N-1)(N-1)$ channels present in the feedback loop back into the $N(N-1)$ inter-node wave variables required for reinsertion and further propagation through the system.

Since the WGW makes use of essentially the same scattering operation as the SDN, it is similarly stable [266] regardless of the length of the delay lines connecting the nodes. As a result the addition of losses at the nodes will always result in a stable network.

This section has detailed the design and structure of the WGW. Highlighted are the key points where its design differs from that of the SDN, namely: the separate calculation of first-order reflections as well as the direct path; the extension of the structure at each node to accommodate directionally dependent filtering; and the accordant changes made to the scattering operation.

7.4 Evaluation

The performance of the WGW will now be evaluated using the following case studies:

- A comparison of IR simulations for a shoebox room using the SDN and the WGW. The purpose of this comparison is to validate the WGW and see where the differences in design between the two manifest themselves in the resultant rendered impulse responses.
- A simulation of a forest environment using filters designed according to Spratt and Abel's Treeverb design [281] and Morse's solution to acoustic scattering from a rigid

cylinder [298]. The forest environment used is formed of 25 trees arranged in a semi-regular grid pattern. It is evaluated with reference to Chobeau's results regarding the sound propagation in forests [299], and Wiens' MATLAB implementation of Treeverb [300].

- A simulation of an urban courtyard where impulse responses have been previously measured [47], [215]. These measurements were used to inform the structure of the WGW model as an approximation of the space, and to compare with the resultant simulated impulse response.

All of the IRs generated using the WGW here were calculated at a sample rate of 48 kHz, and have been made available online as part of the OpenAIR Library [301]. They are also on the attached data CD.

7.4.1 SDN Shoebox Comparison

In order to validate the design of the WGW, presented here is a comparison of a simulation for a $9\,\mathrm{m} \times 7\,\mathrm{m} \times 4\,\mathrm{m}$ shoebox room (with an absorption coefficient $\alpha = 0.2$ defined for all surfaces) made using the SDN [271], and the same simulation made using the WGW. Figure 7.10 (a) shows the two results overlaid on one another (with the WGW results marked in red), and Figure 7.10 (b) shows the remainder following subtraction of the WGW result from that of the SDN.

The results presented in Figure 7.10 validate the design of the WGW as they indicate identical reproduction of the direct sound path and first-order reflections (i.e. at those sample instances the difference between the SDN and WGW simulations is 0). This figure shows the difference in second-order reflection amplitude between the WGW and SDN results, where the WGW calculates them correctly according the the $\frac{1}{r}$ law. This difference in second-order reflection amplitude accordingly leads to further small differences for the reflections beyond second-order, and as such the reverberant tail of the response, these results are confirmed by Figure 7.11 which shows reverberation time for octave bands from 125 Hz to 16 kHz, based on derivation of the T30 room acoustic parameter [302], [303].

The results shown in Figure 7.10 and Figure 7.11 validate the WGW results as they are close to those obtained from the SDN, but differ according to the novel design elements

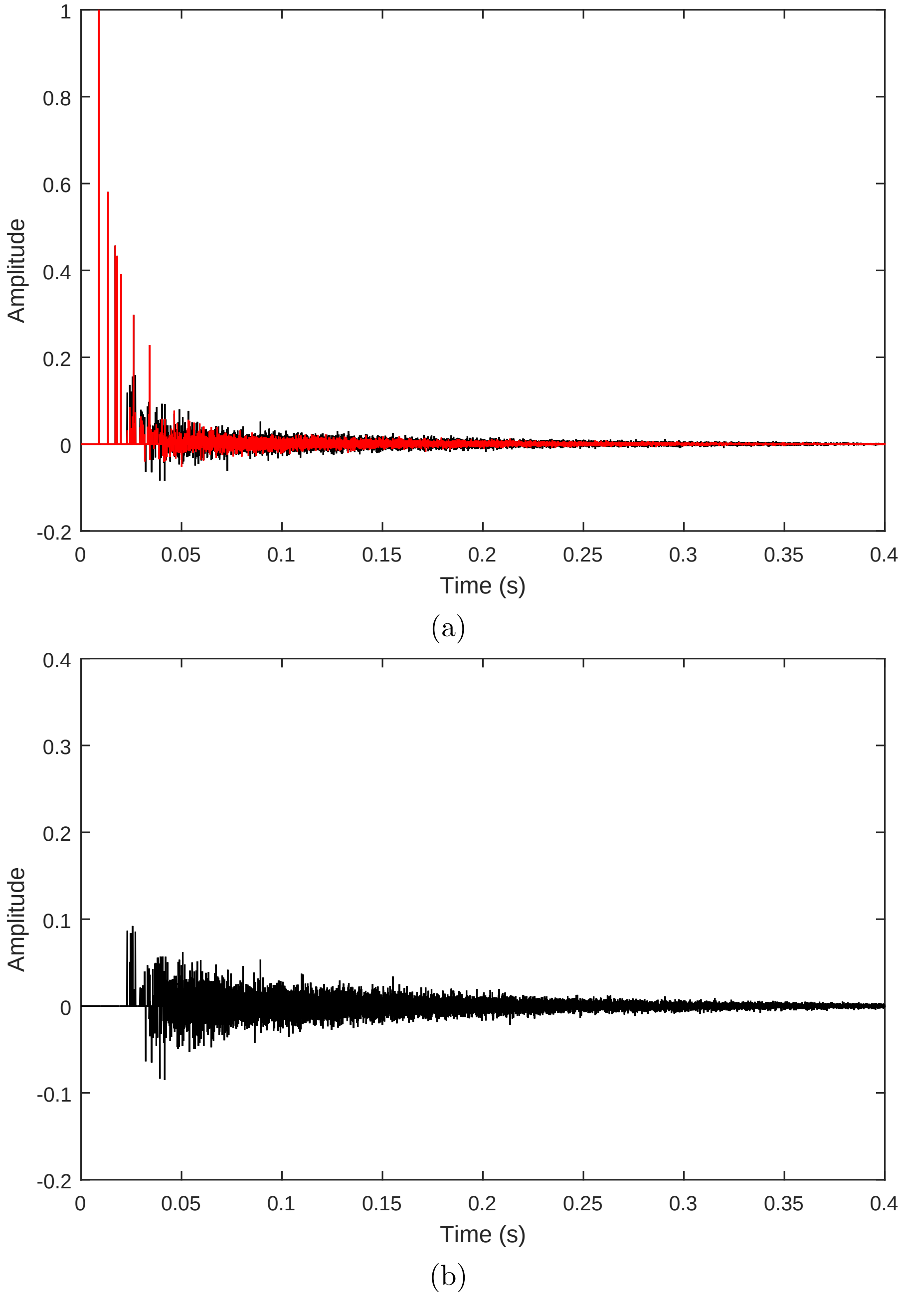


FIGURE 7.10: (a) Comparison of SDN result (black) with WGW (red). (b) Difference between the two.

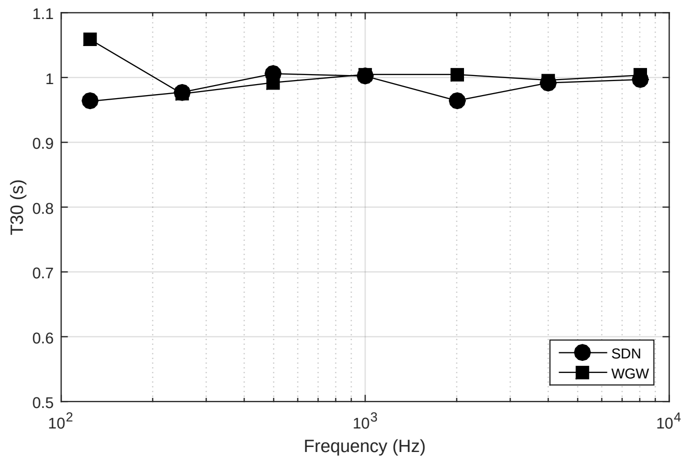


FIGURE 7.11: Octave band reverberation time results for SDN and WGW derived impulse responses obtained from a 9 m × 7 m × 4 m shoebox-like room simulation.

implemented in the WGW (i.e. the attenuation values defined by equations (7.26)-(7.28)). The results shown in Figure 7.11 are clearly similar, and mostly fall within the 5% just-noticeable-difference (JND) associated with T30 measurement [304][2], indicating the results from the SDN and the WGW to be perceptually alike. On this basis the WGW can be used to obtain simulation results for further case studies.

7.4.2 Forest Environment

This section includes a comparison of results from the WGW with those from two previous studies regarding the acoustic properties of forest environments: Spratt and Abel's Treeverb design as mentioned in Section 7.2.3; and Chobeau's modelling of forest environments using the transmission line matrix (TLM) method [299].

Chobeau's work was chosen for comparison as its results contain a set of impulse responses based on two-dimensional reflection and scattering from regular arrays of tree positions. This method therefore represents an approximate acoustic model of a forest that is suitable for comparison with results gathered from the WGW.

[2]Although other research has suggested the JND for T30 to be higher than this [305].

Filter Design

As suggested by Spratt and Abel [281], the filtering used at each node when simulating a forest acoustic is designed to emulate scattering from a rigid cylinder (representing a tree trunk). As detailed in Section 7.2.3 when a plane wave is incident upon such a cylinder the result is formed of two parts, with scattered propagation paths travelling in each direction around the cylinder.

In Figure 7.5, $z^{-\tau_\theta}$ represents a delay equivalent to the path distance between the two parts of the scattered signal, $\alpha_1(z)$ represents the filtering action associated with the shorter of the two scattering paths, and $\alpha_2(z)$ represents the filtering associated with the longer path.

In the current design of the WGW, these filters are implemented using first-order IIR filters that can be used to perform both high and low pass operations. Two of these filters are used in the configuration shown in Figure 7.5 to emulate the two scattered paths shown in Figure 7.4. A high pass filter is applied to the shorter of the two scattering paths with cut-off frequency f_c defined by:

$$f_c = \frac{c}{(\pi + \theta)r} \tag{7.46}$$

where f_c has a wavelength equal to the total length of the longer scattering path around the tree. For small scattering angles the longer of the two paths has a high pass characteristic with the same break frequency. For large scattering angles it changes to a low pass filter with cut-off frequency:

$$f_c = \frac{c}{r \cdot \frac{\theta}{\pi}} \tag{7.47}$$

where f_c has a wavelength equal to the proportion of half of the tree's circumference represented by the scattering angle.

This filtering is applied at each tree node together with the additional application of a reflectance factor determined by the total amount of scattered energy according to Morse's solution. Figure 7.12 shows how the total scattered energy changes with the scattering angle. At $\theta = 0$ there is effectively complete transmission, but at $\theta = \pi$ the reflectance value is reduced to about 0.25. This represents an average tree trunk with radius $r = 0.2\,\mathrm{m}$, and is representative of values of r between 0.1 and 0.5 m.

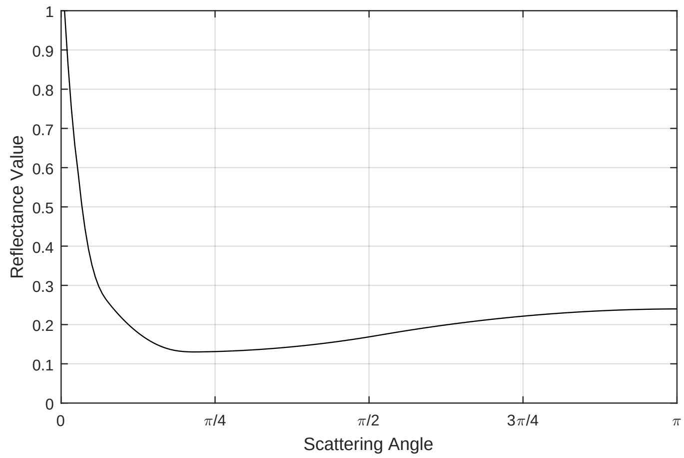


FIGURE 7.12: Plot of reflectance against scattering angle as used in the forest reverb example. These values are generated using Morse's solution to scattering from a rigid cylinder.

Figure 7.13 compares Morse's solution to acoustic scattering from a rigid cylinder with results made using the design detailed here at three different scattering angles. These plots indicate a good match at high frequencies with greater discrepancies appearing at lower frequencies below around 100 Hz. Whilst more sophisticated (i.e. higher order) filters could be implemented to match more closely Morse's solution, the results presented in Figure 7.13 are sufficient for use here.

Treeverb Comparison

In order to compare WGW results with those made using Treeverb, a MATLAB implementation of the image-source based Treeverb algorithm [300] was used to generate a forest environment (shown in Figure 7.14 where 25 trees with radii between 0.2 m and 0.5 m have been distributed over a 30×30 m region. This layout of trees was input to this algorithm and used to generate the IR shown in Figure 7.15, as obtained when considering up to fifth-order scattering paths.

This same tree layout is also used to generate the WGW IR shown in Figure 7.16. Figure 7.17 shows the same IR with a focus on the first 200 ms. Comparison with the Treeverb result shows this to be more plausible with distinct initial reflections followed by a much less reverberant tail.

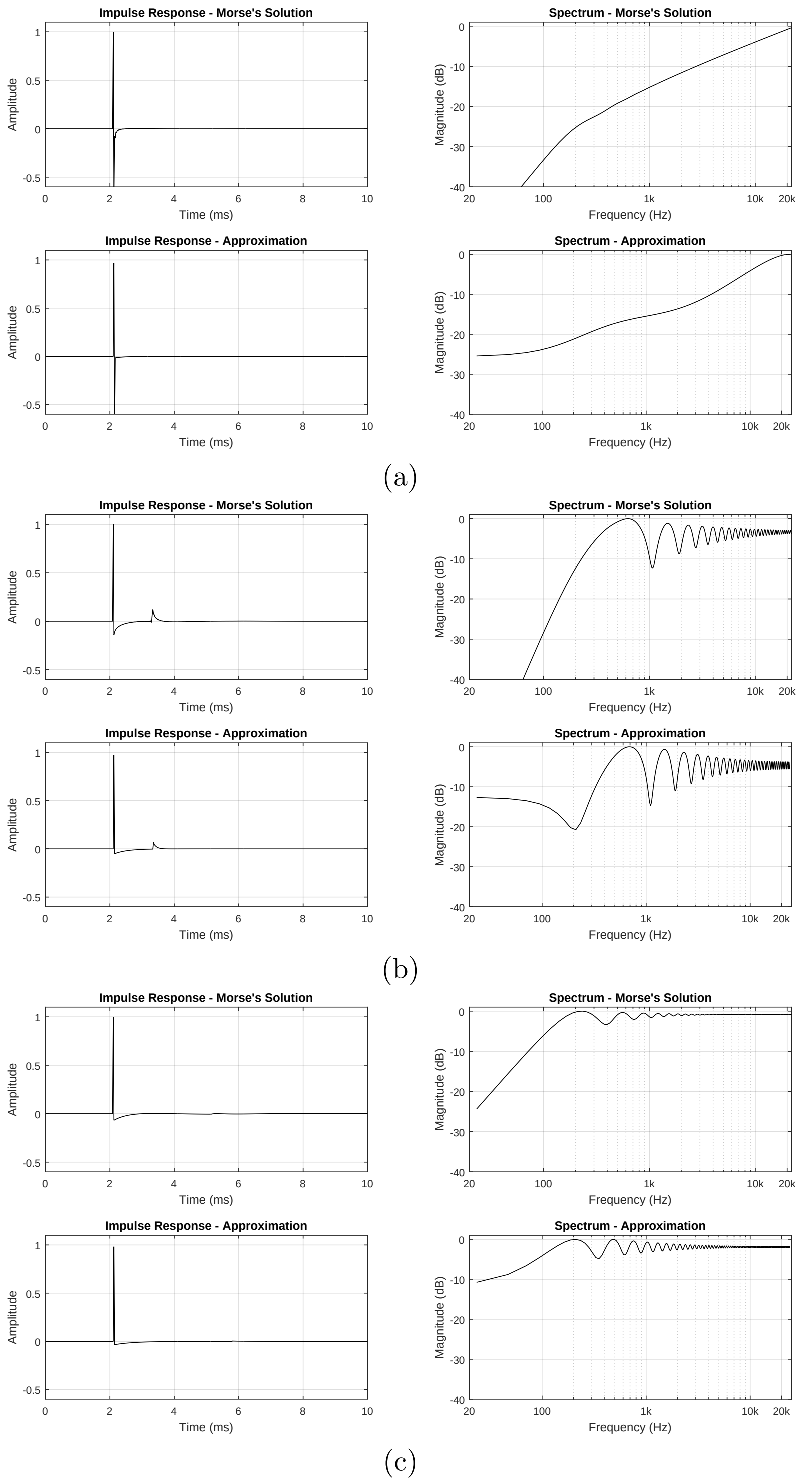


FIGURE 7.13: A comparison of Morse's solution to acoustic scattering from a rigid cylinder with the approximation formed of first-order filters as used in the WGW. For all examples $r = 0.2\,\text{m}$, the three examples represent the following reflection angles: (a) $\theta = 0^\circ$ (b) $\theta = 60^\circ$ (c) $\theta = 180^\circ$.

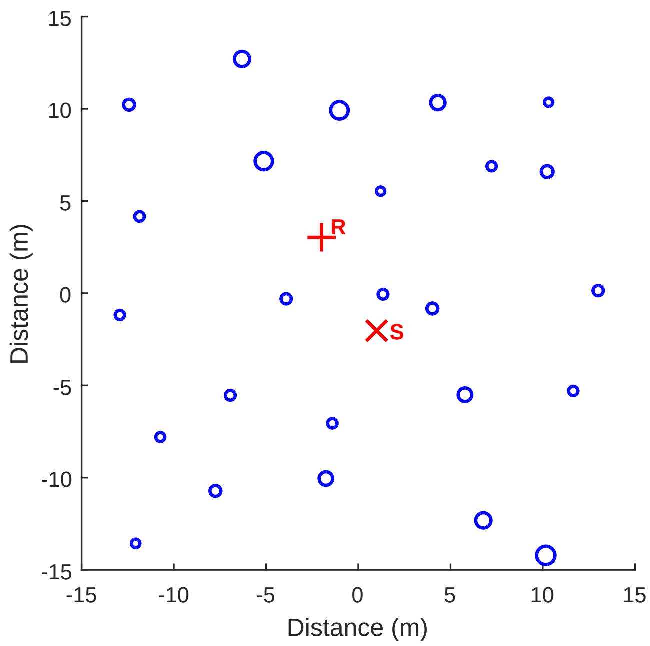


FIGURE 7.14: Forest configuration formed of 25 trees with radii between 0.2-0.5m, one source, S, and one receiver, R, distributed over a 30 × 30m region, used to generate the impulse response shown in Figure 7.15.

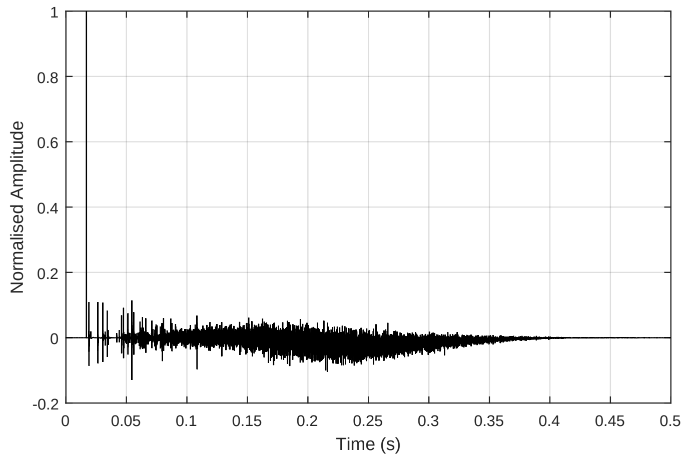


FIGURE 7.15: Impulse response for the forest configuration shown in Figure 7.14 generated using Spratt and Abel's Treeverb methodology based on the MATLAB implementation presented in [300].

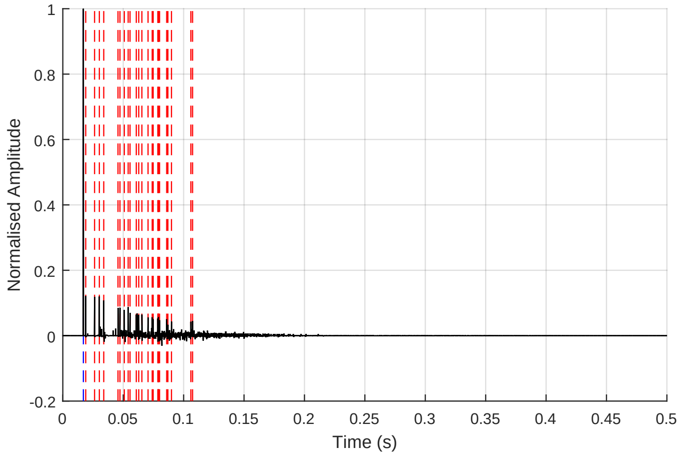


FIGURE 7.16: An IR generated using the WGW for the forest layout shown in Figure 7.14. The red lines indicate the timing of first-order reflections.

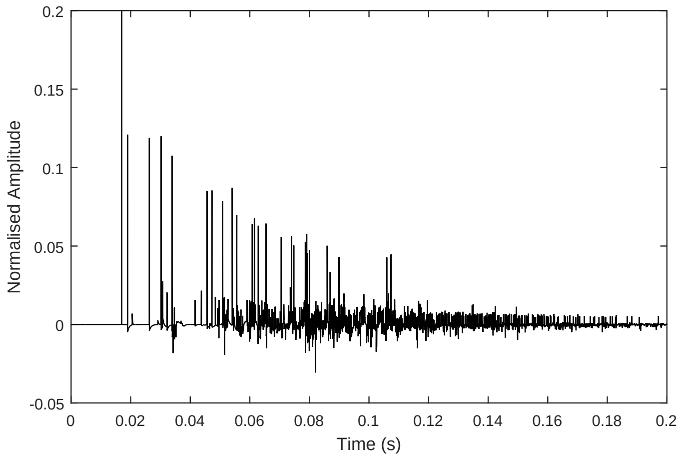


FIGURE 7.17: A closer view of the first 200 ms of the IR shown in Figure 7.16.

The lack of more objective analysis of the results obtained in the Treeverb paper make more detailed comparisons somewhat difficult to make. As such, in the next section results made using the WGW method will be compared with results from Chobeau's use of the Transmission Line Matrix (TLM) method to simulate a forest acoustic.

Chobeau Comparison

Chobeau made use of multiple forest layouts to determine the effect that tree placement has on sound attenuation level. In Chobeau's thesis it was determined that a major factor affecting the acoustic properties of a forest environment is the filling fraction, i.e. the fraction of the space occupied by the trees (represented by cylinders). For an aligned square grid of cylinders the filling fraction F is given by:

$$F = \frac{\pi d^2}{4a^2} \tag{7.48}$$

where d is the diameter of the cylinders and a is the 'lattice constant', or the distance between the cylinders' central points [306]. The filling fraction has an affect on the bandwidth of the spectral gaps produced by the environment. According to Bragg's law, the centre frequency of such a band gap can be approximated using:

$$f_c = \frac{c}{2 \cdot a} \tag{7.49}$$

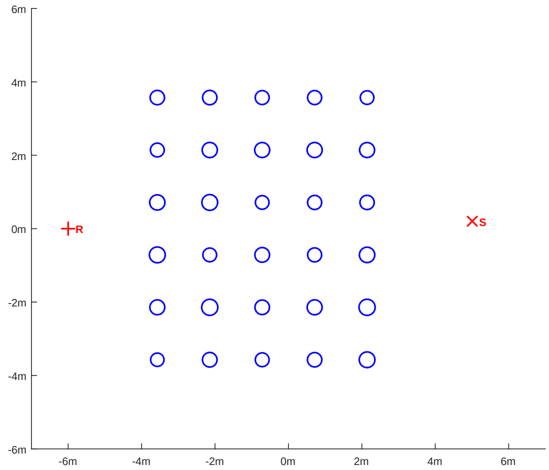
where c is the speed of sound in air and a is, again, the lattice constant [24].

The three distributions used to generate WGW results are shown in Figure 7.18, and were chosen to emulate the different distributions used by Chobeau. For all distributions the trees have a radius of 0.2m. In D1 the lattice constant is 1.42m, for D2 it is 2m, and for D3 it is 1m. The filling fractions for these distributions therefore range from about $F = 0.15$ to about $F = 0.03$.

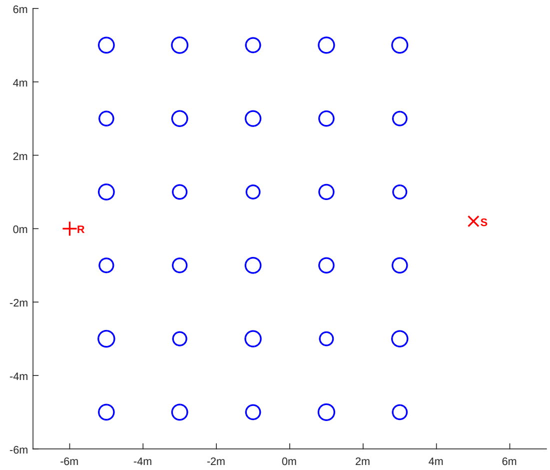
Using (7.49), for each of the regular distributions presented here, the centre frequency of an expected band gap can be calculated: for D1, $f_c = 121\,\text{Hz}$; for D2, $f_c = 86\,\text{Hz}$; for D3, $f_c = 172\,\text{Hz}$. These band gaps would typically be expected to be observed in topologies with filling fraction $0.4 < F < 0.6$ [104]. However, one would still expect to see pseudo-band gaps for low filling fractions [299] corresponding to the distributions used in this work. Figure 7.19 shows a spectral comparison between results for these four distributions.

The results shown in Figure 7.19 do line-up with these predictions somewhat, in that the first dip in each spectra for each distribution is near its predicted f_c value. As predicted, however, the low filling fraction associated with the distributions used means that these spectral band gaps are not very prominent. This is in accord with Chobeau's key findings, where no pronounced band-gaps were observed in simulations of similar forest environments.

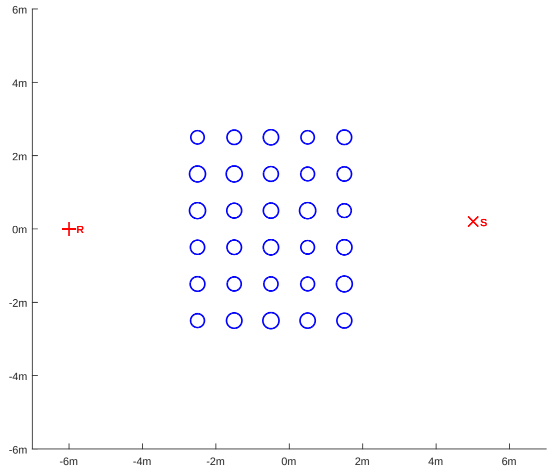
Also in accord with Chobeau's results are the global attenuation levels associated with each of the three distributions ($-7.57\,\text{dB}$ for D1,$-6.8\,\text{dB}$ for D2, $-10.1\,\text{dB}$ for D3), where the greatest attenuation is associated with the distributions showing the highest filling fraction. These attenuation levels are calculated as the average of the calculated sound



(a) D1



(b) D2



(c) D3

FIGURE 7.18: Schematics of the three distributions used in the forest simulations, each with 30 'trees' of radius $r = 0.2$ m: (a) D1 aligned distribution with 1.42 m spacing; (b) D2 Aligned distribution with 2 m spacing; (c) D3 aligned distribution with 1 m spacing.

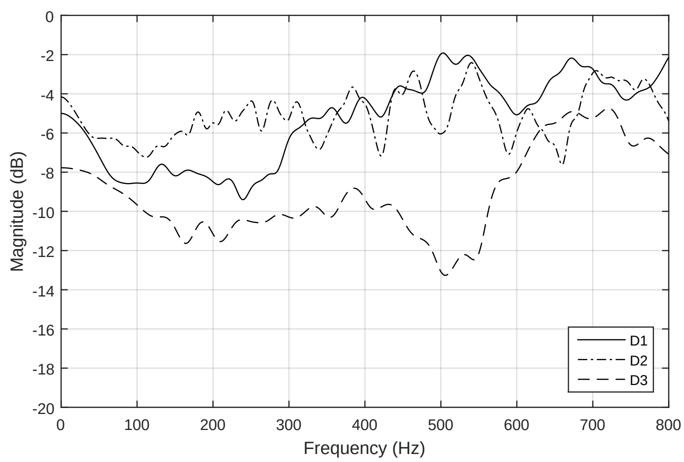


FIGURE 7.19: Frequency responses of simulations made using the three distributions shown in Figure 7.18.

level across all frequencies relative to an environment where no trees are present. The absolute values of these attenuation values are less important than how they relate to one another: they show that attenuation increases with the density of the distribution (i.e. the filling fraction), which is one of two main factors identified by Chobeau as having an effect on attenuation levels.

7.4.3 Urban Courtyard

While the WGW's design lends itself to the simulation of a forest acoustic, it is also intended for the modelling of more general sparse outdoor spaces. This section presents a comparison of results with sparsely reflecting IRs as obtained from acoustic measurements in a semi-enclosed courtyard [47]. The node positioning is determined from a full 3D model (shown in Figure 7.20[3]) used to calculate the main reflection paths based on a geometrical acoustics approach [215], with nodes placed at a selection of identified main first- and second-order reflection points. The resultant node layout is shown in Figure 7.21.

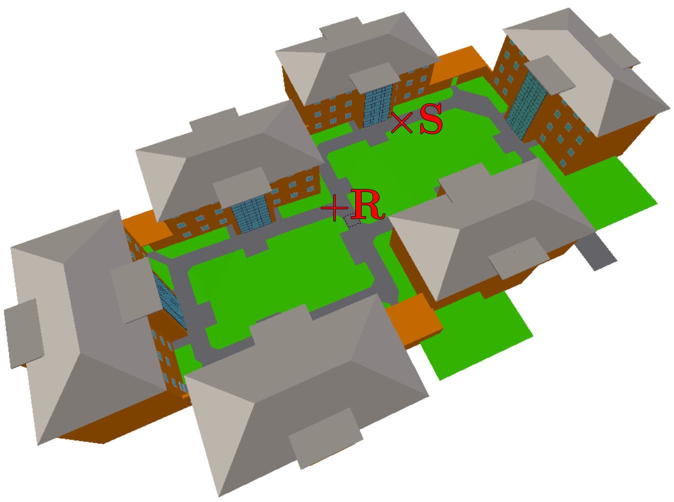


FIGURE 7.20: The 3D model of the urban courtyard considered here, based on actual measurements, and also used to identify the main reflecting surfaces. The labels S and R denote the position of the source and the receiver respectively.

Initial simulations were made using acoustic absorption coefficient (α) values for the materials observed in the internal courtyard buildings (predominately brick) with no filtering applied at each node (due to the relatively minimal frequency dependence in the acoustic absorption coefficients for these materials). However, it was found that these

[3]This 3D model was made in Blender using the building's plans and on-site measurement.

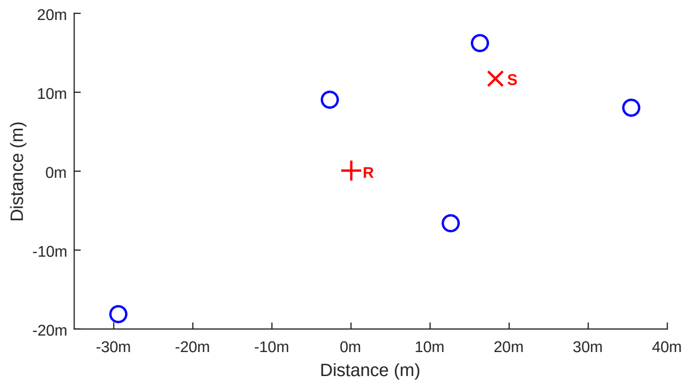


FIGURE 7.21: WGW node layout used for the simulations of the urban courtyard, corresponding to the main reflecting surfaces in Figure 7.20.

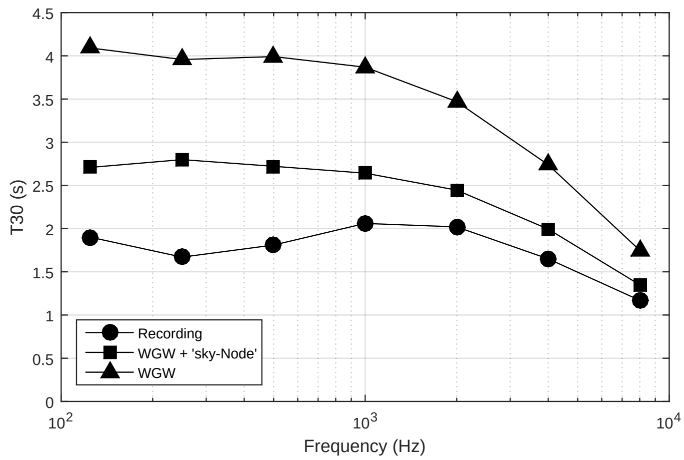


FIGURE 7.22: Octave band reverberation time results for the IR measurement/recording obtained from the urban courtyard and from two WGW simulations made using the same 5-node layout with one also including an absorptive 'sky-node'.

simulations produced overly reverberant IRs due to the location's open nature not being compensated for. As such an extra, totally absorptive, node was added as a 'sky-node' to compensate for the open nature of the courtyard.

The 'sky-node' is implemented simply by adding a node at some arbitrary position and assigning it an α value of 0. In terms of Figure 7.8 this in practice means all of the filter centered at that node will now include a coefficient such that all incoming samples are multiplied by zero.

Figure 7.22 presents reverberation time for octave bands from 125 Hz to 16 kHz based on T30 for two WGW simulations compared with reverberation times obtained from the measured IR[4]. In this figure the effect of adding a sky-node to the simulation can clearly be seen. Both WGW simulations in this case also had air absorption added using an analytical solution as presented in [307].

Figure 7.22 shows general good agreement in reverberation time results between the recorded and simulated IRs in octave bands from 500 Hz to 4 kHz but a larger deviation in the octave bands centered at 1 kHz and below. There remains considerable scope for further investigation into how best to model such a space, given the large number of possible node positions.

7.4.4 Computational Requirements

Table 7.1 shows the run time and memory required to run a WGW simulation of a forest environment with different numbers of nodes. For each simulation the source and receiver were positioned 10 m apart, and the nodes were positioned at random within a 10 m^2 space between the source and receiver. Each simulation was computed to give 1s of audio output at a sampling rate of 48 kHz. From this table it can be seen there is a very large increase in memory usage with an increasing number of nodes. This exponential increase is due to the implementation of directionally dependent filtering at each node, which requires the implementation of N^2 filters at each of the N nodes—resulting in the number of filters required for implementation therefore increasing with N^3.

Table 7.1: Run time and memory required for different numbers of nodes.

Nodes	Times (s)	Memory (MB)
5	4.35	5.65
10	9.58	18.32
15	39.65	672.10
20	155.89	5 508.98
25	358.38	27 350.03
30	667.23	102 404.16

[4] It was decided not to make use of the SDN method for comparison here. This was in part due to the high degree of similarity between the SDN and WGW shoebox simulations, and the fact that the SDN method was not designed with sparsely reflecting, outdoor scenes (such as the courtyard example) in mind.

7.5 Conclusion

This chapter has presented the Waveguide Web, a new reverberator design for outdoor, or sparsely reflecting, environments, which offers an extension to De Sena's SDN reverberator to include accurate second-order reflections and directionally dependent filtering at each node. Following the presentation of the WGW's design, several case studies were presented comparing WGW results with other examples. Firstly, a comparison was made with the SDN shoebox example as a validation of WGW operation. Secondly, a forest acoustic was modelled and compared with results from Spratt and Abel's Treeverb, which showed that the WGW produced more realistic results. This was further verified by considering the band gap absorption effect of a regular, grid-based, arrangement of trees. The results showed some agreement with prior work by Chobeau, indicating how acoustic attenuation due to the presence of such a regular arrangement of trees varies with tree/grid spacing. Spacings based on a higher filling fraction were also shown to increase global attenuation levels to some extent. The final case study was a comparison with impulse response measurements obtained from a semi-enclosed urban courtyard. Reverberation time results were in fairly good agreement following the incorporation of air absorption effects and the addition of a totally absorptive 'sky-node'.

The are several avenues for further research. The reflection/absorption filtering function for a node is well defined in the ideal specular case, and for the forest environment considered here. The boundary interactions observed in more general spaces, such as the urban courtyard example, are not so well defined in the context of this approach. The more formal use of acoustic bidirectional reflectance distribution functions to categorize reflecting nodes, as outlined in [258], may help in this regard. It would also be beneficial to implement analytical air absorption compensation directly as part of the WGW itself. The filtering currently used to represent a tree-node could also be extended to include filters of higher order, to better approximate Morse's solution to acoustic scattering from a rigid cylinder. Alternatively, behaviour of each tree node could also be modified to consider the sound-scattering properties of a single tree [107]. It is also important to note that a limitation of the current forest model is that, like Treeverb, it does not include ground reflections or foliage. As such future work could also consider the addition of these aspects to the WGW by using results from [308]. In the case of acoustic scenes, such as the courtyard case study, where there are large surfaces resulting in diffuse acoustic reflections, the use of radiosity modelling could also be considered [309].

This work represents an important step in the general development of reverberation algorithms more capable of modelling open acoustic scenes. Although this work was in part inspired by the simulation of reverberant forest environments, the results can be applied in more general circumstances, with the WGW algorithm additionally offering improvements in terms of accuracy over other existing reverb algorithms. The WGW offers a new simulation tool for researchers interested in the perception of environmental acoustics and the associated effects that sound and our built or natural environments can have on human health and well-being.

Chapter 8

Conclusion

This thesis has presented a portfolio of research into various aspects of environmental sound research. This has included the measurement of soundscape stimuli, panoramic visual capture, IR measurement, soundscape evaluation, and the acoustic modelling of sparsely reflecting acoustic scenes (as might be found in typical outdoor urban environments). This chapter will draw this thesis to a conclusion, including a summary of the work presented, and a re-statement of the hypothesis introduced in Chapter 1. Potential avenues for future research will then be considered, after which this thesis is concluded by some final comments from the author.

8.1 Summary

The thesis opened with Chapters 2, 3, and 4 which presented the background detail required in order to contextualise the thesis work. This began with a summary of the fundamentals of relevant acoustic and psychoacoustic theory, describing the behaviour of sound waves and the mechanism of the human hearing system. Following this introduction, the concept of environmental noise was presented, alongside a summary of its effects (auditory and non-auditory), and how it can be measured and quantified. Having then established the negative effects of noise on health and wellbeing, the field of soundscape research was introduced, including methodologies for the evaluation of environmental sound.

Chapter 5 then presented the measurement work conducted as part of this thesis. Firstly, a program of soundscape capture work was shown, including B-format audio recording, and panoramic visual capture. IRs recorded at Creswell Crags, a limestone gorge containing a series of caves, were also examined. This section showed how IR measurement techniques developed for use in indoor space can be applied in outdoor spaces, and then made use of SIRR analysis in order to demonstrate the acoustic difference between acoustic scenes inside the caves and in the sparsely reflecting gorge.

The soundscape material captured in Chapter 5 was then used in a series of listening tests, which were presented in Chapter 6. The first of these listening tests made use of surround-sound playback of the B-format material in order to compare SD pairs (an established evaluation tool) with the SAM. Results showed the SAM to be a viable, intuitive, alternative to SD pairs. This result warranted the use of the SAM in further testing. The second listening test then made use of stereo UHJ renderings of the B-format soundscape recordings. Playback of the soundscape material in this format was shown to be ecologically valid, where the test results indicated that the stereo UHJ rendering evoked statistically similar emotional states when compared with surround-sound presentation.

The third and fourth listening tests then introduced visual elements. The third listening test made use of the same stereo UHJ renderings of the soundscape material, presented alongside still panoramic images of the recording locations. Results from this test (which also introduced soundscape categorisation, and visual/aural element identification to the subjective evaluation process) were used to identify which soundscape scenes were to be used in the final listening test utilising full-motion spherical video. This final test presented the chosen soundscape scenes (in audio-only and audiovisual conditions) using YouTube as a VR platform, allowing for dynamic binaural rendering of the B-format recordings.

The final chapter included in the main body of this thesis was Chapter 7 which introduced the Waveguide Web (WGW), a novel method for the simulation of sparsely reflecting outdoor acoustic scenes. This chapter included a summary of extant acoustic modelling methods before explaining the design of the WGW, and an evaluation of that design in the form of three case studies: comparison with an SDN (a similar acoustic modelling method) simulation of a shoebox room, modelling of a forest environment with reference to Spratt and Abel's 'Treeverb' modelling method, and modelling an urban courtyard with comparison made to real-world recorded IRs.

Taken together the work presented in this thesis is an example of how a multi-faceted approach making use of soundscape methodologies can allow for the development of the understanding of environmental sound, and describe the physical components of a soundscape (sound sources, sound propagation) alongside the subjective components of the human experience of that soundscape. The contributions to the field will now be summarised, after which the hypothesis stated in Chapter 1 will be restated and evaluated.

8.2 Contributions to the Field

The contributions to the field of soundscape research made by this thesis are highlighted below:

- This thesis shows the success of applying soundscape recording (in Ambisonic B-format), visual capture, and IR measurement techniques to a variety of outdoor environments. These techniques have previously been used primarily in indoor spaces (particularly in the case of IR measurement), and little research has been conducted making concurrent audio and video recording in surround sound with spherical video. This thesis has shown these techniques to work very well in outdoor spaces, where Chapter 5 evidences the suitability of using a starter pistol for IR measurement (in certain scenarios. The success of the soundscape and visual capture approach is demonstrated by the listening tests in Chapter 6.
- The Self-Assessment Manikin (SAM), which is a pictorial evaluation tool for measuring emotional state, is validated for use in soundscape evaluation. This is shown in the first listening test in Chapter 6, a test based on Ambisonic playback of Ambisonic soundscape recordings over a 16-speaker listening rig. This test makes use of the SAM alongside a set of semantic differential (SD) pairs. Use of SD pairs, and this playback format, is well established in soundscape research, and therefore provides a suitable framework within which the validity of the SAM can be determined. Whilst the SAM has been used a small number of times in previous soundscape research, there is no extant research that has made an attempt to validate by comparison with established subjective evaluation tools. The results of this listening test have previously been presented in [189].

- The ecological validity of stereo UHJ renderings of B-format Ambisonic soundscape recordings is established in the results of the second listening test presented in Chapter 6. This is demonstrated where these renderings are shown to evoke statistically similar emotional states to the surround-sound renderings. The novelty of this work makes a significant contribution to the field as it affords the possibility of conducting online listening tests making use of spatial acoustic data (albeit limited to the horizontal plane) beyond standard stereo recordings, or binaural renderings of Ambisonic audio. This contribution was previously presented in [310].

- The remaining listening tests in Chapter 6 show how soundscape category ratings and aural/visual feature identification can be utilised as evaluation tools alongside the SAM. These use of these two approaches to subjective evaluation together is novel, and shows that they are complementary to one another and allow for the intuitive subjective evaluation of both the contents of a soundscape and the emotional state it evokes. The category rating question used in these tests was developed according to the soundscape categories identified in previous soundscape research, and the use of the categorisation question as an evaluation tool is a novel contribution to the field (as presented in [149]).

- The third listening test in Chapter 6 also introduces the Perceptual Noise Impact Rating (PNIR) to this thesis. The PNIR is a rating scale combining the valence and arousal dimensions of the SAM into a single metric. The PNIR was first introduced in [250] as part of this research, and this thesis includes the first application of the PNIR to listening test data. The PNIR is used in the fourth listening test where it is shown to allow for the identification of distinct emotional states not apparent in the valence or arousal results alone.

- The third and fourth listening tests presented in Chapter 6 also contain novel research into the phenomenon of cross-modal perception, which in the context of this thesis is the consideration of the presence of certain visual and aural features that affect the subjective evaluation of a soundscape scene. Examples of this are identified through the use of still panoramic images and full-motion spherical video, including the potential positive impact of the presence of green infrastructure in a highly developed environment. Conversely the potential negative impact of the presence of traffic in otherwise ‘natural’ environments is also demonstrated. Whilst these results may not be entirely unprecedented, or indeed, surprising, the use of listening tests to determine the existence of this effect through the use of VR technologies is a novel contribution to the wider field. Results from these two listening tests have been published in [311].

- This thesis presents a novel formalised workflow for the collection, handling, and presentation of panoramic, Ambisonic, audiovisual soundscape stimuli. This workflow brings together existing recording and data handling steps into a novel VR content creation scheme. This workflow has been presented in [311][1].
- Chapter 7 presents the WGW, a method for the modelling of sparsely reflecting acoustic scenes. This design draws on elements from the extant Scattering Delay Network and Treeverb modelling methods, and develops novel inter-node connections and filtering structure methodologies. This work has been published as a journal paper [287].

8.2.1 Restatement of Hypothesis

Chapter 1 summarised the aims of this thesis through the introduction of this hypothesis:

> *The measurement, modelling, and evaluation of environmental sound are effective soundscape methodologies for the identification and assessment of positive sonic environments.*

The program of research presented in this thesis ultimately supports and confirms this hypothesis, as it has shown definitively how soundscape methodologies, including soundscape recording, auralisation, and evaluation, as well as acoustic measurement and modelling, can be used to develop an understanding of multiple aspects environmental sound. These aspects include the identification of sound sources comprising a soundscape, and (more significantly) the emotional state evoked, as well as the behaviour of sound in sparsely reflecting acoustic scenes. This thesis has shown support of this hypothesis in the following main ways:

[1]Note that, as mentioned in Chapter 6, this workflow is missing SN3D normalisation of the 4-channel audio.

- State-of-the-art Ambisonic audio and spherical video recording techniques have been used to gather VR soundscape recordings. The success of these recordings has been demonstrated by their use in the conducted listening tests.

- The sound evaluation work has shown the validity of the SAM, and the use of soundscape categorisation and sound source identification as complementary evaluation tools.

- The results of these listening tests have allowed for the successfully identification of positive soundscapes and sound sources, and for the distinction between emotional sates evoked by different soundscape stimuli.

- The IR measurement and acoustic modelling work has allowed for the exploration and understanding of the characteristics of sparsely reflecting acoustic scenes, including the limestone gorge at Creswell Crags, forest environments, and an urban courtyard. The novel acoustic modelling method shown here, the Waveguide Web, makes novel contributions to the wider community in terms of how such acoustic scenes can be simulated.

These points, alongside the contributions to the field as outlined in the previous section, demonstrate the value of this program of research and clearly confirm this hypothesis. Possibilities for future work to develop these findings will now be considered.

8.3 Future Work

There are many areas for future research building upon the findings in this thesis, in particular in the areas of soundscape evaluation and environmental sound modelling.

The listening tests presented in Chapter 6 represent a base on which to build further experimentation regarding the effect of visual features on soundscape evaluation. The tests presented in this thesis have made use of still images and full-motion spherical video footage of real, unmodified environments, presented with the 'correct' soundscape recording (i.e. the visual and aural stimuli were captured at the same location). The successful test methodologies presented in Chapter 6 could therefore be applied to several possible future tests:

- The soundscape data could be presented with multiple combinations of visual and aural stimuli, for example showing the video recorded at Location 1 with the audio recorded at Location 3. A rigorous approach to such a 'cross-presentation' of aural and visual stimuli would provide interesting further data regarding the cross-modal impact resulting from the presence of certain visual features, although the resultant asynchrony between aural and visual elements might interfere with the results.

- Augmented reality and audio editing/synthesis techniques could be used to modify the stimuli. For example the video recorded at Location 1 could be presented with its original audio, and then with modified version of the original recording with other sound sources added (for example a single car driving by, multiple cars driving by, conversation etc.). Alternatively, the original audio could be presented with changing visual stimuli, for example adding concrete paths, reducing the amount of green infrastructure etc. This approach would allow for the impact of particular sound sources and visual features to be identified, although care would need to be taken to ensure the added/modified stimuli 'fits' the environment.

- This thesis has made use of the PNIR which is a combination of the valence and arousal scales of the SAM. Initial results indicate this to be a useful metric, although there is further research to be conducted comparing PNIR results generated from valence and arousal responses with a direct measure of the perceptual noise impact of soudnscape stimuli.

- There is room for further testing regarding the ecological validity of VR soundscape presentation, whether through the use of YouTube as in this thesis or another similar VR technique. Given the rapidity of recent hardware and software development in this area, a meaningful comparison methodology will become very important in order to compare the competing approaches.

In Chapter 7, the three case studies evaluating the WGW showed the algorithm's design to be 'successful' in terms of the objective metrics applied. The obvious limitation of this approach is that it gives no real indication of the aural qualities of the resultant IRs (that is, what the results 'actually sound like'). As such future work would make use of perceptual listening tests, in order to make a subjective evaluation of the similarity between the WGW results and other recorded/simulated IRs. However, as shown in Chapter 7, in the case of the forest environment very little prior work has been conducted and there is a distinct lack of suitable IRs recorded in real forest environments available for testing. Future work could therefore make use of the IR measurement techniques used at Creswell Crags (presented in Chapter 7) to capture a set of IRs for some real-life forests. Providing care was taken to record the position and diameter of a suitable number of trees near the source and receiver, those parameters could be used to develop a WGW simulation of each forest environment.

8.4 Final Comments

This thesis has presented a portfolio of research that has deliberately covered a wide range of environmental sound research areas. This approach has reflected the position of soundscape research, representing as it does a confluence of acoustics, psychology, acoustic ecology, sociology, musicology, architecture, and environmental research. There is lack of a unified approach to many issues, and therefore a need for these disciplines to come together in order to identify, design, and improve positive sonic environments.

The fact that soundscape research draws on elements from such disparate areas shows the extreme degree to which our acoustic surroundings affect our lives, and in particular the pervasive nature of noise pollution and the multi-faceted approach that is required in order to effect change. In a fairly modest way this thesis represents an example of such an approach, showing as it does how a variety of techniques can be used individually and together in order to develop an understanding of environmental sound.

Certain aspects of this research, were they to be made use of in a real-world context, could present the possibility of economic, societal, and cultural benefits beyond the immediate research community. This could include:

- The application of the IR recording techniques used at Creswell to other real-world environments. In particular this could include the recording of IRs at places of historical, cultural, and archaeological interest. This would allow for the acoustic signatures of these places to be recorded for posterity (an example of 'digital heritage' [312]), and for creative/musical use.

- The VR workflow presented in Chapter 6 could also be used in those same environments as another form of digital heritage. This would allow people to experience those environments through virtual reality, including individuals that might otherwise be unable to experience them (due to disability, for example).

- The methods for the subjective evaluation of soundscapes developed in Chapter 6 could be integrated into existing noise monitoring/measurement schemes, both as part of the auralisation process sometimes applied to planned building project, and as a monitoring method for existing noisy locations. An approach making use of some form of subjective evaluation alongside objective metrics would ultimately result in the design of more positive acoustic environments, and the improvement of extant ones.

- The WGW acoustic modelling method could be developed and used to simulate IRs of sparsely reflecting locations for creative use, and in the auralisation of planned structures likely to have a sparsely reflective acoustic profile.

It is the personal view of the author that an approach making used of multiple techniques is imperative in order to develop a comprehensive understanding of sound, particularly the use of both objective and subjective evaluation methods. Similarly in order for the negative impact of noise pollution to be addressed effectively, a collaboration between experts in many of these fields is required. Unless this collaboration is established from the first instance (e.g. in the planning process of a new building) then soundscape research will forever be relegated to damage limitation, rather than the development of positive, restorative, sonic environments.

Appendix A

North York Moors and Leeds Recording Details

A.1 Locations

FIGURE A.1: Map of the North York Moors National Park with locations 1-5 indicated. Map from [213].

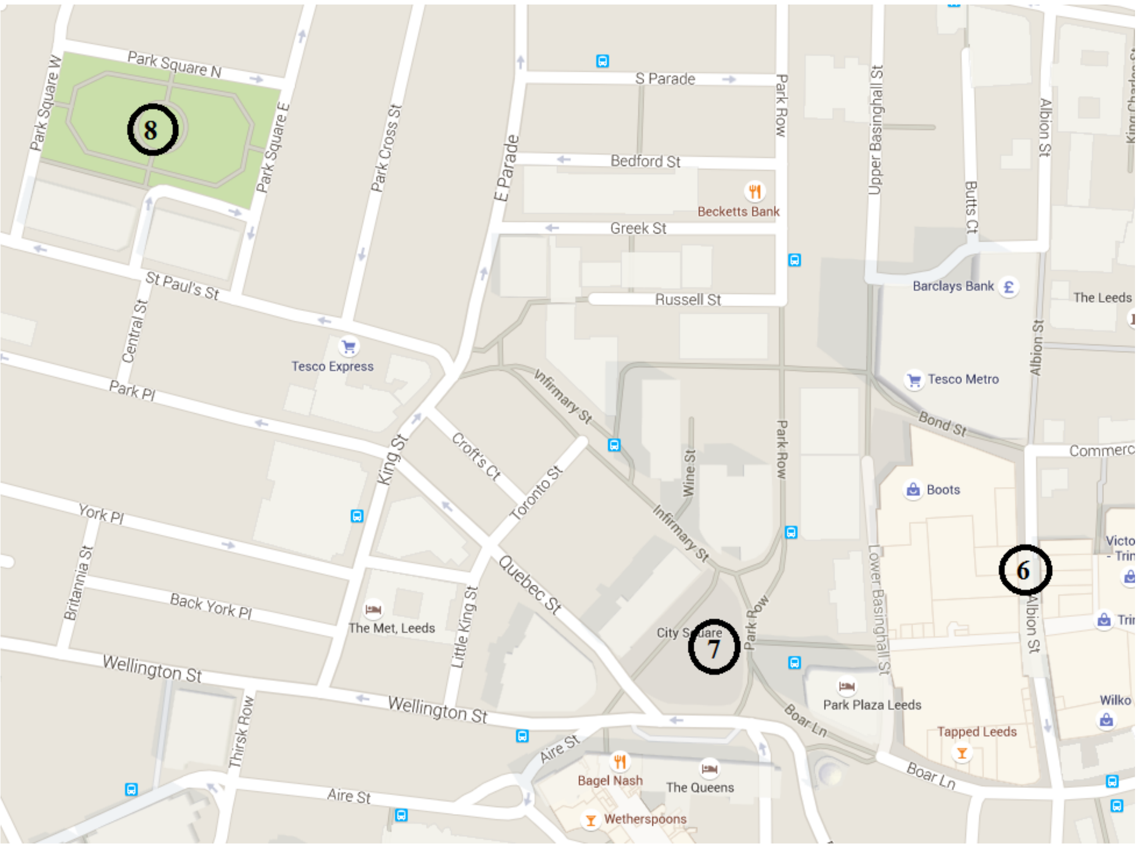


Figure A.2: Map of Leeds with locations 6-8 indicated. Map from [213].

(a)

(b)

FIGURE A.3: Two views of location 1, Dalby forest.

(a)

(b)

Figure A.4: Two different views at location 2. From [213].

(a)

(b)

Figure A.5: Two different views at location 3, the A169 at the Hole of Horcum, facing (a) west and (b) east. From [213].

(a)

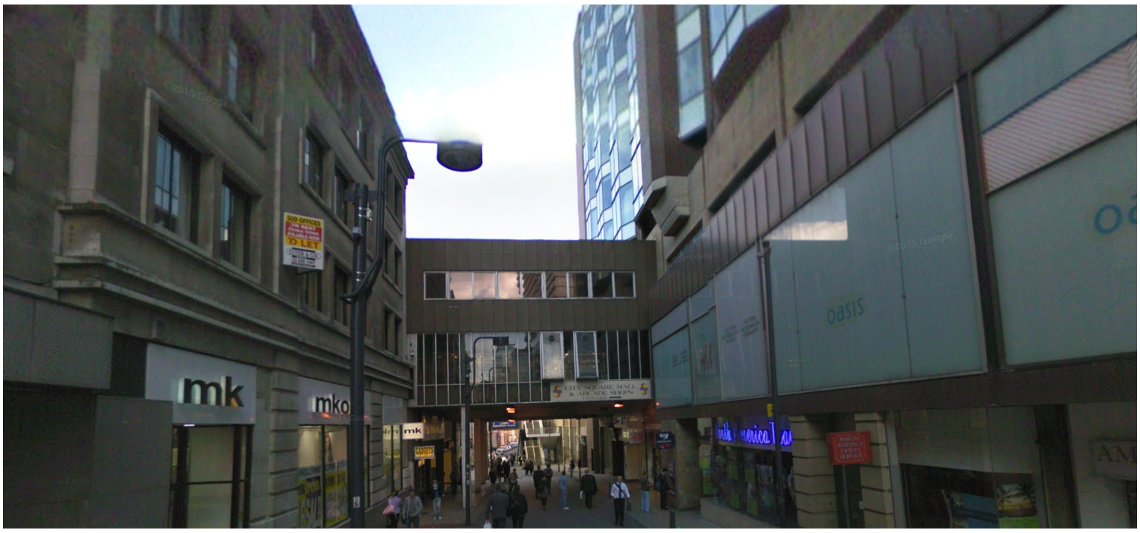
(b)

FIGURE A.6: Two different views at location 4, the A169 at the Fox and Rabbit Inn showing (a) the Fox and Rabbit Inn and car park, and (b) western view from the same position. From [213].

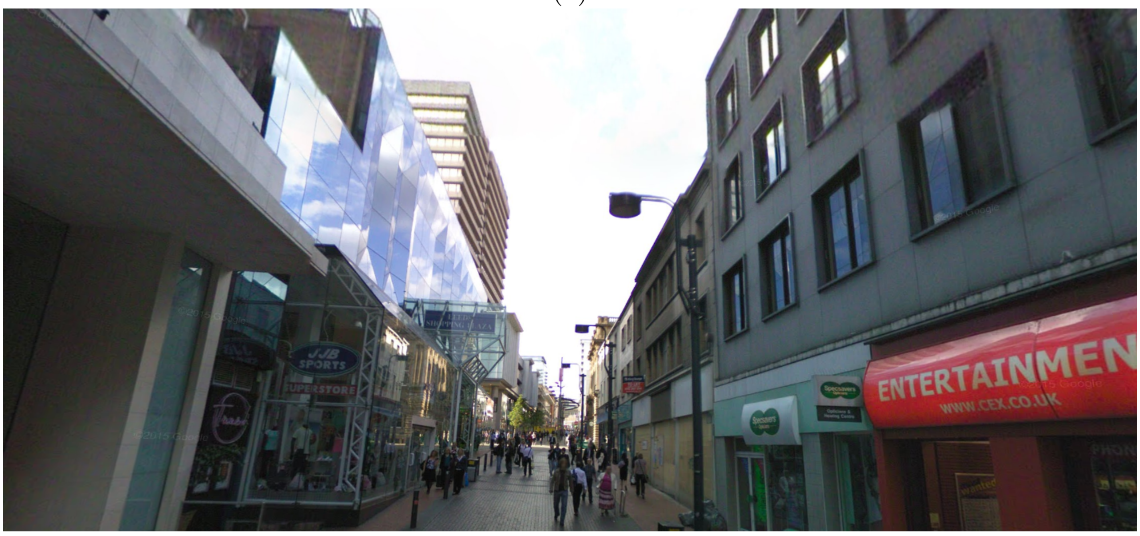
(a)

(b)

FIGURE A.7: Two different views at location 5, Smiddy Hill in Pickering, looking (a) to the east and (b) to the west. From [213].



(a)



(b)

FIGURE A.8: Two different views at location 6, Albion Street in Leeds, (a) to the south and (b) to the north. From [213].

(a)

(b)

FIGURE A.9: Two different views at location 7, Park Row in Leeds(a) to the west (showing the Black Prince statue) and (b) to the east. From [213].

(a)

(b)

FIGURE A.10: Two different views at location 8, Park Square in Leeds showing (a) from the south (From [313]), and (b) from the centre of the square looking north (From [213]).

A.2 Recordings

North York Moors Locations	
Site 1: Dalby Forest	
dB(LA_{eq}):	44
Soundscape:	A variety of birds (including some owls) singing and tweeting, some insects buzzing around the microphone. There is background wind noise and occasional aeroplane noise. There are some footsteps and quiet conversation also present.
Visual features:	Dense forest including a variety of trees and foliage. Various gravel footpaths.
Site 2: Dalby Forest Lake	
dB(LA_{eq}):	47
Soundscape:	Similar soundscape to site 1, replace aeroplane noise and footsteps/conversation with some passing traffic. Fewer, more distinct bird calls and some water noise.
Visual features:	Fairly large lake surrounded by trees and bushes. Footpaths and benches around the edge of the lake. Nearby road with occasional traffic. Hills surrounding on all sides.
Site 3: Hole of Horcum	
dB(LA_{eq}):	50
Soundscape:	Soundscape dominated by constant traffic noise. Some distant birdsong and the occasional buzzing insect. Some footsteps and quiet conversation as part of the later recording, plus some bleats from nearby sheep.
Visual features:	A large section of the valley of Levisham Beack, a hollow c. 400 ft deep and 3/4 of a mile across, with a footpath running around the top. Behind is the A169 and a car park.

Site 4: Fox & Rabbit Inn	
dB(LA_{eq}):	50
Soundscape:	Constant traffic noise from a variety of vehicles, including farm traffic, large trucks and cars. Some low level, distant birdsong and occasional opening/closing of car doors.
Visual features:	Public house on the A169 with its own car park. Surrounded by countryside/farmland.
Site 5: Smiddy Hill, Pickering	
dB(LA_{eq}):	55
Soundscape:	Constant traffic in the distance, some closer traffic noise an other car noises - opening/closing of doors, turning on engine. Some birdsong and distant conversation and footsteps (as well as the occasional sneeze).
Visual features:	Small open green area near Pickering high street. Grass and benches surrounding a war memorial, with some parking spaces on the road opposite and some small shops.
Leeds Locations	
Site 6: Albion St., Leeds	
dB(LA_{eq}):	66
Soundscape:	Some constant distant traffic noise. Soundscape dominated by human activity, primarily footsteps and conversation. Occasional noise from nearby workmen. Busker present offering scintillating renditions of *My Way* and *Ghost Riders In The Sky*.
Visual features:	Highly developed, pedestrianised, shopping area in Leeds. Surrounded on all sides by various shops and shoppers. Nearby busker. Some traffic in the distance.

Site 7: Park Row, Leeds	
dB(LA_{eq}):	70
Soundscape:	Constant high level traffic noise, including lots of buses. Some pedestrians and luggage bag wheel noises. Distant busking flute player throughout.
Visual features:	An open square near Leeds train station, next to a statue of the Black Prince. Some nearby traffic, included several buses. Some trees and bushes present next to some benches, opposite a restaurant with outside seating.
Site 8: Park Square, Leeds	
dB(LA_{eq}):	50
Soundscape:	More constant traffic noise, accompanied by occasional bird-song, noise from workmen, and conversation and footsteps from people using the park during their lunch break.
Visual features:	A small urban park with several different types of flower present, some trees, well kept lawns, and bushes. Coffee vendor nearby with several people present eating their lunch.

Table A.1: Tabulated results from the recording work - for each location the max A-weighted sound level is included, as well as a description of the soundscape and visual features in each case.

A.3 Equipment

FIGURE A.11: The recording equipment in situ, including the Freedom360 GoPro rig and STM450 Soundfield microphone.

Figure A.12: Another view of the recording equipment, here indicating the portable recorder and Soundfield microphone pre-amp as well as the GoPro rig and the Soundfield microphone itself.

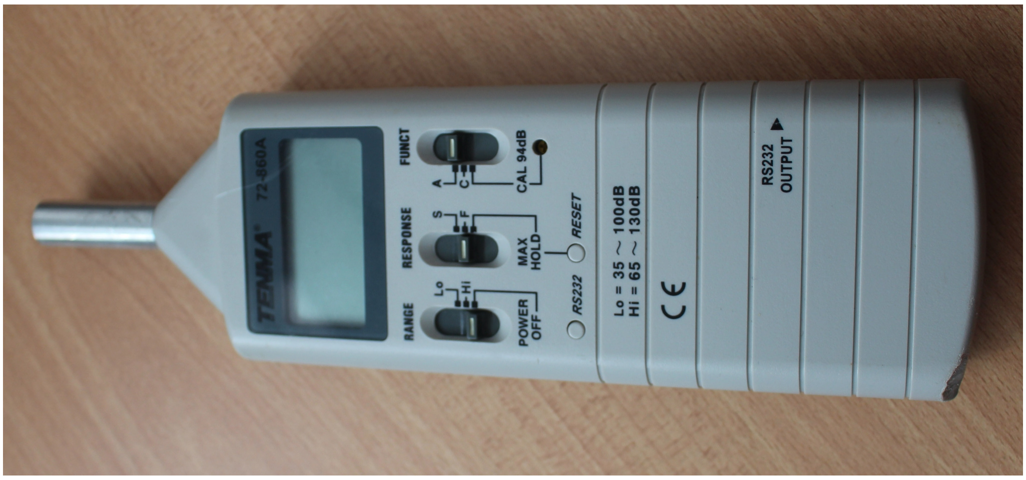


Figure A.13: The TENMA 72-860A sound level meter used in the recording work. The unit has a frequency range 31.5 Hz - 8 kHz [314].

Appendix B

Digital Assets

the following items are included on the data disc accompanying this thesis. The sections headings here match those of the file folders on the disc. Each of these folders also contains a `readme.txt` file to give further information on its contents.

B.1 Environmental Sound Measurement

This folder contains the supporting material relevant to Chapter 5.

B.1.1 Soundscape Recording Data

This folder contains a link to an online repository containing the recorded visual data and the raw B-format soundscape recordings (due to size limitations). The clips edited from these recordings for use in the conducted listening tests are included in the relevant folders on the disc.

B.1.2 Creswell Crags

The folder contains the IRs recorded at Creswell Crags, the MATLAB analysis script used to evaluate them, and a short report written to summarise the recording work.

B.2 Environmental Sound Evaluation

This folder contains the supporting material relevant to Chapter 6.

B.2.1 Ethical Approval Forms

This folder include the two ethical approval forms which were written in order to allow the listening tests to take place.

B.2.2 Listening Test 1 Material

This folder contains the introductory statement and questionnaire supplied to the participants of the first listening test. It also include the B-format audio clips used as stimuli in this experiment and the MaxMSP patch used to present the stimuli.

B.2.3 Listening Test 2 Material

This folder contains the questionnaire supplied to the participants of the second listening test. It also includes the stereo UHJ audio clips used as stimuli in this experiment, as well as the MATLAB script used to convert the B-format clips to this format.

B.2.4 Listening Test 3 Material

This folder contains the questionnaire supplied to the participants of the third listening test. It also includes the still panoramic images used as stimuli in this experiment (the stereo UHJ audio clips used in this test are the same as those used in the second listening test).

B.2.5 Listening Test 4 Material

This folder contains the questionnaire supplied to the participants of the fourth listening test. It also includes links to the YouTube based audiovisual content used as stimuli in this experiment.

B.2.6 Listening Test Results and Analysis

This folder contains the results of the four listening tests, and the analysis scripts used to evaluate the results and generate the plots used in this study.

B.3 Environmental Sound Modelling

This folder contains the supporting material relevant to Chapter 7.

This folder includes the MATLAB code written to implement the Waveguide Web, as well as audio files of example IRs generated using it.

Bibliography

[1] T. Cox, *Sonic wonderland: A scientific odyssey of sound*. Bodley Head, 2014, ISBN: 9781847922106.

[2] K. McElhearn, *John Cage and the anechoic chamber*, 2016. [Online]. Available: https://www.kirkville.com/john-cage-and-the-anechoic-chamber/.

[3] D. Revill, *The roaring silence: John cage: A life*. Skyhorse Publishing, Inc., 2014.

[4] G. Berkeley, *A treatise concerning the principles of human knowledge*. 1734.

[5] B. Berglund, T. Lindvall, and D. Schwela, "Guidelines for community noise," *World Health Organisation, Geneva, Switzerland*, 1999. [Online]. Available: http://apps.who.int/iris/handle/10665/66217.

[6] "Specialist services: Health Technical Memorandum 08-01: Acoustics," *Department of Health*, [Online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/144248/HTM_08-01.pdf.

[7] M. Goldsmith, *Discord: The story of noise*. OUP Oxford, 2012, ISBN: 9780199600687.

[8] D. Hendy, *Noise: A human history of sound and listening*. Profile, 2013, ISBN: 9781781250891.

[9] European Commission, "Directive 2002/49/EC of the European Parliament and the Council of 25 June 2002 relating to the assessment and management of environmental noise," *Official Journal of the European Communities, L*, vol. 189, no. 18.07, pp. 12–25, Jul. 2002.

[10] L. Goines, L. Hagler, *et al.*, "Noise pollution: A modern plague," *Southern Medical Journal - Birmingham Alabama*, vol. 100, no. 3, p. 287, 2007.

[11] J. Treasure, "Building in sound," *Biamp Systems*, 2012. [Online]. Available: http://67aa6fee3b112cf7b085-a4daa72d047cd5cf1107a27466ad39b3.r75.cf1.rackcdn.com/Biamp_Whitepaper_Building_in_Sound.pdf.

[12] T. Ten Wolde, "European commission's future noise policy," in *INTERNOISE*, NOISE CONTROL FOUNDATION, vol. 3, 2000, pp. 1719–1722.

[13] R. Schafer, *The soundscape: Our sonic environment and the tuning of the world*. Inner Traditions/Bear, 1993, ISBN: 9781594776687. [Online]. Available: http://books.google.co.uk/books?id=ltBrAwAAQBAJ.

[14] A. Arteaga, *Klangumwelt ernst-reuter-platz: A project of the auditory architecture research unit*. Errant Bodies Press, 2017, ISBN: 9780988937574. [Online]. Available: https://books.google.co.uk/books?id=RCNWvgAACAAJ.

[15] B. Truax, *Acoustic communication*. Greenwood Publishing Group, 1984.

[16] S. Payne, W. Davies, and M. Adams, *Research into the practical policy applications of soundscapes concepts and techniques in urban areas. defra report nanr200, june 2009*, 2009.

[17] D. Blackstock, *Fundamentals of physical acoustics*. John Wiley & Sons, 2000.

[18] D. Howard and J. Angus, *Acoustics and psychoacoustics*. Focal Press, 2009.

[19] T. Rossing, *The science of sound*. Addison Wesley, 2009.

[20] L. Kinsler, A. Frey, A. Coppens, and J. Sanders, "Fundamentals of acoustics," *Fundamentals of Acoustics, 4th Edition, by Lawrence E. Kinsler, Austin R. Frey, Alan B. Coppens, James V. Sanders, pp. 560. ISBN 0-471-84789-5. Wiley-VCH, December 1999.*, vol. 1, 1999.

[21] C. E. Speaks, *Introduction to sound: Acoustics for the hearing and speech sciences*. Plural Publishing, 2017.

[22] E. Pedersen and P. Larsman, "The impact of visual factors on noise annoyance among people living in the vicinity of wind turbines," *Journal of Environmental Psychology*, vol. 28, no. 4, pp. 379 –389, 2008, ISSN: 0272-4944. DOI: http://dx.doi.org/10.1016/j.jenvp.2008.02.009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0272494408000224.

[23] *Table of typical SPL levels*, 2015. [Online]. Available: http://www.gcaudio.com/resources/howtos/loudness.html.

[24] S. Harriet, "Application of auralisation and soundscape methodologies to environmental noise," PhD thesis, University of York, 2013.

[25] A. Jahn and J. Santos-Sacchi, *Physiology of the ear*. Raven Press, 1988.

[26] G Von Békésy, *Experiments in hearing*. McGraw-Hill, 1960.

[27] D. Levitin, *This is your brain on music*. Dutton, 2006.

[28] H. Fletcher and W. Munson, "Loudness, its definition, measurement and calculation," *Bell System Technical Journal*, vol. 12, no. 4, pp. 377–430, 1933.

[29] M. Goldsmith, *Sound: A very short introduction*, ser. Very short introductions. Oxford University Press, 2015, ISBN: 9780198708445. [Online]. Available: https://books.google.co.uk/books?id=ZHLNCgAAQBAJ.

[30] E. L. Teng, "Dichotic ear difference is a poor index for the functional asymmetry between the cerebral hemispheres," *Neuropsychologia*, vol. 19, no. 2, pp. 235–240, 1981.

[31] R. S. Woodworth and H. Schlosberg, *Experimental psychology*. Oxford and IBH Publishing, 1954.

[32] E. R. Hafter, "Spatial hearing and the duplex theory: How viable is the model," *Dynamic aspects of neocortical function. Wiley, New York*, pp. 425–448, 1984.

[33] M. Carson, "Surround sound impulse response measurement with the exponential sine sweep: Application in convolution reverb," Master's thesis, University Of Victoria: Department of Electrical and Computer Engineering, 2009.

[34] A. Southern, "The synthesis and auralisation of physically modelled soundfields," PhD thesis, University of York, 2010.

[35] *About - the open access impulse response library*. [Online]. Available: http://www.openairlib.net/about.

[36] *Music by Skeleton worm*. [Online]. Available: https://skeletonworm.bandcamp.com/.

[37] J. Abel, N. Bryan, P. Huang, M. Kolar, and B. Pentcheva, "Estimating room impulse responses from recorded balloon pops," in *Audio Engineering Society Convention 129*, Nov. 2010.

[38] *ISO 3382-1: Measurement of room acoustic parameters in performance spaces*, ISO, Geneva, Switzerland, 2009.

[39] A. Farina, "Advancements in impulse response measurements by sine sweeps," in *122nd AES Convention*, May 2007.

[40] G. Stan, J. Embrechts, and D. Archambeau, "Comparison of different impulse response measurement techniques," *Journal of the Audio Engineering Society*, vol. 50, no. 4, pp. 249–262, 2002.

[41] S. Shelley, D. Murphy, and A. Chadwick, "B-format acoustic impulse response measurement and analysis in the forest at koli national park, finland," in *Proceedings of the 16th Int. Conference on Digital Audio Effects (DAFx-13)*, Sep. 2013.

[42] A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Audio Engineering Society Convention 108*, Feb. 2000.

[43] C. Goussios, G. Kalliris, C. Dimoulas, G. Papanikolaou, and S. Charalampidis, "Improvements of a horn-loaded omni-directional sound source," in *Audio Engineering Society Convention 114*, Mar. 2003.

[44] M. Jambrošić K. Horvat and M. Bogut, "Comparison of impulse sources used in reverberation time measurements," in *Proceedings of the 51st international symposium ELMAR*, 2009.

[45] V. Välimäki, J. D. Parker, L. Savioja, J. O. Smith, and J. S. Abel, "Fifty years of artificial reverberation," *IEEE Trans Audio, Speech, and Lang. Process.*, vol. 20, no. 5, pp. 1421–1448, Jul. 2012.

[46] M. Lopez, S. Pauletto, and G. Kearney, "The application of impulse response measurements techniques to the study of the acoustics of stonegate, a performance space used in medieval english drama," *Acta acustica united with acustica*, vol. 99, no. 1, pp. 98–109, 2013.

[47] F. Stevens and D. T. Murphy, "Spatial impulse response measurement in an urban environment," in *Audio Engineering Society Conference: 55th International Conference: Spatial Audio*, Helsinki, Finland, Aug. 2014. [Online]. Available: `http://www.aes.org/e-lib/browse.cfm?elib=17355`.

[48] T. Embleton, "Sound propagation in homogenous deciduous and evergreen woods," *Journal of the Acoustical Society of America*, vol. 35, no. 8, 1963.

[49] *Acoustic absorption coefficients*, 2015. [Online]. Available: http://www.acousticalsurfaces.com/acoustic_IOI/101_13.htm.

[50] H. Kuttruff, *Room acoustics*. CRC Press, 2009.

[51] M. Schroeder, "A new method of measuring reverberation time," *The Journal of the Acoustical Society of America*, 1964.

[52] W. C. Sabine and M. D. Egan, *Collected papers on acoustics*, 1994.

[53] C. Eyring, "Reverberation time in "dead" rooms," *The Journal of the Acoustical Society of America*, vol. 1, no. 2A, pp. 168–168, 1930.

[54] S. Shelley, A. Foteinou, and D. Murphy, "Openair: An online auralization resource with applications for game audio development," in *Audio Engineering Society Conference: 41st International Conference: Audio for Games*, Feb. 2011.

[55] *Openair - the open acoustic impulse response library*, 2015. [Online]. Available: http://www.openairlib.net/.

[56] A. Farina, "Impulse response measurements," in *23rd Nordic Sound Symposium, Bolkesjø (Norway)*, 2007, pp. 27–30.

[57] M. Barron, "Interpretation of early decay times in concert auditoria," *Acta Acustica united with Acustica*, vol. 81, no. 4, pp. 320–331, 1995.

[58] D. Halliday, *Fourier transforms: Convolution and correlation. university of york*, 2013.

[59] K. Bijsterveld, *Mechanical sound: Technology, culture, and public problems of noise in the twentieth century*. MIT Press, 2008, ISBN: 9780262026390.

[60] T. Cox, "Bad vibes: An investigation into the worst sounds in the world," *19th ICA Madrid, PPA-09*, vol. 3, 2007.

[61] G. Keizer, *The unwanted sound of everything we want: A book about noise*. PublicAffairs, 2010, ISBN: 9781586488628. [Online]. Available: https://books.google.co.uk/books?id=yZ44DgAAQBAJ.

[62] L. Steg, A. van den Berg, and J. de Groot, *Environmental psychology: An introduction*, ser. BPS textbooks in psychology. Wiley, 2012, ISBN: 9780470976388. [Online]. Available: http://books.google.co.uk/books?id=RFHmw57kiNwC.

[63] D. Juvenal and R. Humphries, *The satires of juvenal*, ser. A Midland book, MB20. Indiana University Press, 1958, ISBN: 9780253200204. [Online]. Available: `https://books.google.co.uk/books?id=HMd1OV7gK4oC`.

[64] *New york city noise survey*, 1930. [Online]. Available: `http://cityroom.blogs.nytimes.com/2013/10/25/listening-to-the-roar-of-1920s-new-york/`.

[65] E. U., "Directive 2002/49/ec of the european parliament and the council of 25 june 2002 relating to the assessment and management of environmental noise," *Official Journal of the European Communities*, vol. 189, no. 12, 2002.

[66] P. A. Franken and G. Jones, "On response to community noise," *Applied Acoustics*, vol. 2, no. 4, pp. 241–246, 1969.

[67] H Godfrey, "The city's noise," *Atlantic Monthly*, pp. 609–610, Sep. 1909.

[68] M. S. Hammer, T. K. Swinburn, and R. L. Neitzel, "Environmental noise pollution in the united states: Developing an effective public health response," *Environmental health perspectives*, vol. 122, no. 2, p. 115, 2014.

[69] B. Berglund, T. Lindvall, W. H. Organization, *et al.*, *Community noise*. Center for Sensory Research, Stockholm University and Karolinska Institute, 1995.

[70] W. Ward, A. Glorig, and D. Sklar, "Temporary threshold shift from octave-band noise: Applications to damage-risk criteria," *The Journal of the Acoustical Society of America*, vol. 31, no. 4, pp. 522–528, 1959.

[71] *Tinnitus*, 2015. [Online]. Available: `http://www.nhs.uk/conditions/tinnitus/Pages/Introduction.aspx`.

[72] S. Stansfeld and M. Matheson, "Noise pollution: Non-auditory effects on health," *British Medical Bulletin*, vol. 68, no. 1, pp. 243–257, 2003.

[73] R. Cantrell, "Pathophysiological effects of noise," in *Audio Engineering Society Convention 51*, May 1975. [Online]. Available: `http://www.aes.org/e-lib/browse.cfm?elib=2416`.

[74] *Cerebrum anatomy*, 2015. [Online]. Available: `http://www.edoctoronline.com/medical-atlas.asp?c=4&id=21706`.

[75] N. Weinstein, "Individual differences in reactions to noise: A longitudinal study in a college dormitory.," *Journal of Applied Psychology*, vol. 63, no. 4, p. 458, 1978.

[76] K. Hume, H. Barrett, A. Ip, T. McDonagh, W. Davies, M. Adams, N. Bruce, R. Cain, P. Jennings, G. Czanner, *et al.*, "Physiological responses and subjective estimates of sounds: Initial results of pilot study," *Proc. I. o. A.*, vol. 30, no. 2, 2008.

[77] J. Alvarsson, S. Wiens, and M. Nilsson, "Stress recovery during exposure to nature sound and environmental noise," *International Journal of Environmental Research and Public Health*, vol. 7, no. 3, pp. 1036–1046, 2010, ISSN: 1660-4601. DOI: `10.3390/ijerph7031036`. [Online]. Available: `http://www.mdpi.com/1660-4601/7/3/1036`.

[78] M. Annerstedt, P. Jönsson, M. Wallergård, G. Johansson, B. Karlson, P. Grahn, Å. Hansen, and P. Währborg, "Inducing physiological stress recovery with sounds of nature in a virtual reality forest—results from a pilot study," *Physiology & behavior*, vol. 118, pp. 240–250, 2013.

[79] M. Bradley and P. Lang, "Affective reactions to acoustic stimuli," *Psychophysiology*, vol. 37, no. 02, pp. 204–215, 2000.

[80] P. Gomez and B. Danuser, "Affective and physiological responses to environmental noises and music," *International Journal of Psychophysiology*, vol. 53, no. 2, pp. 91–103, 2004.

[81] W. Davies, M. Adams, N. Bruce, M. Marselle, R. Cain, P. Jennings, J. Poxon, A. Carlyle, P. Cusack, D. Hall, *et al.*, "The positive soundscape project: A synthesis of results from many disciplines," in *InterNoise*, 2009.

[82] A. B. Knol, A. C. Petersen, J. P. Van der Sluijs, and E. Lebret, "Dealing with uncertainties in environmental burden of disease assessment," *Environmental Health*, vol. 8, no. 1, p. 21, 2009.

[83] "ANSI S1.42-2001 design response of weighting networks for acoustical measurements," *Acoustical Society of America*, 2001.

[84] R. M. Aarts, "A comparison of some loudness measures for loudspeaker listening tests," *J. Audio Eng. Soc*, vol. 40, no. 3, pp. 142–146, 1992.

[85] "ISO 1996-1:2016: Description, measurement and assessment of environmental noise," 2016. [Online]. Available: `https://www.iso.org/standard/59765.html`.

[86] M. Raimbault, C. Lavandier, and M. Bérengier, "Ambient sound assessment of urban environments: Field studies in two french cities," *Applied Acoustics*, vol. 64, no. 12, pp. 1241–1256, 2003.

[87] G. Pennington, N. Topham, and R. Ward, "Aircraft noise and residential property values adjacent to manchester international airport," *Journal of Transport Economics and Policy*, pp. 49–59, 1990.

[88] *Noise and Number Index*, 2015. [Online]. Available: `http://www.sfu.ca/sonic-studio/handbook/Noise_and_Number_Index.html`.

[89] *The noise and number index*, 1981.

[90] H. Eldien, "Noise mapping in urban environments: Application at suez city center," in *Computers & Industrial Engineering, 2009. CIE 2009. International Conference on*, IEEE, 2009, pp. 1722–1727.

[91] *York noise map*, 2018. [Online]. Available: `http://www.extrium.co.uk/noiseviewer.html`.

[92] C. Rosas and S. Luna-Ramirez, "Binaural recording system and sound map of malaga," in *Audio Engineering Society Convention 142*, Audio Engineering Society, 2017.

[93] *The Widenoise App*, 2015. [Online]. Available: `http://www.widetag.com/widenoise/`.

[94] *The Noise Nuisance App*, 2015. [Online]. Available: `http://noisenuisance.org/the-app/`.

[95] S. Marry and J. Defrance, "Analysis of the perception and representation of sonic public spaces through on site survey, acoustic indicators and in-depth interviews," *Applied Acoustics*, vol. 74, no. 2, pp. 282–292, 2013.

[96] N. Garg, O. Sharma, V. Mohanan, and S. Maji, "Passive noise control measures for traffic noise abatement in delhi, india," *Journal of Scientific and Industrial Research*, vol. 71, 2012.

[97] *Transmission loss in different frequency regions for a rigid barrier*, 2015. [Online]. Available: `http://personal.inet.fi/koti/juhladude/soundproofing.html`.

[98] B. Smith, R. Peters, and S. Owen, *Acoustics and noise control*. Longman Harlow, 1996.

[99] M. Simons and J. Waters, *Sound control in buildings: A guide to part e of the building regulations*. John Wiley & Sons, 2008.

[100] *Noise barrier design handbook*, 2015. [Online]. Available: `http://www.fhwa.dot.gov/environment/noise/noise_barriers/design_construction/design/design03.cfm`.

[101] R. Martinezsala, J. Sancho, J. Sánchez, V. Gómez, J. Llinares, and F. Meseguer, "Sound-attenuation by sculpture," *Nature*, vol. 378, no. 6554, pp. 241–241, 1995.

[102] *Sonic crystal photo, new scientist*, 2015. [Online]. Available: `http://www.newscientist.com/gallery/sonic-doom-noise-in-pictures/5`.

[103] J. Sánchez-Pérez, D. Caballero, R. Martinez-Sala, C. Rubio, J. Sánchez-Dehesa, F. Meseguer, J. Llinares, and F. Gálvez, "Sound attenuation by a two-dimensional array of rigid cylinders," *Physical Review Letters*, vol. 80, no. 24, p. 5325, 1998.

[104] T. Miyashita, "Sonic crystals and sonic wave-guides," *Measurement Science and Technology*, vol. 16, no. 5, R47, 2005.

[105] W. Bragg, "The diffraction of short electromagnetic waves by a crystal," *Proceedings of the Cambridge Philosophical Society*, vol. 17, no. 43, p. 4, 1913.

[106] *Bragg's law diagram*, 2015. [Online]. Available: `http://hyperphysics.phy-astr.gsu.edu/hbase/quantum/bragg.html`.

[107] H. Yang, J. Kang, C. Cheal, T. Van Renterghem, and D. Botteldooren, "Quantifying scattered sound energy from a single tree by means of reverberation time," *The Journal of the Acoustical Society of America*, vol. 134, no. 1, pp. 264–274, 2013.

[108] M. Martens, "Foliage as a low-pass filter: Experiments with model forests in an anechoic chamber.," *The Journal of the Acoustical Society of America*, vol. 67, no. 1, pp. 66–72, 1980.

[109] D. Aylor, "Noise reduction by vegetation and ground," *The Journal of the Acoustical Society of America*, vol. 51, no. 1B, pp. 197–205, 1972.

[110] R. Martínez-Sala, C. Rubio, L. García-Raffi, J Sánchez-Pérez, E. Sánchez-Pérez, and J. Llinares, "Control of noise by trees arranged like sonic crystals," *Journal of sound and vibration*, vol. 291, no. 1, pp. 100–106, 2006.

[111] M. Price, K. Attenborough, and N. Heap, "Sound attenuation through trees: Measurements and models," *The Journal of the Acoustical Society of America*, vol. 84, no. 5, pp. 1836–1844, 1988.

[112] W. Huisman and K. Attenborough, "Reverberation and attenuation in a pine forest," *The Journal of the Acoustical Society of America*, vol. 90, no. 5, pp. 2664–2677, 1991.

[113] E. Thalheimer and C. Shamoon, "New york city's new and improved construction noise regulation," in *Noise-Con 2007*, Reno, USA, Oct. 2007.

[114] N. Pierpont, *Wind turbine syndrome: A report on a natural experiment.* K-Selected Books Santa Fe, NM, 2009.

[115] E. Thompson, *The soundscape of modernity: Architectural acoustics and the culture of listening in america, 1900-1933.* MIT Press, 2004, ISBN: 9780262701068. [Online]. Available: `http://books.google.co.uk/books?id=7jvtvGbatv4C`.

[116] P. Lercher and B. Schulte-Fortkamp, "The relevance of soundscape research to the assessment of noise annoyance at the community level," in *Proceedings of the Eighth International Congress on Noise as a Public Health Problem*, 2003, pp. 225–231.

[117] V. Lamme, "The neurophysiology of figure-ground segregation in primary visual cortex," *The Journal of Neuroscience*, vol. 15, no. 2, pp. 1605–1615, 1995.

[118] E. Wilson, *Biophilia.* Harvard University Press, 1984.

[119] J. Lockard, *The evolution of human social behavior.* New York NY Elsevier 1980., 1980.

[120] R. Thayer, *The biopsychology of mood and arousal.* Oxford University Press, 1989, ISBN: 9780195361759. [Online]. Available: `http://books.google.co.uk/books?id=0RiwiDNqcbEC`.

[121] J. Appleton, "Prospects and refuges re-visited," *Landscape Journal*, vol. 3, no. 2, pp. 91–103, 1984.

[122] R. Kaplan and S. Kaplan, *The experience of nature: A psychological perspective.* Cambridge University Press, 1989, ISBN: 9780521341394. [Online]. Available: `http://books.google.co.uk/books?id=7l80AAAAIAAJ`.

[123] S. Bell, *Landscape: Pattern, perception and process.* Routledge, 2012.

[124] R. Pheasant, M. Fisher, G. Watts, D. Whitaker, and K. Horoshenkov, "The importance of auditory-visual interaction in the construction of 'tranquil space'," *Journal of Environmental Psychology*, vol. 30, no. 4, pp. 501–509, 2010.

[125] A. Léobon, "La qualification des ambiances sonores urbaines," *Natures-Sciences-Sociétés*, vol. 3, no. 1, pp. 26–41, 1995.

[126] S. Viollon, C. Lavandier, and C. Drake, "Influence of visual setting on sound ratings in an urban environment," *Applied Acoustics*, vol. 63, no. 5, pp. 493 –511, 2002, ISSN: 0003-682X. DOI: http://dx.doi.org/10.1016/S0003-682X(01)00053-6. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0003682X01000536.

[127] W. Yang and J. Kang, "Acoustic comfort and psychological adaptation as a guide for soundscape design in urban open public spaces," in *Proceedings of the 17th International Congress on Acoustics (ICA)*, 2001.

[128] L. Anderson, B. Mulligan, L. Goodman, and H. Regen, "Effects of sounds on preferences for outdoor settings," *Environment and Bevior*, vol. 15, no. 5, pp. 539–566, 1983.

[129] G. Watts and R. Pheasant, "Tranquillity in the scottish highlands and dartmoor national park–the importance of soundscapes and emotional factors," *Applied Acoustics*, vol. 89, pp. 297–305, 2015.

[130] B. Krause, *The great animal orchestra: Finding the origins of music in the world's wild places*. Hachette UK, 2012.

[131] G. Milner, *Perfecting sound forever: The story of recorded music*. Granta Books, 2010, ISBN: 9781847081407.

[132] N. Kaplanis, S. Bech, S. Jensen, and T. van Waterschoot, "Perception of reverberation in small rooms: A literature study," in *Audio Engineering Society Conference: 55th International Conference: Spatial Audio*, Aug. 2014. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=17348.

[133] S. Bech and N. Zacharov, *Perceptual audio evaluation - theory, method and application*. Wiley, 2007, ISBN: 9780470869246. [Online]. Available: http://books.google.co.uk/books?id=1WGPJai1gX8C.

[134] F. Figueiredo and F. Iazzetta, "Comparative study of measured acoustic parameters in concert halls in the city of são paulo," in *Proc. Int. Congress and Exposition on Noise Control Engineering, Rio de Janeiro, Brazil*, 2005.

[135] M. Rand and A. Clarke, "The environmental and community impacts of wind energy," *International Journal of Energy, Environment, Economics*, vol. 1, no. 1, 1991.

[136] *Roughness - an introduction to sound quality testing - university of salford*, 2015. [Online]. Available: `http://www.salford.ac.uk/computing-science-engineering/research/acoustics/psychoacoustics/sound-quality-making-products-sound-better/sound-quality-testing/roughness-fluctuation-strength`.

[137] P. Daniel and R. Weber, "Psychoacoustical roughness: Implementation of an optimized model," *Acta Acustica united with Acustica*, vol. 83, no. 1, pp. 113–123, 1997.

[138] K. Waye and E. Öhrström, "Psycho-acoustic characters of relevance for annoyance of wind turbine noise," *Journal of sound and vibration*, vol. 250, no. 1, pp. 65–73, 2002.

[139] S. Viollon and C. Lavandier, "Multidimensional assessment of the acoustic quality of urban environments," in *Conf. proceedings "Internoise", Nice, France, 27-30 Aug*, vol. 4, 2000, pp. 2279–2284.

[140] J. Kang and M. Zhang, "Semantic differential analysis of the soundscape in urban open public spaces," *Building and environment*, vol. 45, no. 1, pp. 150–157, 2010.

[141] A. Lindau, V. Erbes, S. Lepa, H. Maempel, F. Brinkman, and S. Weinzierl, "A spatial audio quality inventory (saqi)," *Acta Acustica united with Acustica*, vol. 100, no. 5, pp. 984–994, 2014.

[142] M. Bradley and P. Lang, "Measuring emotion: The self-assessment manikin and the semantic differential," *Journal of behavior therapy and experimental psychiatry*, vol. 25, no. 1, pp. 49–59, 1994.

[143] W. Wundt, *Grundriss der psychologie (outlines of psychology)*. Leipzig: Entgelann, 1896.

[144] P. Lang, M. Bradley, and B. Cuthbert, *International affective picture system (iaps): Technical manual and affective ratings*, 1999.

[145] M. Bradley and P. J Lang, *The international affective digitized sounds (iads)[: Stimuli, instruction manual and affective ratings.* NIMH Center for the Study of Emotion and Attention, 1999.

[146] R. Cain, P. Jennings, and J. Poxon, "The development and application of the emotional dimensions of a soundscape," *Applied Acoustics*, vol. 74, no. 2, pp. 232–239, 2013.

[147] T. Herzog and P. Bosley, "Tranquility and preference as affective qualities of natural environments," *Journal of environmental psychology*, vol. 12, no. 2, pp. 115–127, 1992.

[148] J. Russell, "A circumplex model of affect.," *Journal of personality and social psychology*, vol. 39, no. 6, p. 1161, 1980.

[149] F. Stevens, D. T. Murphy, and S. L. Smith, "Soundscape categorisation and the self-assessment manikin," in *Proceedings of the 20th International Conference on Digital Audio Effects (DAFx-17)*, Edinburgh, UK, 2017.

[150] D. Handayani, A. Wahab, and H. Yaacob, "Recognition of emotions in video clips: The self-assessment manikin validation," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 13, no. 4, pp. 1343–1351, 2015.

[151] M. K. Højlund, "Overhearing: An attuning approach to noise in Danish hospitals," PhD thesis, Aarhus University, 2017.

[152] T. C. Andringa and K. A. Van Den Bosch, "Core effect and soundscape assessment: Fore-and background soundscape design for quality of life," in *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, Institute of Noise Control Engineering, vol. 247, 2013, pp. 2273–2282.

[153] K. Lynch, *The image of the city.* MIT press, 1960, vol. 11.

[154] M. Hunter, S. Eickhoff, R. Pheasant, M. Douglas, G. Watts, T. Farrow, D. Hyland, J. Kang, I. Wilkinson, K. Horoshenkov, *et al.*, "The state of tranquility: Subjective perception is shaped by contextual modulation of auditory connectivity," *Neuroimage*, vol. 53, no. 2, pp. 611–618, 2010.

[155] H. Fastl *et al.*, "Audio-visual interactions in loudness evaluation," in *Proceedings of the 18th International Congress on Acoustics*, Citeseer, 2004, pp. 1161–1166.

[156] C. Patsouras, T. Filippou, and H. Fastl, "Influences of color on the loudness judgement," in *Proc. Forum Acusticum*, vol. 2, 2002.

[157] Z. Bangjun, S. Lili, and D. Guoqing, "The influence of the visibility of the source on the subjective annoyance due to its noise," *Applied Acoustics*, vol. 64, no. 12, pp. 1205–1215, 2003.

[158] R. Ulrich, "View through a window may influence recovery," *Science*, vol. 224, no. 4647, pp. 224–225, 1984.

[159] E. Bjork, "Psychophysiological responses to some natural sounds," *Acta acústica*, vol. 3, no. 1, pp. 83–88, 1995.

[160] W. Yang and J. Kang, "Acoustic comfort evaluation in urban open public spaces," *Applied Acoustics*, vol. 66, no. 2, pp. 211–229, 2005.

[161] J. Fields, "Effect of personal and situational variables on noise annoyance in residential areas," *The Journal of the Acoustical Society of America*, vol. 93, no. 5, pp. 2753–2763, 1993.

[162] K. Zimmer and W. Ellermeier, "Psychometric properties of four measures of noise sensitivity: A comparison," *Journal of Environmental Psychology*, vol. 19, no. 3, pp. 295–302, 1999.

[163] H. Bregman, *Development of a noise annoyance sensitivity scale*. National Aeronautics and Space Administration, 1972.

[164] E. Verona, C. Patrick, J. Curtin, M. Bradley, and P. Lang, "Psychopathy and physiological response to emotionally evocative sounds.," *Journal of abnormal psychology*, vol. 113, no. 1, p. 99, 2004.

[165] N. Moreira and M. Bryan, "Noise annoyance susceptibility," *Journal of Sound and Vibration*, vol. 21, no. 4, pp. 449–462, 1972.

[166] M. Raimbault and D. Dubois, "Urban soundscapes: Experiences and knowledge," *Cities*, vol. 22, no. 5, pp. 339 –350, 2005, ISSN: 0264-2751. DOI: http://dx.doi.org/10.1016/j.cities.2005.05.003. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0264275105000557.

[167] C. Watson, "Trent falls to spurn point," 2017. [Online]. Available: https://www.hull2017.co.uk/whatson/events/trent-falls-to-spurn-point/.

[168] L. Barclay, "Listening to communities and environments," *Contemporary Music Review*, vol. 36, no. 3, pp. 143–158, 2017.

[169] *The Overheard*, 2017. [Online]. Available: `https://www.overheard.dk/the-overheard/`.

[170] *Liminal - cochlea unwound*. [Online]. Available: `http://www.liminal.org.uk/portfolio/cochlea_unwound/`.

[171] *Liminal - organ of corticochlea unwound*. [Online]. Available: `http://www.liminal.org.uk/portfolio/organ-of-corti/`.

[172] P. Cusack *et al.*, "Favourite sounds project," 2001. [Online]. Available: `http://favouritesounds.org/`.

[173] "Listening to the commons," 2017. [Online]. Available: `https://www.york.ac.uk/history/listening-to-the-commons/about/`.

[174] *Voice and vote: Women's place in parliament exhibition*, 2018. [Online]. Available: `https://www.parliament.uk/get-involved/vote-100/voice-and-vote/`.

[175] J. Hughes, *Refugium*, 2017. [Online]. Available: `https://refugium2017.wordpress.com/`.

[176] *AudioMoth*. [Online]. Available: `https://www.openacousticdevices.info/audiomoth`.

[177] A. P. Hill, P. Prince, E. P. Covarrubias, C. P. Doncaster, J. L. Snaddon, and A. Rogers, "Audiomoth: Evaluation of a smart open acoustic device for monitoring biodiversity and the environment," *Methods in Ecology and Evolution*, 2017.

[178] M. Kleiner, B. Dalenbäck, and P. Svensson, "Auralization-an overview," *Journal of the Audio Engineering Society*, vol. 41, no. 11, pp. 861–875, 1993.

[179] F. Rumsey, *Spatial audio*. CRC Press, 2012.

[180] D. Hammershøi and H. Møller, "Methods for binaural recording and reproduction," *Acta Acustica united with Acustica*, vol. 88, no. 3, pp. 303–311, 2002.

[181] W. G. Gardner and K. D. Martin, "Hrtf measurements of a kemar," *The Journal of the Acoustical Society of America*, vol. 97, no. 6, pp. 3907–3908, 1995.

[182] P. Mokhtari, H. Takemoto, R. Nishimura, and H. Kato, "Computer simulation of hrtfs for personalization of 3d audio," in *Universal Communication, 2008. ISUC'08. Second International Symposium on*, IEEE, 2008, pp. 435–440.

[183] C. Yue and T. de Planque, "3-d ambisonics experience for virtual reality," *Space*, vol. 6, p. 7,

[184] R. Furness, "Ambisonics - an overview," in *Audio Engineering Society Conference: 8th International Conference: The Sound of Audio*, May 1990.

[185] D. G. Malham and A. Myatt, "3-d sound spatialization using ambisonic techniques," *Computer music journal*, vol. 19, no. 4, pp. 58–70, 1995.

[186] E. Benjamin and T. Chen, "The native b-format microphone," in *Audio Engineering Society Convention 119*, Oct. 2005.

[187] *B-format microphone response*, 2015. [Online]. Available: `http://www.surround-library.com/images/white-papers/wp-b-format.jpeg`.

[188] J. Daniel, S. Moreau, and R. Nicol, "Further investigations of high-order ambisonics and wavefield synthesis for holophonic sound imaging," in *Audio Engineering Society Convention 114*, Audio Engineering Society, 2003.

[189] F. Stevens, D. T. Murphy, and S. L. Smith, "Emotion and soundscape preference rating: Using semantic differential pairs and the self-assessment manikin," in *Sound and Music Computing conference, Hamburg, 2016*, Hamburg, Germany, 2016.

[190] B. Wiggins, *Youtube, ambisonics and vr*, 2016. [Online]. Available: `https://www.brucewiggins.co.uk/?p=666`.

[191] M. C. Green and D. Murphy, "Eigenscape: A database of spatial acoustic scene recordings," *Applied Sciences*, vol. 7, no. 11, p. 1204, 2017.

[192] A. J. Horsburgh, K. B. McAlpine, and D. F. Clark, "A perspective on the adoption of ambisonics for games," in *Audio Engineering Society Conference: 41st International Conference: Audio for Games*, Audio Engineering Society, 2011.

[193] C. Guastavino, B. Katz, J. Polack, D. Levitin, and D. Dubois, "Ecological validity of soundscape reproduction," *Acta Acustica united with Acustica*, vol. 91, no. 2, pp. 333–341, 2005.

[194] N. Bruce and W. Davies, "The effects of expectation on the perception of soundscapes," *Applied Acoustics*, vol. 85, pp. 1–11, 2014.

[195] S. Chegg, *Soundscape recording using a Beef-format microphone*, 2015. [Online]. Available: https://twitter.com/WiganIOA/status/614484686044327936.

[196] S. Berge and N. Barrett, "High angular resolution planewave expansion," in *Proc. of the 2nd International Symposium on Ambisonics and Spherical Acoustics May*, 2010, pp. 6–7.

[197] D. Aylor and L. Marks, "Perception of noise transmitted through barriers," *The Journal of the Acoustical Society of America*, vol. 59, no. 2, pp. 397–400, 1976.

[198] G. Watts, A. Miah, and R. Pheasant, "Tranquillity and soundscapes in urban green spaces - predicted and actual assessments from a questionnaire survey.," *Environment and Planning B: Planning and Design*, 2013.

[199] G. Watts and R. Pheasant, "Factors affecting tranquillity in the countryside," *Applied Acoustics*, vol. 74, no. 9, pp. 1094–1103, 2013.

[200] G. Watts, R. Pheasant, and K. Horoshenkov, "Validation of tranquillity rating method," *Proceedings of the Institute of Acoustics and Belgium Acoustical Society: Noise in the built environment, Ghent*, vol. 32, no. 3, 2010.

[201] R. Pheasant, K. Horoshenkov, G. Watts, and B. Barrett, "The acoustic and visual factors influencing the construction of tranquil space in urban and rural environments tranquil spaces-quiet places?" *The Journal of the Acoustical Society of America*, vol. 123, no. 3, pp. 1446–1457, 2008.

[202] R. Gifford and C. Ng, "The relative contribution of visual and auditory cues to environmental perception," *Journal of Environmental Psychology*, vol. 2, no. 4, pp. 275–284, 1982.

[203] F. Ruotolo, L. Maffei, M. Gabriele, T. Iachini, M. Masullo, G. Ruggiero, and V. P. Senese, "Immersive virtual reality and environmental noise assessment: An innovative audiovisual approach," *Environmental Impact Assessment Review*, vol. 41, no. 0, pp. 10 –20, 2013, ISSN: 0195-9255. DOI: http://dx.doi.org/10.1016/j.eiar.2013.01.007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0195925513000188.

[204] *The Oculus Rift*, 2015. [Online]. Available: http://www.oculus.com/.

[205] *Oculus rift headset photos*, 2015. [Online]. Available: http://en.wikipedia.org/wiki/Oculus_Rift.

[206] *Kolor autopano*, 2015. [Online]. Available: http://www.kolor.com/autopano-video/#start.

[207] *Freedom 360 mount*, 2015. [Online]. Available: http://freedom360.us/shop/freedom360/.

[208] Ö. Axelsson, "How to measure soundscape quality," in *Euronoise2015, Maastricht*, 2015.

[209] O. Rummukainen, J. Radun, T. Virtanen, and V. Pulkki, "Categorization of natural dynamic audiovisual scenes," *PloS one*, vol. 9, no. 5, e95848, 2014.

[210] M. Raimbault, "Qualitative judgements of urban soundscapes: Questionning questionnaires and semantic scales," *Acta acustica united with acustica*, vol. 92, no. 6, pp. 929–937, 2006.

[211] M. Southworth, "The sonic environment of cities," *Environment and Behavior*, vol. 1, no. 1, p. 49, Jun. 1969, Last updated - 2013-02-22.

[212] P. Fröhlich, S. Egger, R. Schatz, M. Mühlegger, K. Masuch, and B. Gardlo, "Qoe in 10 seconds: Are short video clip lengths sufficient for quality of experience assessment?" In *Quality of Multimedia Experience (QoMEX), 2012 Fourth International Workshop on*, IEEE, 2012, pp. 242–247.

[213] *Google maps*, 2015. [Online]. Available: https://www.google.co.uk/maps?source=tldsi&hl=en.

[214] A. Pike, M. Gilmour, P. Pettitt, R. Jacobi, S. Ripoll, P. Bahn, and F. Muñoz, "Verification of the age of the palaeolithic cave art at creswell crags, uk," *Journal of Archaeological Science*, vol. 32, no. 11, pp. 1649–1655, 2005.

[215] F. Stevens and D. T. Murphy, "Acoustic source localisation in an urban environment using early reflection information," in *Euronoise2015*, Maastricht, The Netherlands, 2015.

[216] J. Merimaa and V. Pulkki, "Spatial impulse response rendering i: Analysis and synthesis," *J. Audio Eng. Soc*, vol. 53, no. 12, pp. 1115–1127, 2005.

[217] F. Stevens, *Creswell crags impulse responses*, 2017. [Online]. Available: http://www.openairlib.net/auralizationdb/content/creswell-crags.

[218] C. Osgood, "The nature and measurement of meaning.," *Psychological bulletin*, vol. 49, no. 3, p. 197, 1952.

[219] T. Hashimoto and S. Hatano, "Effects of factors other than sound to the perception of sound quality," *17th ICA Rome, CD-ROM*, 2001.

[220] W. Davies, N. Bruce, and J. Murphy, "Soundscape reproduction and synthesis," *Acta Acustica United with Acustica*, vol. 100, no. 2, pp. 285–292, 2014.

[221] A. Zeitler and J. Hellbrück, "Semantic attributes of environmental sounds and their correlations with psychoacoustic magnitude," in *Proc. of the 17th International Congress on Acoustics [CDROM], Rome, Italy*, vol. 28, 2001.

[222] B. Schulte-Fortkamp, "The quality of acoustic environments and the meaning of soundscapes," in *Proc. of the 17th international conference on acoustics*, 2001.

[223] M. Bradley, B. Cuthbert, and P. Lang, "Picture media and emotion: Effects of a sustained affective context," *Psychophysiology*, vol. 33, no. 6, pp. 662–670, 1996.

[224] K. Hume and M. Ahtamad, "Physiological responses to and subjective estimates of soundscape elements," *Applied Acoustics*, vol. 74, no. 2, pp. 275–281, 2013.

[225] R. Hull and A. Harvey, "Explaining the emotion people experience in suburban parks," *Environment and behavior*, vol. 21, no. 3, pp. 323–345, 1989.

[226] K. Hanyu, "Visual properties and affective appraisals in residential areas in daylight," *Journal of Environmental Psychology*, vol. 20, no. 3, pp. 273–284, 2000.

[227] J. Jot and O. Warusfel, "Spat: A spatial processor for musicians and sound engineers," in *CIARM: International Conference on Acoustics and Musical Research*, 1995.

[228] J. Jot, "The spat processor and its library of max objects-introduction," Technical report IRCAM, Tech. Rep., 2006.

[229] R. A. Fisher, *Statistical methods for research workers*. Genesis Publishing Pvt Ltd, 1925.

[230] *Calculate r critical*, 2015. [Online]. Available: http://blog.excelmasterseries.com/2014/05/pearson-correlation-r-critical-and-p.html.

[231] *Jury test analysis*, 2015. [Online]. Available: http://www.salford.ac.uk/computing-science-engineering/research/acoustics/psychoacoustics/sound-quality-making-products-sound-better/sound-quality-testing/assessment-methods/jury-testing/10.

[232] *Shimmer GSR+ module*, 2015. [Online]. Available: http://www.shimmersensing.com/images/uploads/docs/GSR_Revision_1.2.pdf.

[233] B. K. Wiederhold, D. Jang, M. Kaneda, I. Cabral, Y. Lurie, T May, I. Kim, M. D. Wiederhold, and S. Kim, "An investigation into physiological responses in virtual environments: An objective measurement of presence," *Towards cyberpsychology: Mind, cognitions and society in the internet age*, 2001.

[234] J. G. Peatman, *Introduction to applied statistics.* Harper & Row, 1963.

[235] R. Elen, "Ambisonics: The surround alternative," in *Proceedings of the 3rd Annual Surround Conference and Technology Showcase*, 2001, pp. 1–4.

[236] M. A. Gerzon, "Ambisonics in multichannel broadcasting and video," *Journal of the Audio Engineering Society*, vol. 33, no. 11, pp. 859–871, 1985.

[237] ISO, *Mechanical vibration and shock: Evaluation of human exposure to whole-body vibration. part 1, general requirements: International standard iso 2631-1: 1997 (e).* ISO, 1997.

[238] A Farina, *Conversion between B-format and UHJ*, [Online] Available: `http://pcfarina.eng.unipr.it/Aurora/B-Format_to_UHJ.htm`, 2007.

[239] J. Snow and M Mann, *Qualtrics survey software: Handbook for research professionals*, 2013.

[240] T. Cox, "The effect of visual stimuli on the horribleness of awful sounds," *Applied acoustics*, vol. 69, no. 8, pp. 691–703, 2008.

[241] F. Stevens, D. T. Murphy, and S. L. Smith, "Ecological validity of stereo uhj soundscape reproduction," in *In Proceedings of the 142nd Audio Engineering Society (AES) Convention*, Berlin, Germany, 2017.

[242] K. Pearson, "Note on regression and inheritance in the case of two parents," *Proceedings of the Royal Society of London*, vol. 58, pp. 240–242, 1895.

[243] J. Macdonald and H. McGurk, "Visual influences on speech perception processes," English, *Perception & Psychophysics*, vol. 24, no. 3, pp. 253–257, 1978, ISSN: 0031-5117. DOI: `10.3758/BF03206096`. [Online]. Available: `http://dx.doi.org/10.3758/BF03206096`.

[244] K. Tzoulas, K. Korpela, S. Venn, V. Yli-Pelkonen, A. Kaźmierczak, J. Niemela, and P. James, "Promoting ecosystem and human health in urban areas using green infrastructure: A literature review," *Landscape and urban planning*, vol. 81, no. 3, pp. 167–178, 2007.

[245] Arup, *Cities Alive: Rethinking green infrastructure*, 2014. [Online]. Available: https://www.arup.com/-/media/Arup/Files/Publications/C/Cities_Alive_booklet.pdf.

[246] D. T. Murphy, A. Southern, and F. Stevens, "Sounding our smart cities: Soundscape design, auralisation and evaluation for our urban environment," in *Sound + Environment 2017*, Hull, UK, 2017.

[247] *FSPViewer*, 2017. [Online]. Available: http://www.fsoft.it/FSPViewer/.

[248] S. Shapiro and M. Wilk, "An analysis of variance test for normality (complete samples)," *Biometrika*, pp. 591–611, 1965.

[249] H. Mann and D. Whitney, "On a test of whether one of two random variables is stochastically larger than the other," *The annals of mathematical statistics*, pp. 50–60, 1947.

[250] A. Southern, F. Stevens, and D. T. Murphy, "Sounding out smart cities: Auralization and soundscape monitoring for environmental sound design," *The Journal of the Acoustical Society of America*, vol. 141, no. 5, pp. 3880–3880, 2017.

[251] *FFMPEG*. [Online]. Available: https://www.ffmpeg.org/.

[252] *Google support: Upload 360-degree videos*. [Online]. Available: https://support.google.com/youtube/answer/6178631\\?hl=en-GB.

[253] *Spatial media metadata injector*, 2016. [Online]. Available: https://github.com/google/spatial-media/releases.

[254] D. Swart, *Equirectangular perspective lines*, 2016. [Online]. Available: https://www.flickr.com/photos/dmswart/26363697850.

[255] C. Nachbar, F. Zotter, E. Deleflie, and A. Sontacchi, "Ambix-a suggested ambisonics format," in *Ambisonics Symposium, Lexington*, 2011.

[256] F. Stevens, *Final test audio stimuli YouTube playlist*, 2017. [Online]. Available: https://www.youtube.com/playlist?list=PL-3kCuZ4n30QM5zUhzqfn9vkiwZPl2QzD.

[257] ——, *Final test visual stimuli YouTube playlist*, 2017. [Online]. Available: https://www.youtube.com/playlist?list=PL-3kCuZ4n30TIn40XSXdz5Y-sPr5brNet.

[258] L. Savioja and U. P. Svensson, "Overview of geometrical room acoustic modeling techniques," *J. Acoust. Soc. Am.*, vol. 138, no. 2, pp. 708–730, Aug. 2015.

[259] A. Krokstad, S. Strom, and S. S., "Calculating the acoustical room response by the use of a ray tracing technique," *J. Sound Vibr.*, vol. 8, no. 1, 118—125, Jan. 1968.

[260] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Amer.*, vol. 65, no. 4, pp. 943–950, 1979.

[261] J. Borish, "Extension of the image model to arbitrary polyhedra," *J. Acoust. Soc. Amer.*, vol. 75, no. 6, pp. 1827–1836, Jun. 1984.

[262] J. O. Smith, "A new approach to digital reverberation using closed waveguide networks," in *Proc. Int. Computer Music Conf.*, Vancouver, BC, Canada, Aug. 1985, pp. 47–53.

[263] P. Huang, M. Karjalainen, and J. Smith, "Digital waveguide networks for room response modeling and synthesis," in *Proc. Audio Eng. Soc. 118thth Conv.*, preprint no. 6394, Barcelona, Spain, May 2005.

[264] J.-M. Jot and A. Chaigne, "Digital delay networks for designing artificial reverberators," in *Proc. Audio Eng. Soc. 90th Conv.*, preprint no. 3030, Paris, France, Feb. 1991.

[265] S. J. Schlecht and E. A. P. Habets, "Feedback delay networks: Echo density and mixing time," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 25, no. 2, pp. 374–383, Feb. 2017.

[266] S. J. Schlecht and E. A. Habets, "On lossless feedback delay networks," *IEEE Trans. Signal Process.*, vol. 65, no. 6, pp. 1554–1564, Mar. 2017.

[267] L. Savioja, T. Rinne, and T. Takala, "Simulation of room acoustics with a 3-D finite difference mesh," in *Proc. Int. Computer Music Conf.*, Aarhus, Denmark, Sep. 1994, pp. 463–466.

[268] K. Kowalczyk and M. van Walstijn, "Room acoustics simulation using 3-D compact explicit FDTD schemes," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 1, 34—46, Jan. 2011.

[269] S. Bilbao, "Modeling of complex geometries and boundary conditions in finite difference/finite volume time domain room acoustics simulation," *IEEE Trans. Audio Speech and Lang. Process.*, vol. 21, no. 7, pp. 1524–1533, Jul. 2013.

[270] S. Bilbao, B. Hamilton, J. Botts, and L. Savioja, "Finite volume time domain room acoustics simulation under general impedance boundary conditions," *IEEE/ACM Trans. Audio Speech and Lang. Process.*, vol. 24, no. 1, pp. 161–173, Jan. 2016.

[271] E. De Sena, H. Hacihabiboglu, and Z. Cvetkovic, "Scattering delay network: An interactive reverberator for computer games," in *Audio Engineering Society Conference: 41st International Conference: Audio for Games*, Feb. 2011. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=15751.

[272] F. M. Wiener, C. I. Malme, and C. M. Gogos, "Sound propagation in urban areas," *J. Acoust. Soc. Amer.*, vol. 37, no. 4, pp. 738–747, Apr. 1965.

[273] H. Sakai, S. Sato, and Y. Ando, "Orthogonal acoustical factors of sound fields in a forest compared with those in a concert hall," *J. Acoust. Soc. Amer.*, vol. 104, no. 3, pp. 1491–1497, Sep. 1998.

[274] H. Sakai, S. Shibata, and Y. Ando, "Orthogonal acoustical factors of a sound field in a bamboo forest," *J. Acoust. Soc. Amer.*, vol. 109, no. 6, pp. 2824–2830, Jun. 2001.

[275] T. Lokki, A. Southern, S. Siltanen, and L. Savioja, "Acoustics of Epidaurus—Studies with room acoustics modelling methods," *Acta Acust. united Ac.*, vol. 99, no. 1, pp. 40–47, Jan. 2013.

[276] J. Kang, "Numerical modeling of the sound fields in urban squares," *J. Acoust. Soc. Amer.*, vol. 117, no. 6, pp. 3695–3706, Jun. 2005.

[277] R. E. Collecchia, J. S. Abel, S. Coffin, E. Callery, Y. H. Yeh, K. Spratt, and J. O. Smith, III, "On the acoustics of alleyways," in *Proc. Audio Eng. Soc. 137th Conv.*, Los Angeles, USA, Oct. 2014.

[278] J. Stienen and M. Vorländer, "Auralization of urban environments—Concepts towards new applications," in *Proc. EuroNoise 2015*, Maastricht, The Netherlands, Jun. 2015.

[279] R. Pieren and J. M. Wunderly, "A model to predict sound reflections from cliffs," *Acta Acust. united Ac.*, vol. 97, no. 2, pp. 243–253, Mar. 2011.

[280] S. Harriet and D. T. Murphy, "Auralisation of an urban soundscape," *Acta Acustica united with Acustica*, vol. 101, no. 4, pp. 798–810, Jul. 2015.

[281] K. Spratt and J. S. Abel, “A digital reverberator modeled after the scattering of acoustic waves by trees in a forest,” in *Proc. Audio Eng. Soc. 125th Conv.*, San Francisco, CA, Oct. 2008.

[282] E. D. Sena, H. Hacihabiboglu, Z. Cvetkovic, and J. O. Smith, “Efficient synthesis of room acoustics via scattering delay networks,” *IEEE/ACM Trans. Audio, Speech, and Language Process.*, vol. 23, no. 9, pp. 1478–1492, Sep. 2015.

[283] S. Van Duyne and J. O. Smith, “The 2-D digital waveguide mesh,” in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust. (WASPAA)*, New Paltz, NY,USA, Oct. 1993, pp. 177–180.

[284] F. Fontana and D. Rocchesso, “Signal-theoretic characterization of waveguide mesh geometries for models of two-dimensional wave propagation in elastic media,” *IEEE Trans. Speech and Audio Process.*, vol. 9, no. 2, pp. 152–161, Feb. 2001.

[285] D. T. Murphy, A. Kelloniemi, J. Mullen, and S. Shelley, “Acoustic modeling using the digital waveguide mesh,” *IEEE Signal Process. Mag.*, vol. 24, no. 2, pp. 55–66, Mar. 2007.

[286] B. Hamilton, “Sampling and reconstruction on a diamond grid and the tetrahedral digital waveguide mesh,” *IEEE Signal Process. Lett.*, vol. 10, no. 20, pp. 925–928, Oct. 2013.

[287] F. Stevens, D. T. Murphy, L. Savioja, and V. Välimäki, “Modeling sparsely reflecting outdoor acoustic scenes using the waveguide web,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 8, pp. 1566–1578, Aug. 2017, ISSN: 2329-9290. DOI: 10.1109/TASLP.2017.2699424.

[288] J. Van Mourik and D. Murphy, “Geometric and wave-based acoustic modelling using blender,” in *Audio Engineering Society Conference: 49th International Conference: Audio for Games*, Audio Engineering Society, 2013.

[289] H. Lee and B. H. Lee, “An efficient algorithm for the image model technique,” *Applied Acoustics*, vol. 24, no. 2, pp. 87–115, 1988.

[290] T. Funkhouser, N. Tsingos, I. Carlbom, G. Elko, M. Sondhi, J. E. West, G. Pingali, P. Min, and A. Ngan, “A beam tracing method for interactive architectural acoustics,” *The Journal of the acoustical society of America*, vol. 115, no. 2, pp. 739–756, 2004.

[291] A. Ondet and J. Barbry, "Modeling of sound propagation in fitted workshops using ray tracing," *The Journal of the Acoustical Society of America*, vol. 85, no. 2, pp. 787–796, 1989.

[292] C. Schissler and D. Manocha, "Gsound: Interactive sound propagation for games," in *Audio Engineering Society Conference: 41st International Conference: Audio for Games*, Audio Engineering Society, 2011.

[293] S. Oxnard, "Efficient hybrid virtual room acoustic modelling," PhD thesis, University of York, 2016.

[294] S. Oxnard and D. T. Murphy, "Room impulse response synthesis based on a 2d multi-plane fdtd hybrid acoustic model," in *Applications of Signal Processing to Audio and Acoustics (WASPAA), 2013 IEEE Workshop on*, IEEE, 2013, pp. 1–4.

[295] M. Vorländer, *Auralization: Fundamentals of acoustics, modelling, simulation, algorithms and acoustic virtual reality.* Springer Berlin Heidelberg, 2007.

[296] D. T. Murphy, *Advanced music technology systems: Modelling 2-d systems*, University Lecture Notes, University of York, 2014.

[297] ——, *Physical modelling synthesis 2d fdtd membrane modelling exercise*, University Laboratory Exercise notes, University of York, 2014.

[298] P. Morse, *Vibration and sound.* McGraw-Hill New York, 1948.

[299] P. Chobeau, "Modeling of sound propagation in forests using the transmission line matrix method," PhD thesis, Université du Maine, Le Mans, France, Nov. 2014.

[300] T. Wiens, *Forest reverb generator – file exchange – MATLAB central*, [Accessed: 03-11-2015], Jul. 2008. [Online]. Available: `http://www.mathworks.com/matlabcentral/fileexchange/20764-forest-reverb-generator`.

[301] F. Stevens, *Waveguide web example audio*, 2017. [Online]. Available: `http://www.openairlib.net/auralizationdb/content/waveguide-web-example-audio`.

[302] ISO 3382-1, "Acoustics, measurement of room acoustic parameters. part 1: Performance spaces," *The International Organization for Standardization*, 2009.

[303] ISO 3382-2, "Acoustics, measurement of room acoustic parameters. part 2: Reverberation time in ordinary rooms," *The International Organization for Standardization*, 2009.

[304] Z. Meng, F. Zhao, and M. He, "The just noticeable difference of noise length and reverberation perception," in *Proc. Int. Symp. Communications and Info. Technologies*, Bangkok, Thailand, Oct. 2006, pp. 418–421.

[305] A. Buck, M. G. Blevins, L. M. Wang, and Z. Peng, "Measurements of the just noticeable difference for reverberation time using a transformed up–down adaptive method," *The Journal of the Acoustical Society of America*, vol. 132, no. 3, pp. 2060–2060, 2012.

[306] W. Robertson and J. Rudy III, "Measurement of acoustic stop bands in two-dimensional periodic scattering arrays," *J. Acoust. Soc. Amer.*, vol. 104, no. 2, pp. 694–699, Aug. 1998.

[307] H. Bass, H.-J. Bauer, and L. Evans, "Atmospheric absorption of sound: Analytical expressions," *J. Acoust. Soc. Amer.*, vol. 52, no. 3B, pp. 821–825, Sep. 1972.

[308] K. Attenborough, "Ground parameter information for propagation modeling," *J. Acoust. Soc. Amer.*, vol. 92, no. 1, pp. 418–427, Jul. 1992.

[309] J. Kang, "Numerical modelling of the sound fields in urban streets with diffusely reflecting boundaries," *Journal of Sound and Vibration*, vol. 258, no. 5, pp. 793–813, 2002.

[310] F. Stevens, D. Murphy, and S. Smith, "Ecological validity of stereo uhj soundscape reproduction," in *Audio Engineering Society Convention 142*, Audio Engineering Society, 2017.

[311] F Stevens, D. T. Murphy, and S. L. Smith, "Soundscape auralisation and visualisation: A cross-modal approach to soundscape evaluation," in *DAFx 2018*, Aveiro, Portugal, Sep. 2018.

[312] R. Parry, "Digital heritage and the rise of theory in museum computing," *Museum management and Curatorship*, vol. 20, no. 4, pp. 333–348, 2005.

[313] *Park square leeds*, 2015. [Online]. Available: `http://en.wikipedia.org/wiki/Park_Square,_Leeds#/media/File:Park_Square,_Leeds_001.jpg`.

[314] *Tenma 72-860a instruction manual*, 2015. [Online]. Available: `http://www.farnell.com/datasheets/317251.pdf`.