The Affective Evaluation of Immersive Soundscapes Including Physiological Measures

Simon Durbridge

PhD

University of York

Physics, Engineering and Technology

May 2023

Abstract

This thesis is focussed on evaluating affective and physiological responses to soundscape stimuli, building towards understanding the importance of ecological validity on ex-situ soundscape evaluation. The primary focus of soundscape research is the subjective evaluation of environmental sound, furthering our understanding of our relationship with our acoustic environment. Although relationships between cognitive and physical measures of soundscape quality have been explored in the literature, there is a lack of research that has focussed on the relationship between physiological and subjective responses.

Two of the experiments presented in this thesis focus on developing a methodology for identifying if changes in physiological behaviour can be used to identify changes in the subjective experience of soundscapes. Results of the first experiment indicate that using more accurate heart rate measurement methods may be necessary to identify the effects of soundscape stimuli on modulation of the autonomic nervous system. The results of a second experiment provide evidence that heart rate variability does not vary in a systematic manner with changes in the subjective experience of soundscape stimuli. It is hypothesized that a lack of presence and immersion in the experimental environment contribute to these results. The results of a third experiment show that control over listener orientation does not influence the subjective experience of soundscape stimuli.

Finally, a set of software tools are described which facilitate the design of soundscape evaluation experiments that utilise virtual reality, offering a greater sense of presence and immersion than those provided in a typical listening test. Future work will test this toolset and attempt to validate its use for the ex-situ evaluation of soundscapes.

Declaration

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 a degree or other qualification at this University or elsewhere. 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:

S. Durbridge and D.T. Murphy, "Assessment of soundscapes using self-report and physiological measures," Acta Acust., vol.7, pp.6, Feb. 2023, doi: 10.1051/aacus/2022059

S. Durbridge and D. T. Murphy, "Soundscape Evaluation In The Virtual Reality: Tools for the creation of soundscape studies", in Proc. AES. AVAR. Conf., Aug. 2022, pp. 13, Permalink: `http://www.aes.org/e-lib/browse.cfm?elib=21843`

Contents

List of Figures

List of Tables

Acknowledgements

No man is an island, and no PhD student is either.

I would like to thank many people for their help and support during the course of this work.

First and foremost I would like to thank Prof. Damian Murphy for his help, support, guidance, kindness and patience. Prof. Murphy was a constant source encouragement despite all of the roadblocks, distractions and setbacks that I encountered.

I would like to thank my fellow students and the staff of the AudioLab for their friendship and support. In particular I would like to thank Dan Turner for always maintaining a positive attitude and for his help and advice.

I was very privileged to be supported by an EPSRC doctoral studentship during the course of this work, and if studentships were not available I would not have been able to take the opportunity to work towards a PhD. For this, I am very grateful to the EPSRC.

I would like to thank my undergraduate and MSc supervisors, Dr. Adam Hill, Dr. Bruce Wiggins and Dr. Simon Lewis for supporting me during my time at Derby.

I would like to thank Dr. William Draffin for helping me get started in the world of sound and audio, I wouldn't be here today if he hadn't given me the opportunity to work with him.

I would like to thank my family for being there for me when I needed them, and for their constant encouragement and support.

And finally I would like to thank Bethany, my rock, for her love, support and unrelenting encouragement. I couldn't have done it without you.

Chapter 1

Introduction

The sensation of a musical tone is due to a rapid periodic motion of the sonorous body; the sensation of noise to non-periodic motions.

Herman Von Helmholtz [1]

1.1 Background

Soundscapes underlie our perception of the world around us, and are an integral part of our daily lives, playing a fundamental role in shaping our experiences. In this thesis the term soundscape is used to refer to the acoustic environment as perceived by people, in context, following the definition given by the ISO standards committee for soundscape [2]. This definition is based on the work of R. Murray Schafer [3] who popularised the concepts of soundscape and acoustic ecology. Unlike research that focusses solely on how some sounds have a negative impact on our health and well-being, soundscape research often approaches sound as a resource that can be used to improve our quality of life. This thesis follows that philosophy by exploring how a variety of soundscapes have an effect on the human body and mind. By investigating the relationship between soundscapes, environmental context and human perception, we aim to increase our understanding of how these factors influence each other. The intention being that improved understanding of these relationships can be leveraged to improve the quality of life for individuals and communities by designing acoustic environments that promote improved health and well-being.

The impact of soundscapes on well-being extends further than aesthetics or preference, and the burden of disease from environmental noise contributes to 12,000 premature deaths and 48,000 new cases of Ischemic Heart Disease each year in the European region [4]. Further research from the World Health Organization (WHO) has shown that exposure to excess noise pollution is a factor in the increased risk of cardiovascular disease, cognitive impairment in children, sleep disturbance, tinnitus, and annoyance [5]. These results come from the WHO's environmental noise guidelines for the European region, a comprehensive meta analysis of evidence on the effects of environmental noise on health in an epidemiological context. However, the WHO guidelines and much environmental noise policy is focussed on identifying and mitigating the negative impact of noise pollution from key noise sources in particular circumstances.

In contrast, soundscape research is focussed on understanding the relationship between people and their acoustic environment in a way that is more holistic, taking contextual factors into account and prioritising the perception of the listener. This approach centres on the investigation of soundscape quality through the subjective evaluation of soundscapes, and the identification of the acoustic and contextual factors that influence the perception of soundscapes. Researchers in the field continue to drive towards a strategy for modelling and predicting soundscape quality [6], an endeavour that is complicated by the subjective and ephemeral nature of soundscapes as a perceptual phenomenon. However, soundscape as a field of research is still in its relative infancy, and there is much to be understood about the relationship between soundscapes and human perception before a robust and generalisable model can be established.

One of the key challenges associated with developing soundscape as a field of research is the creation of a complete theory of soundscape perception that can underlie a model of soundscape quality estimation [7]. This challenge is significant and far beyond the scope of even several PhD theses. However, an important step towards achieving this goal would be to develop a better understanding of the role physiological mechanisms play in the perception of soundscapes. A first step on this path would be to develop a methodology for investigating the presence of physiological responses to soundscapes, and to identify the acoustic and contextual factors that influence these responses. The work presented in this thesis constitutes an investigation into the perception of soundscapes by attempting to identify physiological and emotional responses evoked by participants in laboratory, or controlled ex-situ listening tests. By exploring the connections between soundscapes, emotion and physiology, this body of work aims to contribute to the understanding and development of knowledge in the field of soundscape research.

1.2 Hypothesis

The work presented in this thesis is intended to test the following hypothesis:

The experience of soundscapes classified as featuring natural and mechanical sound sources evoke physiological and affective responses in listeners.

Attempting to prove this statement will provide a first step towards achieving a more complete understanding of the relationship between soundscapes, human perception and the physiological mechanisms that underpin cognitive processes. Understanding how soundscapes can be used in the context of physiological measurement can improve our understanding of the relationship between soundscapes and human perception, and have further impact on applications of soundscape design such as in therapeutic interventions [8].

1.3 Aims & Objectives

To answer the research question stated above, the work presented in this thesis is divided into three main aims that are supported by the objectives that follow:

1. *To explore the background and mechanisms present in soundscape.*

    (a) This will be achieved by performing a comprehensive literature review that introduces soundscape as a field of research, and discusses the physical and psychological processes that are active in the perception of the acoustic environment.

    (b) The main body of this discussion is given in Chapter 3, and is supported by the background information presented in Chapter 2.

2. *To identify measurable physiological responses to the presence of soundscapes.*

    (a) This will be achieved through the design and implementation of a series of experiments that explore the relationship between physiological and subjective responses to soundscape stimuli. These experiments are described in Chapters 4, 5.

    (b) The first experiment described in Chapter 4 explores the feasibility of using physiological measures in soundscape evaluation, building on work identified in the literature and discussed in Sections 3.10 and 4.2 of this thesis.

    (c) The second experiment described in Chapter 5 builds upon the first experiment by improving several factors including experimental control and instrumentation.

    (d) A fourth experiment was planned to investigate the influence of head mounted virtual reality (VR) displays and dynamic sound field orientation on the perception of soundscapes as observed through physiological measures. However, this experiment was not performed due to the COVID-19 pandemic, and an online experiment was performed instead that is discussed in Chapter 6. The tools developed for performing the fourth experiment are described in Chapter 7.

3. *To identify how contextual factors in the experimental methods used in soundscape evaluation might influence the perception of soundscapes.*

   (a) This will be achieved by assessing the difference in the quality of soundscapes in several scenarios with variations in soundscape presentation quality and experimental method.

   (b) Both of the experiments discussed in Chapters 4 and 5 feature the same quality of soundscape presentation, but vary in the protocol used for the presentation and survey of the stimuli.

   (c) The third experiment described in Chapter 6 varies the quality of soundscape presentation by comparing the perception of soundscapes that are presented with static and user controlled dynamic sound field orientation.

   (d) The fourth experiment described in Chapter 7 would further vary the quality of soundscape presentation by including the use of a head mounted display (HMD) to present the soundscapes. The results from this fourth experiment would have been compared to the results of the third experiment in order to further meet this objective.

1.4 Novel Contributions to the Field

Publication of the work presented in this thesis has provided several novel contributions to the field of soundscape research. The primary novel contribution of this thesis is the experimental methodology that was developed for the evaluation of soundscapes using physiological measures. The methodology described in this thesis is novel in that it combines the evaluation of naturally occurring soundscapes that were recorded in the Ambisonic format and presented to participants using binaural rendering. The discussion of the experimental methodology, data collection and analysis presented in this thesis provides a basis for future work in the field of soundscape evaluation using physiological measures. The experiment described in Chapter 4 was to be presented in the proceedings of the Forum Acusticum 2020 conference, however, the conference was cancelled due to the COVID-19 pandemic. The developed methodology was subsequently published with the results of the experiment described in Chapter 5 in the European Acoustics Association journal Acta Acustica [9].

Another novel contribution of this thesis is the development of a toolset for the evaluation of soundscapes using virtual reality technologies. The toolset described in Chapter 7 provides a means of presenting soundscapes to participants in a manner that is theorised to elicit a higher degree of presence and spatial immersion than a traditional listening test that does not take advantage of immersive technologies. This toolset was developed in order to facilitate the evaluation of soundscapes presented via interactive virtual reality technologies such as a head mounted display, and was presented as part of the proceedings of the Audio Engineering Society International Conference on Audio for Virtual and Augmented Reality 2022 [10].

Another novel contribution of this thesis is evidence pertaining to the influence of interactivity on the perception of soundscapes. The results presented in Chapter 6 indicate that in the context of an online listening test, the ability to interactively change the orientation of a soundscape does not significantly influence the perceived affective quality of the soundscape. This finding allows researchers to consider the importance of interactivity in the context of the wider experimental design when performing soundscape evaluations as part of a listening test.

1.5 Structure

To provide context for the work presented in this thesis, Chapters 2 and 3 provide a brief overview of important background information. Chapter 2 introduces the foundations of acoustics and sound perception relevant to this thesis, including a description of spatial hearing and the measurement, recording and reproduction of sound field content. Chapter 3 first provides a background to environmental noise management, before introducing soundscape as an alternative philosophy to sound management. This chapter provides the context for the use of soundscape evaluation in the experiments described in the work that follows, including a discussion of literature that explores the use of physiological measures in soundscape research. The following chapters then describe the experimental work undertaken to investigate the research question described above. Chapter 4 describes the pilot experiment, intended to investigate the feasibility of using physiological measures in soundscape evaluation. A subsequent experiment is described in Chapter 5 that builds upon this experimental design. Following this, Chapter 6 describes a third experiment intended to investigate the importance of dynamic sound field orientation in soundscape quality evaluation. Finally, Chapter 7 describes a system designed for the evaluation of soundscapes using virtual reality (VR) technologies including a head mounted display. The thesis is then summarised and concluded in Chapter 8.

Chapter 2

Fundamentals of Sound & Sound Sensing

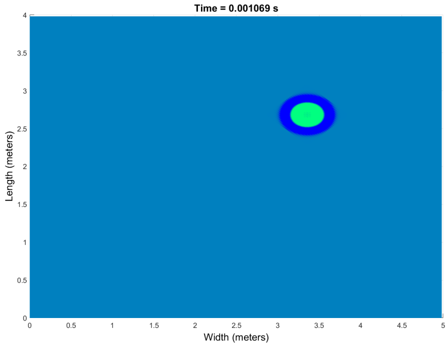
Chapter 1 introduced the background motivation for this thesis, as well as a statement of hypothesis in Section 1.2 and a set of aims and objectives in Section 1.3. Aim 1 of this thesis is *To explore the background and mechanisms present in soundscape*, and this will be predominantly achieved by the literature review that is presented in this and the following chapter. Chapter 3 will focus on the conceptual background from the perspective of environmental noise and soundscape, giving context to the experiments that are presented in the following chapters. This chapter will focus on the fundamentals of sound and sound sensing, providing a fundamental basis for the discussion and experiments that follow. The chapter starts with a description of sound and how sound travels, followed by a summary of the process of hearing including spatial hearing. This is drawn in parallel with a discussion of sound sensing technology including spatial sound sensing by including a brief introduction to Ambisonics.

2.1 Introduction

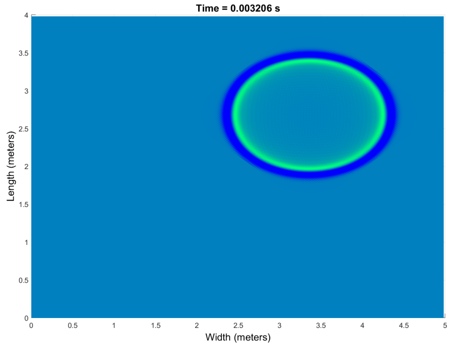
Soundscape, or the perception of the acoustic environment, starts with the vibration of air molecules and resolves with the actions of the listener. The acoustic environment consists of sound sources, sound sensors and the physical environment that the sound waves propagate through. The aim of this chapter is to describe the physical and physiological basis for sound perception, describing the transit of sound from source to perception. The chapter starts with a description of sound and how sound travels, followed by a summary of the process of hearing. The anatomy of the hearing system is described, including the neural pathways involved in processing sound. The chapter concludes a brief discussion of spatial sound perception and one approach for handling spatial sound in signal processing.

2.2 Sound

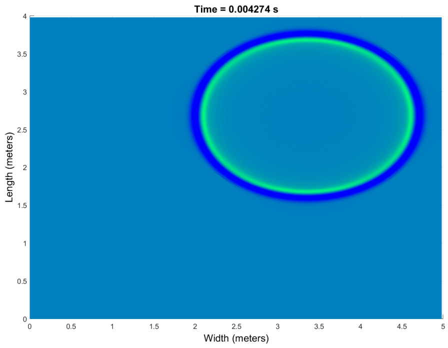
The behaviour of acoustic pressure waves as they travel through the air from a source outwards is the basis for the soundscape we hear, underpinning our understanding of the properties of the environment. Sound can be defined as fluctuations of pressure through an elastic medium such as air or water, resulting in the excitation of the hearing system [1]. Sound waves travel outward from an ideal point source in a spherical motion as presented in Figure 2.1 [11]. Figure 2.1 presents three sub-figures that are two dimensional visualisations of the propagation of acoustic waves from an ideal point source, and one sub-figure that presents the stimulus signal that is being output by the point source. The visualisation was calculated using an implementation of the simplified acoustic wave equation modelled via the finite-difference time domain method [11], and each sub-figure presents the state of the pressure of the simulated space at different points in time. The different colour regions in this figure represent the simulated pressure across the domain, with the x and y axes representing coordinates in two dimensional space. The sinusoidal point source fluctuates the pressure at the centre of the modelled area, and the sound waves propagate outwards in a circular pattern over time. Each wave front can be seen



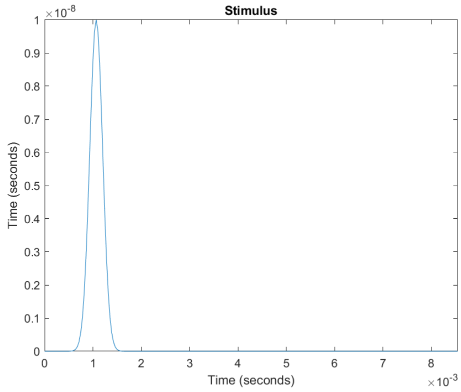
(a) Point source acoustic wave propagation after 30 iterations.



(b) Point source acoustic wave propagation after 90 iterations.



(c) Point source acoustic wave propagation after 120 iterations.



(d) Stimulus signal output by the point source.

Figure 2.1: Point source acoustic wave propagation modelled with the finite difference time domain method.

in the figure as a ring around the centre of the diagram where the pressure is uniform, and as time progresses this ring continues outwards. The speed at which sound travels is dependent on the mechanical properties of the medium and its temperature, and for air is summarised by the following equation:

$$c = \sqrt{\frac{\gamma RT}{M}} \tag{2.1}$$

Where:

- c is the speed of sound measured in meters per second.

- γ represents the ratio of specific heats, also known as the adiabatic index, and is typically expected to be 1.410 for air at room temperature.
- R represents the molar gas constant which is the molar equivalent to the Boltzmann constant, a constant of proportionality .between the thermal energy of the particles in the air and the air temperature, and is approximately $8.314\,\mathrm{J\,mol^{-1}\,K^{-1}}$.
- T is the temperature of the air measured in degrees Kelvin.
- M is the mean molar mass of the air, which is approximately $28.96\,\mathrm{g\,mol^{-1}}$ for dry air.

The speed of sound is typically quoted as being $343.3\,\mathrm{m\,s^{-1}}$ at 20°C at one atmosphere of pressure. Any energy in a sound wave that is not lost via the viscoelastic properties of the medium is distributed across the surface area of the wave front following the inverse square law. The inverse square law states that the amount of pressure on the wave front is inversely proportional to the square of the distance the wave has travelled from the source [12]. The inverse square law can be expressed using the following equation:

$$I = \frac{P}{4\pi r^2} \tag{2.2}$$

Where:

- I is the intensity of sound, the distribution of pressure over area measured in units of Watts per square meter $\mathrm{W\,m^{-2}}$.
- P is the pressure of the radiated sound at the source measured in units of Pascals.
- r is the distance from the epicentre of the source measured in meters (the radius of a sphere).

Sound pressure and intensity is often transformed onto the decibel dB scale, a logarithmic scale that is used to express the ratio of two values of a physical quantity [12]. The convenience of the decibel scale is that the extremely large range of values that occur when computing with the physical quantities of sound pressure and intensity can be expressed in a more manageable range.

When measured using the decibel scale, the proportional relationship of the size of r equates to the reduction of I such that every doubling of r reduces I by 6dB or half as much energy. Equation 2.3 presents the equations for the conversion of intensity between the linear scale to the decibel scale.

$$\begin{aligned} y_{\text{dB}} &= 20 \log 10x \\ y_{lin} &= 10^{\frac{x}{20}} \end{aligned} \tag{2.3}$$

Eventually as the r^2 in Equation 2.2 increases and the denominator becomes sufficiently large, I will become smaller than the average background noise level and the wave front will no-longer be audible. Examples of true point sources are rare in reality, but a specially designed loudspeaker such as a subwoofer can provide a good approximation. A collection of several sound sources with appropriate phase relationships can behave like collections of several sound sources that are arrayed to form a line source such as that presented in Figure 2.2. Figure 2.2 presents a visualisation of a simulation of three sound sources that are arranged in a line, with the centre source being in phase with the outer sources. These sources are producing a sinusoidal signal at the same time or *in phase* with the centre source emulating a line source, and as the sound waves from each source meet they constructively interfere. An ideal line source has a cylindrical radiation pattern, and the modification of the inverse square law equation to represent a cylinder would result in an I of -3dB for each doubling of r [12].

$$I = \frac{P}{2\pi r} \tag{2.4}$$

Line sources can be found in many realistic situations, with line array loudspeaker systems being a prime example. However, in some situations sound sources appear to behave as line sources at one distance, point sources at further distances and plane wave behaviour can be observed at very close distances. One example of a sound source with

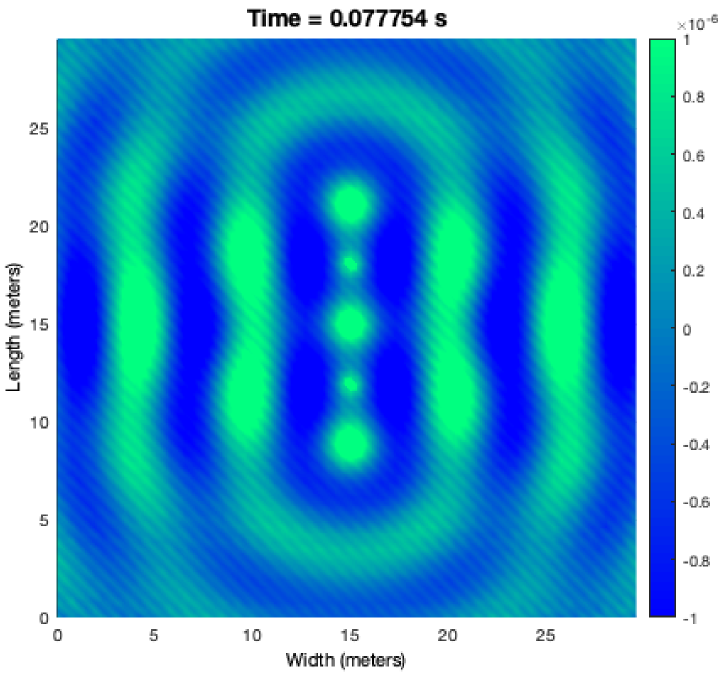


Figure 2.2: Line source acoustic wave propagation modelled with the finite difference time domain method.

this kind of behaviour is a high speed train, which behaves as a line source at closer distances and a point source at far distances [13].

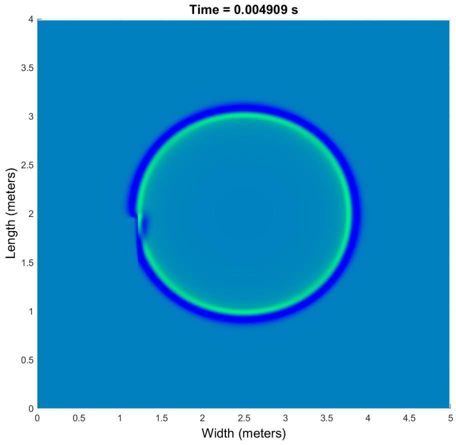
Sound waves often hit an interface between two different mediums that have different mechanical characteristics, and under these circumstances sound waves are absorbed, reflected and diffracted. Some of the energy of the sound wave travels through into the new medium, some is reflected back into the original medium and the remaining energy is absorbed via material deformation and dissipated as heat. The visualization in Figure 2.3 depicts the reflection of a sound wave off of a rigid surface that is highly reflective.

The reflections presented in Figure 2.3 show that sound waves continue to spread in a spherical pattern after meeting the boundary of an obstacle. The amount of energy that is reflected, refracted and absorbed is dependent on the material properties of the medium, and the angle of incidence of the sound wave. Sound waves are signals with an amplitude that changes over time, and the periodic components of sound waves can be analyzed on the basis of frequency i.e. the rate at which the periodic behaviour repeats over time, usually measured relative to a one second period also known as Hertz (Hz). The reflection, refraction and absorption of sound waves as they hit the interface between materials is often frequency dependent, and different materials may have different absorption characteristics at different frequencies. Another behaviour of sound waves is diffraction which is the bending of sound waves around obstacles that are small relative to the wavelength of the sound wave. The wavelength of a signal is the physical distance over which a full cycle of a particular frequency would span, and can be calculated with the following equation:

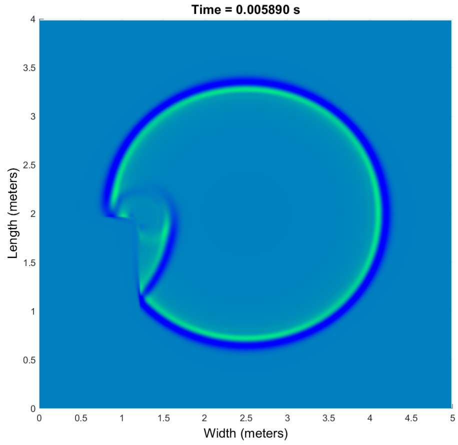
$$\lambda = \frac{c}{f} \tag{2.5}$$

Where:

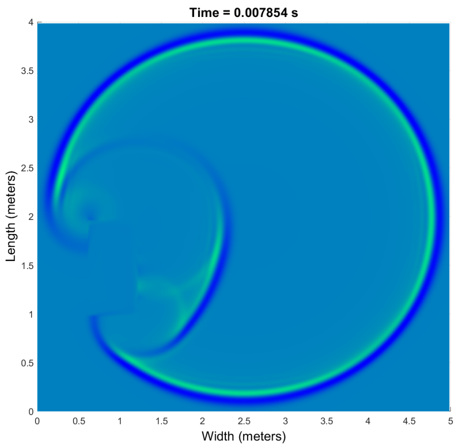
- λ is the wavelength.
- f is the frequency of the wave in Hz.
- c is the speed of sound described in equation 2.1.



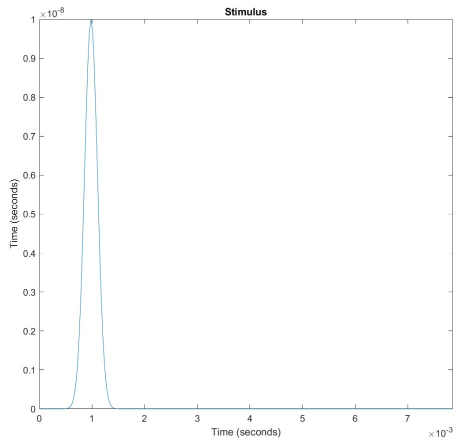
(a) Acoustic wave propagation after 150 iterations.



(b) Acoustic wave propagation after 180 iterations.



(c) Acoustic wave propagation after 240 iterations.



(d) Stimulus signal output by the point source.

Figure 2.3: Point source acoustic wave propagation reflecting off rigid object, modelled using the finite difference time domain method.

The range of frequencies that are audible to humans is often quoted as being between 20Hz and 20kHz for a relatively young person with ideal hearing, though this range is often smaller for older people and people with hearing damage. The range of audible wavelengths is therefore between 17.15m and 0.017m, and for 4kHz, a critical frequency for human hearing, the wavelength is 0.085m which is small relative to the size of the average human head. High frequency sounds do not diffract well around obstacles as the wavelengths of these sounds are short, leading to a phenomenon of acoustic shadowing where sounds are blocked by obstacles.

2.3 Sensing Sound

Sound waves can be measured using a variety of different sensors, many of which are based on the process of transduction, converting one form of energy into another. The most common form of sound sensor is the microphone, which is a transducer that converts sound pressure waves into an electrical signal. There are a variety of microphone technologies with different characteristics, and the choice of microphone is dependent on the application. A typical type of microphone used in sound measurement is the condenser microphone, which consists of a diaphragm that is spaced a small distance away from a fixed backplate. Both the backplate and the diaphragm are made of conductive materials and electrically charged, forming a capacitance between the two plates known as polarization. Condenser microphones are typically polarized either by supplying a voltage from an external source to the backplate, or by applying a thin layer of permanently charged material to the back plate known as electret. Figure 2.4 presents an illustration of a condenser microphone from [14]. The illustration of a condenser microphone in Figure 2.4 show an electrical circuit that terminates at the diaphragm and backplate at one end, and two of the legs of the circuit are connected via a biasing resistor. In this illustration the microphone is polarized by an internal battery instead of an electret material or an external source. The diaphragm is mechanically compliant and as it moves in response to sound pressure waves the capacitance between the diaphragm and the backplate changes, resulting in a change in voltage across the two plates. This change in voltage is amplified and conditioned to produce an electrical signal that is proportional to the sound pressure wave, and can be recorded and analyzed. The sensitivity of a microphone is a measure of the output voltage of the microphone for a known sound pressure level, and is typically measured in $\mathrm{V\,Pa^{-1}}$. A microphone with a high sensitivity will produce a higher output voltage for a given sound pressure level than a microphone

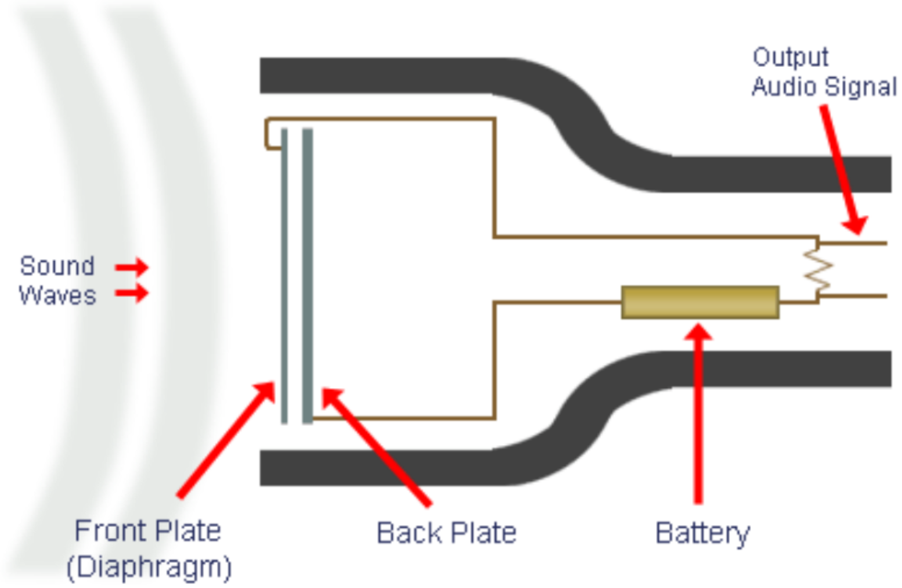


Figure 2.4: Diagram of a condenser microphone from [14].

with a lower sensitivity, and the sensitivity of the microphone is also dependent on the angle of incidence at which the sound wave hits the diaphragm. The direction dependent sensitivity of a microphone is known as the polar pattern of the microphone, and this direction dependent sensitivity is tuned via the mechanical design of the microphone. A microphone in which the rear of the diaphragm is baffled or insulated from external sound waves would be equally sensitive to sound waves from all directions, and would be described as having an omnidirectional polar pattern. A microphone in which the rear of the diaphragm is completely open to external sound waves is equally sensitive to sound waves that travel in a direction perpendicular to the plane of the diaphragm, and would be least sensitive to sounds that travel in a direction parallel to the plane of the diaphragm. This because the force of the sound waves acting upon each side of the diaphragm are unbalanced in the case that sound waves are travelling perpendicular to the plane of the microphone, and are balanced in the case that sound waves are travelling parallel to the plane of the microphone. Sound waves that are travelling at angles between parallel and perpendicular to the plane of the diaphragm will result in a

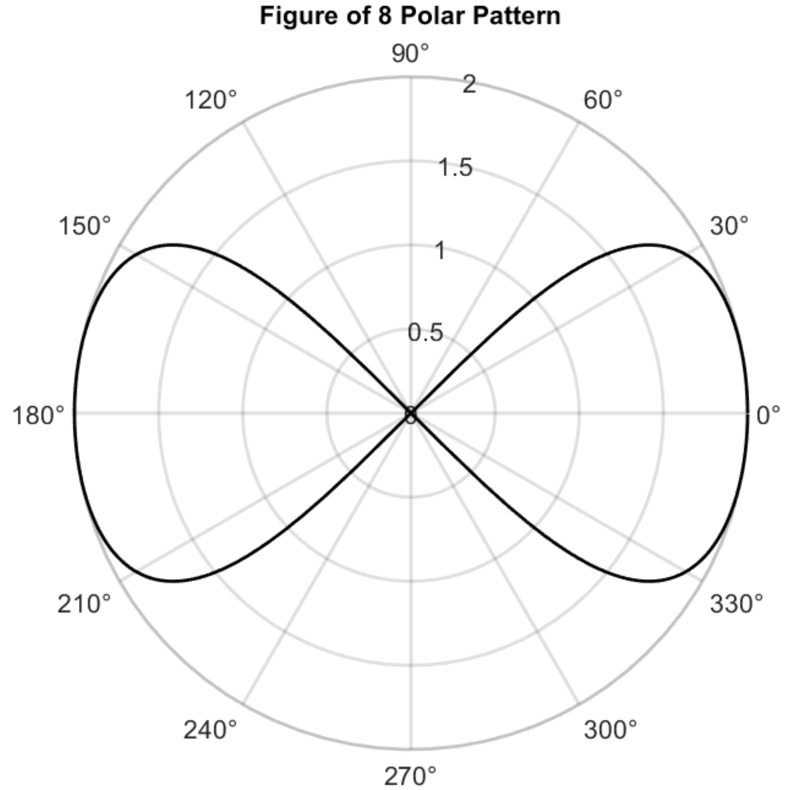
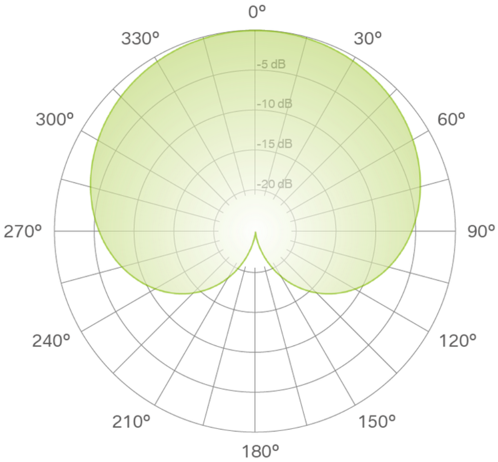
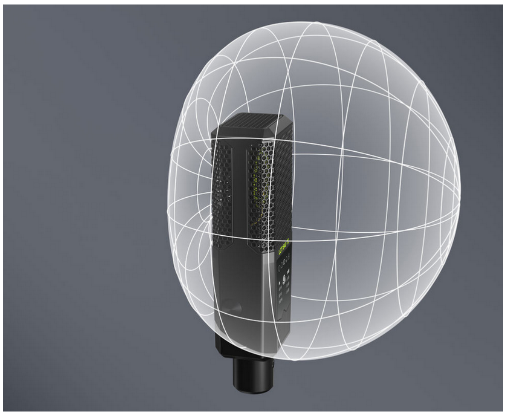


Figure 2.5: Polar plot of a figure-of-eight polar pattern.

force that is partially balanced, and the sensitivity of the microphone will be dependent on the angle of incidence of the sound wave. This behaviour results in a polar pattern that is shaped like a figure-of-eight, and an example of this polar pattern is presented in Figure 2.5. In Figure 2.5 the angle represents the angle of incidence of the sound wave relative to the plane of the diaphragm, and the radius represents the sensitivity of the microphone at that angle. However, the mechanism required to allow sound to reach the diaphragm from the side of the capsule with the plate requires porting in the assembly and the backplate. This porting can be tuned to change the polar pattern of the microphone to be more sensitive to sounds coming from the diaphragm side of the assembly, resulting in a cardioid polar pattern. This direction dependent sensitivity to sound is an important characteristic of microphones is important in the context of



(a) The polar plot of a cardioid polar pattern from [15].



(b) A three-dimensional cardioid polar response superimposed onto a microphone body from [15].

Figure 2.6: Two diagrams of a cardioid polar response from [15].

recording and measuring sound, as it allows the microphone to be positioned to record sounds from a particular direction and reject sounds from other directions. Figure 2.6 presents two illustrations of the cardioid polar pattern of a microphone from [15]. The cardioid polar pattern is named after the shape of the pattern, which is similar to a heart. Cardioid microphones are often popular because they allow the microphone to be positioned to record sounds from a particular direction and reject sounds from other directions. The plot presented in Figure 2.6a illustrates the shape of a typical cardioid polar response, with the sensitivity of the microphone being highest at the front of the microphone and lowest at the back. The illustration in Figure 2.6b shows the same cardioid polar response as Figure 2.6a but is rendered as a three-dimensional balloon that is superimposed on top of a microphone body, with the front of the microphone being perpendicular to the plane of the diaphragm. This balloon is intended to represent the three-dimensional nature of the polar response of the microphone, and the surface of the balloon represents the sensitivity of the microphone at different angles of incidence by the distance of the balloon to the centre of the diaphragm of the microphone. This figure has no units, and is intended to be a qualitative representation of the polar response of the microphone.

2.4 The Auditory Pathway

Sound waves are perceived by people through a process of transduction, passing through the mechanisms of the ear and ascending the auditory pathway, eventually being subconsciously processed and consciously understood. Figure 2.7 presents a cross sectional diagram of the anatomy of the ear. Acoustic waves first reach the pinnae and the su-

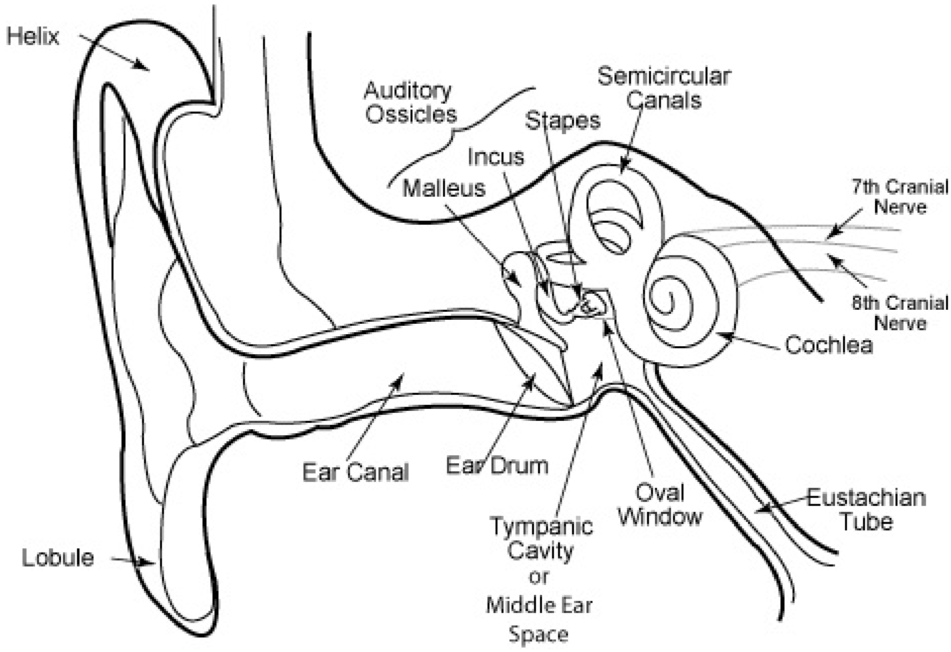


Figure 2.7: Diagram of the anatomy of the human ear from [16].

perficial structure of the ear, travelling down the ear canal and actuating the tympanic membrane that is also known as the ear drum. The reflections of sound waves off the superficial surfaces of the outer ear and the resonance of the ear canal influence the impression of the sound [17]. The tympanic membrane is attached at one end to a mechanism of three bones known as the ossicles. The ossicles are also connected to the oval window, a section on the surface of the fluid filled sack known as the cochlea in the inner ear. Inside the cochlea is the basilar membrane that supports an array of sets of hair cells, including one set of inner hair cells and up to five sets of outer hair cells. Each hair cell supports several hair-like nerves known as the stereocilia. The length and stiffness of the stereocilia are tuned such that the hair cells are stimulated on a frequency dependent basis, with hair cells nearer the the oval window resonating to high frequency vibrations, and hair cells towards the end of the basilar member vibrating sympathetically to low frequency vibrations. Because of this frequency dependent translation of sound energy, sounds are processed by the brain on a tonotopic basis (ordered by frequency), and the amplitude of a sound wave influences the amount of hair cells that are stimulated [18]. Figure 2.8 presents an illustration of the tonotopy or frequency dependent mapping of the basilar membrane. In this diagram the basilar membrane is shown as if it has been unrolled and flattened, with the oval window at the left hand side of the expanded cochlea and the apex of the cochlea at the right hand side of the diagram.

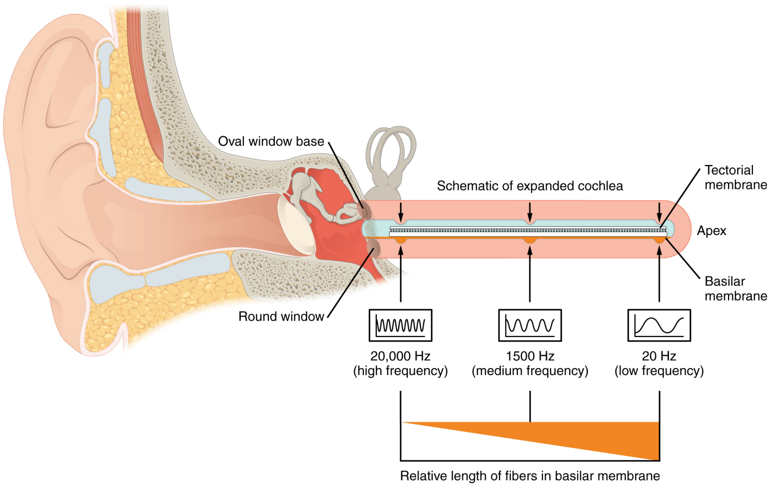


Figure 2.8: A diagram of the tonotopic nature of the basilar membrane, from [19].

The stereocilia feature mechanically gated ion channels that are opened and closed as the nerve resonates, allowing the transit of ions that depolarize the cells and trigger spikes in the electrical potential of the nerve that ascend the auditory pathway [20]. These spikes are the rapid electrical depolarizations that occurs when sodium is allowed to flood the nerve, increasing the electrical potential of the nerve from -55mV to $+40$mV The spiking hair cells project to the Spiral Ganglion, which then projects up the auditory pathway through the vestibulocochlear nerve. The auditory pathway consists of two prominent sections, the (primary) lemniscal pathway and the non-lemniscal pathway.

The lemniscal pathway leads to the primary auditory cortex, and the non-lemniscal pathway diverges towards other parts of the brain such as the auditory thalamus and the amygdala [21]. The auditory cortex is the section of the brain that deals with the primary audio processing tasks such as classification and localisation, and it is split into the core, belt and parabelt sections. The auditory cortex is interconnected with other areas such as the cerebral cortex and the thalamus. The thalamus is the primary connection centre between the cerebral cortex and other centre of the brain, and the cerebral cortex is the outer section of the brain that is involved in higher level processes such as reasoning, thought, emotion and consciousness. The combination of processes and associations of the auditory cortex and other parts of the brain allow the listener to actively perceive and place a sounds within the conscious frame of reference. The neural coding of interconnected spiking neurones provides the physiological basis for sound perception. Figure 2.9 shows a diagram of the human ascending auditory pathway, from [19].

The diagram presented in Figure 2.9 shows a simplified version of the human ascending auditory pathway, and only one of the two pathways from one of the two ears is presented. The ascending auditory pathway as presented in Figure 2.9 integrates with several parts of the brain that are involved multiple neural systems including the limbic system which is involved in behavioural and emotional responses and the regulation of the autonomic nervous system.

2.5 Auditory Perception

The auditory system is a complex system that is capable of processing a wide range of sounds, and the perception of sound is a complex process that is not fully understood. Psychoacoustics is the study of the perception of sound, and is a branch of psychophysics that is concerned with the features of the sensation of hearing [22]. Though there are several psychoacoustic phenomena that involved in the perception of the acoustic environment, four phenomena that have been studied in the context of soundscape assessments

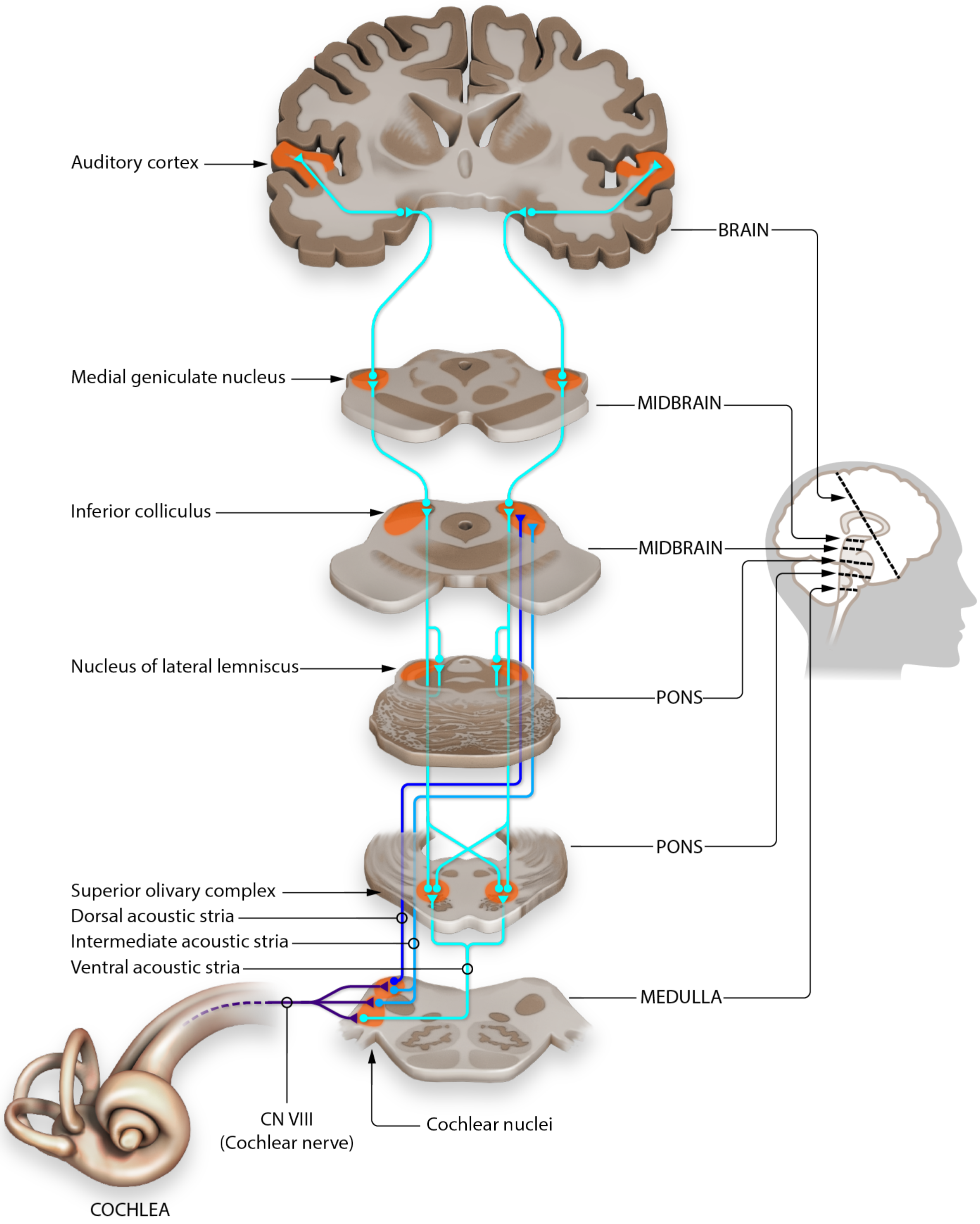


Figure 2.9: Diagram of the human ascending auditory pathway from [19].

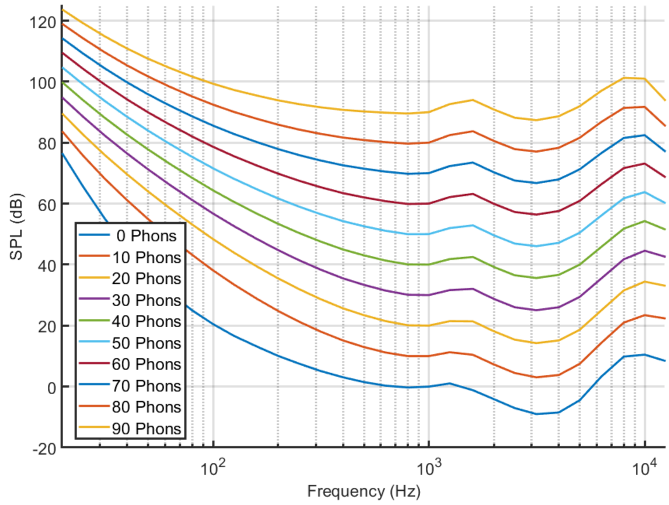


Figure 2.10: ISO226 Equal Loudness Contours, calculated using [24].

are introduced below [23].

2.5.1 Loudness

One psychoacoustic phenomenon that is important in the perception of soundscapes is perception of loudness and the frequency dependent sensitivity of the human ear. Loudness is the perception of the intensity of a sound, and the loudness of a sound is both a function of the sound pressure level and the frequency of the sound. The unit of measure for Loudness is the Phon which is a subjectively driven unit based on the the perception of loudness of a tone at a given frequency, relative to the perceived loudness of a 1kHz tone at 40dB_{SPL}. The frequency and intensity dependent sensitivity of the human ear is summarised by equal-loudness contours, and is shown in Figure 2.10. In Figure 2.10 each curve represents the perceived loudness level of a pure tone at each frequency that is perceived to be equally loud to a 1kHz pure tone at a given sound pressure level. The equal loudness contours in Figure 2.10 show that the human ear is most sensitive to sounds in the frequency range between 2kHz and kHz, and that the sensitivity of the human ear is dependent on the sound pressure level. This characteristics of hearing was pioneered by Fletcher and Munson in 1933 [25], and refined by Robinson and Dadson in

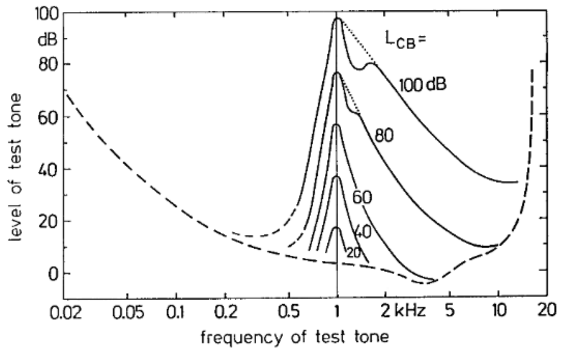


Figure 2.11: The amplitude of a pure tone that is masked by critical-band-wide noise which was centred at 1kHz, from [28].

1956 [26], eventually being standardised as ISO 226 [27].

2.5.2 Masking

Although the perceived loudness of a sound is a function of both sound pressure level and frequency, the perception of multiple sounds at once can influence the perceived loudness of each individual sound. This psychoacoustic phenomenon is known as the masking of sounds, in which the presence of one sound can make another sound inaudible [17]. This phenomenon occurs when sounds excite overlapping region of the basilar membrane [28], and the region of potential overlap increases with the level of the sound. A diagram of the tonotopy of the basilar membrane is presented in Figure 2.8. Figure 2.11 presents the level of a pure tone that is masked by critical-band-wide noise, from [28]. A critical band described the relative bandwidth of audible frequencies on the basilar membrane that is excited by a sound with a given frequency [25]. In Figure 2.11 each of the functions represent the threshold of the level at which a pure tone with a given frequency was masked by critical-band-wide noise that was centred at 1kHz, and each function represents the level if the critical-band-wide noise. Figure 2.11 shows the effective bandwidth in which a tone can be masked by band-limited noise increases with the level of the masking noise, and the shape of this effect increases at a greater rate for frequencies above the centre frequency of the masking noise than below.

The effects of masking requires that a mathematical model of loudness includes critical bands, and several models for calculating the loudness of sounds including critical bands exist including the Zwicker and Moore & Glassberg models [28, 29]. Loudness is computed using these models by aggregating and weighting sound levels over critical bands of the frequency spectrum, although different models of loudness may use different weighting functions and frequency bands. The ISO standardized Zwicker method was identified as the most commonly used method for the measurement of loudness in a study of the use of psychoacoustic parameters in soundscape assessment [23].

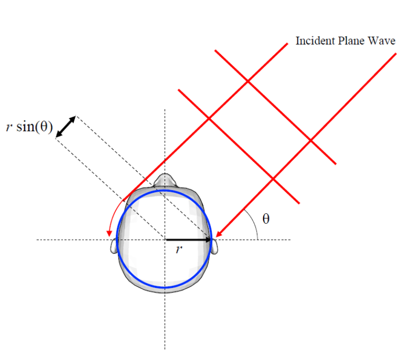
2.5.3 Sharpness, Roughness & Fluctuation Strength

Sharpness, roughness and fluctuation strength are three further psychoacoustic phenomena that have been investigated in the context of soundscape quality assessment [23]. Sharpness is a psychoacoustic measure that reflects the timbre and spectral balance of a signal, which Zwicker identifies as the inverse to pleasantness [28] Sharpness reflects that the perception of narrow band sounds is based on the loudness of that sound, which becomes increasingly sensitive with increasing frequency [28]. Several models for the calculation of sharpness exist, and the Aures model was found to be most used [30], followed by the DIN 45692 model [31] and the Fastl and Zwicker model [28].

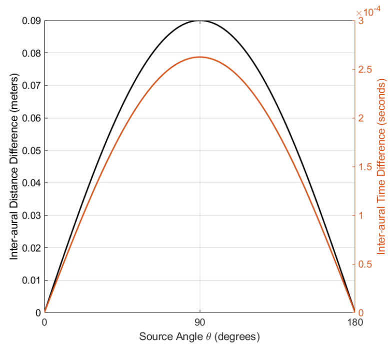
Roughness and Fluctuation Strength are two measures of the perception of the amplitude modulation of a signal [28]. Both Roughness and Fluctuation Strength have an intensity characteristic that is weaker at higher and lower relative frequencies, with a peak at 4Hz modulation rate for Fluctuation Strength and a peak 70Hz modulation rate for Roughness for a 1Hz tone. Roughness is quantified in units of asper, and one asper is the roughness produced by a 60dB 1000Hz tone which is 100% amplitude modulated at 70Hz. Fluctuation strength is quantified in units of vacil, and one vacil is the fluctuation strength produced by a 60dB 1kHz tone which is 100% amplitude modulated at 4Hz. The review found that the Sottek model was most commonly used for the calculation of Roughness and Fluctuation Strength [32], followed by the Daniel and Weber model [33].

2.6 Binaural Hearing

The block diagram presented in Figure 2.9 highlights that the auditory cortex has two mirrored paths that interconnect, with each pathway on each side of the brain processing information from both cochlea. The discrimination of combinations of differences in signals from each ear including time differences and level differences allow humans to localise the direction of a sound source. The localisation of sounds based on these binaural cues of inter-aural time differences (ITD) and inter-aural level differences (ILD) is known as the duplex theory of sound localization which was pioneered by Lord Rayleigh [34, 35].Localization is the ability of the auditory system to determine the direction of a sound source, and is an important aspect of the perception of sound. Figure 2.12 presents a diagram and a plot illustrating how ITDs change with source angle. Sub-Figure a of



(a) Diagram illustrating binaural localization cues from [34].



(b) Plot of Inter-aural distance and time differences as for [34].

Figure 2.12: Figure and plot illustrating the effect of source angle on inter-aural time difference.

Figure 2.12 presents an illustration of the difference in distance a plane wave must travel to reach both ears, when travelling towards a listener at an angle that is non-coincident with the direction the listener is looking in. The illustration presented in Sub-Figure a shows that the difference in distance that a plane wave must travel to reach both ears is a function of the angle (θ) between the direction of travel of the plane wave and the direction the listener is looking in, and the radius of the head (r). The increased distance that the plane wave must travel to one ear relative to the other results in differences in

ITD and ILD cues that are received by the listener. Sub-Figure b of Figure 2.12 presents a plot of the estimated time and distance difference for the plane wave reaching both ears, plotted as a function of the angle between the direction of travel of the plane wave and the direction the listener is looking in. This plot assumes that the plane wave is travelling at $343\mathrm{m\,s^{-1}}$, and that the radius of the head is 0.09m. The maximum time delay between the plane wave reaching both ears is when the relative angle is $90°$, the angle at which the plane wave would be travelling directly towards the side of the listeners head. Inter-aural time differences (also known as inter-aural phase differences) are the difference in time between a sound arriving at each ear due to the relative direction and distance of the sound source to the position and look direction of the listener. ITDs are important for localising sounds with wavelengths that are larger than the geometry of the head and hearing system, as these waves diffract around the head and reach both ears. Inter-aural level differences refer to the differing level of a sound arriving at each ear due to the shadowing of the head, the lack of diffraction of higher frequency sounds instead blocking and reflecting sound waves. This causes the sound to be louder in one ear than the other, and the brain can use this difference in loudness to calculate the angle of the sound source relative to the head. There is a crossover range in which neither ITDs or ILDs provide reliable cues for sound localisation, between 800Hz and 1.6kHz [35]. Further, the use of ITDs and ILDs for sound localisation beyond the azimuth plain is geometrically limited as similar ITDs and ILDs can be produced by sounds sources at several different positions with similar respective angles and distances to the listener. The cone of confusion is a conceptual representation of the positions of greatest ambiguity in localisation cues due to the geometric symmetry of the head and ears, and is illustrated in Figure 2.13. The cone of confusion extends outwards along the median plane, and sound sources that are located on the surface of the cone cannot be accurately localised using ITDs and ILDs alone. In Figure 2.13 ITDs and ILDs at points A and B are identical, and the listener cannot distinguish between these two positions using ITDs and ILDs alone. Monaural cues, cues that are not dependent on two ears, are used to further assist sound source localisation. These cues include the effects of caused by the

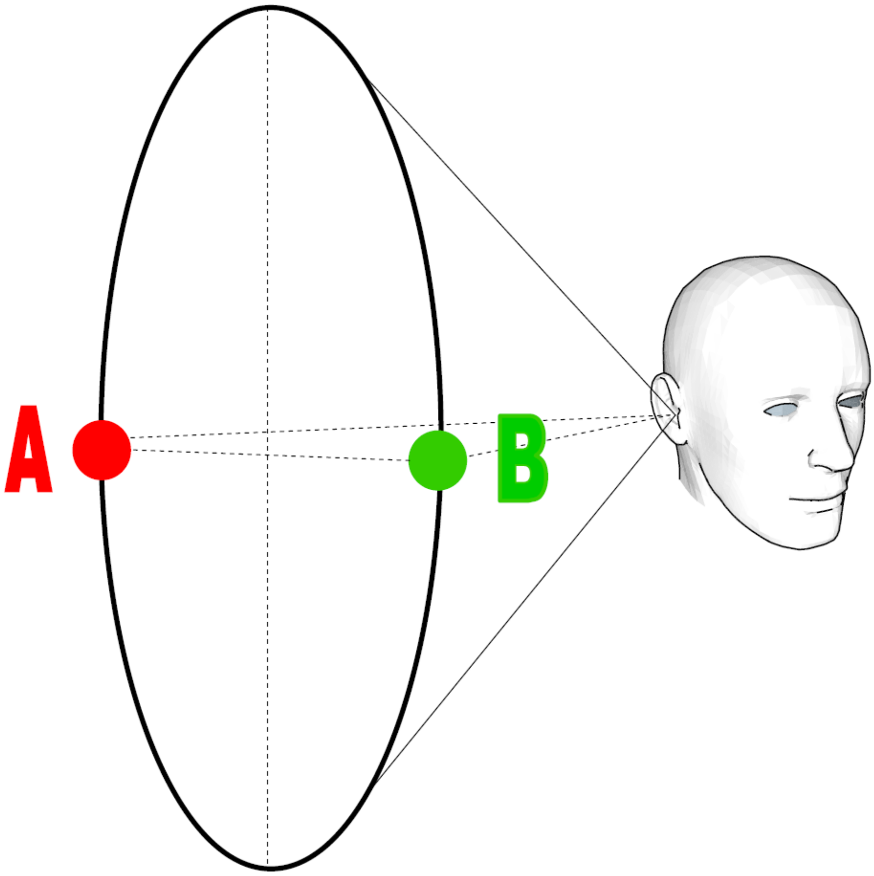


Figure 2.13: Diagram illustrating the cone of confusion representing uncertainty in binaural localisation cues, from [34].

interaction of sound waves with the head, torso, and pinnae, absorbing and reflecting sound resulting in changes to the frequency spectrum of the sound that reaches the inner ear. Spectral cues contribute to the accuracy of sound source localization in the azimuth plane, and are critical for localisation above and below the horizontal plane as well as behind and in front of the listener. Further, spectral cues from the pinnae are important in both sound localisation and maintaining the impression of sound originating from outside of the head [36, 35]. The effects of the path of travel from sound source to inner ear on sound waves can be thought of as a filter[1], and the transfer function of this filter

[1] For an introduction to filters, see [37]

is called the head related transfer function (HRTF) [34]. The HRTF is a function of the angle and distance of the sound source relative to the listener and the geometry of the head, torso, and pinnae. HRTFs can be used to simulate the effect of positioning a sound source in three dimensional space external to the listener, and are used in binaural audio systems and the spatial reproduction of sound-fields to create the impression of sound originating from outside of the head [38]. However, the HRTF is unique to each individual and sound source position, and the creation of generalized HRTFs that give an adequate perception of sound quality, externalization and sound source localization are still an open problem [39]. Binaural and monaural cues are static, remaining relatively constant over time. However, the position of a sound source relative to a listener is rarely static in reality, and the movement of the head and body is often used to further improve the accuracy of sound source localization. Moving the head and altering the HRTF between a sound source and a listener is a key strategy for sound source localization by humans with only one ear [36].

An important binaural psychoacoustic phenomenon is the precedence effect which describes the influence of the time and amplitude of reflected sounds on the localization of the sound source [40]. The precedence effect identifies that the localization of a sound source is dominated by the first arriving sound, and this phenomenon allows for accurate sound source localization to be maintained in highly reverberant environments.

2.7 Spatial Sound Reproduction & Recording

The recording and reproduction of sound-fields with spatial information has seen continual development with improvements in the quality of sound recording and reproduction. The breadth of sound recording and playback techniques is vast, and this section will focus on the techniques that are relevant to the research presented in this thesis. All of the experiments presented in this thesis involved the reproduction of sound over headphones, and discussion of the reproduction of sound over loudspeakers is not included.

Headphones are electrodynamic transducers that convert an electrical signal into the mechanical actuation of a diaphragm, which in turn causes the air to vibrate and produce sound waves. Two transducers are usually packaged into a chassis including a headband and earcups, and when placed over the ears of the listener are refereed to as circumaural headphones [41]. Supra-aural headphones are similar to circumaural headphones, but the earcups are smaller and rest on the ears of the listener. Earcups may typically include a soft padding material where the cup meets the head and ears of the listener, and this padding material is often designed to provide some comfort and isolation from external sound sources. Miniaturized transducers can be packaged into earbuds that are placed directly into the ear canal of the listener, using a rubber moulding to form a tight seal between the earbud and the ear canal. Off ear headphones have been developed which do not form a seal between the earcups and the ears of the listener, instead the transducers are held off of the ears of the listener. Finally, bone conduction headphones rely on the transfer of vibrations through the bones and soft tissue of the skull to the inner ear of the listener [42].

2.8 Monophonic, Stereophonic & Binaural Sound Reproduction Over Headphones

The most simple form of sound reproduction in the headphone context is monophonic, with each transducer of the headphones reproducing the same signal. This signal is made with one or several signals that are mixed together into a single signal or channel that contains intensity information and no directional information or spatial cues. In Stereophonic sound reproduction each transducer of the headphones reproduces a different signal, stereo sound typically having two channels representing left and right. Stereophony was pioneered by Alan Blumlein [43, 44], who developed several concepts for the recording and reproduction of sound-fields using two microphones and two loudspeakers respectively. One example is of these innovations is the Blumlein pair, a stereo recording technique that uses two microphones placed in a coincident pair to capture a stereo image of a sound-field [45]. Figure 2.14 shows an illustration of the arrangement

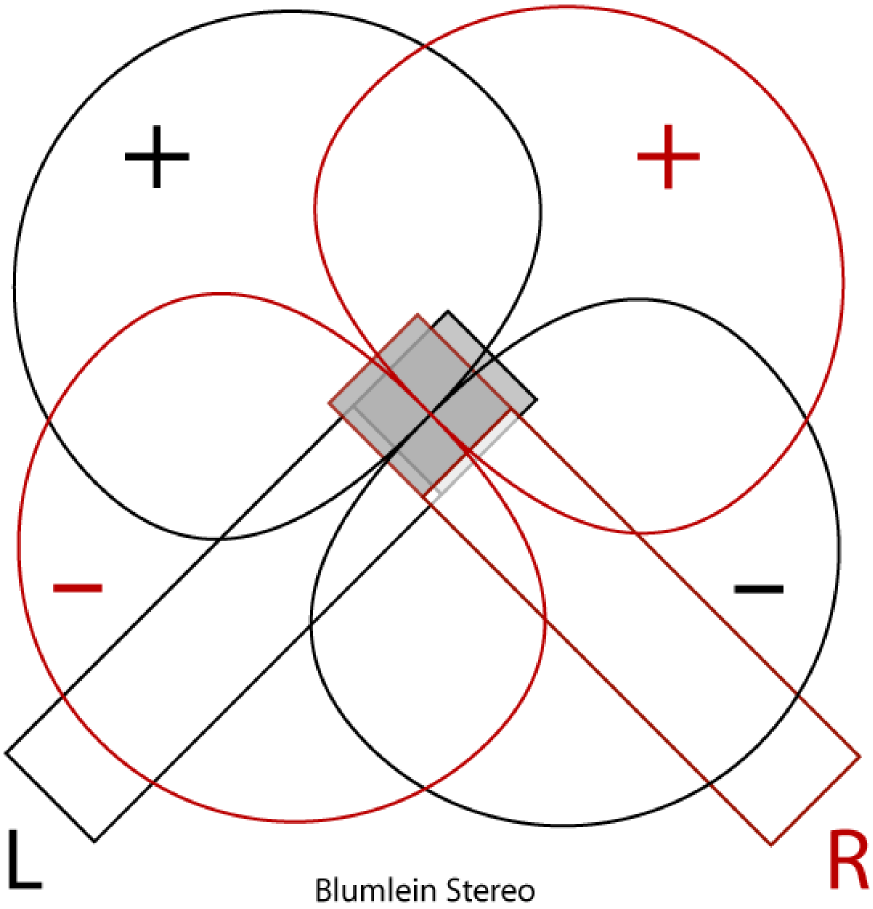


Figure 2.14: Illustration of the arrangement of two microphones into a Blumlein pair from [46].

of two microphones into a Blumlein pair from [46]. The illustration in Figure 2.14 shows two microphones with figure-of-eight polar patterns placed at a 90 degree angle to each other, with the capsules of the microphones placed one above the other. This method takes advantage of the spatial sensitivity characteristic of each microphone to capture a stereo image of the sound-field, as the position of each sound source being recorded will influence the amplitude of the sounds source that is recorded. When reproduced over loudspeakers or headphones, the two signals captures by the microphones will be reproduced as a sound-field that maintains an impression of spatial information in the form of a stereo image [47].

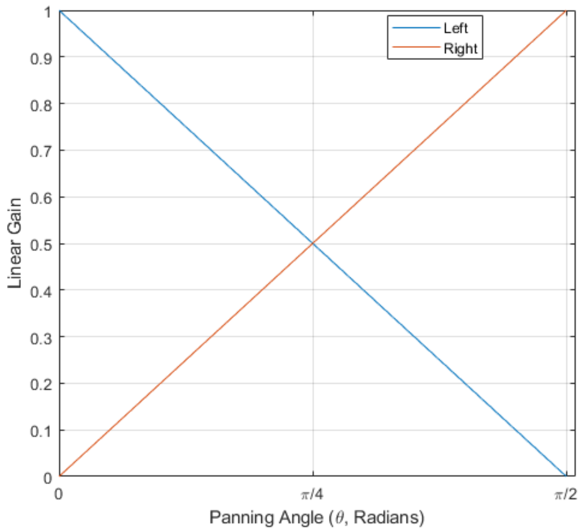


Figure 2.15: A plot of the linear gain functions for stereo panning from Equation 2.6.

Monophonic sounds can be encoded into a stereo signal by taking advantage of the level difference characteristic of stereo recording. This encoding can be achieved by taking a monophonic signal and duplicating it into two channels, and then applying a gain to each channel to simulate the level difference that would be inherent in a stereo recording [48]. A pan or balance law is often used to determine the gain that should be applied to each channel of the signal. Equation 2.6 shows a formula for calculating the left and right gains of a linear pan law from [48].

$$\begin{aligned} L(\theta) &= \left(\frac{\pi}{2} - \theta\right)\left(\frac{\pi}{2}\right)^{-1} \\ R(\theta) &= \theta\left(\frac{\pi}{2}\right)^{-1} \end{aligned} \tag{2.6}$$

In Equation 2.6 L and R are the stereo channels, and θ is the angle of the pan. As the panning angle θ increases from 0 to $\frac{\pi}{2}$, the gain applied to the left channel L decreases from 1 to 0, and the gain applied to the right channel R increases from 0 to 1. The resulting gain functions from Equation 2.6 are presented in Figure 2.15. The functions in Figure 2.15 show a linear increase and decrease in the left and right gains respectively as the panning angle progresses from 0 to $\frac{\pi}{2}$, with the sum of the two gains always being equal to 1. The gains calculated with Equation 2.6 can be used to determine the gain that should be applied to each channel of a stereo signal to simulate the level difference that would be inherent in a stereo recording. One drawback of this panning law is that despite having balanced gains, the perceived change in loudness of the signal is not linear with the angle of the pan. The presentation of this panning law is included to illustrate the most simple encoding of a signal with spatial information into a signal with two channels. There are several other pan laws that can be used to determine the gains that should be applied to each channel of a stereo signal, and these pan laws are discussed in [48]. Stereo recordings may be the most common form of spatial sound reproduction, but they do not include the spatial cues described in Section 2.6 and thus don't inherently give the impression that the sound-field is external to the listeners head when experienced over headphones [49].

Binaural recording is one approach to capturing a sound-field with spatial information that can be reproduced over headphones. Binaural recording can refer to the recording techniques pioneered by Blumlein [44] who positioned microphones at either side of a block of wood to represent the head shadowing effect on high frequency sound captured by each microphone. Modern binaural recordings are typically made using a dummy head that includes microphones that have been moulded into the ears of the head, or using a pair of binaural microphones that are placed in the ears of a listener. A typical dummy head used in binaural recordings is the Neumann KU 100 [50] that shown in Figure 2.16. The Neumann KU 100 presented in Figure 2.16 is a dummy head that features a replica of a human head that includes a pair of anatomically inspired artificial ears and a pair of condenser microphones that are coupled into the ears of the head [51]. When recording with a dummy head, the HRTFs between the dummy head and the sound sources are encoded within the recording of each channel, providing the necessary cues for externalisation and sound source localization when the recording is reproduced over headphones. Binaural recordings are most easily reproduced over headphones, though it is possible to reproduce binaural recordings over loudspeakers if

Figure 2.16: Neumann KU100 from [50].

one of several special techniques are used to ensure that the spatial cues are not lost [52, 34]. One drawback of this technique of spatial sound encode, recording binaural sound, and decoding, reproducing binaural sound, is that the method of encoding is fixed. Because the orientation of the binaural cues is fixed within the recording, listeners are unable to use head movement to improve spatial cues when listening to binaural recordings. Further, the HRTFs that are used in the recording are not matched to the listener, and may lead to a degraded externalization and sound source localization [39].

2.9 First Order Ambisonics

Ambisonics is another strategy for encoding spatial information, decomposing a soundfield into a signal that includes several channels. Ambisonics was pioneered by Michael Gerzon [53, 54], and is reputed as being a logical extension of the work of Blumlein [52, 34]. Ambisonics is based on the idea of the spherical harmonic decomposition of a soundfield into a pressure field and several acoustical velocity vectors. Spherical harmonics are a set of special functions that are defined on the surface of a sphere [55]. Each order of spherical harmonic function represents a different spatial frequency, and the functions are orthogonal to each other. Spherical harmonics can be combined to represent any

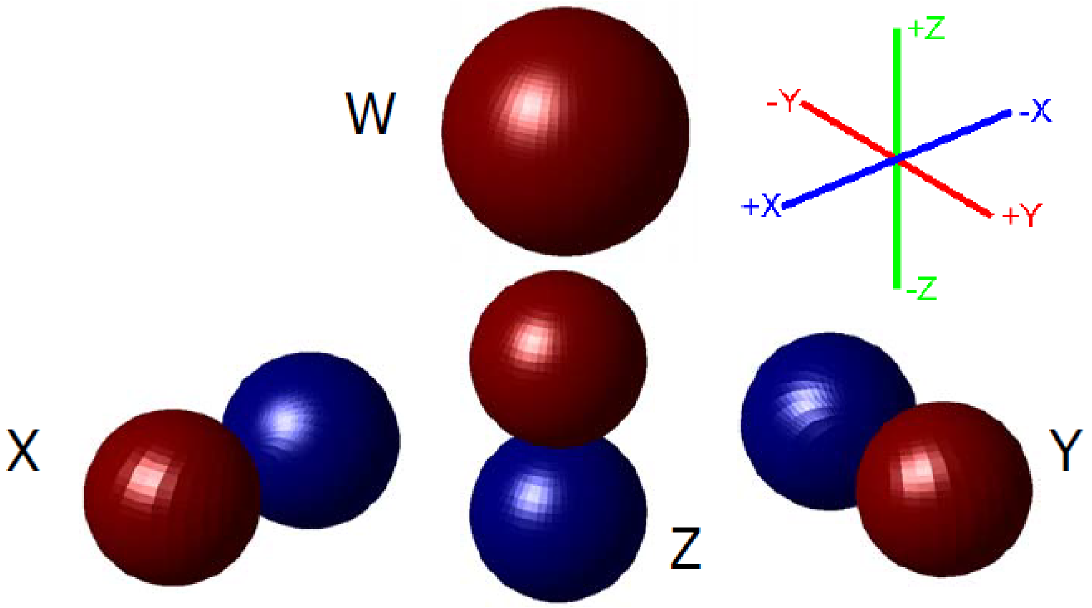


Figure 2.17: First-order spherical harmonics from [52].

function on the surface of a sphere, and the combination of higher order harmonics allows for the creation of highly complex three dimensional functions.

The work described in this thesis only utilized 1$^{\text{st}}$ order Ambisonics and this will be the focus of the following section. For a more in depth discussion of higher order Ambisonics see [56, 52, 34]. Figure 2.17 shows the 0$^{\text{th}}$ and three harmonics of the 1$^{\text{st}}$ order of spherical harmonics from [52]. The first four spherical harmonics illustrated in Figure 2.17 are referred to as the W, X, Y and Z channels in the context of Ambisonics, and are used as gains functions to encode a sound-field into a set of four 1$^{\text{st}}$ order Ambisonic channels. This configuration of ambisonic channels is referred to as *B-format*. The 0$^{\text{th}}$ order or W^{th} channel is an omnidirectional signal that encodes the intensity of the sound-field. The subsequent X, Y and Z channels encode the velocity component of the sound-field in each of the principal orthogonal directions in three-dimensional space. Each of the X, Y and Z channels appear as functions that have similar characteristics to a figure-of-eight polar pattern as described in Section 2.3. The formula for computing the angular gain function for each of the 0$^{\text{th}}$ and 1$^{\text{st}}$ order Ambisonic channels are shown in Equation 2.7, from [52].

$$
\begin{aligned}
W &= \frac{1}{\sqrt{2}} \\
X &= \cos(\theta)\cos(\alpha) \\
Y &= \sin(\theta)\cos(\alpha) \\
Z &= \sin(\alpha)
\end{aligned}
\tag{2.7}
$$

Where:

- α is the elevation angle of the source.
- θ is the azimuth angle of the source.

The velocity of sound waves propagating through a sound-field can be expressed as functions of the velocity vectors encoded in the X, Y and Z channels, and the intensity of the sound-field can be expressed as a function of the W channel. This strategy of encoding a sound-field into Ambisonic channels decouples the recording of the sound-field from both the recording and the reproduction methods, unlike stereo or multichannel sound formats that are arranged for specific surround sound loudspeaker configurations which require a specific channel for each speaker [57].

Using the formulas described in Equation 2.7 it is possible to encode monophonic signals into a 1^{st} order Ambisonic sound-field by calculating the gain of each Ambisonic channel and applying the gain to one of four copies of the monophonic signal, one for each Ambisonic channel. This newly encoded Ambisonic signal can be mixed with other Ambisonic signals, though a process of normalization [58] and re-ordering may be required to ensure the signal meets the specification of a particular Ambisonic format such as the Ambix format [59].

Another benefit of the Ambisonic format is that it is possible to rotate the relative orientation of a sound-field by applying a rotation matrix to the Ambisonic channels. A rotation matrix for 1^{st} order ambisonic signals can be calculated using the formulae in

Equations 2.8a to 2.8h from [52].

$$W^{'} = W \tag{2.8a}$$

$$X^{'} = X\cos(\theta) + Y\sin(\theta) \tag{2.8b}$$

$$Y^{'} = Y\cos(\theta) - X\sin(\theta) \tag{2.8c}$$

$$Z^{'} = Z \tag{2.8d}$$

$$W^{'} = W \tag{2.8e}$$

$$X^{'} = X \tag{2.8f}$$

$$Y^{'} = Y\cos(\theta) - Z\sin(\theta) \tag{2.8g}$$

$$Z^{'} = Z\cos(\theta) + Y\sin(\theta) \tag{2.8h}$$

Where:

- θ is the angle of rotation in the given axis.

Equations 2.8a, 2.8b, 2.8c, and 2.8d describe a rotation about the Z axis, and Equations 2.8e, 2.8f, 2.8g, and 2.8h describe a rotation about the X axis. These rotations are equivalent to panning, turning your head to the left or right, and tilt, elevating and lowering your head. A rotation matrix calculated using one of the formulae in Equations 2.8a to 2.8h will apply a weighted mix of the X, Y and Z channels, resulting in a change in the relative orientation of the sound-field.

Ambisonic channels can be decoded using one of several strategies to reproduce the sound-field over a loudspeaker array or to create binaural signals for reproduction over headphones. There are several different approaches for decoding Ambisonic signals that attempt to optimize the reproduction of the sound-field for loudspeaker arrays of different sizes, shapes and degrees of regularity, see [54, 52, 34, 60, 61].

The rendering of an Ambisonic recording into a two channel binaural recording can be achieved by first decoding the Ambisonic channels into a set of virtual loudspeaker signals, and then convolving each of these virtual loudspeaker signals with an appropriate HRTF [52]. The decoding of Ambisonic signals into virtual loudspeaker signals is achieved by applying a set of gains to each Ambisonic channel, and the gains applied to each channel are determined by the decoding strategy. Each decoding strategy attempts to ensure that the pressure field and velocity vectors produced by each loudspeaker sum correctly, but this not always possible due to the non-ideal nature of loudspeakers and the physical constraints of the loudspeaker array. However, virtual loudspeakers are not as constrained and the rendering of binaural signals from Ambisonic recordings can be achieved with any number of virtual loudspeakers using as few as three pairs of HRTFs as described by [52]. Each Ambisonic channel can be decoded by scaling each channel with the static gain values that represent the relative position of each virtual loudspeaker, resulting in one decoded channel for each virtual loudspeaker [62]. The anechoic HRTF for each virtual loudspeaker position can then be applied to each virtual loudspeaker channel for each ear. The computational complexity of this approach is proportional to the number of virtual loudspeakers, and though a high resolution of virtual loudspeakers is desirable for better spatial accuracy, this comes at increased computational cost. This process can be simplified as described in [52] and [34] by summing the HRTFs for each ear, Ambisonic channel and virtual loudspeaker such as in Equations 2.9a to 2.9d.

$$W^{\mathrm{hrtf}} = \sqrt{2}\sum_{k=1}^{N}(S_k^{\mathrm{hrtf}}) \tag{2.9a}$$

$$X^{\mathrm{hrtf}} = \sum_{k=1}^{N}(\cos(\theta)\sin(\phi)S_k^{\mathrm{hrtf}}) \tag{2.9b}$$

$$Y^{\mathrm{hrtf}} = \sum_{k=1}^{N}(\sin(\theta)\sin(\phi)S_k^{\mathrm{hrtf}}) \tag{2.9c}$$

$$Z^{\mathrm{hrtf}} = \sum_{k=1}^{N}(\cos(\phi)S_k^{\mathrm{hrtf}}) \tag{2.9d}$$

Where:

- N is the number of virtual loudspeakers.
- θ is the source azimuth.
- ϕ is the source elevation.

- S_k^{hrtf} is the HRTF for the k^{th} virtual loudspeaker.

Using equations Equations 2.9a to 2.9d results in two sets of four Ambisonic filters, one set for each channel of the binaural signal. Each channel of the binaural signal can then be rendered by convolving the Ambisonic channels of the recording with each set of filters as presented in Equations 2.10a and 2.10b:

$$\text{Left} = (W \circledast W_{\text{left}}^{\text{hrtf}}) + (X \circledast X_{\text{left}}^{\text{hrtf}}) + (Y \circledast Y_{\text{left}}^{\text{hrtf}}) + (Z \circledast Z_{\text{left}}^{\text{hrtf}}) \tag{2.10a}$$

$$\text{Right} = (W \circledast W_{\text{right}}^{\text{hrtf}}) + (X \circledast X_{\text{right}}^{\text{hrtf}}) + (Y \circledast Y_{\text{right}}^{\text{hrtf}}) + (Z \circledast Z_{\text{right}}^{\text{hrtf}}) \tag{2.10b}$$

This results in a two channel binaural signal that can be reproduced over headphones to create the impression of a sound-field external to the listener. There are several further optimizations that can be made to the process of decoding Ambisonic signals into binaural signals, such as synthesizing binaural cues to improve the performance of a particular decoding strategy [63], or optimizing the amplitude of the virtual loudspeakers in the decoding process to more closely reflect inter-aural level differences when using low order Ambisonic reproduction [64].

In order to decode Ambisonic signals into binaural signals as described above, it is necessary to have access to a set of HRTFs. One way to acquire HRTFs is to use a publicly available dataset such as the SADIE II dataset [65], a dataset including thousands of HRTFs from twenty subjects including the Neumann KU100 [50] and KEMAR [66] mannequins. This collection of HRTFs is widely used and has been adopted by Google and YouTube [67]. The procedure for measuring as described in [65] is difficult, tedious and time consuming, requiring several pieces of specialized equipment and several hours of time per subject. Several strategies have been developed for improving the availability and speed of HRTF measurement [68], and for improving the availability of personalised HRTF measurement [69].

Figure 2.18: The head of a Rode NT-SF1 SoundField Microphone from [70].

2.10 Recording First Order Ambisonics

A sound-field can be captured into a set of Ambisonic channels using a microphone array, though this may be practically challenging as an assumption of Ambisonics is that the pressure and velocity vectors are collocated and coincident. Positioning several microphones in the same location is not practical, and it is necessary to use a microphone placement strategy that allows for practical correction to be applied to each recorded signal. One solution for recording 1^{st} order Ambisonic sound-fields is to use a tetrahedral microphone array, an example of which is the Rode NT-SF1 SoundField Microphone, pictured in Figure 2.18. The tetrahedral microphone array was patented by Craven & Gerzon [71], and is often referred to as a SoundField microphone. The microphone array in Figure 2.18 consists of four condenser microphones, each with a cardioid polar pattern. Each microphone is positioned at the centre of a face of of a tetrahedron, resulting in sitting at opposing angles of azimuth and elevation. The advantage of this 1^{st} order Ambisonic microphone array design is that the positions of each microphone capsule are equally non-coincident, allowing for simplified correction of the non-coincident microphone positions [71]. The angle and elevation of each microphone in the tetrahedral array is presented in Table 2.1.

Capsule	Azimuth	elevation
A	$45°$	$35.3°$
B	$135°$	$-35.3°$
C	$-45°$	$-35.3°$
D	$-135°$	$35.3°$

Table 2.1: The orientation of each microphone capsule in a tetrahedral microphone array as for [71] and [52].

The Ambisonic channel configuration of the tetrahedral microphone is referred to as A-format, and the signals captured with each microphone in the array can be converted to B-format with formula presented in Equations 2.11a to 2.11d, from [52], that are adapted from [71]:

$$W = 0.5(A+B+C+D) \tag{2.11a}$$

$$X = (A+C)-(B+D) \tag{2.11b}$$

$$Y = (A+B)-(C+D) \tag{2.11c}$$

$$Z = (A+D)-(B+C) \tag{2.11d}$$

Where A, B, C and D are the signals captured by each microphone in the array and W, X, Y and Z are the Ambisonic channels. Due to the geometry of the array resulting in a phase offset between the capsules, and the polar pattern of each capsule being non-ideal, the signals captured by each microphone require a correction filter to be applied as discussed in the original patent [71]. However, as Wiggins describes, only the average correction filters can be applied to each signal [52]. At a frequency determined by the spacing of the microphone capsules the correction filtering required will be non-constant with angle, resulting in spatial aliasing. Despite this inconsistent discolouration of the recording, the SoundField microphone has several advantages over other microphone techniques, including consistent placement of the microphone capsules, the ability to generate any 1st order microphone pattern from the recorded signals [52], and the ability to arbitrarily change the orientation and zoom of the recorded signals [72]. In this instance, to zoom refers to the ability to change the effective polar pattern of the microphone array to a more directional pattern, allowing the listener to focus on a particular direction in the sound-field.

Summary

This chapter gives a simplified description of the pathway of sound from source to perception, describing the pathways that are involved in the processing of sound. This description is interwoven with a description of the principles of sound sensing, Ambisonics, and binaural audio. The description of all of the topics in this chapter are intended to give an overview of the core concepts as related to the work presented later in this thesis, and it is not intended to be a comprehensive description of any of these topics. Although the topics are described in a simplified manner, the description is sufficient to understand the work presented in this thesis. Further, there are several psychoacoustic principals that are relevant to the perception of sound that are not discussed in this chapter, including the perception of loudness, the perception of pitch and timbre, the perception of spatial attributes and spatial effects such as precedence and auditory scene analysis. Some of these will be briefly mentioned in following chapters.

Chapter 3

Environmental Noise & Soundscape

Chapter 2 introduced the fundamentals of sound, including the physical properties of sound waves and the human auditory system. The foundation of acoustics described in Chapter 2 underpins the principals that will be discussed in the rest of this thesis. Sound is an integral and almost inescapable part of our daily lives, and it has a significant impact on our health, well-being and ultimately our quality of life. At a larger scale, undesirable noise is another form of pollution, and it is one of the most pervasive forms of pollution in modern society. This chapter will build on the concepts discussed in Chapter 2 by introducing environmental noise and modern European environmental noise policy, followed by an introduction to soundscape and soundscape evaluation. This chapter will frame the form of evaluation used in the experimental work that is presented in the subsequent chapters of this thesis, and will further introduce the concepts of soundscape and soundscape evaluation in the context of environmental noise management.

3.1 Introduction

In order to discuss environmental noise, it is first necessary to define what is meant by the term *noise*. As with the epigraph in Chapter 1, Helmholtz described periodic sounds as music and aperiodic sounds as noise [73]. Although this quote is taken out of context, it is useful for illustrating that the definition of noise is subjective and context dependent. In the context of physics, electronics and signal processing, noise is defined as any error or undesirable components in a signal that are not part of the desired information [17, 12]. In the context of acoustics and environmental acoustics, noise is defined as any unwanted sound, and Everest suggests that tackling the unwanted sound from a Heating, Ventilation, and Air Conditioning system is simply a case of determining a criteria for what is considered to be unwanted and then engineering a solution to meet this criteria [17]. This philosophy of identification and mitigation as at the heart of environmental noise management, and it is the basis of the European Union's environmental noise policy.

Environmental noise is defined by the European Commission as *unwanted or harmful outdoor sound created by human activities, including noise emitted by means of transport, road traffic, rail traffic, air traffic, and from sites of industrial activity such as those defined in Annex I to Council Directive 96/61/EC of 24 September 1996 concerning integrated pollution prevention and control* [74]. This verbose and specifically targeted definition of environmental noise is sourced directly from the European ComMission Directive 2002/49/EC, commonly referred to as the Environmental Noise Directive (END) [74]. This directive is the basis of environmental noise policy in the European Union has been passed into the legislation of each member state [2]. More generally environmental noise has been described as the accumulation of noise pollution caused by transport, industrial and recreational activities [75].

[2] and an ex-member state

3.2 The Effects of Environmental Noise Exposure

Environmental noise pollution is highly prevalent in modern society, and is associated with a high socio-economic cost [76, 77]. Environmental noise pollution is considered by the European Commission's Directorate General for the Environment to be the second biggest environmental health threat in Europe after bad air quality [78]. A body of systematic literature reviews and meta-analysis conducted by the World Health Organization (WHO) and published in 2018 found that environmental noise levels above 45dBL_{den} have been associated with various diseases, a reduced quality of life and increased mortality [5]. The measure of noise L_{den} used by the WHO is defined in Equation 3.1 and is further discussed in the following section. As described in Section 1.1, the burden of disease from environmental noise contributes to 12,000 premature deaths and 48,000 new cases of Ischemic Heart Disease each year in the European region [4]. Ischemic Heart Disease, often referred to as Coronary Heart Disease, is a condition where blood flow around the heart is restricted [79]. Further research from the WHO has shown that exposure to excess noise pollution is a factor in the increased risk of cardiovascular disease, cognitive impairment in children, sleep disturbance, tinnitus, and annoyance [5]. These results come from the WHO's environmental noise guidelines for the European region, a comprehensive meta analysis of evidence on the effects of environmental noise on health in an epidemiological context. The WHO generate guideline noise exposure limits for key noise sources of environmental noise based on the weight of evidence from a wide pool of systematic reviews of epidemiological studies of the impact of noise on populations. The current WHO guideline noise exposure level limits are summarised in Table 3.1, which is adapted from [5].

Noise Source	**Context**	**Benchmark Level**	**Strength of Recommendation**
Road Traffic	Total	53dBL$_{\text{den}}$	Strong
Road Traffic	Night	45dBL$_{\text{night}}$	Strong
Railway	Total	54dBL$_{\text{den}}$	Strong
Railway	Night	44dBL$_{\text{night}}$	Strong
Aircraft	Total	45dBL$_{\text{den}}$	Strong
Aircraft	Night	40dBL$_{\text{night}}$	Strong
Wind Turbine	Total	45dBL$_{\text{den}}$	Conditional
Wind Turbine	Night	None	Conditional
Leisure	Total	70dBL$_{\text{Aeq,24h}}$	Conditional

Table 3.1: WHO guideline noise exposure limits for key noise adapted from [5].

The WHO guidelines adapted in Table 3.1 give a benchmark exposure level for each noise source that is associated with negative impacts of health due to overexposure. The impact of night noise has been shown to be highly severe, and guideline levels for each noise source are given for both total exposure and night time exposure. The strength of a recommendation based on the weight of evidence from the meta-analysis associated with each noise source is also given, with two levels. A strong recommendation is based on evidence that the impact of not adhering to the defined limit is more undesirable than the consequences of exceeding it, and the weight of evidence is high. Further, this type of recommendation is suggested to be implementable in most circumstances. A conditional recommendation is based on evidence that is less certain than a strong recommendation, and there might be significant challenges in implementing or adhering to the recommended benchmark level.

3.3 Environmental Noise Quantification

Each of the benchmark levels in Table 3.1 are given in terms of a noise metric, which is a method of describing the noise exposure level in a single value. The choice of noise metric was chosen to ensure that the recommendations were compatible across EU member states, and those noise metrics were dBL$_{\text{den}}$, dBL$_{\text{night}}$ and dBL$_{\text{Aeq,24h}}$. The metric dBL$_{\text{den}}$ is a weighted sum of average A-weighted equivalent noise levels, collected over four contiguous time periods, over a 24 hour period. The formula for L_{den} is presented in Equation 3.1, and is adapted from [74].

$$L_{den} = 10 \log \frac{1}{24}\left(12\left(10^{\frac{L_{day}}{10}}\right) + 4\left(10^{\frac{L_{evening}+5}{10}}\right) + 8\left(10^{\frac{L_{night}+10}{10}}\right)\right) \tag{3.1}$$

Where:

- L_{day} is the A-weighted long-term average sound level determined over *all the day periods of a year.*
- $L_{evening}$ is the A-weighted long-term average sound level determined over *all the evening periods of a year.*
- L_{night} is the A-weighted long-term average sound level determined over *all the night periods of a year.*
- The day is 12 hours, the evening 4 hours and the night 8 hours.

And

$$L_{day,k} = 10 \log\left(\sum_i N_{ref,i} 10^{0,1 L_{E,i,k}}\right) \mathrm{dB} - 10 \log\left(t_{day}\right) \mathrm{dB} \tag{3.2}$$

Where

- k indexes the meteorological window.
- i indexes the sound source category.
- $L_{E,i,k}$ is the average A weighted sound level for sound source and meteorological event.
- $N_{ref,i}$ is the statistical yearly day average number of single events per source category i.
- t_{day} is the number of hours of duration for the period, such as the day.

Each of these metrics are defined in terms of the A-weighted equivalent continuous sound pressure level, L_{Aeq}, which is a measure of the average sound pressure level over a period of time [80, 12]. The procedure for measuring L_{Aeq} is defined in the ISO standard for determination of sound pressure levels ISO 1996-2:2017[E] [81]. The L_{eq} denotes that the averaging scheme is intended to calculate the equivalent average noise exposure over a given time period A-weighting refers to the filtering operation that is performed on the signal being analysed in order to approximate the frequency dependent sensitivity of the human hearing system [82]. The A in A-weighting is indicative of the weighting filter that is applied to the signal being analysed, and several weighting filters have been defined for different scenario including the B, C, D and Z [12]. The spectral shape of these weighting filters are similar to the equal loudness contours that were introduced in Section 2.5, compensating for the frequency dependent sensitivity of the human hearing system at different levels.

Prior to the publication of the END, in the position paper on EU noise indicators, L_{den} was described as the format for the L_{EU} metric, which is essentially identical to L_{den} [83]. The same formula can be used for calculating L_{day}, $L_{evening}$ and L_{night} with adjustment to the value of t_{day} [74]. This formula is derived in the ISO standard for determination of sound pressure levels ISO 1996-2:2017[E] to account for the effect of different noise sources and weather conditions and is specified for determining levels based on long term measurements [81].

3.4 European Environmental Noise Policy

As described in Section 3.1, the Environmental Noise Directive (END) was ratified by the European Parliament in order to mandate and unify the management of environmental noise pollution in European member states [74]. The END has two primary objectives:

1. To define a common approach for environmental noise management at the policy level, intended to avoid, prevent or reduce on a prioritized basis the harmful effects, including annoyance, due to exposure to environmental noise

2. To provide a basis for developing community measures to reduce noise emitted by the major sources

The END defines a procedure that member states of the European Union should undertake in order to identify locations where the population are exposed to excess environmental noise, and create action plans in order to mediate excess environmental noise pollution in the area. The procedure legally obliges member states to determine and publish data pertaining to the exposure of *agglomerations* to critical environmental noise sources on a geographic basis through a process called strategic noise mapping. An agglomeration[3] is defined in the END as an urbanised area with an appropriate population density as defined by the member state, and with a population in excess of 100,000 people. Member states must develop action plans to manage areas with excess levels of environmental noise pollution, these action plans must be submitted to European Commission on a rolling 5 year basis. Member states were required to integrate the END into their native policy, and the UK implemented The Environmental Noise (England) Regulations in 2006 as a national legislative implementation of the END [84].

As highlighted in Section 3.3, the END defines a set of noise metrics that are to be used for the purposes of strategic noise mapping and action planning. The form of data on noise exposure that member states are required to publish is defined in the END as a strategic noise map, which is a map of the noise exposure in a given area due to different noise sources, or for overall predictions for such an area [74]. Strategic noise maps are published by the UK Government and are available via a web view from the Extrium consultancy [85]. Figure 3.1 shows a strategic noise map of agglomerations and noise levels from road traffic noise around the city of York, from [85]. The strategic noise map presented in Figure 3.1 is a map shows a group of agglomerations as the portions of the map that are overlaid with a slightly opaque colour, and the noise levels from road traffic noise are clearly visible as orange, red and blue areas that are concentrated around the main roads and dual carriageways. Figure 3.1 shows that several agglomerations near the main roads encircling the city of York are exposed to noise levels in excess of 53dBL_{den}, which is the recommended limit for road traffic noise described in Table 3.1.

[3] also known as an urban area

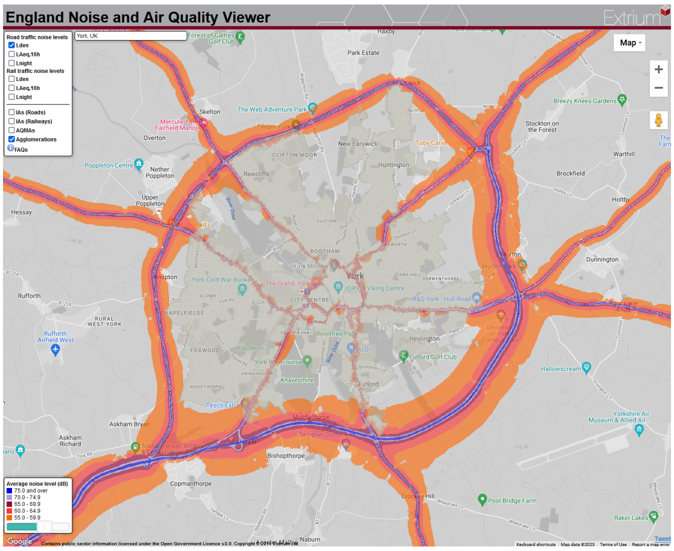


Figure 3.1: Strategic noise map of agglomerations and noise levels from road traffic noise around the city of York from [85].

3.5 Soundscape

When attempting to understand how we perceive and respond to sounds such as environmental noise, it is important to consider those sounds in context and not in isolation. The combination of sounds and the context of their situation come to form a *soundscape*, analogous to a landscape in the visual domain. One of the first scientific publications making reference to soundscape was reputedly written by Michael Southworth and published in 1969 [86, 7]. The popularization of the concept of soundscape is often attributed to R. Murray Schafer who founded The World Soundscape Project. The first publication of the project was *The Book of Noise* in 1967 [87], which was followed by a range of books, papers and compositions. In The Book of Noise an early description of a soundscape is given: *...HOWL! This is a portrait of your city. Listen to it closely. Perhaps you have never really listened before. It is a fascinating and exasperating concert of sounds. Listen... Horns, sirens, motorcycles, trucks, jack hammers, power saws and construction machinery, helicopters and jets. Any attentive listener will conclude that these are the dominant instruments of the orchestra. These sounds are all louder than* *the human voice and they are getting louder*[87]. Schafer later published *The Tuning of the World: Toward a Theory of Soundscape Design* in 1977 [3, 88], a seminal text introducing soundscape and cementing the desire to make environmental acoustics a positive endeavor where sounds are preserved, encouraged and curated. Within this work Schafer introduced several pieces of terminology that are still used today, including the terms *keynote sound*, *soundmark*. A keynote sound is described as the fundamental tone that anchors a soundscape, forming the reference point for the soundscape. A soundmark is a key feature of a place that is acoustically unique or possesses a special quality that distinguishes the soundscape. With a background of composition, writing and teaching, Schafer has approached the study of soundscapes from an artistic and naturalistic perspective and this perspective is present in the video piece about soundscape and listening, *Listen/Ecoute* [89]. Schafer and his contemporaries not only observed and reported the soundscapes as they perceived them, but they shared a vision for how soundscapes could be curated and tuned to identify the place that is coupled with the soundscape. These concepts have gone on to grow a broad tapestry of interdisciplinary research including the field of acoustic ecology [90], as well as approaches to working with modern policy to create better urban environments [91].

3.5.1 Soundscape Definitions

In [3] R. Murray Schafer defined soundscape as *The sonic environment. Technically, any portion of the sonic environment regarded as a field for study.* In 2014, the International Standards Organization published ISO 12913-1:2014 *Acoustics – Soundscape Part 1: Definition and conceptual framework* [2] the first in a series of soundscape standards. The definition of soundscape is given in the standard as:

[the] acoustic environment as perceived or experienced and/or understood by a person or people in context

where the acoustic environment is defined as:

sound at the receiver from all sound sources as modified by the environment

This definition builds on that given by Schafer by adding the perception of the environment as a key factor in soundscape. As defined in Section 1.1 the term Soundscape in the context of this thesis refers to the perception of the acoustic environment, by people, in context. This follows the definition given by the ISO standards committee for soundscape [2], which itself follows on from the definition given by Schafer.

3.6 Soundscape Quality Evaluation

The purpose of soundscape evaluation is to identify the qualities and attributes of a soundscape, drawing together relationships between the two. Quality in relation to an item is defined as *"the standard or nature of something as measured against other things of a similar kind; the degree of excellence possessed by a thing"* [92]. Soundscape exists as a construct that is bound to the perception of the listener, and as such, soundscapes are often evaluated using subjective measures. The quality of a soundscape may refer to a collection of subjective factors that can be used to describe soundscape, or it may be assessed as an entity in its own right on a good-bad scale [93]. However, the subjective factors used to describe a soundscape are not necessarily explanatory in themselves, and these factors may be influenced by other properties of the signal chain.

3.6.1 The Filter Model

It can be useful to consider soundscape evaluation through the lense of the filter model as described by Bech and Zacharov [94, 95] and shown in Figure 3.2. This filter model in Figure 3.2 has three layers, the first layer is the physical layer which encompasses the physical properties of the acoustic environment such as the measured noise level or the reverberation time. The second layer is the perceptual layer which encompasses the intrinsic perceptual properties of the soundscape, reflected by psychoacoustic properties such as Loudness and Fluctuation Strength. Loudness as introduced in Section 2.5 is the perception of the intensity of a sound, and Fluctuation Strength as defined in Section 3.11 is the perception of the intensity of the low frequency amplitude modulation of a

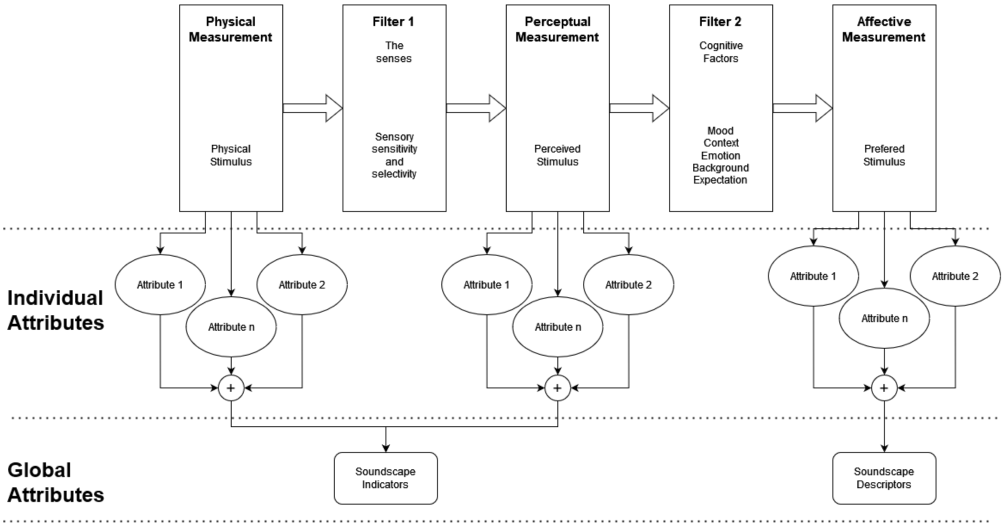


Figure 3.2: A Filter model after that proposed by Bech & Zacharov [95], adapted for the evaluation of soundscape.

signal. The third layer encompasses the aspects of the soundscape that are dependent on cognition such as the affect i.e. emotional response of the listener. The definition of affect will be elaborated upon in the following section of this chapter. The model includes two filter states, the sensory filter and the cognitive filter. The sensory filter encompasses physiological processes including those described in Chapter 1. The cognitive filter encompasses the processes of the brain that are responsible for the perception of the soundscape as well as remembering, conceiving, reasoning, judgement etc. As can be observed in Figure 3.2 each layer of the model feeds into the next, and there is an assumption of dependence between the cognitive, perceptual and physical layers. This assumption of consecutive dependence shows that it is important to consider how all of the aspects of soundscape perception might interact when performing soundscape evaluation.

3.7 Descriptors & Indicators

The filter model of a soundscape is useful in that it allows for the consideration of soundscape as a holistic entity that can be assessed at different stages of this model. Factors at each stage of the filter model can be grouped into one of two sets that can then be cross-referenced for variance, in the language of ISO 12913-2:2018 [96], and these groups are called *descriptors* and *indicators*. A descriptor in this context is a factor that can be used to describe the perception of the acoustic environment, and an indicator is defined as a factor that can be used to predict a descriptor or part thereof [96]. Descriptors of soundscape are typically grouped from the affective layer of the filter model, and indicators are typically grouped from the physical and perceptual layers. However, the classification of soundscapes and sound sources may be considered objective truths, but are dependent on the cognitive filter if gathered via survey. The language of [2] is aimed at the assessment of descriptors and indicators with methods of inference, using statistics to support a causal relationship between indicators influencing a descriptor, and a systematic review of the use of statistical modelling methods for the prediction of soundscape quality has been published [6].

3.8 Types of Assessment

Gathering subjective information about soundscapes is most easily achieved through forms of survey, and a wide variety of survey methods and instruments have been used in soundscape assessment. In a systematic review of surveys used in soundscape studies, Engel *et al.* investigated 52 peer reviewed publications [97]. The review found that the four primary data collection methods used in soundscape studies were soundwalks, interviews, listening tests and focus groups.

Soundwalks are one of the methods of soundscape evaluation discussed in the ISO 12913 standards [96], but are reputed to having been developed independently by two different groups [97]. Soundwalks are typically guided walks through an environment where the participants are asked to listen to the soundscape and record their subjective experience. However, the actual methodology of a soundwalk can very greatly. One of the key benefits of the soundwalk methodology is that it allows for the collection of both qualitative and quantitative data, with practitioners recording the soundscape for later analysis. Soundwalks also benefit from being conducted in-situ, guaranteeing that any covariates that might influence the perception of the soundscape are unfiltered [98]. Variations of the soundwalk methodology have been previously used in laboratory conditions as a *virtual soundwalk* [99, 100].

Guided interviews are another method of soundscape evaluation that is discussed in the ISO 12913 standards [96]. Although it is not clearly defined in [97] which types of interview are present in the literature, the soundscape standard presses the importance of using guidelines when conducting narrative interviews. One of the benefits of using interviews as a method of soundscape evaluation is that they allow for the collection of contextual information about the soundscape.

Listening tests in this instance describe methods of soundscape evaluation that are conducted in a laboratory setting under controlled conditions. These controlled conditions are beneficial for isolating the influence of key factors on the perception of the soundscape, but at the cost of the important contextual factors that are present in-situ [98]. Further, the methods used for the reproduction of soundscapes in a laboratory setting becomes a factor in the perception of the soundscape, and the ecological validity of soundscape reproduction methods varies with the degree of spatiality and the type of soundscape being reproduced [101, 102]. Ecological validity is defined by Bronfenbrenner as "the extent to which the environment experienced by the subjects in a scientific investigation has the properties it is supposed or assumed to have by the investigator" [103], whereas Schmuckler identifies ecological validity as "the extent to which the results of a laboratory study can be generalized to and across other settings" [104]. Gibson [105], and later Guastavino [101] extend this definition to include the need for the experiment environment and context to be ecologically valid, in the sense that in studies of perception, the experiment protocol should be designed such that participants will react to stimuli in a way that is representative of their reaction to the same stimuli in the real world.

Researchers have identified that the degree of spatial information present in the reproduction of the soundscapes can have a direct impact on the cognitive representation of the events that occur within the recording [101]. In the same experiments the researchers identified that the visual environment present during listening tests had an influence on the perception of the soundscapes, remarking that a neutral visual environment and spatial immersion are important to ensure ecological validity.

Focus groups are a method for gathering qualitative information about a soundscape in a moderated setting. In a focus group a researcher will guide a discussion on a specific topic, and the participants will discuss their opinions and experiences on the topic. The advantage of a focus group over other interview techniques is that they allow for the emergence of ideas and opinions that might not have been considered by the researcher [97], also allowing the researcher to capture detailed opinions [106]. Further, focus groups facilitate a reflective state of mind, aiding in the expression of ideas and cultivation of group discussion [97].

3.9 Affect

In psychology, affect refers to the experience of feeling or emotion and it encompasses a range of subjective experiences such as joy, sadness, anger, fear etc [107]. Affect was originally conceived by William Wundt as a state of feeling which is constant, direct and psychologically primitive experience [108]. Affect can be thought of as the way in which people respond emotionally to their environment. Affect is often used interchangeably with emotion, although some researchers make a distinction between the two terms, with affect referring to more basic feelings and emotions referring to more complex experiences involving cognition and behaviour.

One of the pioneers in the study of affective psychology was Charles Osgood who developed the semantic differential technique for assessing affect through the connotative meaning of words [109]. The semantic differential technique involves using pairs of adjectives, semantic differentials, that represent contrasting qualities of the topic under investigation. These semantic differentials are applied to 1 dimensional numerical scales with each member of the pair at an opposing end of the scale. Participants in a study that uses the semantic differential technique will consider the semantic differentials with respect to the topic under investigation and will perform a rating that will result in a numerical representation of the features of the topic under investigation. Statistical analysis can then be applied to these numerical ratings to provide insight into the perception of the topic under investigation. In the 1950s Osgood *et al.* pioneered the basis for the description of emotion as having three fundamental dimensions called *valence*, *arousal* and *dominance*. Valence is the dimension that represents pleasantness and unpleasantness; arousal is the dimension of excitement and relaxation; and dominance is the dimension of control. The description of affect in terms of these base dimensions was refined and extended by Russell [110] and others into the circumplex model of affect. The circumplex model presented in Figure 3.3 represents valence and arousal together on a circular graph, with emotions positioned according to their valence and arousal levels, the horizontal axis representing the dimension of valence, and the vertical axis representing the dimension of arousal.

Emotions that are similar in terms of valence and arousal are located close to each other on the circle, while those that are dissimilar are located farther apart. The circumplex model has been widely used in research on emotion and has been found to provide a useful framework for understanding the structure of affective experiences. Both the circumplex model and semantic differentials have also been applied in soundscape [112, 113, 114, 111].

One of the most prominent instruments for soundscape quality assessment that uses a similar approach is the Swedish Soundscape Quality Protocol [113], which was reputedly used as a basis for the ISO 12913 series of standards [115]. The Swedish Soundscape Quality Protocol is a survey instrument for the assessment of perceived affective quality that includes a good-bad scale and eight Likert like items i.e. items that are rated on a scale of one to five. Perceived affective quality in this context is defined as an

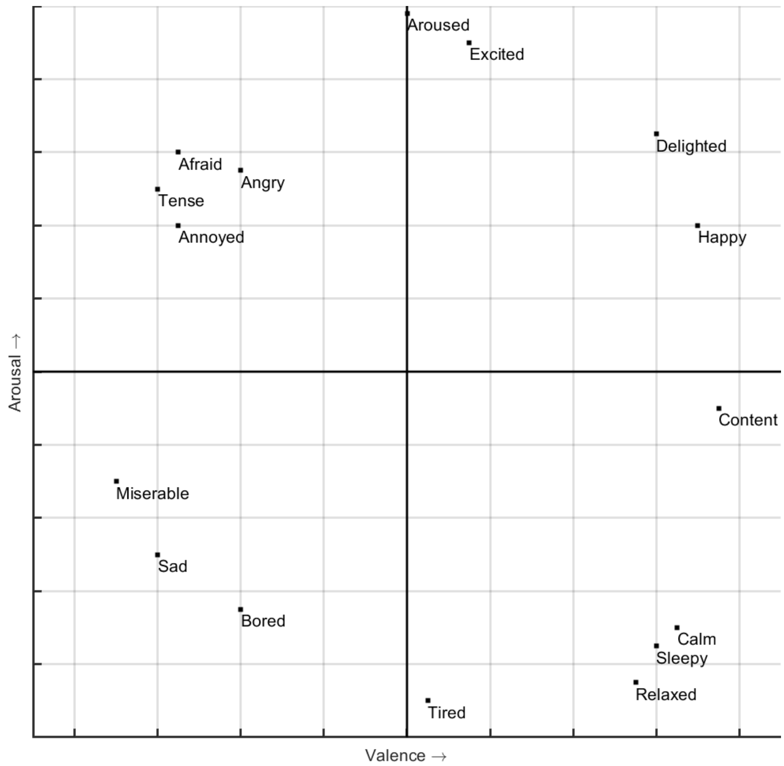


Figure 3.3: Emotions distributed across the circumplex model of affect as for [111].

individuals perception of the ability of a soundscape to change his or her state of affect [116]. The items are intended to describe perceived affective quality and can be asked in the form, “To what extent do you agree with the 8 statements on how you experience the present surrounding sound environment: pleasant, chaotic, exciting, eventful, calm, annoying, uneventful, monotonous” [117]. The one to five scale reflects answers ranging from, “disagree completely”, to, “agree completely”. An update to the protocol added an appropriateness dimension, suggesting estimates of both perceived affective quality and appropriateness may be important in soundscape evaluation [93]. Questions from the Swedish Soundscape Quality Protocol were included in the Soundscape Indices Protocol, a more recently developed protocol for the evaluation of soundscapes [98]. Semantic differentials and the Swedish Soundscape Quality Protocol are both examples of affect assessment that rely on word associations to describe affective experiences.

An evolutionary step was taken by Bradley and Lang who published the self-assessment manikin (SAM) [118], a pictographic instrument for the assessment of affect that surveys for the three key dimensions directly. Figure 3.4 shows the SAM, where each dimension of affect is represented by a five point scale, with the middle point of each scale representing a neutral state. Each position on each dimension is represented by a picture which is intended to be more universally understood than words. The upper scale represents valence, the middle scale represents arousal, and the lower scale represents dominance. Bradley and Lang compared the performance of the SAM to semantic-differential pairs, and for the valence and arousal dimensions they found between 94% and 99% correlation between the SAM and semantic-differential pair factor analysis for students reviewing a set of pictures [119]. The SAM has been used in the assessment of sound and soundscapes in a large number of studies [120, 121, 122], however Stevens *et al.* found the dimension of dominance to have weaker correlation than valence and arousal in a comparison between semantic differentials and the SAM manikin in the context of soundscape assessment [123].

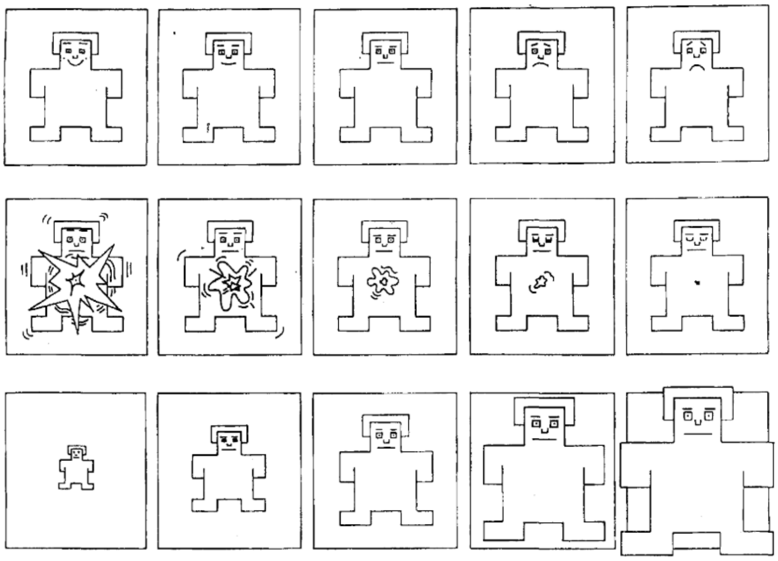


Figure 3.4: The self-assessment manikin, a pictographic instrument for the survey of affect, from [118].

3.10 Physiological sensing in soundscape studies

As well as assessing affect using questionnaires, researchers have also used physiological sensing to assess the impact of soundscape on affect, well-being and behaviour. In the fields of psychology and engineering there is a growing body of evidence identifying that changes in emotional state can be identified through physiological measures [124]. Physiological responses are reactions to situations such as relaxation, stress or danger, leading to changes in the regulation of mechanisms like heart rate, breathing rate, body temperature and sweating. Physiological measures are contributed to by both the sensory and cognitive filters of the filter model in Figure 3.2 and could be considered to be positioned on the layer of affect if it is the case that there is a systematic relationship between emotions and physiological behaviour. Psychophysiology is the study of the relationship between physiological signals and cognitive processes [125]. As discussed in Section 1.2 it is hypothesized that physiological mechanisms are modulated by the state of affect, and it is hypothesized that the state of affect of an individual is modulated by the soundscape that they are placed within.

The use of physiological measures that are influenced by emotional state may provide soundscape researchers with evidence of the psychophysiological basis for changes in affect in response to different soundscapes. Several theories in psychophysiology support the relationship between states of affect and the behaviour of the autonomic nervous system [126], including changes in heart rate variability and the regulation of emotions. The aim of this study is to understand whether participants exhibit changes in physiological behaviour when listening to recordings of soundscapes in the context of a directed listening experiment under laboratory conditions.

Although there is a large body of research that has investigated the relationship between soundscape and physiological responses, there is a lack of consensus between the results of these studies. A systematic review of into research studying psychophysiological factors in the context of soundscape identified only six relevant studies in the literature [127]. All of the studies were conducted in the laboratory, and all of the studies used a stimulus-locked repeated measures design. All of the studies include the survey of affective dimensions [128], although in studies such as [129] affective evaluation is performed separately to physiological evaluation. However, some studies evaluated the subjective quality of the stimuli by surveying with measures that describe the soundscape itself, as opposed to the participants' experience, including measures such as eventfulness, pleasantness and vibrancy [113]. These scales are identified by Erfanian *et al.* as analogous to the affective dimensions of valence and arousal. The studies include a variety of different physiological measures and measurement equipment, including galvanic skin response [130] and functional magnetic resonance imaging [129]. Average heart rate measurement was included in all studies but one, in which high frequency heart rate variability measures were included instead [131].

The outcome of this review was that results from the studies were generally weak or conflicting when relating physiological responses to affect. These conflicting results may be unsurprising, given the range of psychophysiological studies in the music psychology literature that come to similarly disparate conclusions [132]. However, it could also be that the size of the effect of the manipulation was insufficient for the given sample sizes in the presence of the other confounds that are likely to influence physiological measures. Further, some of the experimental conditions such as [129] were not conducted in scenarios that are likely to have high ecological validity.

3.11 Psychoacoustic Parameters

Psychoacoustic parameters are positioned on the perceptual layer of the filter model presented in Figure 3.2, between the sensory and cognitive filters. Just as the weighting filters discussed in Section 3.3 are intended to reflect the frequency dependent sensitivity of the human hearing system, psychoacoustic parameters are intended to reflect properties of sound as filtered and processed by the human auditory system. Publications have discussed how psychoacoustic parameters can assist practitioners as explanatory features in soundscape assessment [133].

A systematic review of the use of psychoacoustic indicators in soundscape studies identified 46 articles that were focussed on testing a variety of hypotheses including (but not limited to) soundscape characterization and relationships between psychoacoustic parameters and perceptual or health outcomes [23]. The review identified the use of *Loudness*, *Sharpness*, *Roughness* and *Fluctuation Strength* as psychoacoustic parameters that are discussed in Section 2.5.

3.12 Soundscape Classification

Soundscape classification is the process of identifying the nature of a soundscape by identifying the dominance of different sound types and properties in the environment. Several different strategies of soundscape classification have been proposed in the literature, most commonly in the form of taxonomies. Taxonomies are ordered systems for the classification and organization of objects or concepts, and several taxonomies have been proposed for the classification of soundscape [134, 135, 136, 96].

It has been suggested that choosing an appropriate taxonomy for urban soundscapes is an important step in the development of data sets that can be used for soundscape analysis and information retrieval [137]. Most taxonomies classify the soundscape based on the dominance of a particular type of sound such as traffic or natural sounds. The taxonomy included in the ISO standard for soundscape [96] is presented in Figure 3.5, and features a hierarchical depiction of nested lists that describe indoor, outdoor, natural and human made sounds . The taxonomy presents a diverging tree structure, with the root node representing the acoustic environment as a whole, the subsequent nodes representing different types of acoustic environment and finally the leaves representing individual classes of sound source. A common strategy for the classification of soundscapes is based on the dominance of natural, human and mechanical sounds [111, 127]. This type of classification is congruent with the taxonomies of environmental sound used by Schafer [3, 111].

Summary

Environmental noise is a form of pollution that is has a significant impact on the health, well-being and quality of life for people living in areas that are exposed to high levels of noise. Areas of housing near to major roads, airports and industrial facilities are particularly susceptible to high levels of environmental noise, and these populations may be at risk of overexposure. However, in the European region, governments are mandated to identify areas that may be at risk of excess noise exposure and mitigate that risk through an action plan. This chapter has highlighted some of the features of this policy, and the metrics that are used to quantify environmental noise exposure. This information gives background context for soundscape, an alternative philosophy to sound and noise. This chapter also introduced soundscape as a field of research that is concerned with the perception of the acoustic environment as a holistic entity. Soundscape evaluation involves assessing the characteristics and quality of the acoustic environment as perceived by the listener though a variety of methods. Descriptors are used to assess the perception of a soundscape, and indicators are used to quantify properties of the soundscape. Statistical relationships between the descriptors and indicators of soundscape are examined in order to understand how the soundscape is perceived, and how changes to the soundscape may influence its perceived quality. This approach to soundscape quality assessment will be used in the following chapters to examine the influence of soundscape on affect and physiological responses.

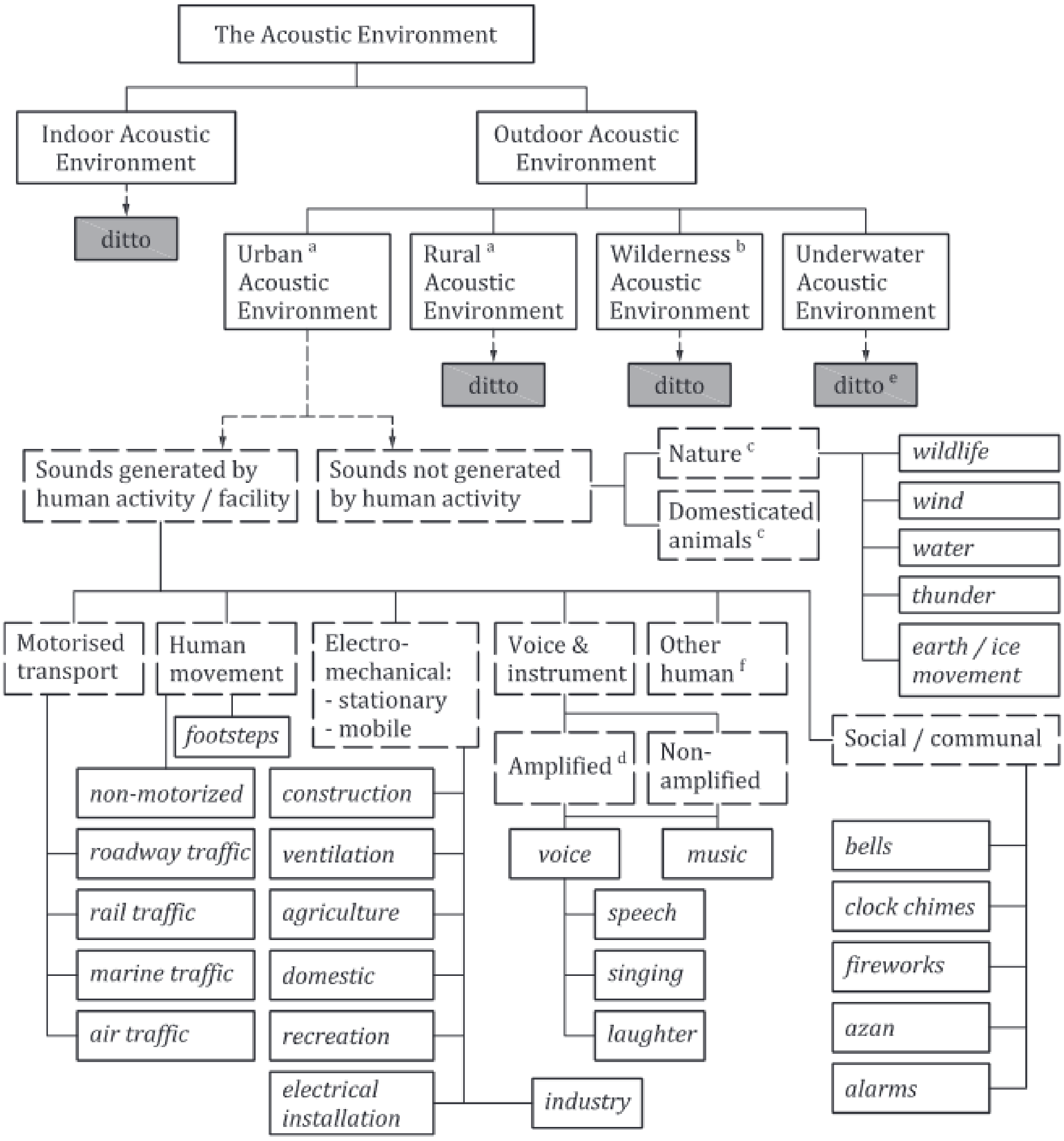


Figure 3.5: Taxonomy of sound sources classes from ISO:12913-1 [2].

Chapter 4

Assessment of soundscapes using self-report and physiological measures: Pilot Study

In Section 1.2 a hypothesis was presented that aimed to identify if soundscapes evoke physiological and affective responses in listeners. That hypothesis stated that the experience of soundscapes including natural and mechanical sounds would evoke physiological and affective responses in listeners. To support this, a foundation for sound propagation and sound perception was established in Chapter 2, followed by a discussion of environmental noise and soundscape evaluation in Chapter 3. This chapter presents the first of two experiments that are designed to test the hypothesis established in Section 1.2. This experiment included the use of a heart rate sensor to measure physiological responses to soundscapes, and a survey to measure subjective responses to soundscapes.

4.1 Introduction

In order to determine if soundscapes evoke physiological and affective responses in listeners, a pilot study was conducted to test the experimental design and methodology. A pilot study is intended to test the feasibility of the experimental design and methodology, and to identify any issues that may arise in the full study.

As discussed in Section 3.10 a systematic review of psychophysiological studies in soundscape [127] identified six studies that used physiological measures to evaluate the experience of soundscapes. From the set of experimental methodologies that have been used for the evaluation of soundscapes that were discussed in Section 3.8, all of the experiments discussed in [127] were listening tests. Most of these studies used a repeated measures design in which all participants were exposed to the same set of stimuli in a particular order. Repeated measures designs are useful for reducing the number of participants required for an experiment, and for maximizing the amount of data collected from each participant [138]. All of the studies included an evaluation of perceptual attributes related to experiencing the stimulus, and more specifically emotional or affective attributes. Further, five of the six studies included average heart rate as a physiological measure. In this type of experiment design all participants are exposed to a set of stimuli following a particular protocol of exposure, response and recovery. All of the studies included an evaluation of perceptual attributes related to experiencing the stimulus, and more specifically emotional or affective attributes. Some studies included the evaluation of participants' emotional state through the survey of affective dimensions [128], though in studies such as [129] affective evaluation is performed separately to physiological evaluation. Some studies evaluate the subjective quality of the stimuli by surveying with measures that describe the soundscape itself, as opposed to the participants' experience, including measures such as eventfulness, pleasantness and vibrancy [113]. These scales are identified by Erfanian *et al.* as analogous to the affective dimensions of valence and arousal.

The studies include a variety of different physiological measures and measurement equipment, including galvanic skin response [130] and functional magnetic resonance imaging [129]. Average heart rate measurement was included in all studies but one, in which high frequency heart rate variability measures were included [131]. From all of the studies discussed in [127] only one study included stimuli that included spatial information, the study by Alvarsson *et al.* [131] used binaural recordings made with a dummy head. Beyond the use of binaural recordings in [131], and the choice not to normalize the levels of stimuli in [128], none of these studies discuss or appear to consider the importance or impact of the ecological validity of the experiment protocol or stimuli. Ecological validity was not widely discussed in any of the six sources, and a focus on the ecological validity of the experimental protocol is recommended by Erfanian *et al.* [127].

4.2 Methods

In order to test the hypothesis established in Section 1.2, a directed listening experiment was designed to evaluate the affective and physiological response to a set of soundscapes. The methodology used in this experiment follows on from Stevens *et al.* [139, 140] by performing a descriptor indicator comparison of changes in heart rate, affective report and classification score. However, the stimuli selection methodology, the experiment protocol and method of analysis were all developed to suit the aims of this experiment. The experiment was designed to be a pilot study, and as such the methodology was designed to be simple and to be easily repeatable. The results presented by Stevens *et al.* have identified that the classification of a soundscape as being natural or mechanical is a significant predictor of the affective response to that soundscape. Based on this principal, the experiment was designed so that participants would evaluate a set of soundscapes, giving both a classification score to the soundscape and an affective report. The classification score and affect report given by each participant would be compared with the aim of identifying if there was a significant relationship between the two. Further, the physiological measures taken during the experiment were normalised on a by-participant basis, and were then compared across the stimuli to see if systematic changes were observed. Average heart rate was the physiological measure chosen for this initial experiment, as it are theorized to reflect changes in affect [141, 142, 143]. The expectation is that stimuli with differing classification scores would elicit different states of affect, and these would be reflected in the physiological measures. Based on the reports of natural soundscapes being restorative or eliciting positive valence, a further expectation would be that soundscapes classified as being more natural would elicit higher valence [144, 145, 146]. This experiment was reviewed and approved by the University of York Department of Electronics Research Ethics Committee. The documentation used in the application for ethical approval is included in Appendix A.

4.2.1 Participants

Participants were recruited from groups of audio engineering students and staff via email. All participants had training in subjective testing, and could be considered expert listeners (see section 5.4.1 of [95] for a discussion on the definition of expert listeners). A total of 14 participants were recruited to take part in this experiment. Of the participants 11 identified as male, 2 as female. Three participants were in the age range $18-24$, 10 were in the age range $25-30$, and 1 in each of the $31-35$ and $36-40$ age ranges. The participants were screened by self-exclusion for the following criteria:

- Hearing impairment.
- Heart condition or ailment.
- Sensitive or damaged (broken) skin around the wrist or fingers.
- Any injuries around the skin or fingers.
- Any know neurological condition or ailment.
- Younger than 18 years of age.
- Older than 60 years of age.

4.2.2 Experimental Stimuli

The experimental stimuli were selected from the EigenScape dataset, a set of Ambisonic B-format recordings made in various locations around the UK in 2017 [147]. To record this dataset Green used the mh acoustics EigenMike em32, a 32-channel spherical microphone array that can be used to record a sound field in Ambisonics up to fourth-order. In a series of comparisons of spatial audio microphones the EigenMike em32 has been identified to have high quality spatial accuracy and a full but dull sonic characteristic [148, 149]. To record the dataset, Green used the EigenMike em32 in combination with the EigenMike proprietary audio interface and the EigenStudio recording software. Some recordings were made with the assistance of a custom wind shield that was later replaced with a higher quality wind shield, leading to a minor gain discrepancy between some recordings. This dataset was used because it is a set of Ambisonic recordings with a wide range of different environments, including several examples of each environment type and 10 minute long recordings for each environment. The version of the dataset used in this experiment is comprised of first order Ambisonic recordings, and is available from Zenodo [150]. The stimulus lengths used in the literature identified by Erfanian *et al.* [127] ranged from 8 seconds to 4 minutes, however the stimulus length used by Stevens *et al.* was 30 seconds [139]. To maximise the time available for physiological responses to occur and keep the total experiment time to a minimum, a stimulus length of 100 seconds was chosen for this experiment. This stimulus length allows the experiment to be completed within 30 minutes, maximizing the available listening time for each experiment interval while minimizing the impact of listening fatigue. The EigenScape dataset was segmented into contiguous clips 100 from which the test stimuli were then selected. The stimuli were screened for markers that could be used to identify individual people such as clear discernible speech. Stimuli were selected by algorithmically evaluating the proportion of natural and mechanical sounds in each clip, which was calculated using the Normalized Difference Soundscape Index (NDSI). NDSI is an acoustic index that is intended to identify the proportions of natural and mechanical sound sources in a recording by comparing the proportions of spectral energy in two frequency bands, 1kHz to 2kHz for mechanical sounds and 2kHz to 11kHz for natural sounds [151, 152]. Green did not publish the details of the frequency response of the microphone that was used to record the EigenScape dataset, and the only information available from the manufacturer of the microphone identifies that all em32 microphones are band-limited to a maximum frequency of 20kHz [153]. Being a high quality professional microphone, it is assumed that the em32 that was used for recording the EigenScape dataset had a flat frequency response across the region of frequencies that are used to compute NDSI. NDSI scores range from $+1.0$ to -1.0, where a positive score indicates the presence of a higher proportion of natural sounds, and a negative score a higher proportion of mechanical sounds. The NDSI score was calculated for the 0^{th} order omnidirectional W-channel of each soundscape sample by using a Python3 implementation of the R Acoustic Indices package [154].

From the data set, the 4 clips with the most positive and the 4 clips with the most negative NDSI scores were selected as test stimuli, resulting in several clips being selected from the same recording. For training intervals two clips were randomly selected from two classes of soundscape that were not used in the experiment intervals, a beach soundscape and the soundscape of a shopping centre. The selected test stimuli are summarised in Table 4.1. In the dataset recordings are organized by location type i.e. beach, woodland, of which there are eight different location types. Individual recordings and locations can be identified using a map provided by the author [150]. In Table 4.1 the location type and recording number are both from the dataset structure. The clip number identifies which contiguous 100 second slice of the recording number is used. The stimuli were grouped into two groups based on their NDSI score, with the most positive NDSI scores being grouped together and the most negative NDSI scores being grouped together. The group set is provided to identify whether a clip has a higher proportion of natural sounds (a more positive NDSI score) or mechanical sounds (a more negative NDSI score).

Location Type	**Recording Number**	**Clip**	**NDSI**	**Group**	**Stimuli Index**
Woodland	5	5	0.749	High	2
Woodland	5	3	0.709	High	8
Woodland	2	6	0.820	High	6
Woodland	4	6	0.861	High	7
QuietStreet	4	3	-0.884	Low	1
QuietStreet	4	5	-0.777	Low	3
QuietStreet	4	1	-0.882	Low	5
BusyStreet	6	4	-0.869	Low	4

Table 4.1: Stimuli selection NDSI and group summary.

The two groups of stimuli as shown in Table 4.1 have opposed NDSI scores, and have been drawn from five different locations of three different types. All of the stimuli from the high NDSI group were all selected from the woodland location type, and only one of the stimuli was from a location near to an urban centre. The *Woodland5* recording location was in the grounds of the Dalby forest, a large forest located in the North York Moors National Park. The *Woodland4* recording location was in the grounds of the Hagg wood, an area of woodland located within the south west of the city of York near the ring road that encircles the city. The *Woodland2* recording location was in the grounds of the Acomb wood, a small area of urban woodland located to the south east of the city of York. All of the stimuli from the low NDSI group were selected from one of two locations. The *QuietStreet4* recording location was situated directly outside the AudioLab building on the University of York Heslington West Campus. The *BusyStreet6* recording location was situated on the inner ring road in the centre of the city of Huddersfield. Still images of these locations were not published by Green *et al.* as part of the EigenScape dataset [150], but the stimuli used in the experiment are included in the materials published with this thesis that are summarised in Chapter C.

The selected stimuli were converted from B-format first order Ambisonic to 2 channel binaural which could be easily reproduced over headphones. The B-format to binaural conversion was performed using the binaural decoder VST plugin that is part of the IEM plug-in suite [155], which is a selection of free tools that can be used for spatial audio processing. The Binaural Decoder plug-in settings were set to SN3D normalization and 1st order Ambisonics, with headphone equalization disabled. The head related transfer functions used in the Binaural Decoder plug-in were recorded using a Neumann KU 100 dummy head, and the binaural rendering was performed using the magnitude least squares approach that is described in [156].

The stimuli were peak normalized to ensure a consistent peak level between the stimuli, allowing the difference between the peak and average levels of the stimuli to be maintained so that soundscapes maintained their dynamic range while having consistent relative maxima.

4.2.3 Data Collection Instruments

Two forms of data gathering were used in the experiment, physiological sensing and self-report. For each test interval participants were asked to complete a survey based on their experience of the soundscape. The survey data collection instrument was implemented in a web-based user interface using the Qualtrics platform [157]. Qualtrics is a platform designed for performing surveys and data collection using a web interface, allowing researches to design custom surveys and securely collect and store experimental data. Figure 4.1 presents the test page of the user interface. As shown in Figure 4.1, the stimuli was presented to participants at the top of the form using the HTML5 audio element [158]. Below this, the two components of the survey were presented:

- A five-point self-assessment manikin in which participants report their affect
- A classification task in which participants are tasked to report the proportions of sound sources present in the soundscape.

The self-assessment manikin as described in Section 3.9 featured three sets of graphics with 5 elements in each set, representing the three primary dimensions of affect; valence, arousal and dominance. The participants were instructed to select which of the five levels of each dimension affect reflected their state when listening to each soundscape. The dimensions were described to participants as follows:

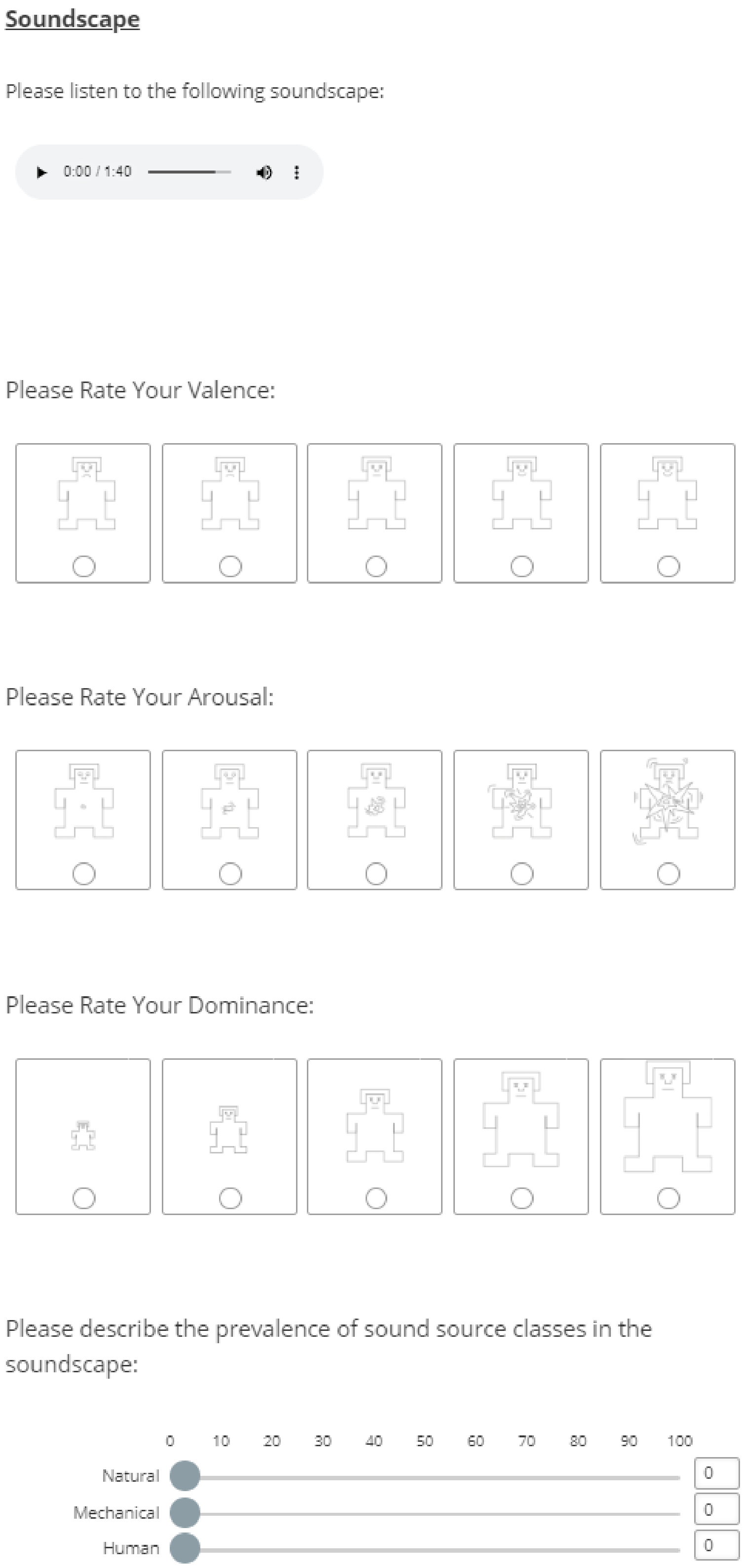


Figure 4.1: Listening test user interface.

- *Valence*: positive or negative emotions, analogous to feelings of pleasantness and unpleasantness, happiness and sadness.
- *Arousal*: analogous to excitement, restfulness, agitation and relaxation.
- *Dominance*: analogous to the participants feeling or control or presence within the situation.

These descriptions were based on those from Stevens *et al.* [123], and were suggested to participants including several descriptors with positive and negative connotations where possible to direct the participant to the nature of each dimension. The classification task included three sound source classes from a typically used taxonomy of sound sources [111, 127]:

- *Natural* e.g. animal sounds, bird song and environmental sounds such as the wind or the ocean.
- *Mechanical* e.g. air, rail and road traffic, as well as construction and industrial sounds.
- *Human* e.g. foot fall, masked speech and laughter.

Each class was represented by a continuous linear slider, as opposed to the 5 point scale used by Stevens *et al.* [139]. The participants were instructed to consider the collection of sounds they could identify when listening to a soundscape, and position the three sliders to indicate the general proportionality of the different classes of sounds. Participants were given the following example scenario for a soundscape. A soundscape that is primarily composed of mechanical sounds, with no natural sounds and few human sounds, participants should move the slider associated with mechanical sounds toward the top position, the natural slider to the bottom position, and the human slider to somewhere between the bottom and middle position. Participants were directed to disregard any numerical value associated with the sliders in order to avoid any confusion or association with a particular number and the presence of sounds in the soundscape.

Participants were further instructed to use the slider's position to indicate the proportionality of sounds from each class, avoiding participants rating a soundscape as being 4 on the natural scale because they identified 4 sounds that they considered to be natural sounds.

4.2.4 Physiological Measurement

The heart rate monitor used in this experiment was the Shimmer GSR+ Figure 4.2 shows the Shimmer GSR+ sensor [159]. The shimmer GSR+ presented in Figure 4.2 is

Figure 4.2: The Shimmer Sensing Shimmer GSR+ with PPG and GSR electrodes [159].

a wearable sensor that can be used to measure galvanic skin response (GSR) and heart rate [159]. The shimmer GSR+ can be used to estimate heart rate by measuring changes in blood volume through the skin using a photoplethysmograph (PPG) sensor, pictured in Figure 4.2 as the sensor with the white cable. PPG sensors measures the change in light absorption by the skin as blood volume changes in response to the heart's pumping action [160]. Figure 4.3 shows an illustration of a PPG waveform from [161]. The PPG signal presented in Figure 4.3 has three distinct sections, the first peak is caused by

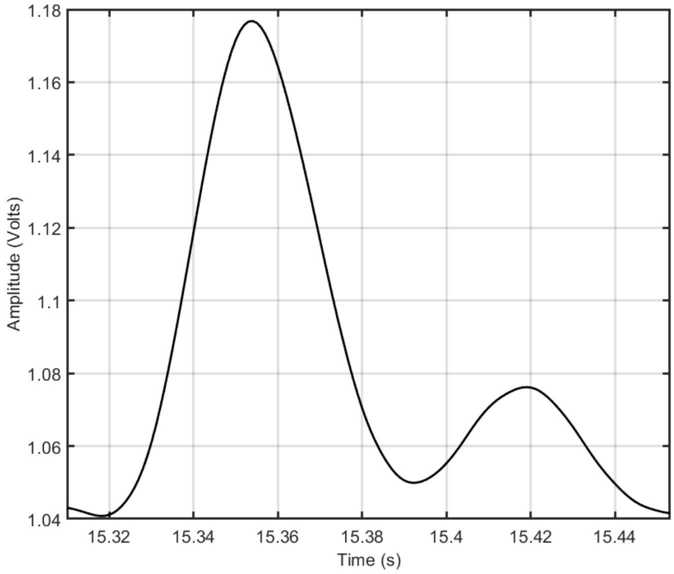


Figure 4.3: An annotated PPG waveform of a heart beat that was captured during the experiment. The waveform captured by the Shimmer GSR+ PPG sensor is recorded in full scale and the y-axis is representative of the density of blood flowing underneath the PPG sensor.

the systolic pressure, the following depression is caused by the dicrotic notch, and the following peak is caused by diastolic pressure [162]. The systolic peak is caused by the contractions of the left ventricle of the heart, forcing blood out of the heart and into the aorta. The dicrotic notch is caused by the reversal of blood flow at the end of the systole, resulting in the closure of the aortic valve. The diastolic peak is caused by the relaxation of the ventricles, allowing the heart to refill with blood. Tracking the changes in blood volume allows signals from the sensor to be used in the estimation of heart rate, but the nature of the sensing method means that the data is not suitable for measures of heart rate variability or more complex information about circulatory performance [163]. The heart rate signal was measured via a PPG sensor that was secured to the finger of participant via a Velcro ring. The sensor was attached to one of the analogue inputs of the Shimmer GSR+.

Data from the shimmer sensor was streamed to a computer via Bluetooth, and was the processed and stored using a custom MATLAB application that wrapped the Shimmer MATLAB Instrument Driver version 2.8 [164, 165]. The instrument driver includes an algorithm that converts PPG data to heart rate, and this was used by the custom app to calculate heart rate. The measured data was stored in a CSV file including the absolute and relative timestamps, estimated heart rate, raw PPG signal as well as the differential analogue signals reported by the shimmer.

4.2.5 Experimental Design and Procedure

The experiment was conducted in a listening room with damped walls and a carpeted floor. Participants were seated in a chair near the centre of the room at a desk with a computer. Participants were instructed to read the information sheet presented on the computer, and fill out a participation consent form that followed. Following this initial screening stage, when participants had given informed consent to participate, the participants were invited to put on the headphones and then fit the shimmer GSR+ to their non-dominant hand. Once fitted and tested, participants were encouraged to continue with the experiment.

The experiment followed a repeated measures design as used in the literature [139, 127]. Participants blindly listened to the randomly ordered stimuli, after which they reported their affect and performed the classification task. The experiment procedure featured a set of 2 training intervals followed by a set of 10 test intervals. In the first 8 test intervals all test stimuli were presented to participants. Two of the 8 test stimuli were randomly selected and were re-played in the last 2 test intervals. The test process was managed by the design of the survey flow in the Qualtrics environment.

The test was designed to take approximately 30 minutes to complete in order to minimize listener fatigue. The main body of test took approximately 21 minutes to complete, with a further 10 minute period for preparing the participant for the test including the training intervals. Each interval of the test was set to a duration of 130 seconds, allowing participants an extra 30 seconds to complete the survey on top of the 100 seconds of listening. Participants were free to fill in the survey while listening to the soundscape, and were instructed to complete the survey as soon as possible after the soundscape had finished playing.

Throughout the experiment a safe playback system level was maintained. The experiment was always performed in the same listening room and with the same equipment. The playback system level was set to 90dB sound pressure level (dB SPL) at full scale. The system was calibrated using a Behringer ECM8000 measurement microphone, which itself was calibrated using a Tenma 72-7260 sound level meter calibrator. The audio playback system signal chain is presented in Figure 4.4.

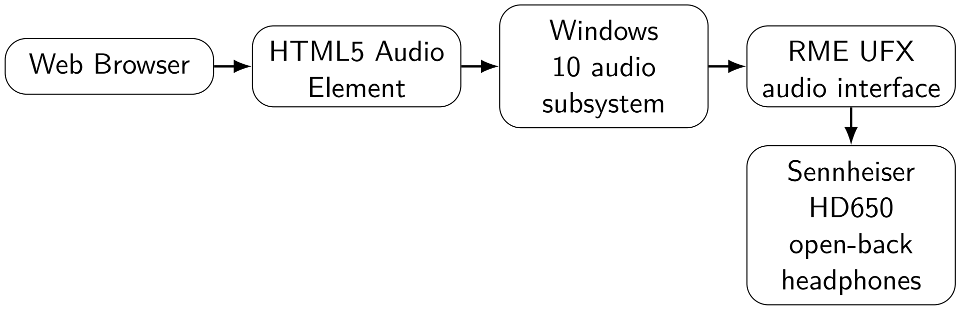


Figure 4.4: Signal chain of playback system.

4.3 Data Preparation

As the stimuli were presented in a random order via the Qualtrics web interface, it was necessary to identify which order the stimuli were presented in. During each experiment session, the audio signal played to the participant was also recorded. This file was then manually inspected, removing the first 2 training intervals and the last 2 repeat intervals. The shortened audio file was then cross-correlated against the original stimuli to identify the order of the stimuli. The maximum correlation coefficient and the lags indices were used to identify the position at which each stimuli occurred.

Data retrieved from the Qualtrics survey was downloaded as a CSV file, which was manually cleaned and prepared for analysis. This cleaning procedure involved removing unnecessary columns and rows, renaming the data columns, and converting the data set to the long format. Long format is an arrangement of data for repeated measures

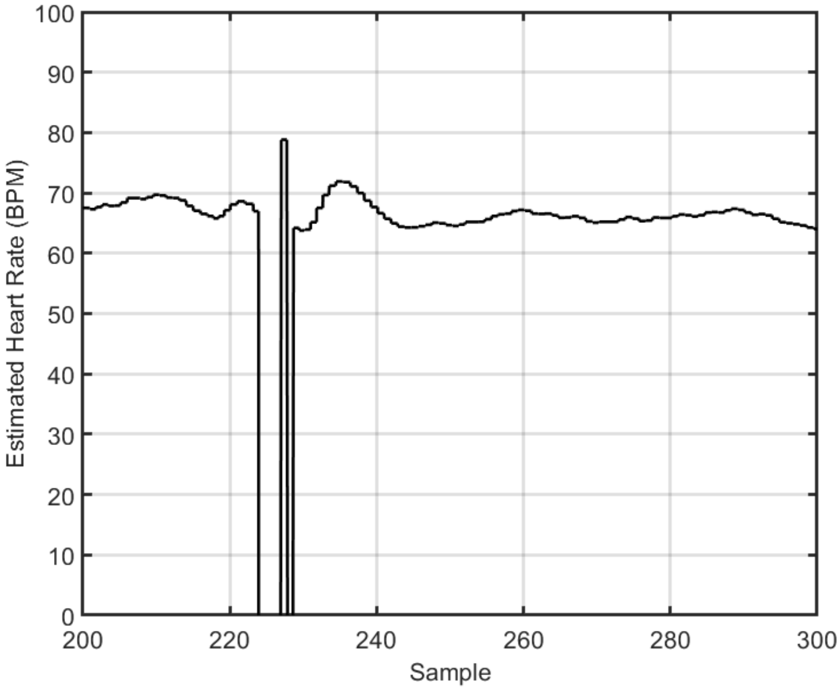


Figure 4.5: A sample of heart rate estimation from the Shimmer instrument driver.

experiments such that each row is an individual response by a participant, and each column is a feature that can be attributed to that response. The transformation was used because it is the form of data expected by some of the statistical analysis tools that were used in the analysis described below. Any personal or identifying data such as start and end times were removed from the data, leaving only an anonymised participant ID. The data was kept in the numeric format, and columns were added to indicate the stimulus type and the trial number. Finally, the practice trials and the repeat trials were removed from the data, leaving only the 8 test trials.

The quality of the physiological data that was collected via the MATLAB app was assessed. Unfortunately, the heart rate estimates from the Shimmer instrument driver for MATLAB [164] were prone to large errors and complete loss. An example of this loss is presented in Figure 4.5. Figure 4.5 shows an example of a period from a test interval in which noisy data has resulted in the Shimmer instrument driver returning null results. This localised period of estimated heart rate from within a test includes several estimates of 0 beats per minute, interspersed with a BPM estimate that is above the average of the rest of the sample. The HeartPy python library developed by van Gent *at al.* [161, 166] was used to gather heart rate estimates from the raw PPG data. Only 9 participants data contained complete sets of heart rate data for all 8 test trials once outliers has been removed. The make the heart rate, inter-beat interval and breathing rate estimates comparable across participants and intervals, the data for each participant was normalized using the following formula:

$$\hat{x} = \frac{x - \mu}{\sigma} \tag{4.1}$$

Where:

- μ is the mean of the data.
- σ is the standard deviation of the data.
- x is the raw data.
- $\hat{x}$ is the normalized data.

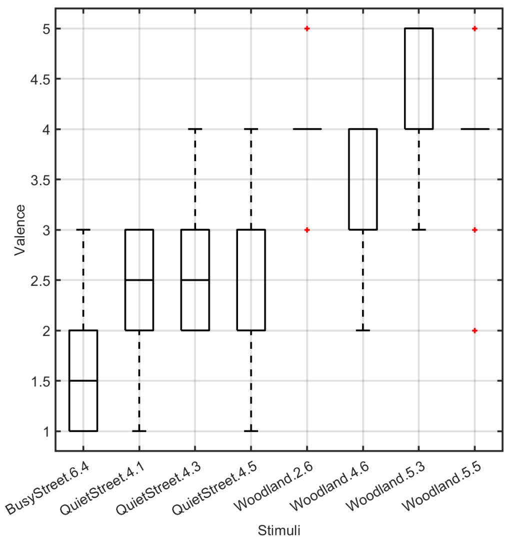
The full set of features returned by the HeartPy library included sets of heart rate variability measures. Heart rate variability measures assess the variation in the time between heart beats, and can used to assess the activity of the autonomic nervous system. However, these variables were not included in the analysis of this study because these measures require accurate estimates of features described in Section 5.2.4 that are not available from PPG data.

4.4 Results

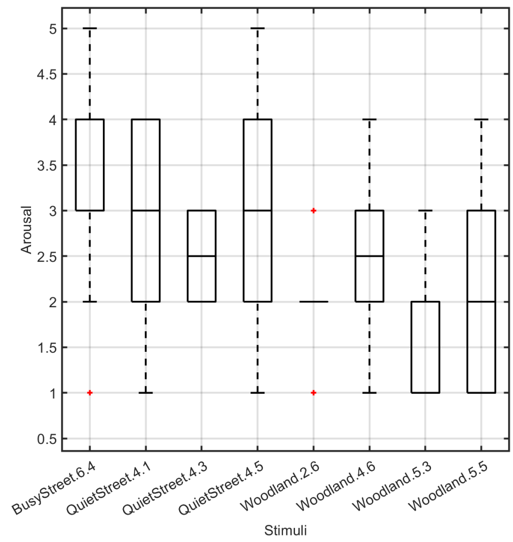
Due to the imbalanced nature of the data set, with complete physiological data from 9 participants and subjective data from 14 participants, it was not possible to use mixed models to analyse physiological and subjective data together. Therefore, the analysis of the physiological and subjective data was performed separately. Statistical analysis was performed using MATLAB [165].

4.4.1 Subjective Estimates

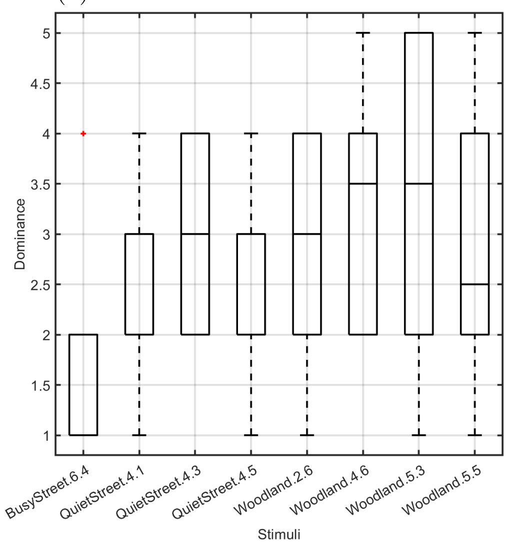
The distribution of the subjective responses of participants with respect to the stimuli are summarised by the series of box plots in Figure 4.6. Each box plot in Figure 4.6 is a graphical representation that summarises the distribution of the data of the set of responses for a particular parameter, for example self-reported Valence in Figure 4.6a, from all participants for all of the stimuli. Each box on the box plot represents the distribution of samples for one stimulus, and the top and bottom edges of the box represent the interquartile range or the range between the 25th percentile and the 75th percentile of the data. The line inside each box represents the median of the data which separates the lower and upper 50% of the data, showing how the middle 50% of the data is skewed compared to the median value. The dashed *whisker* lines extend from each box to the minimum and maximum values within 1.5 times the interquartile range, showing how data points outside the middle 50% are distributed. Outliers that are not within the box and whiskers and may not be representative of the general distribution of the dataset are presented as red crosses. Figures 4.6a, 4.6b and 4.6c summarise the distribution of affect scores from all participants across the stimuli. Figures 4.6d, 4.6e and 4.6f summarise the distribution of classification scores from all participants across the stimuli. The box plots of Figures 4.6a, 4.6b and 4.6c indicate that the state of affect experienced by the participants was varied across the stimuli. The *QuietStreet* and *BusyStreet* stimuli evoked reports of lower valence and higher arousal than the *Woodland* soundscapes. The self-reported dominance scores presented in Figure 4.6c do not appear to have as clear a difference across the stimuli as the valence and arousal scores. The natural and mechanical scores presented in Figures 4.6d and 4.6e indicated that the *QuietStreet* and *BusyStreet* stimuli were perceived as having a higher degree of mechanical sounds than the *Woodland* stimuli which were reported as having a much higher degree of natural sounds. The human scores presented in 4.6f are generally low



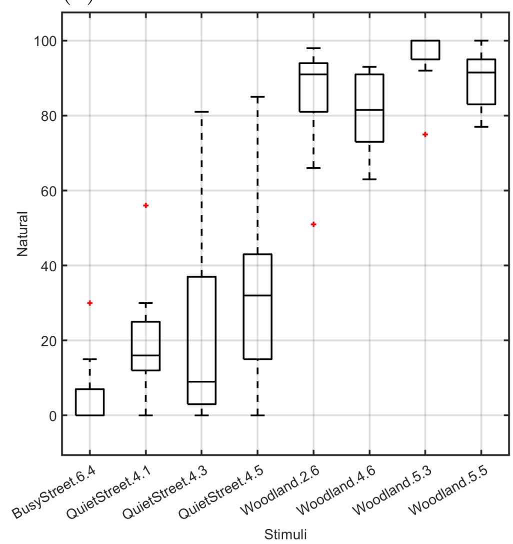
(a) Valence Scores Across Stimuli.



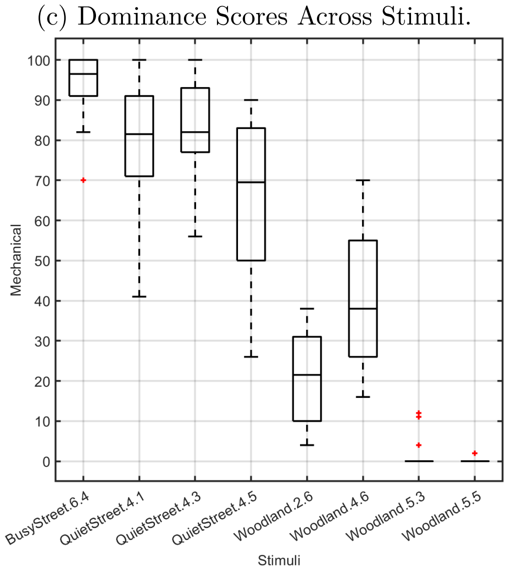
(b) Arousal Scores Across Stimuli.



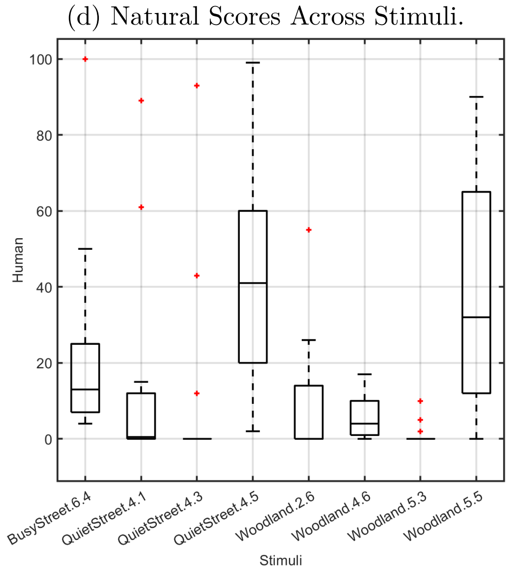
(c) Dominance Scores Across Stimuli.



(d) Natural Scores Across Stimuli.



(e) Mechanical Scores Across Stimuli.



(f) Human Scores Across Stimuli.

Figure 4.6: A series of box plots displaying the distributions of subjective estimates for each stimuli across all participants. Each box plot represents the distribution of a subjective descriptor across all of the stimuli.

with the exception of the *QuietStreet.4.5* and *Woodland.5.5* stimuli which were reported as having a higher degree of human sounds than the other stimuli. To accommodate for the ordinal nature of the subjective data, the Kruskal-Wallis test was used to assess the statistical significance of the differences between the subjective and psychoacoustic factors. Table 4.2 shows the results of the Kruskal-Wallis test for the subjective data between all of the stimuli.

Variable Type	**Significance p**	**Size χ^2**
Valence	$3.3246e^{-12}*$	68.2554
Arousal	$6.0973e^{-06}*$	36.398
Dominance	0.0516	13.9751
Natural	$1.3654e^{-16}*$	89.7915
Mechanical	$5.6949e^{-18}*$	96.498
Human	$1.1702e^{-08}*$	50.4657

Table 4.2: Results of the Kruskal-Wallis test for the subjective data between all of the stimuli.

Table 4.2 shows the result for the Kruskal-Wallis test for each of the subjective factors surveyed in the experiment, on a per-stimuli basis. This comparison was performed to determine if there was a significant difference between the stimuli for each of the subjective factors. Each of the significance levels in Table 4.2 is highlight with an asterisk $*$ to indicate if it reaches the significance level of $\alpha = 0.05$. The results presented in Table 4.2 show that there was a significance difference in almost all of the subjective factors between the stimuli, indicating that the stimuli had a statistically significant effect on the state of affect experienced by the participants when experiencing each soundscape. The only subjective factor that didn't vary between the stimuli was the dominance score from the SAM.

In Table 4.1 each stimuli are allocated to one of two groups, indicating the if the NDSI score of that stimuli was positive or negative. To compare these two groups, the Kruskal-Wallis test was used to assess the statistical significance of the differences between the subjective and psychoacoustic factors. Table 4.3 shows the results of the

Kruskal-Wallis test for the subjective data between the sets high and low NDSI stimuli groups.

Variable	**Significance**	**Size**
Type	**p**	χ^2
Valence	0.00086657∗	11.0929
Arousal	0.0065252∗	7.3992
Dominance	0.0875	2.9197
Natural	$1.3988e^{-05}$∗	18.8708
Mechanical	0.001145∗	10.577
Human	$3.675e^{-06}$∗	21.4273

Table 4.3: Results of the Kruskal-Wallis test for the subjective data between the high and low NDSI groups.

Table 4.3 shows the result for the Kruskal-Wallis test for each of the subjective factors surveyed in the experiment, on a per-stimuli basis. This comparison was performed to determine if there was a significant difference between the two groups of stimuli for each of the subjective factors. Each of the significance levels in Table 4.3 is highlight with an asterisk∗ to indicate if it reaches the significance level of $\alpha = 0.05$. The results presented in Table 4.3 show that all of the subjective factors, except for dominance, had a significant difference between the high and low NDSI groups of stimuli.

To identify if each of the factors in the the survey were correlated, Kendall's rank correlation was used to test each combination of the subjective estimates. The results of the Kendall's rank correlation test are summarised in Figure 4.7 which presents a graphical representation of the correlation matrix of the subjective estimates. Each subplot in Figure 4.7 shows the correlation of two subjective estimates that are labelled on each axis at the left and bottom of the figure. The subplots on the diagonal of the matrix show the distribution of each subjective estimate as a histogram. Each of the off-diagonal subplots show the correlation between the two subjective estimates as a scatter plot with a least-squares reference line. Each subplot shows the correlation coefficient in the upper left corner, and statistically significant results, results within the standard confidence interval of $p < 0.05$ are coloured in red. The correlations presented in Figure

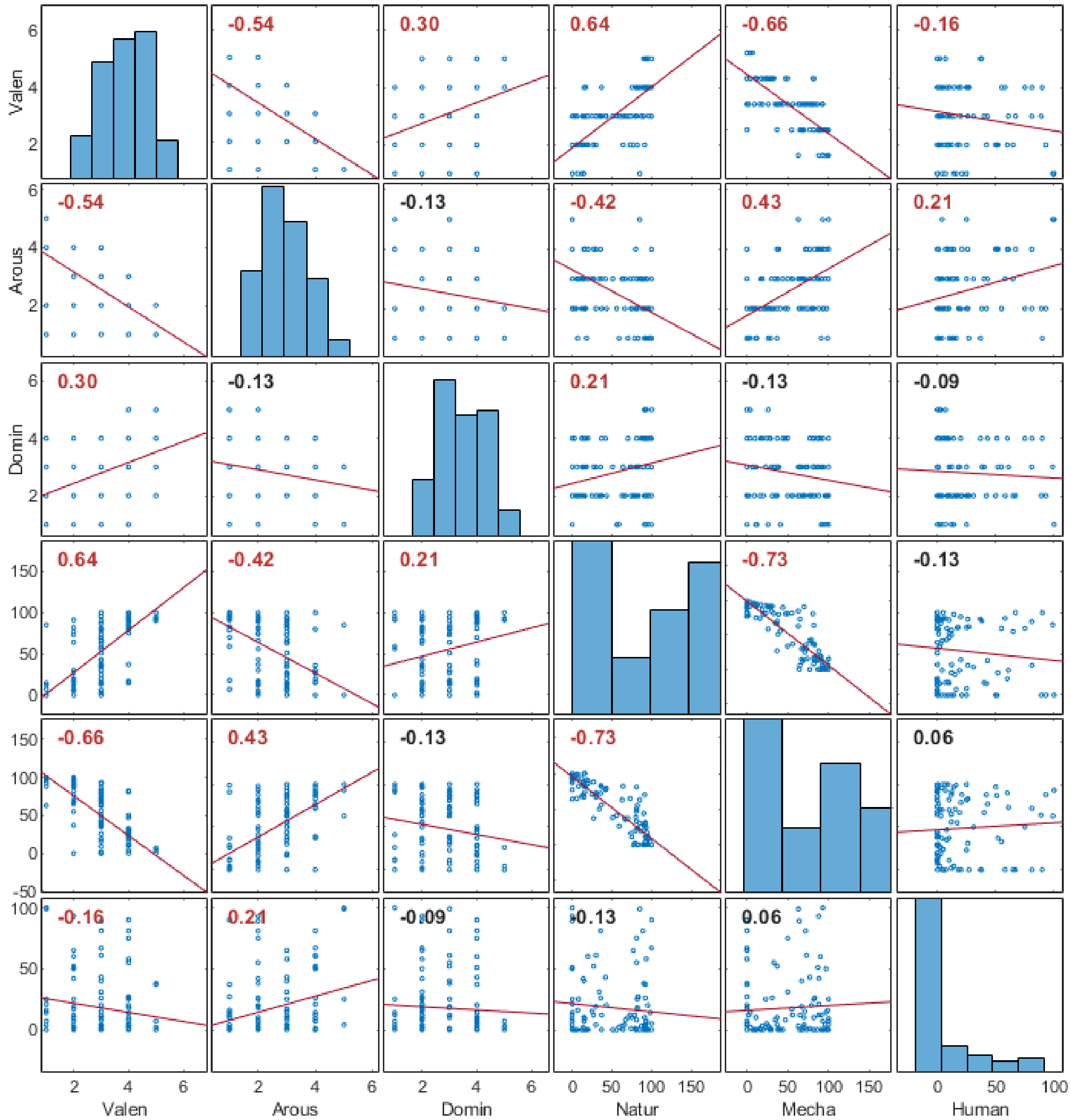


Figure 4.7: A matrix of correlation plots for each of the subjective estimates that were surveyed.

4.7 show that the valence, arousal and dominance estimates are correlated with each other and with each of the natural and mechanical scores. However, dominance scores were not correlated with arousal, mechanical or human scores. Neither were human scores correlated with dominance, natural or mechanical scores.

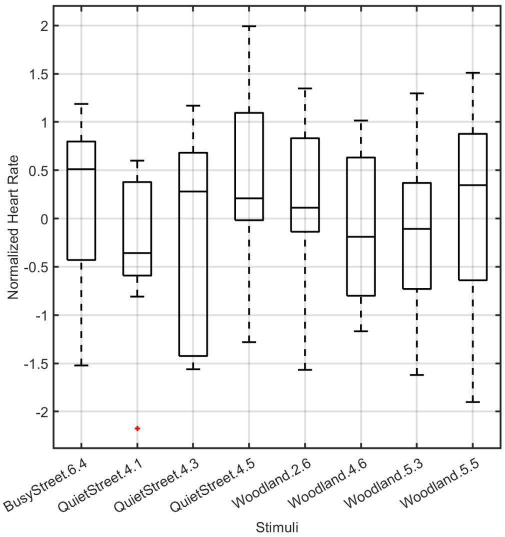
4.4.2 Physiological Responses

Figure 4.8 presents a boxplot of the distribution of physiological measures across all stimuli. The boxplot presented in Figure 4.8a shows the distribution of the normalized average heart rate of all participants for each stimuli. The data presented in Figure 4.8a shows that although the distributions of heart rate are different for each stimuli, the median heart rate is similar for several of the stimuli. The boxplot presented in Figure 4.8b shows the distribution of the mean absolute deviation of heart rate for an experiment interval for each stimuli. That data presented in Figure 4.8b shows that the distribution of the mean absolute deviation of heart rate is similar for each stimuli, with differences in the inter-quartile range between the stimuli. The boxplot presented in Figure 4.8c shows the distribution of the normalized average heart rate of all participants for each stimuli. That data presented in Figure 4.8c shows that six of the stimuli evoked similar distributions of breathing rate with median values around or slightly above zero, while the other two have median values between -0.5 and -1.0. Table 4.4 shows the results of the Kruskal-Wallis test for the physiological data between all of the stimuli.

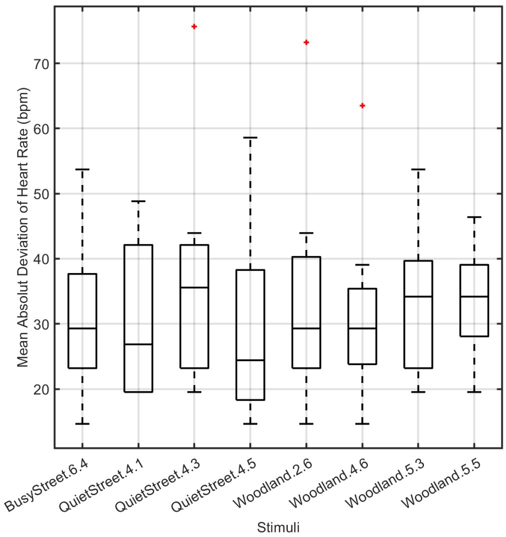
Variable	**Significance**	**Size**
Type	**p**	χ^2
Normalized Average Heart Rate	0.71432	4.553
Mean Absolute Deviation of Heart Rate	0.93843	2.3432
Normalized Average Breathing Rate	0.51947	6.175

Table 4.4: Results of the Kruskal-Wallis test for the physiological data across all stimuli.

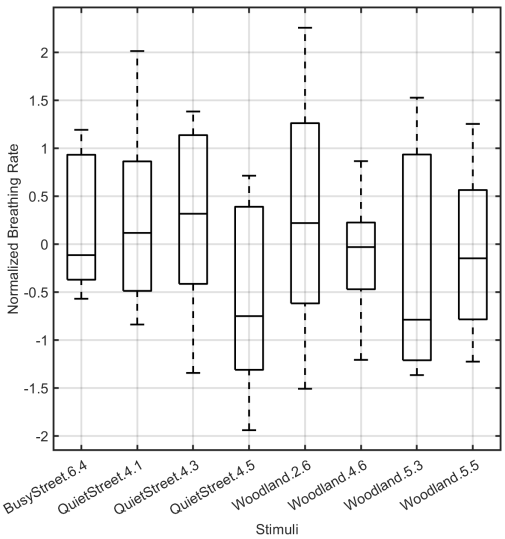
None of the results presented in Table 4.4 are statistically significant, suggesting that there is no significant difference in average average heart rate between all of the stimuli on an individual basis. Table 4.5 shows the results of the Kruskal-Wallis test for the



(a) Distribution of normalized mean heart rate across all stimuli.



(b) Distribution of mean absolute deviation of heart rate across all stimuli.



(c) Distribution of normalized mean estimated breathing rate rate across all stimuli.

Figure 4.8: Distribution of heart rate for each stimuli across all participants.

physiological data between the two stimuli groups.

Variable	**Significance**	**Size**
Type	**p**	χ^2
Normalized Average Heart Rate	0.018265	0.8925
Mean Absolute Deviation of Heart Rate	0.27113	0.60257
Normalized Average Breathing Rate	0.15538	0.69345

Table 4.5: Results of the Kruskal-Wallis test for the physiological data between stimuli groups.

The results presented in Table 4.5 show that there is a statistically significant difference in the normalized average heart rate between the two groups of stimuli.

4.5 Discussion

In this experiment 14 participants evaluated a randomly ordered set of 8 soundscapes, performing a survey of classification scores and self-reported affect. The heart rate of the participant was measured throughout the experiment used a PPG based sensor that was strapped to the index finger of the participants. From the PPG data, the average heart rate, mean absolute deviation of heart rate and average breathing rate were estimates for each experimental interval.

Statistically significant relationships between soundscape classification and affect self-report were identified, with higher natural and lower mechanical scores reflecting higher reported valence and lower reported arousal. These results were observed with both correlation and analysis of variance, and were observed between the individual stimuli as well as the aggregated stimuli groups. Valence and arousal scores were negatively correlated, suggesting that for this set of stimuli no soundscapes evoked a positive valence and arousal response. Further, natural sounds were positively correlated with valence scores, and negatively correlated with arousal scores. Conversely mechanical sounds were negatively correlated with valence scores, and positively correlated with arousal scores. These results are consistent with previous research that has identified a negative correlation between valence and arousal scores [123, 139]. The results from this experiments show that the experimental method and the selection of stimuli was successful in eliciting varying states of affect in the participants. However, the results from analysing the dimension of dominance from the SAM survey were not statistically significant, despite the dimensions of valence and arousal being exercised. This result shows that dominance may not be a suitable dimension to use when evaluating soundscapes in the context of a laboratory based directed listening experiment where participants passively listened to the stimuli without interacting with them. This makes sense as dominance is a measure of control and action potential, and in this experiment the participants had no control over the quality of the stimuli. This is the same finding as Stevens *et al.* [123] who suggested that including dominance in assessments of affect might in the context of soundscape assessment might not provide much more information about the emotional state of the participants than excluding the dimension of dominance.

The results for the physiological data show that there were statistically significant differences in average heart rate between the two larger groups of stimuli. This is a promising result as it shows that even with a small sample size and a relatively short experiment, differences in heart rate can be observed between different types of soundscapes. However, the fact that this effect was not observed between stimuli on an individual basis may suggest that the effect was both not large between the groups of stimuli, and weak enough to not be observed with the current experimental design and with this small sample size. This result may be similar to that observed by Hume and Ahtamad [130] how observed statistically significant changes in heart rate in participants who were exposed to a wide variety of sounds. However, none of the other 5 experiments described in [127] observed statistically significant changes in average heart rate, suggesting that the effect observed in this experiment may be anomalous.

Overall, the results of this experiment support the hypothesis describe in Section 1.2 by showing that soundscapes with a variety of classification scores evoked different states of affect in the participants, and differences in average heart rate. The strength of this evidence is strong for the affect self-report data, as several results exploring the variance of soundscape classification and self-reported affect were statistically significant.

However, the evidence supporting the hypothesis that soundscapes evoke physiological responses was not as strong, as only one statistically significant effect was observed with average heart rate between the groups of aggregated stimuli.

4.5.1 Limitations

This experiment had several limitations that should be addressed when designing further studies of physiological responses to soundscapes. The first limitation is the small sample size of 14 participants. This sample size was chosen due to the time constraints of the experiment, and the fact that the experiment was conducted in a laboratory environment, though the repeated measures design did maximize the amount of data that was gathered per participant. However, this sample size is small compared to other studies of physiological responses to soundscapes, and may have contributed to the lack of statistically significant results for the physiological data.

The second limitation is the quality of the physiological data. There were several issues and challenges when attempting to capture physiological data in this experiment. Data gathered from PPG sensors can be very noisy, and the quality of the data can be affected by the placement of the sensor on the participant [166, 162]. Further, the data gathered from the PPG sensor was not validated against a higher quality sensor such as an Electrocardiogram (ECG). There were two instances where the physiological data that was gathered was not usable, and one where a participants physiological data was successfully not captured at all. Figure 4.9 presents an example of momentary corruption during the recording of PPG data. The data plotted in this examples was filtered by the Shimmer instrument driver for MATLAB prior to being stored as the data was being recorded. The plot presented in Figure 4.9 shows a PPG signal with large peaks that features a sudden low frequency ramp and disturbance to the signal. This disturbance is likely due to the participant moving their finger, and the sensor losing contact with the skin. As described by Van Gent *et al.* incorrectly estimating a false peak in PPG data can result in the incorrect interpretation of the behaviour of the heart which could be more detrimental than not estimating a peak at all [166]. The presence of erroneous

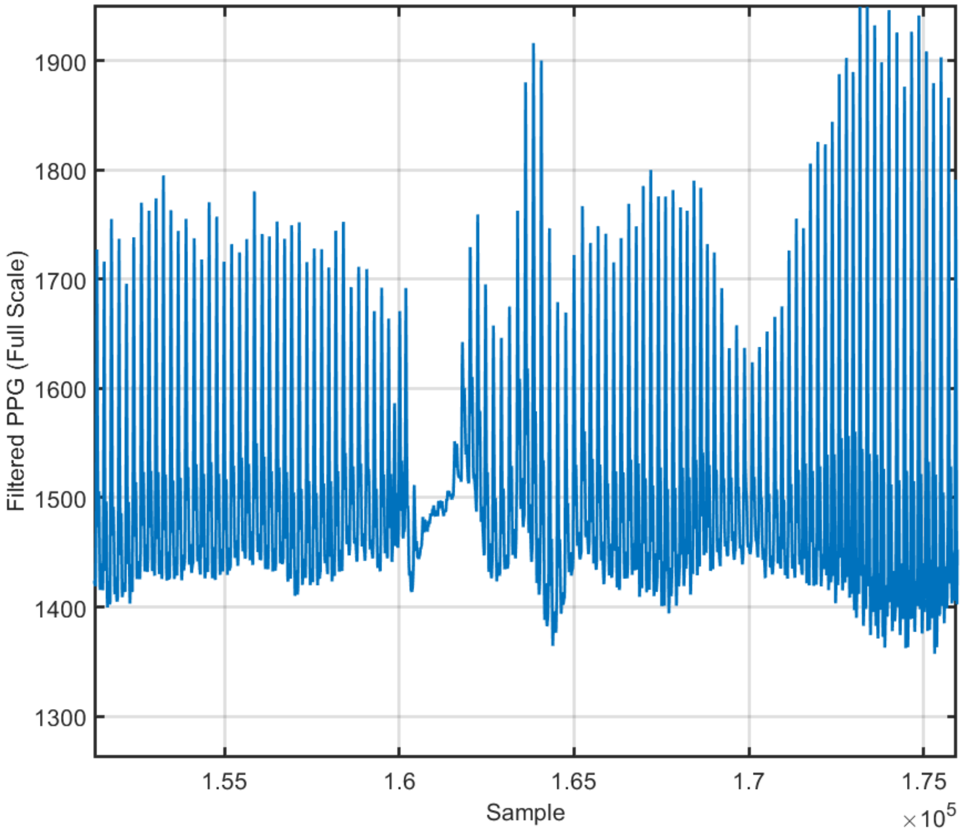


Figure 4.9: An Example of corruption in a recording of a PPG signal that was gathered during the experiment.

peak estimates may have contributed to the lack of statistically significant results for the physiological data, as outliers in the data were removed prior to analysis.

A third limitation is the lack of a wider range of physiological measures. The three measures used in this experiment were chosen due to their ease of use, but the study could have taken advantage of the galvanic skin response sensors that are part of the Shimmer GSR+. Galvanic skin response or electrodermal activity (EDA) is a measure of the electrical conductance of the skin, and is a measure of the activity of the sweat glands [167]. This measure is often used as a measure of arousal, and has been widely used in psychology research [168]. The original intention of this experiment was to include GSR as a measure, but due to the time constraints of the experiment it was decided to focus on heart rate measures and to consider using GSR in future experiments.

Another limitation in the design of this experiment was the lack of control over the timing of the experiment. Each experiment interval was limited to 130 seconds and the stimuli were 100 seconds long. There was no structure or control to the timing of the experiment interval, and the participants were not given a specific rest or recovery time between intervals, and this may have led the participants to transition from a state of recall when filling in the survey into a state of anticipation when listening to the next stimulus. In order to more accurately capture the physiological responses to the stimuli, future experiments should include a more structured timing of the experiment intervals, and should include a rest period between intervals.

Finally, the experiment included a limited set of stimuli, only including three types of environment, some of which came from the same location. By not including a wider variety of soundscapes with a broader range of natural and mechanical sounds and contexts, the range of experiences and therefore the emotions that may have been experienced by the participants may also have been limited. This could have resulted in less generalisable results than if a wider variety of stimuli were included. As described in Section 4.2 the stimuli used in this experiment were selected in order to maximize the proportions of natural and mechanical sounds within the given stimuli, in order to elicit greater subjective and affective responses from the participants. Future experiments should include stimuli that represent a wider range of environments and sound classes in order to identify if the results of this experiment generalise across a greater variety of environments and emotional states. Further experiments should also identify if a three member soundscape classification scheme is appropriate across a wider variety of environments, or if a more granular classification scheme is beneficial.

Summary

In this experiment 14 participants took part in a directed listening experiment that was intended to support the hypothesis described in Section 1.2. The soundscape stimuli used in this experiment were selected in order to maximize the range of natural and mechanical sounds, and were presented to the participants in a random order. The results of this experiment were encouraging, showing support for the hypothesis that soundscapes evoke affective and physiological response. To develop this further, a subsequent experiment described in Chapter 5 was performed that improves on the experiment methodology introduced in this chapter.

Chapter 5

Assessment of Soundscapes using Self-Report and Physiological Measures

In Chapter 4 an experiment was described that aimed to assess the relationship between soundscape perception and average heart rate. The selection of stimuli elicited reports of affect and soundscape class that varied with stimuli, and there was some evidence that changes in heart rate were present in response to the stimuli. One of the primary challenges in the execution of that study was in the gathering and analysis of physiological data, as the quality of the data was poor and the sample size of the data was very small. This chapter describes a second that was intended to improve on the first by improving the experiment methodology and using a higher quality heart rate sensor.

5.1 Introduction

In the experiment described in Chapter 4 fourteen participants evaluated a randomly ordered set of 8 soundscapes. The participants performed a survey of classification scores and self-reported affect. The results of the experiment showed that there were statistically significant differences in the self-reported affect scores across the stimuli, showing that the experiment methodology was sufficient for showing that soundscapes evoke states of affect. This result was observed both for the individual stimuli and for the aggregated stimuli groups, showing that selecting stimuli on the basis of NDSI score was effective for eliciting varying states of affect and selecting soundscapes classified as being predominantly natural or mechanical. Further, the results showed that natural sounds were positively correlated with valence scores, and negatively correlated with arousal scores, and mechanical sounds were negatively correlated with valence scores, and positively correlated with arousal scores. These results are consistent with previous research [123, 139], further supporting the experimental methodology and the hypothesis established in Section 1.2. The hypothesis stated that the experience of soundscapes classified as featuring natural and mechanical sound sources evoke physiological and affective responses in listeners.

The results for the physiological data from the experiment described in in Chapter 4 were less conclusive. The results showed that there was statistically significant difference in average heart rate between the two aggregated groups of stimuli, but not between the individual stimuli. Because this difference in average heart rate was not observed between the individual stimuli, it may have been that the effect was too small to be accurately observed with the small sample size used, and this may have been a type 2 error [169]. A type 2 error is a false negative error, falsely failing to reject the null hypothesis when an effect is present but the evidence is insufficient to confidently reject the null. Further, many of the studies highlighted in [127] that include the measurement of average heart rate in soundscape studies failed to find an effect between average heart rate and soundscape perception, even when including much larger sample sizes.

The goal of the experiment described in this chapter is to test for evidence that further supports the hypothesis established in Section 1.2 by improving on the experiment methodology described in Chapter 4. The first limitation in the experimental method that will be addressed in this experiment will be to improve control over the timing of the experiment. It is suggested in Section 4.5.1 that the timing of the experiment intervals may have influenced the results of the experiment, as the time allowed for participants to listen, recall and rest was not controlled.

The second limitation in the experiment described in Chapter 4 that will be addressed in this experiment will be to improve the selection of stimuli by including a wider range of soundscape types and locations. Three of the stimuli from the first experiment were recorded in the same location, and this may have influenced the results of the experiment.

The third limitation that will be addressed in this experiment is the quality of the physiological data. The data gathered from PPG sensors was noisy and prone to artefacts from movement. Further, there was not a lot of evidence in the soundscape literature that average heart rate alone is a good indicator of affective state. In order to address these limitations, this experiment will use a higher quality heart rate sensor that measures ECG data, and will include measures of heart rate variability in the analysis. Heart rate variability (HRV) metrics are prevalent in affective computing and psychological research, in part thanks to the improvements in the availability and affordability of electrocardiogram (ECG) heart rate monitors [170, 171]. Heart rate variability measures are interpretations of a series of inter-beat intervals, the time periods between successive heart beats.

There are several theories that describe how HRV might reflect the psychophysiological state of an individual, suggesting a causal link between changes in affect and the behaviour of the systems that regulate functions of homeostasis including heart rate variability. This causal link is facilitated by the balance between the sympathetic and the parasympathetic nervous systems (PNS), and it is theorized that the periodicity of the activation of the vagus nerve also known as vagal tone is an indicator of the activity of the parasympathetic nervous system. The parasympathetic nervous system is theorized to be representative of cognitive and emotional function [172].

Two key theories that support the connection between emotion regulation and the function of the autonomic nervous system are the Polyvagal and Neurovisceral Integration theories. The Polyvagal theory was proposed by Porges [173] as a model of the neural regulation of the autonomic nervous system that also supports the relation between autonomic function and primary emotions [174]. This theory suggests that the adaptive mechanism of heart rate regulation are mediated by neurological mechanisms that are influenced by the environment and associated with behaviours include fight-or-flight and social interactions. The Neurovisceral Integration theory proposed by Thayer and Lane provides a model of a network of neurological systems that is theorized to be important for system regulation. This network is used to control several systems including the regulation of heart rate including heart rate variability [172]. Shaffer *et al.* suggest that healthy natural heart beat regulation is both periodic and highly complex [126]. Alternatively, under stress heart rate increases and HRV decreases as the vagal breaking mechanisms releases, allowing the heart rate to increase. Decreased HRV has also been associated with reduced PNS activation and has been observed in people suffering from stress, anxiety and panic [126]. There are several HRV metrics that reflect different statistical properties of the inter-beat interval series. Time domain measures are more suited to very-short-term and ultra-short-term HRV analysis (generally considered less than 5 minutes periods under analysis), as spectral methods require longer measurement periods to accurately reflect low frequency and very low frequency phenomena [175]. A committee report on heart rate variability recommended measurement periods of at least ten times the period of the lowest frequency component that is intended to be observed through spectral analysis, and at least 1 minute for high frequency HRV [176].

The systematic review of research studying psychophysiological factors in the context of soundscape published by Erfanian *et al.* [127] highlighted one published article that included HRV measures. Average heart rate measurement was included in all studies but one, in which high frequency heart rate variability measures were included [131]. The study published by Alvarsson *et al.* [131] aimed to investigate the restorative potential between nature sounds and environmental noise at different sound pressure levels. The results of the study show that the presence of natural soundscape improved the rate of stress recovery over environmental noise, regardless of the sound pressure level of the stimuli. However, this result was not supported by the analysis of HRV which showed no significant variance based on the type of soundscape stimuli that was presented. The authors also noted their results disagreed with prior work in this respect, arguing that the measures they used were only indicative of parasympathetic activity. Alvaarson *et al.* reported that they used spectral analysis following methodologies described in [177] and [178].

5.2 Methods

This experiment included many of the same assumptions as the prior experiment, and including the following:

- Soundscapes reported as having a higher proportion of mechanical or natural sound sources would be reputed as having different subjective qualities, such as the presence or absence of environmental noise.
- The expectation is that stimuli with differing classification scores would elicit different states of affect which would reflect the physiological and subjective measures.
- Based on the reports of natural soundscapes being restorative or eliciting positive valence, a further expectation would be that soundscapes classified as being more natural would elicit higher valence and higher HRV.
- Another assumption was that ultra-short-term time domain heart rate measures are representative of changes in mood and emotion elicited by experiencing soundscapes.

This experiment was approved by the University of York Department of Electronics Research Ethics Committee, and the documentation used in the application for ethical approval is included in Appendix A.

5.2.1 Participants

Participants were recruited from groups of audio engineering students via email. All participants had training in subjective testing, and could be considered expert listeners. A total of 15 participants were recruited to take part in this experiment. Of the participants 12 identified as male, 1 as female and 2 preferred not to say. The average age of participants was 26 and the standard deviation of the age was 5. The participants were screened by self-report for the following exclusion criteria:

- Abnormal or damaged hearing.
- Skin damage or known reactions to materials on the heart rate monitor.
- Known heart conditions, ailments or using medications that might directly effect heart rate.

5.2.2 Experimental Stimuli

The experimental stimuli were selected from the EigenScape dataset as described in Section 4.2.2. A stimulus length of 40 seconds was chosen for this experiment in order to maximize the length of stimulus and the proportion of recall and recovery time for each test interval, while keeping the length of the test to within 30 minutes to avoid listener fatigue. The EigenScape dataset was segmented into contiguous 40 second long clips, from which the test stimuli were then selected. All of the stimuli were screened for markers that could be used to identify individual people such as clear discernible speech. From this set the stimuli were then selected by algorithmically evaluating the proportion of natural and mechanical sounds in each clip, which was calculated using the Normalized Difference Soundscape Index described in Section 4.2.2. The NDSI score was calculated for the 0^{th} order omnidirectional W-channel of each soundscape sample.

The dataset contains soundscapes from 8 location types, and this was reduced down to only include clips from the 4 locations with the greatest variance in NDSI. From each of these four locations, the 2 clips at the limits of the centre quartiles of the NDSI scores were selected as test stimuli. From each location type the clip at the limit of the positive quartile of NDSI score is an example from the set that is maximally natural and minimally mechanical and is therefore considered a high NDSI example. Conversely the clip at the limit of the negative quartile of NDSI scores is an example of a soundscape from that location type that is maximally mechanical and minimally natural, this clip is therefore a low NDSI example of that location type. The two training stimuli were the two samples in the data set that aren't in an environment type used in the test set, and with an NDSI score closest to zero. The selected test stimuli are summarised in Table 5.1. In the dataset recordings are organized by location type i.e. beach, woodland. Individual recordings and locations can be identified using a map provided by the author [150]. In Table 5.1 the location type and recording number are both from the dataset structure. The clip number identifies which contiguous 40 second slice of the recording number is used. The NDSI group is provided to identify whether the clip has a higher proportion of natural sounds (a more positive NDSI score) or mechanical sounds (a more negative NDSI score).

Table 5.1: Stimuli Selection Summary.

Location Type	**Recording Number**	**Clip**	**NDSI Group**	**NDSI Score**	**Stimuli**
Woodland	5	5	High	0.90	1
Woodland	7	1	Low	-0.86	5
Train Station	4	5	High	-0.27	2
Train Station	6	3	Low	-0.81	6
Park	3	3	High	0.41	3
Park	2	2	Low	-0.75	7
Beach	7	4	High	-0.03	4
Beach	2	3	Low	-0.85	8

Still images of these locations were not published by Green *et al.* as part of the EigenScape dataset [150]. However, the stimuli were published as part of the audio article publication of the experiment detailed in this chapter at [9]. The selected stimuli were converted from B-format first order Ambisonic to 2 channel binaural which could be easily reproduced over headphones. The B-format to binaural conversion was performed using the binaural decoder VST plugin that is part of the IEM plug-in suite [155] as used in the prior experiment and described in Section 4.2 .

The stimuli were peak normalized to ensure a consistent relative level is maintained. No further processing was performed, and though there was some low frequency rumble and wind noise in the recordings this was not compensated for. The loudness of each stimuli is presented in Figure 5.1. The loudness of each stimuli is given in loudness units full scale (LUFS) which is a standardized time-weighted and gated unit of the loudness of a signal relative to digital full-scale representation. The standard suite of loudness meters is calculated for each stimulus according to the algorithm defined in ITU-R BS.1770-4 [179], using the loudness meter built into MATLAB [165]. Figure 5.1 presents each of the loudness metrics for the stimuli in the form of a bar chart. The

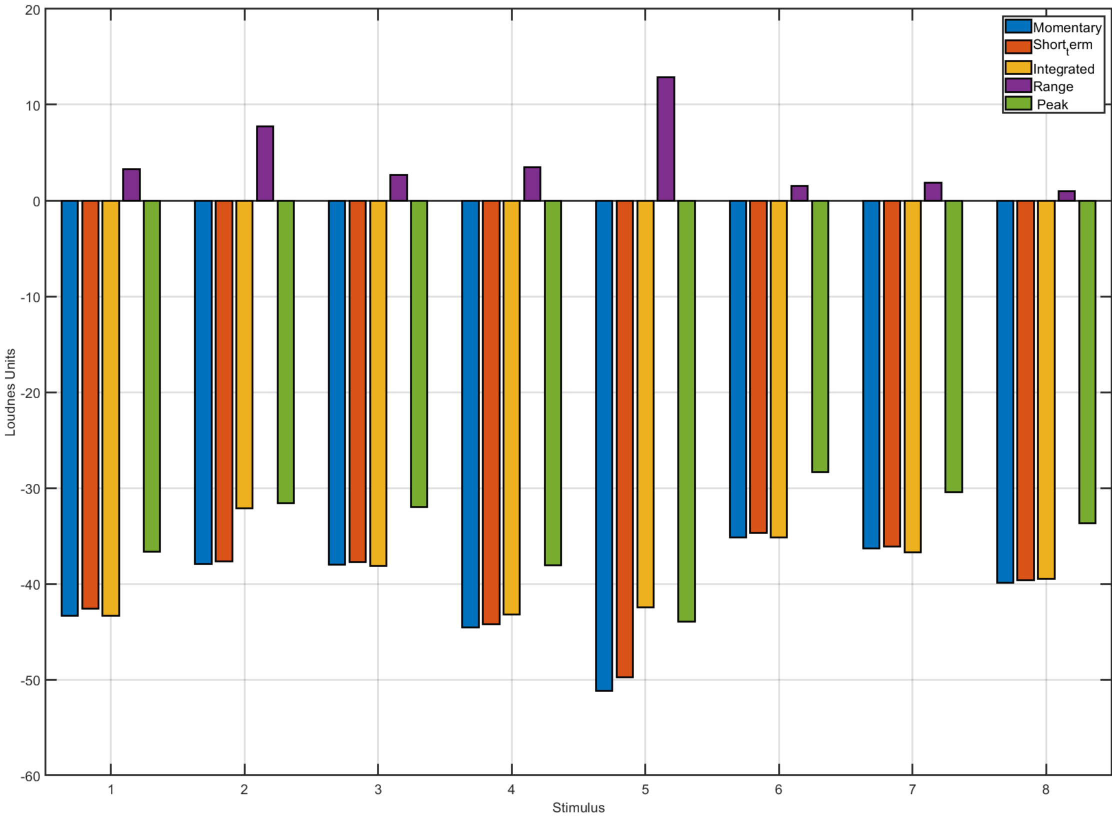


Figure 5.1: Bar chart of the EBU loudness metrics of the stimuli.

momentary, short-term and integrated loudness values in Figure 5.1 are presented in full

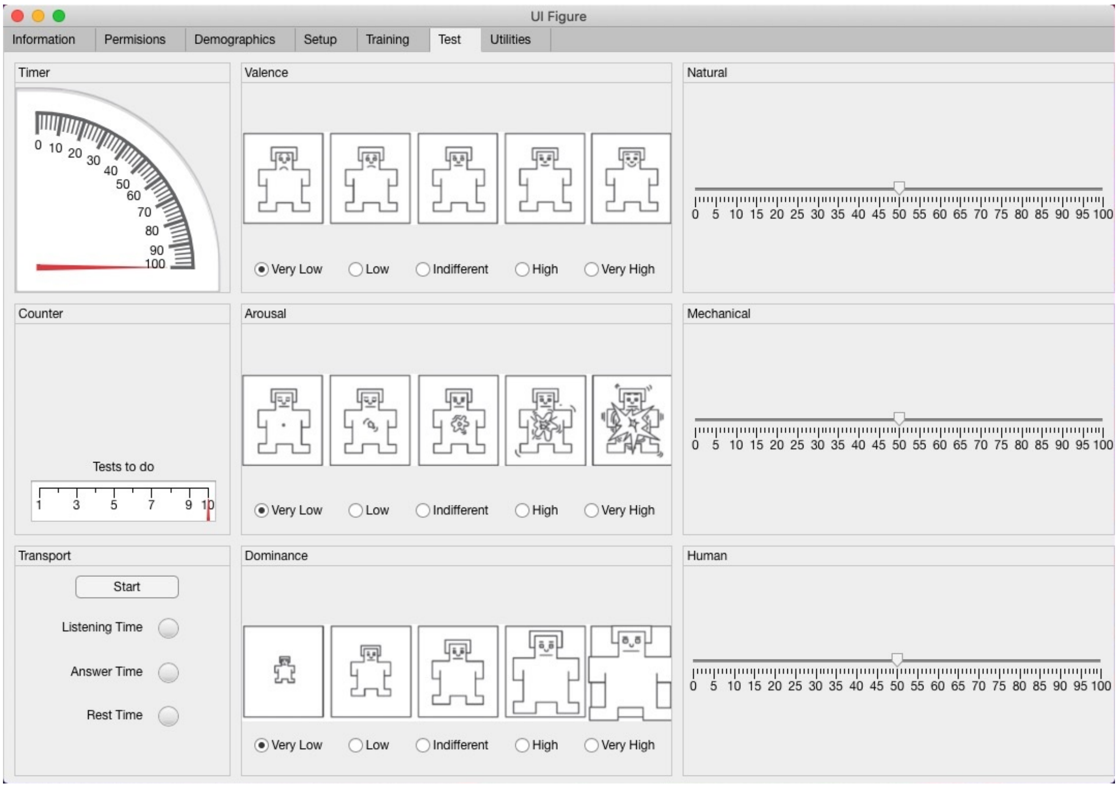


Figure 5.2: Listening test user interface.

scale loudness units. The range of the loudness of the stimuli are presented in Figure 5.1 in loudness units. The peak of the loudness of the stimuli are presented in Figure 5.1 in dB true peak, the algorithm for calculating this metric is described in [179].

5.2.3 Data Collection Instruments

As in the previous experiment, two forms of data gathering were used in the experiment, physiological sensing and self-report. For each test interval participants were asked to complete a survey based on their experience of the soundscape. To have greater control over the participants timing in this experiment, the survey was implemented as a MATLAB application [165]. Figure 5.2 presents the test page of the user interface. The survey featured two components as represented in Figure 3.4:

- A five-point self-assessment manikin in which participants report their affect.

Figure 5.3: The Polar H10 Heart Rate Monitor and chest strap, from [184].

- A classification task in which participants are tasked to report the proportions of sound sources present in the soundscape.

These are the same components used in the prior experiment, and are described in more detail in Section 4.2. Participants were asked to disregard the numerical values associated with the three classification score sliders as in the experiment described in Chapter 4, but at the time of making this graphical user interface the numerical display for this type of slider could not be hidden from the user.

5.2.4 Physiological Measurement

The heart rate monitor used in the experiment was a Polar H10 Electrocardiography (ECG) based chest strap monitor [180]. Electrocardiography is the process of measuring the electrical activity of the heart, and is a common method of measuring heart rate [181]. The H10 is popular, robust, relatively inexpensive, and has been shown to provide data of a quality similar to Holter style monitors [182, 183]. Figure 5.3 presents an image of the Polar H10 and chest strap, from [184]. One of the primary advantages of using the Polar H10 is that the full QRS complex of each heart beat is recorded. The QRS complex is the collection of electrical signals generated by activation of the different muscles of the heart as it beats. Figure 5.4 presents an example of a QRS complex that was measured

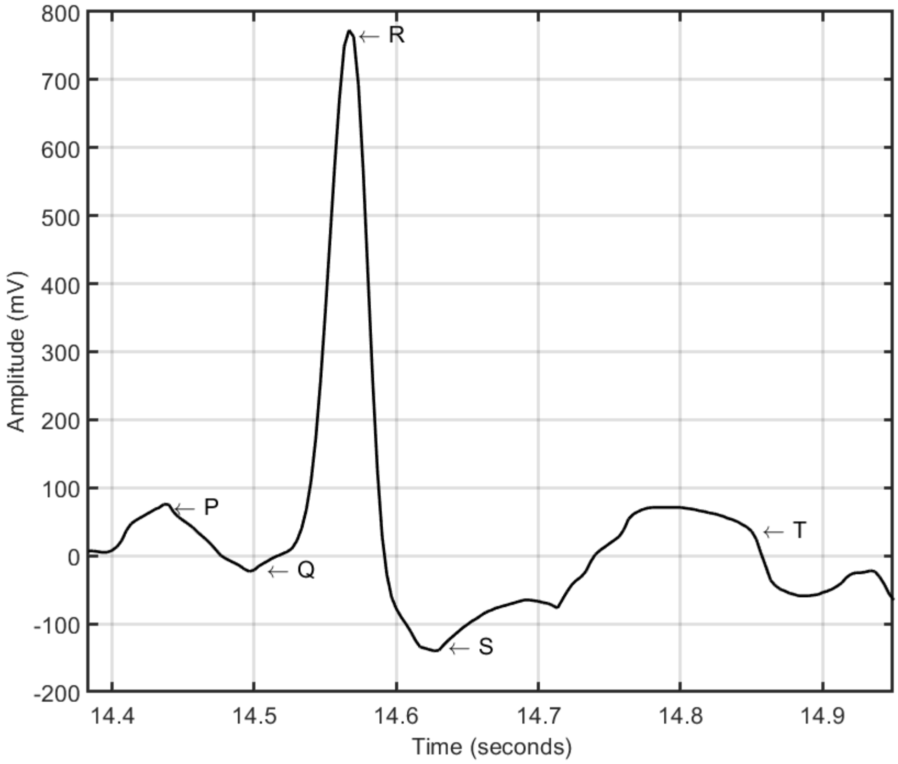


Figure 5.4: Plot of the QRS complex of a heart beat measured with an electrocardiogram.

with an electrocardiogram. The data used for this plot was published as part of the PhysioNet2017 challenge for classifying a heart beat using statistics, signal processing and machine learning [185]. The R wave labelled in the QRS complex in Figure 5.4 is the large spike that occurs in the ECG due to the almost simultaneous depolarization of the left and right ventricle muscles. The time between R waves is used to determine a person's heart rate, and before the proliferation of digital electronics in medicine this was achieved by counting the number of squares between R waves that were printed onto paper at a known speed by an ECG machine [181]. The ability to accurately measure the R-R interval allows for the analysis of heart rate variability (HRV), the variance in R-R intervals over time. Polar have published a white paper article suggesting that the H10 has an overall 95% accuracy in reporting R-R intervals during sports, which is an improved error rate in comparison to the group of Holter monitors tested in the same sporting activities [180]. Although the exact processing used in the Polar H10 is not disclosed, it could be inferred that some correction processes are used to ensure that the R-R series remains stable and accurate during these sporting activities [180, 186]. This monitor was used in this experiment because it is inexpensive, robust, and it was expected to output higher quality data that the PPG based monitor used in the prior experiment.

5.2.5 HRV recording

Heart rate data was streamed from the Polar H10 via Bluetooth to an Android mobile device (A Google Pixel 4 XL using version 10.0 of the Android operating system [187]), which was running the Polar Sensor Logger App (version 6) [188]. The app records time-stamped heart rate estimates and inter-beat interval time in milliseconds into comma-separated value (CSV) files. In post-processing of the heart rate data, sections of recorded heart rate that correspond to each interval of the test were indexed and sliced by the timestamps that were recorded by the listening test user interface. As the sensor sent data to the mobile device at a regular 1 second interval, a variable number of inter-beat intervals were reported with each message step. The inter-beat intervals were subsequently rearranged in post-processing before HRV analysis was performed.

In HRV analysis the inter-beat intervals need to be pre-processed to remove anomalous artefacts and ectopic beats [175]. All processing of inter-beat interval data was performed in MATLAB using the HRVTool toolbox developed by Marcus Vollmer [189]. The data was first filtered to remove artefacts using the default thresholds provided by the toolbox. Where several false beats had been replaced by empty values, linear interpolation was performed to recover an estimate of the missing inter-beat intervals. The test data was subsequently sliced into the periods that align with each of the test intervals. Five HRV measures were computed using the HRVTool toolbox [190]:

- root mean square of successive differences (RMSSD).
- standard deviation normal to normal inter-beat intervals (SDNN).
- percentage of successive normal intervals of more than 50ms (pNN50).

- the median distance to the centre of the RR interval return map (rrHRV).
- the triangular interpolation of the NN interval histogram (TINN).

RMSSD has been indicated as a representative of vagal tone and is reputed to have good correlation with high frequency HRV [170]. SDNN is representative of the median variability in inter-beat intervals [191]. pNN50 is reported to be closely correlated with vagal tone and the the activity of the PNS [175]. rrHRV is a robust geometric measure of HRV that can be applied to short measurements [190]. TINN represents the spread of the histogram of inter-beat intervals by approximating the spread of the data with a triangle, greater variance is represented with a larger triangle [192]. Spectral analysis was not used in this study because the timing of each interval in the experiment was too short to guarantee the validity of the analysis, only time domain measures were used.

Time domain measures can be assessed using the return map as a visual representation of a period of the inter-beat interval series. A return map shows the change in successive inter-beat interval timing from one beat to the next. Figure 5.5 shows an example of the return map for the inter-beat interval series of a participant in the pre-test condition in the experiment. The axes of Figure 5.5 represent the ratio of change of inter-beat interval timing as a percentage, with the X axis representing the change in timing of the previous beat, and the Y axis representing the change in timing of the current beat. The spread of points in the return map indicates the variability of the inter-beat interval series, including the presence of outlier intervals. In intervals where participants are under stress, the return map will show a reduced spread of points, indicating that the vagal breaking mechanism is in effect. The plot in Figure 5.5 shows a high density cluster of inter-beat intervals centred around a return rate of 0, and several outlying points in the plot. The density of the cluster of points in the return map indicates that most of the participants inter-beat intervals are within a variance of up to 10% of the interval period. The shape of this cluster can be compared against the cluster recorded during another activity to identify changes in the condition of the participant. Performing this comparison is similar to comparing HRV metrics between experiment

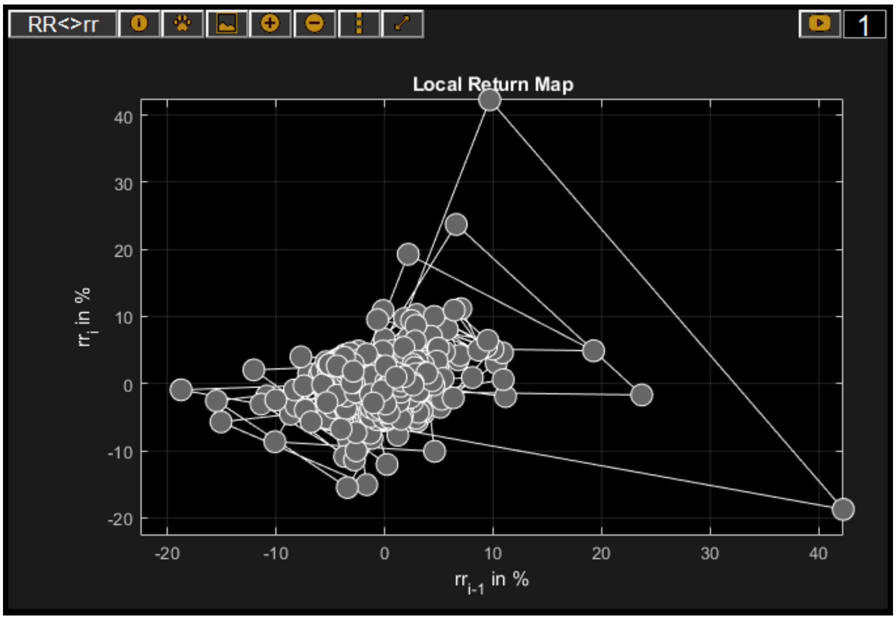


Figure 5.5: Example of the return map for the inter-beat interval series of a participant in the experiment, calculated using the HRV toolbox published by Vollmer [190].

intervals.

5.2.6 Experimental Design and Procedure

After the initial screening and survey stage, participants were invited to fit the heart rate monitor as per the manufacturer's instructions, in privacy. Once fitted and tested, participants were guided to a waiting area to acclimatize to wearing the heart rate monitor. After the acclimatization period of 10 minutes participants were guided back to the formal test environment. The experiment followed a repeated measures design as used in the literature [139, 127]. Participants blindly listened to the randomly ordered stimuli, after which they reported their affect and performed the classification task. The experiment procedure featured a set of 2 training intervals followed by a set of 10 test intervals. In the first 8 test intervals all test stimuli were presented to participants. Two of the 8 test stimuli were randomly selected and were re-played in the last 2 test intervals. The test process was managed by a participant facing user interface that was presented via computer. The timing of the test was kept uniform across all test intervals, following the format:

- 40 seconds of listening to a soundscape.
- 30 seconds of reporting via the user interface.
- 60 seconds of rest.

As the test progressed a timestamp was recorded at each step in the procedure to allow for later synchronization between the stimuli playback and heart rate measurements. The listening portion of each test took approximately 26 minutes to complete. A further 15 minutes was required for the preparation and debrief stages.

Throughout the experiment a safe playback system level was maintained. The experiment was always performed in the same listening room and with the same equipment. The audio playback system signal chain is presented in Figure 5.6. The playback system

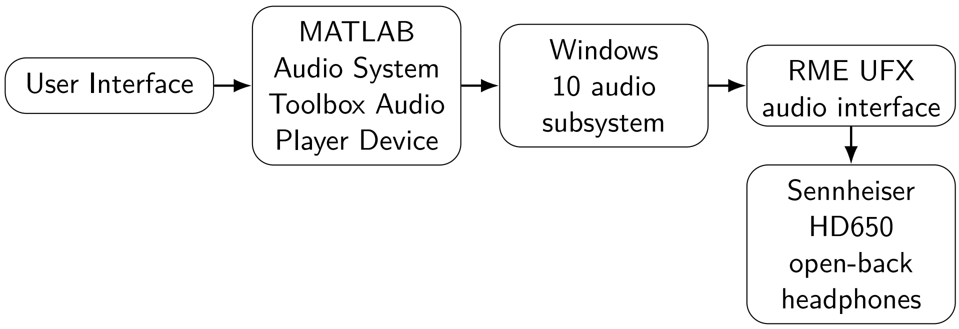


Figure 5.6: Signal chain of playback system.

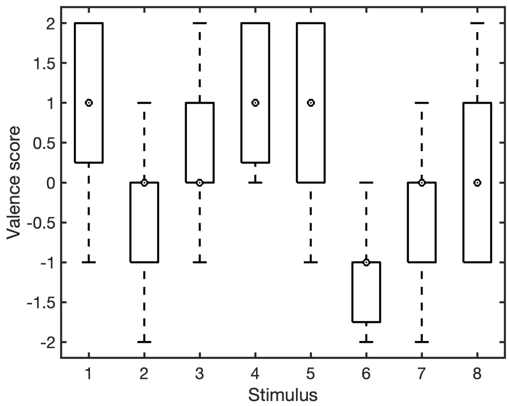
level was set to 90dBSPL at full scale. The system was calibrated using a Behringer ECM8000 measurement microphone, which itself was calibrated using a Tenma 72-7260 sound level meter calibrator.

5.3 Results

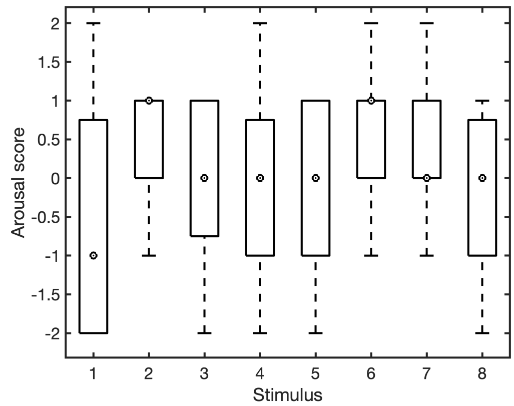
To accommodate for the repeated measures design and the expected individual differences in responses, linear mixed effects models were used in the analysis below. The models included participant and stimuli as random intercepts to satisfy the assumptions of independence between samples. Statistical analysis was performed using MATLAB [165], including the statistics and machine learning toolbox, the econometrics toolbox, the curve fitting toolbox, the *Normality Test Package* [193] and the *HRV toolbox* [189]. The fitting method used in the model was the restricted maximum likelihood estimate, meaning the models are less sensitive to outliers and biased estimates of the random effects terms. The restricted maximum likelihood estimate is a method of parameter estimation for linear mixed-effects models that estimates only the variance components of a model while removing the fixed effects from the likelihood function before maximizing it. This modelling method leads to less biased and more efficient estimates of the variance components [194].

5.3.1 Subjective Estimates

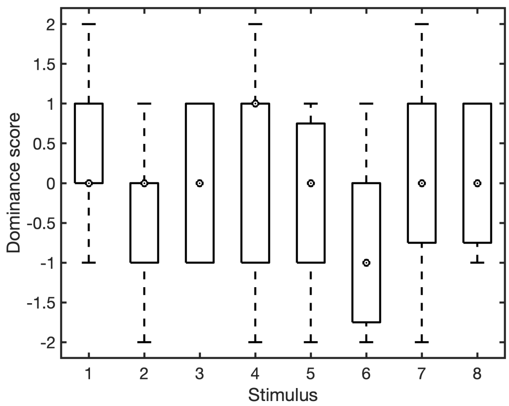
The distribution of the subjective responses of participants with respect to the stimuli are summarised by the series of box plots in Figure 5.7. Figures 5.7a, 5.7b and 5.7c summarise the distribution of affect scores from all participants across the stimuli. Figures 5.7d, 5.7e and 5.7f summarise the distribution of classification scores from all participants across the stimuli. The box plots of Figures 5.7a, 5.7b and 5.7c indicate considerable similarities in the state of affect elicited by the different stimuli, particularly for the dimensions of arousal and dominance. Figures 5.7b and 5.7c give little evidence of a significant relationship between the stimuli and self-reported arousal or dominance, with the median of affect scores being generally 0 (indifference). However, there was a larger range in valence scores across the stimuli leading to an identifiable positive relationship between the stimuli and self-reported valence $F(1, 118) = 12.099, p << 0.001, \eta_p^2 = 0.46$. Comparing Figures 5.7a and 5.7b it appears that decreased valence scores are associated with increased arousal scores, and a significant interaction was identified $F(1, 118) = 9.125, p = 0.003, \eta_p^2 = 0.08$. Figures 5.7d and 5.7e highlight that as mechanical scores increased natural scores decreased, with the notable exception of stimuli 5 which was considered to have a high proportion of both natural and mechanical sounds. Natural scores varied significantly between stim-



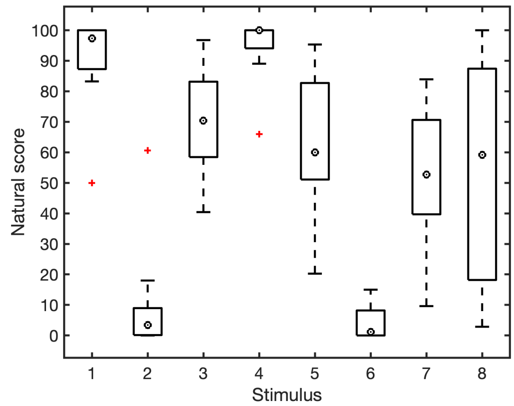
(a) Valence Scores Across Stimuli.



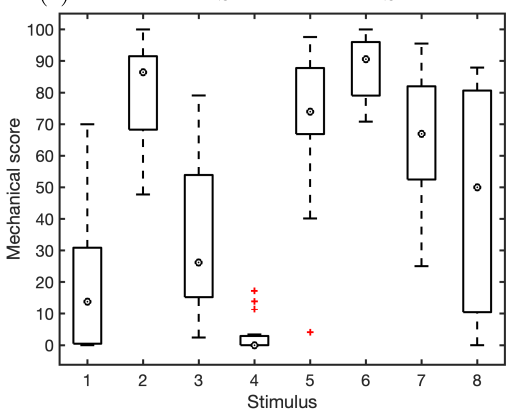
(b) Arousal Scores Across Stimuli.



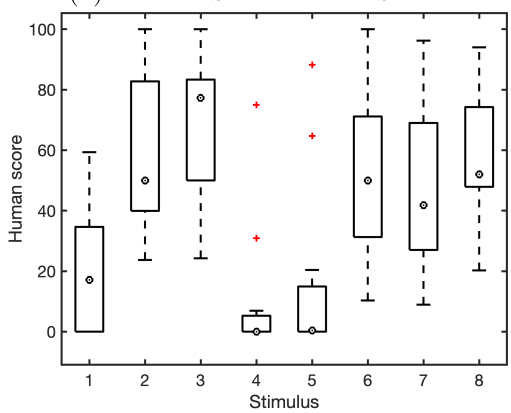
(c) Dominance Scores Across Stimuli.



(d) Natural Scores Across Stimuli.



(e) Mechanical Scores Across Stimuli.



(f) Human Scores Across Stimuli.

Figure 5.7: Distribution of subjective estimates for each stimuli across all participants.

uli $F(7,98) = 56.72, p << 0.001, \eta_p^2 = 0.8$ as did mechanical scores $F(7,98) = 35.18, p << 0.001, \eta_p^2 = 0.72$ and human scores $F(7,98) = 19.91, p << 0.001, \eta_p^2 = 0.59$, suggesting that the selection of stimuli included a reasonable variety of content.

Of the classification scores, two stimuli have elicited results of particular interest, stimuli 8 and 5. Stimulus 8 elicited a wide range of natural and mechanical scores, suggesting that in this context participants disagreed on whether the soundscape was explicitly highly mechanical or highly natural sounding. Stimulus 8 was a beach soundscape with the sound of people walking and talking, the jangle of metal and the rolling of a pram. Stimuli 5 was a woodland soundscape with the sound of a steam locomotive, eliciting both highly mechanical and highly natural scores. The participants reported positive valence for this stimuli, giving a context in which highly mechanical sounds in a highly natural setting elicit positive valence. This provides a counter point for the expectation that mechanical sounds in the context of highly natural settings would lead to reports of negative valence.

5.3.2 Physiological Responses

Heart rate measurements were averaged across each test interval. The average resting heart rate of each participant was calculated by averaging across the heart rate recorded in the pre-test rest condition. Each participant's average resting heart rate was then subtracted from each respective test interval in order to normalize the data, giving the average heart rate change (Δ). Figure 5.8 presents the mean and standard deviation of average heart rate change across all participants per stimuli. It is clear from Figure 5.8 that the mean normalized average heart rate is similar across the stimuli, and there is a clear overlap in the distribution of heart rates across the stimuli. It is clear from Figure 5.8 that there are no instances where a stimulus has elicited a consistent large change in average heart rate from rest. Figure 5.9 presents boxplots of the distribution of heart rate metrics per stimuli across all test intervals. It is clear from all of the subplots of Figure 5.9 that there are no large differences in the effect of any stimuli for any HRV measure. Although there appears to be differences in the medians for several of the

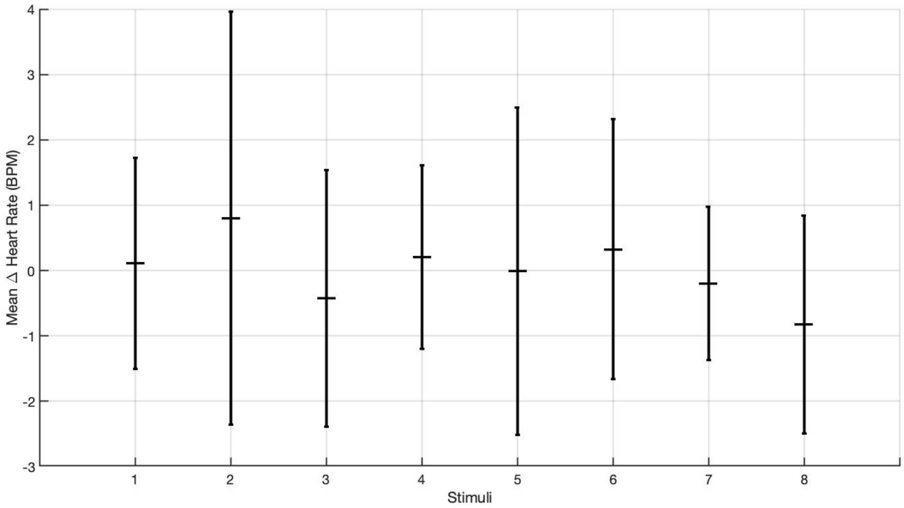
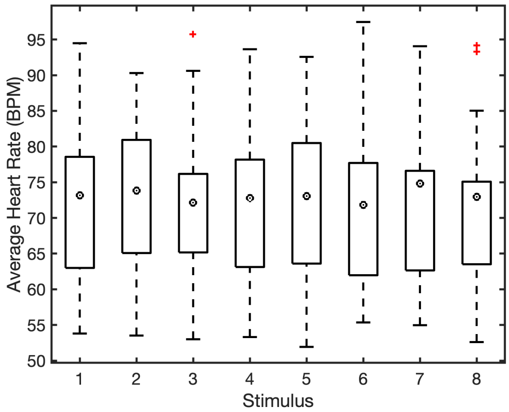
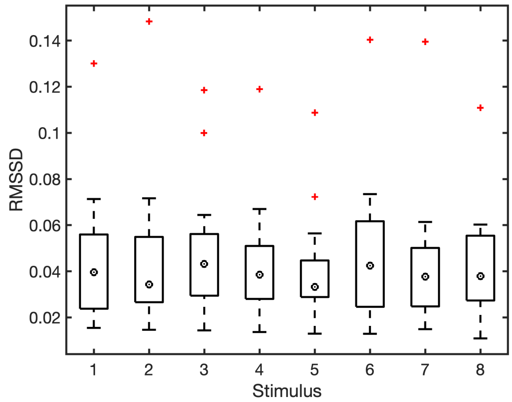


Figure 5.8: Average heart rate change across all participants for each stimulus.

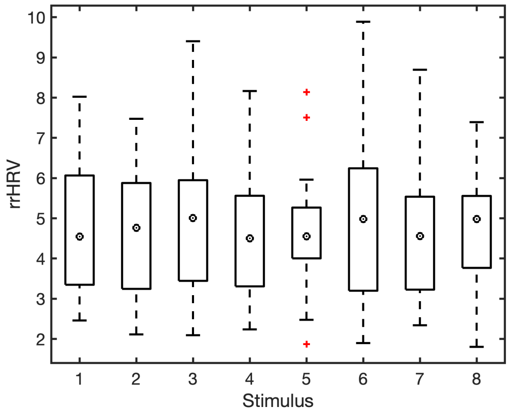
measures, the differences appear very small suggesting any effect size would in turn be very small. The analysis of HRV measures are often performed on data recorded over much larger periods of time. To identify if any experimental effect may be present at all, a comparison is presented for an HRV measure between the pre-test and test conditions for all the participants. Figure 5.10 presents the absolute mean log RMSSD for each participant across the test and rest conditions. The height of the bars presented the absolute mean, and the size of the error bars represents the standard deviation. The data presented in Figure 5.10 shows there are no large differences between these HRV values between the pre-test and test condition. This indicates that the test condition is unlikely to have have had a consistent, systematic or distinguishable effect on heart rate variability compared to the pre-test condition of rest. However, the lengths of the pre-test and test periods used in the data from Figure 5.10 are not the same, the test period was significantly longer than the pre-test period. To improve the visibility of differences on the chart given the small scale of the RMSSD data, a log transform was applied to the data. A log transform can be used to rescale data, making the data more suitable for analysis and visualisation [195]. Figure 5.11 presents the mean and standard deviation



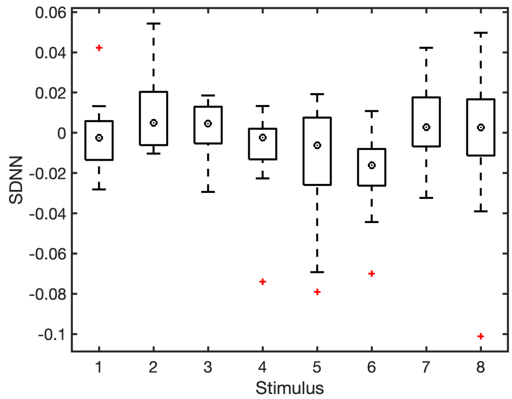
(a) Average Heart Rate.



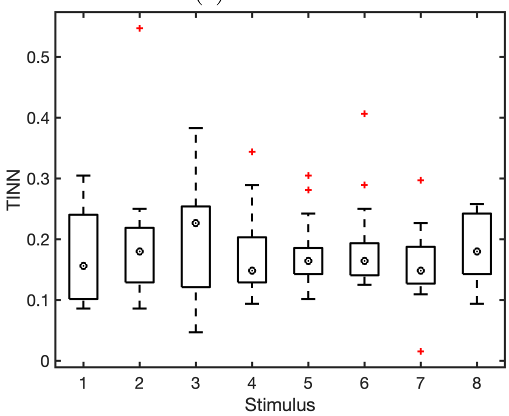
(b) RMSSD.



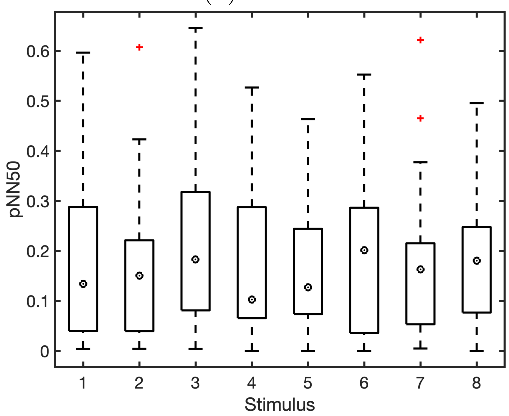
(c) rrHRV.



(d) SDNN.



(e) TINN.



(f) pNN50.

Figure 5.9: Distribution of heart rate metrics for each stimuli across all participants.

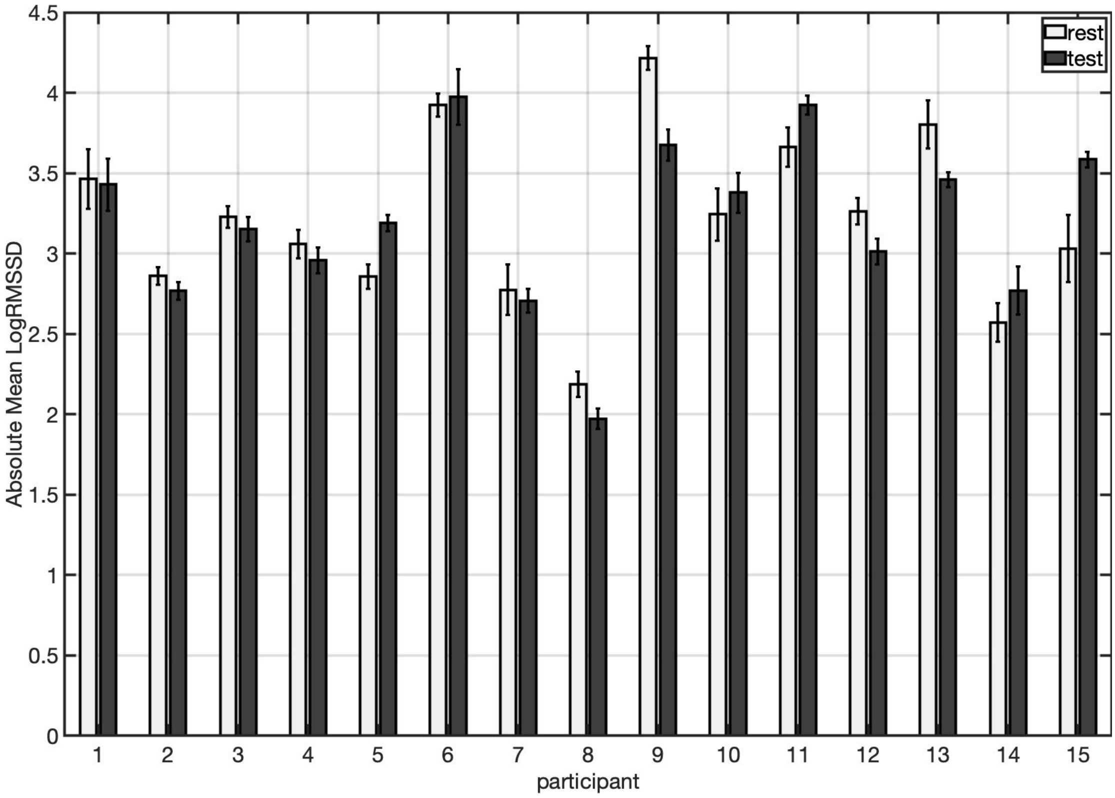


Figure 5.10: Absolute Mean LogRMSSD for each participant in the test and pre-test rest conditions.

of the difference of average heart rate between test intervals and the pre-test period for each participant. The plot in Figure 5.11 presents the difference of the mean heart rate between the rest and test conditions for each participant as a circle, with the standard deviation of the differences as bars that extend out from the marker of the mean. The data presented in Figure 5.11 shows a large variation in the change of heart rate from pre-test and test intervals across the participants. However, most participants' heart rates decreased in the test compared to the pre-test condition. Although no statistically significant effect of the stimuli on heart rate was detected, in most cases participants may have relaxed when sitting down to perform the test.

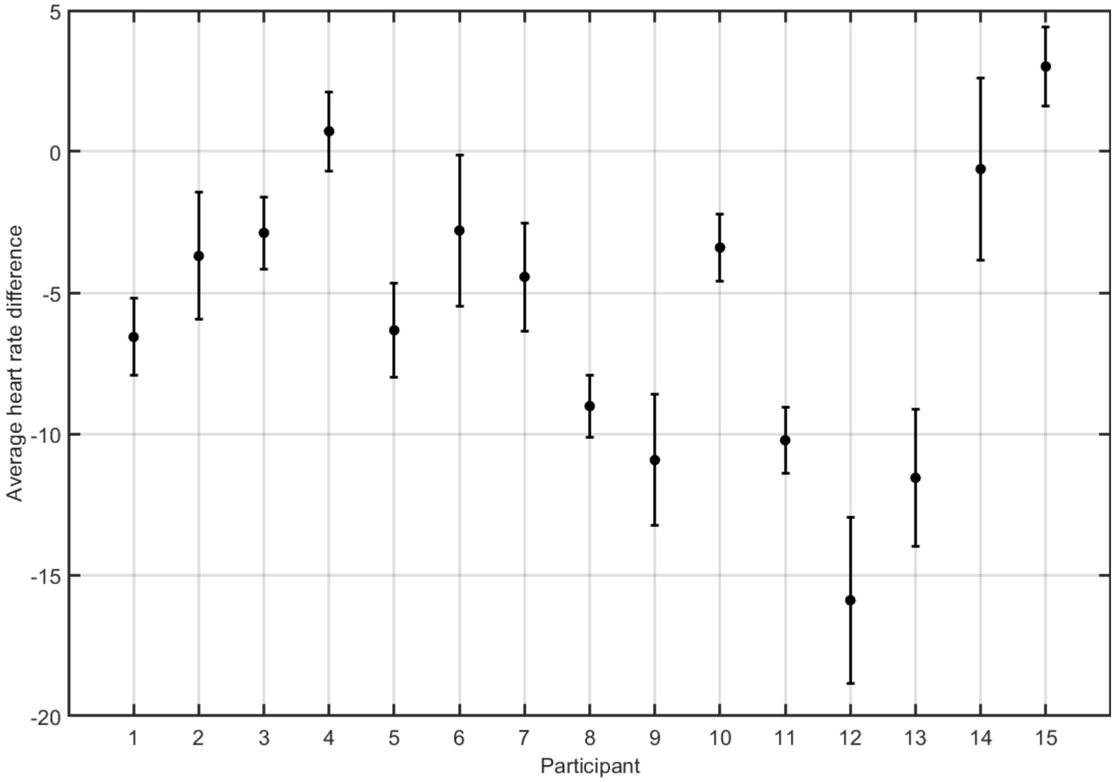


Figure 5.11: The average difference in mean heart rate between pre-test and test conditions.

5.3.3 Correlation

A correlation table comparing measures is presented in Table 5.2. The correlation coefficients were calculated using Spearman's rank correlation, which was chosen because some of the data being correlated is ordinal but not interval data, and so non-parametric statistical analysis is more suitable that parametric statistics such as Pearson correlation. The correlation coefficients with a p value of less than 0.05 are presented in the table in a bold typeface, and the results not presented in bold are non-significant results. A p value represents the probability that the result of a statistical test is truly representative of a statistical relationship between two groups of data. In the case of Spearman's rank correlation, the p value is a representation of the probability that the magnitude of a correlation would be obtained based on the size of the data and the correlation coefficients that were calculated. P values are used in hypothesis testing to provide a value of confidence that can be attributed to the results of a statistical test, and the p value is compared to a significance value α that is a given criteria that the confidence value has to meet or exceed. A typical significance threshold is 0.05, and in the instance that the p value is at or below this threshold, it is 95% likely that the result of the statistical test is not the result of random chance.

Table 5.2: Correlation Coefficients Between Subjective Estimates & HRV Measures.

Factor	Interval	Stimulus	Valence	Arousal	Dominance	Natural
Arousal	0.06	0.06	**-0.36**	**1.0**	**-0.34**	0.35
Dominance	-0.07	-0.05	**0.44**	**-0.34**	**1.0**	0.03
Natural	**0.19**	**-0.23**	**0.63**	**-0.31**	**0.35**	**1.0**
Mechanical	**-0.24**	**0.25**	**-0.57**	**0.28**	**-0.38**	**-0.76**
Human	-0.06	0.16	**-0.38**	**0.22**	-0.03	**-0.38**
Heart Rate	0.03	-0.03	0.12	0.12	-0.09	0.02
rrHRV	-0.04	0.003	**-0.22**	0.04	-0.14	-0.05
SDNN	**0.23**	-0.07	0.08	-0.06	0.02	0.01
TINN	0.05	0.01	-0.09	0.18	-0.15	-0.03
pNN50	-0.04	0.03	-0.11	0.09	-0.07	0.0006

The correlation coefficients presented in Table 5.2 show that natural, mechanical and human scores are correlated with valence, arousal and dominance reports respectively. Natural scores are correlated with mechanical scores and share collinearity, and as such different models are presented below to evaluate these indicators independently. Only two correlations appear between HRV measures and other variables. SDNN is positively correlated with test interval number, suggesting this value may increase over the course of the experiment. rrHRV is positively correlated with valence report, suggesting that the variance of inter-beat intervals might increase when experiencing more pleasant soundscapes.

5.3.4 Linear Mixed-Effects Models

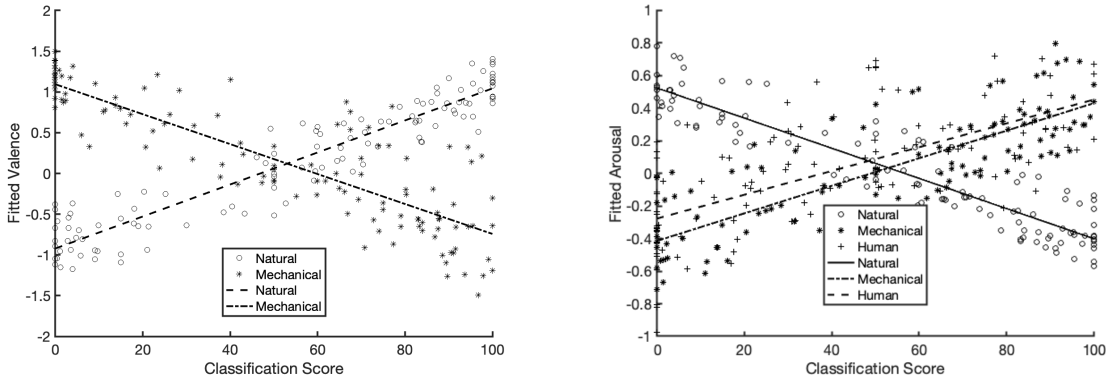
Given the correlations present within the sets of subjective indicators, descriptors and HRV measures, single factor linear mixed-effects models were computed for pairwise descriptors, indicators and HRV measures. The statistics of these models are presented in Table 5.3, including the model estimates, standard errors for the primary effects and the upper and lower bounds of the 95% confidence intervals. The model residuals were tested for normality using the Shapiro-Wilk test, and plots were evaluated for heteroscedasticity. The Shapiro-Wilk test is a statistical test for normality that tests the null hypothesis that the data being tested comes from a normally distributed population [169]. The normality of the residuals of a generalised linear model is one of the assumptions of the statistics applied in a generalised linear model. The heteroscedasticity residuals is another assumption of a generalised linear model, and it is an assumption that the variance of the residuals is consistent across all levels of the independent variables being modelled [169].

Table 5.3: Linear Mixed-effects Models.

Factor	Valence			
	Estimates	Std. Error	95% CI	
Natural	0.0189***	0.0028	-1.289	-0.471
Mechanical	−0.0162***	0.0032	-0.022	-0.0098
Human	-0.0062	0.0035	-0.013	0.7e-3
Heart Rate	0.0059	0.0096	-0.013	0.025
RMSSD	-3.329	3.75	-10.82	4.034
rrHRV	−0.123*	0.055	-0.232	-0.146
SDNN	5.313	3.49	-1.613	12.24
TINN	-0.2189	1.145	-2.486	2.048
pNN50	-0.708	0.613	-1.924	0.506
	Arousal			
	Estimates	Std. Error	95% CI	
Natural	−0.0091***	0.0024	-0.0140	-0.0043
Mechanical	−0.0089***	0.0025	0.0038	0.0140
Human	−0.0094**	0.0032	0.0029	0.0158
Heart Rate	-0.0099	0.0097	-0.0291	0.0092
RMSSD	2.837	3.98	-5.04	10.72
rrHRV	0.0039	0.065	-0.124	0.1327
SDNN	-1.305	3.97	-9.23	6.5
TINN	2.188	1.26	-0.32	4.7
pNN50	0.367	0.674	-0.968	1.7
Model Information				
Observations	120			
Fixed Effects Coeffs	2			
Random Effects Coeffs	23			

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Analysis of variance was performed on the fixed effects of the models presented in Table 5.3 that resulted in statistically significant estimates. From these results we find that valence varied significantly with natural scores $F(1,118) = 44.8, p << 0.001$ and mechanical scores $F(1,118) = 25.05, p << 0.001$. Arousal varied significantly with natural scores $F(1,118) = 14.2, p < 0.001$, mechanical scores $F(1,118) = 12.1, p < 0.001$ and human scores $F(1,118) = 8.37, p < 0.01$. These relationships are summarised in the subplots of Figure 7. Figure 5.12 presents the classification scores with respect to the fitted values of valence and arousal across the respective models. The modelled coefficients are presented with lines of best fit, showing the trend of the relationship represented within each model. Both Figure 5.12a and 5.12b show clear trends between



(a) Fitted Models for Valence. (b) Fitted Models for Arousal.

Figure 5.12: Distribution of subjective estimates for each stimuli across all participants.

the classification scores and affective report scores for the associated models. The only HRV metric to vary significantly with valence was rrHRV $F(1, 118) = 5.04, p = 0.026$. No other statistically significant relationships were observed between a HRV measure, a descriptor or an indicator. Figure 5.13 presents the rrHRV measures with respect to the fitted values of valence, including a line of best fit to make the data more clear. There

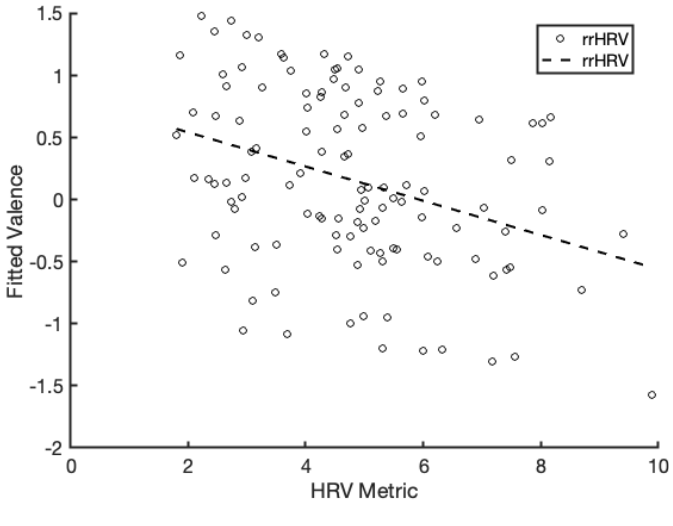


Figure 5.13: Distribution of rrHRV with respect to valence scores.

appears to be a trend in the rrHRV data presented in Figure 5.13 with respect to the fitted values of the model for rrHRV and valence. However, as no other relationships were found between other HRV measures and subjective estimates, this effect should be confirmed in a study with a larger sample size.

5.4 Discussion

In this experiment 15 participants evaluated a randomly ordered set of 8 soundscapes, performing a survey of classification scores and self-reported affect. Heart rate and heart rate variability measures were recorded throughout the experiment. Statistically significant relationships between soundscape classification and affect self-report were identified, with higher natural and lower mechanical scores reflecting higher reported valence and lower reported arousal. The subplots of Figure 5.12 visualize the relationships between classification scores, valence and arousal respectively. These results are similar to those found in the experiment described in Chapter 4 and support the hypothesis that the experience of soundscapes classified as featuring natural and mechanical sound sources evoke effective responses in listeners.

However, unlike the prior experiment, no statistically significant difference in average heart rate was observed between the stimuli. Further, only one relationship between a descriptor and an HRV metric was identified in statistical modelling. This relationship between rrHRV and self-reported valence is presented in Figure 5.13 and reflects a trend of a lower rrHRV with increased valence. This result is contrary to the suggestion that HRV would increase under more relaxed conditions such as listening to a more pleasant soundscape. As rrHRV is intended to be a more suitable and robust HRV measure with respect to outliers and shorter time periods, it could be that the other HRV measures would exhibit similar relationships if the stimuli were longer or the effect size was increased. However, as no other HRV metric presented a similar relationship to any descriptor, this result is likely to be anomalous and warrants further investigation with a larger sample size. These findings are similar to those reported by Alvarsson *et al.* [131] and others in the soundscape literature who similarly reported finding no statistically significant relationships between environmental sounds or soundscapes and heart rate or HRV measures [129, 144, 127].

The results of this experiment support the hypothesis describe in Section 1.2, that he experience of soundscapes classified as featuring natural and mechanical sound sources evoke physiological and affective responses in listeners. However, the support for physiological responses in the scope of that hypothesis is limited to the relationship between rrHRV and self-reported valence. Further work is required to develop stronger evidence in support of the hypothesis that introduced in Section 1.2 of Chapter 1.

5.4.1 Limitations & Further Research

This experiment had several limitations that should be addressed when designing further studies of physiological responses to soundscapes. Although the sample size in this experiment was sufficient to identify the effects of soundscape classification on emotional affect, the sample size was insufficient to develop confident conclusions on the presence of an effect of soundscape on HRV measures. Future work should focus on improving the methodology of the experiment in order to maximize the effect size. Ciuk *et al.* previously identified that physiological measures were significantly weaker than self-reported affect at estimating the influence of attitudes on policy agreement in a study of the influence of affect on policy making decisions of federalism [196]. Although the sample size of 106 undergraduate students used in the study was large compared to those typically used in psychophysiological evaluations of soundscape, the researchers concluded that physiological measurements were not appropriate replacements for self-report in studies of political science, and that the stimuli required for an effective study must elicit a very strong emotional reaction. Although large sample sizes are desirable for robust statistical analysis, a greater focus on managing confounds and improving the stimuli may yield more significant results than larger sample sizes alone.

High frequency HRV measures might not be appropriately sensitive to the effects of soundscape experience with this studies sample size, but as Alvarsson *et al.* [131]

reported similar findings in their study with 40 participants, this may be a more general limitation of the method. Bernston *et al.* suggest the sensitivity of high frequency HRV measures to parasympathetic activity might be modified by breathing rate, and this factor should be included in studies including high frequency HRV [176]. Future studies should include a breathing rate measure to ensure that the effects of breathing rate are not confounding the results of the study.

This experiment was designed to minimize external factors that might elicit changes in heart rate. The experiment took place seated, in a warm insulated environment, with minimal participant engagement. These experimental conditions may have primed participants to be significantly relaxed, indifferent and even bored. Further, in both this experiment and the study published by Bernston *et al.* static binaural rendering was utilized to improve the ecological validity of the soundscape reproduction, but there was little impetus for participants to engage in the evaluation and suspend their disbelief. Perhaps the effect size of soundscape experience on HRV can be boosted by significantly improving the experiment design to be more ecologically valid. Recent studies have suggested that the use of VR technologies can improve the immersive nature of such experiments [197]. Future research should utilize advanced environment rendering technologies to improve ecological validity and participant engagement within the experiment.

Another factor in the quality of the study was the demographics of the participants. Future studies should include a greater diversity of the population being sampled in the study design, with the intention of ensuring the sample of participants is representative of the wider population under consideration. The participants in this experiment were not surveyed for factors that could reflect the prior context of their experience such as their nationality and the type of environment they grew up in. Researchers have reported differences in responses to soundscapes that are related to the nationality of the participants [198]. Further, the participants were not surveyed for their affect and attitudes prior to performing the study. There is evidence that physiological measures might be sensitive to several confounds including a participants' disposition, hydration level and alcohol intake prior to the experiment. Future studies should make appropriate steps to ensure these confounds do not influence the experimental results [199].

The strategy for the selection of stimuli was developed with the intention of avoiding systematic bias. The stimuli were selected through a process of algorithmic evaluation, taking advantage of an established metric for evaluating soundscape ecology [151]. However, researchers using a similar stimuli selection strategy should attempt to compare several metrics that are intended for similar purposes, instead of limiting the range of metrics to one. A further limitation of the stimuli selection procedure was that no further pre-processing was used, and several of the recordings include strong low frequency rumbles that are likely caused by wind noise. This low frequency noise is quite obvious in some playback systems, and a researcher improving on this study should consider using appropriate high pass filtering. However, the noise was not considered to be overly strong in the playback system used in this experiment and extra filtering was not used.

Summary

In this chapter an experiment was described that built on the experiment that was described in Chapter 4. This experiment featured a similar methodology to the first experiment but with several improvements including greater control of the timing of the sections within the experiment intervals, a higher quality heart rate monitor and a slightly larger sample size. The results of this experiment were once again promising but support for physiological responses in response to soundscapes was limited, with only one of the HRV measures showing a statistically significant difference between the stimuli. The results of this experiment show that although the methodology used in this study is effective at influencing the emotional affect of participants, further improvements in the experiment method and a greater sample size are required to identify the effects of soundscape experience on HRV measures. A factor that might improve the effect size of such an experiment is the use of a more immersive experimental environment, such as a virtual reality environment. A subsequent experiment that tests the important of dynamic spatial audio rendering in soundscape evaluation is described in Chapter 6.

Chapter 6

The Influence of Dynamic or Static Binaural Rendering on Soundscape Quality Estimation

The work presented in prior chapters of this thesis has focussed on finding evidence to support the hypothesis laid out in Section 1.2. The results of the experimental work described in Chapters 4 and 5 are promising, and there is evidence that the state of affect of participants is influenced by the presentation of soundscape stimuli. However, the analysis of physiological measures in these experiments yielded few statistically significant results, and the effect sizes of those results was generally very small.

One method of improving the effect size of a results is to increase the number of participants in the experiment, however this is not always possible due to time and resource constraints. Another method of improving the effect size of results is to improve the experimental design, and this chapter is focussed on testing one of the important factors in the experiment design.

Both of the experiments described in Chapters 4 and 5 featured static binaural rendering of soundscape stimuli, meaning that the relative orientation of the soundscape that participants experienced was held constant throughout the duration of the experiment. Participants were free to move their head as they wished, but there was no change in the orientation of the soundscape stimuli, relative to the direction that participant was looking in, when the head moved. This is in contrast to the experience of a real world soundscape, where the relative orientation of the soundscape changes as the head moves. Further, the experiments described in Chapters 4 and 5 did not include any visual stimuli to accompany the soundscapes and provide further context to the participants. This chapter describes a third experiment that was performed to identify the importance of the ability to dynamically change the orientation of soundscape stimuli, on the perceived affective quality of the soundscape, in the context of an online soundscape evaluation experiment.

6.1 Introduction

In both experiments detailed in Chapters 4 and 5 participants were asked to evaluate soundscapes that were pre-rendered into static binaural streams. The definition of static rendering in this instance identifies that the relative position of the soundscape when rendered to a binaural stream was held constant, meaning that the participant was unable to change the orientation of the soundscape relative to the direction that their head was pointing towards. As described in Chapter 2, the ability to change the position of the head is an important cue used in the localization of sound sources and thus the auditory perception of an acoustic environment. Further, the ability to change the orientation of the soundscape relative to the position of the head may be a factor in the sense of immersion experienced by the participant. Immersion in this context is defined as a combination of presence, spatial immersion and absorption, see example [200]. Presence is a term that can be used to describe the sense of feeling present in the particular environment being presented while being situated in a different environment [201]. Spatial immersion is a term that can be used to describe the degree to which a participants' perception is driven and stimulated by the environment being presented, as opposed to the environment in which the participant is situated [202]. One example of this is type of immersion is the use of a head mounted display and headphones to present visual and auditory stimuli from a virtual environment to a participant. Finally, absorption can be used to describe the degree to which a participant is focussed and emotionally and cognitively engaged with the presentation of a virtual environment [203]. Another important form of immersion in virtual and reactive media is the emotional immersion described in [202], though exploring this dimension of immersion is beyond the scope of this thesis.

In the experiments described in Chapters 4 and 5 the degree of presence, spatial immersion and absorption was not measured, but is likely to have been limited by the experimental environment and the use of static binaural rendering. Maximising the sense of presence and spatial immersion might go on to improve the sense of absorption and further the ecological validity of the experiment. Several research papers have highlighted the importance of using appropriate technologies for the rendering of auditory stimuli in the assessment of virtual environments, including a focus on ecological validity. Guastavino *et al.* identified that the method and apparatus used to present soundscape stimuli in a laboratory listening test influenced the ecological validity of soundscape assessment [101, 102]. Later, Tarlo *et al.* identified that several further factors including stimuli reproduction method, survey methods and even time of day were factors in the ecological validity of soundscape quality assessment under laboratory conditions [115]. The evidence presented in these sources suggests that the quality of the auditory presentation method is important in the evaluation of soundscapes ex-situ, and a higher degree of presence and spatial immersion is important for achieving a high degree of ecological validity.

To improve the accessibility of soundscape evaluation, it can be beneficial to perform experiments remotely and online. Experiments that are performed online can take advantage of several online experimentation platforms such as Qualtrics [157], that provide a browser-based interface for the design and hosting of experiments. The presentation of soundscape stimuli in studies that are performed in this way are currently limited to the use of highly accessible methods of stimuli presentation. When presenting stimuli for the assessment of environments, these platforms might be limited to the use of technologies that are widely adopted and supported by most web browsers. Using one such platform, YouTube, stimuli can be embedded into a web page and played back in a web browser. YouTube is a ubiquitous platform for the presentation of audio-visual content, and is supported by most web browsers. YouTube has provided support for the playback of 360 degree videos with binaural rendering of Ambisonic audio for virtual reality content since 2016, though head tracking and stereoscopic visual rendering is only supported for playback in the YouTubeVR app for mobile devices and head mounted displays [204]. Wiggins identified that YouTube supports the rendering of Ambisonic audio to binaural audio by using a set of short anechoic HRTFs in the form of FIR filters, convolving the source Ambisonic audio channels with HRTFs as described in Section 2.9 [205]. The set of HRTFs used by YouTube are from the SADIE II database [67] that was published by the AudioLab at the University of York [65]. Wiggins also identified that YouTube were also making a further optimisation to the rendering of Ambisonic audio by assuming that the listeners' HRTFs were symmetrical, allowing them to halve the number of HRTFs required to render the Ambisonic audio to binaural audio. This optimization is described in Section 4.3 of [52]. Wiggins suggested in [205] that this optimisation may have been necessary for adequate performance of the rendering algorithm on mobile devices.

The use of YouTube videos for the presentation of soundscape stimuli in an online soundscape assessment would allow for the gathering of subjective evaluations from a larger number of participants than would be possible in a laboratory setting. If the degree of presence and spatial immersion is an important factor in the ecological validity of the study, then the use of dynamic binaural rendering may result in a higher degree of presence and spatial immersion, and thus a higher degree of ecological validity. However, if the degree of presence and spatial immersion are not improved by the use of dynamic binaural rendering when compared to static rendering, then the use of dynamic rendering may not be necessary for similar degrees of ecological validity in the assessment of soundscapes. In 2017 Stevens *et al.* compared the results of two experiments where participants evaluated a set of first-order Ambisonic recordings of soundscape stimuli, one in which the stimuli was presented online and one in-person [206]. The in-person study features stimuli that were presented using a 16-channel loudspeaker array, while the online study featured stimuli that were presented over headphones. The stimuli used in the online study were rendered to a format described as Stereo UHJ or *super-stereo*, in which the W, X and Y channels of the Ambisonic audio were used to encode the horizontal plane of the sound field into two channels appropriate for stereo reproduction over a pair of loudspeakers. The results of the study were that there was a significant degree of correlation between the results of the two experiments, suggesting that stereo UHJ is an appropriate format for the presentation of soundscapes. However, the study did not compare the use of binaural rendering with the use of stereo UHJ, and so the rendering method used in the online study will not have included the spatial cues of HRTFs in the presentation of the stimuli. In a further study that included the data from [206], Stevens *et al.* identified that the inclusion of visual stimuli in combination with full dynamic binaural rendering resulted in different self-reported affect results compared to the data from a prior study [207]. However, in this latter study Stevens *et al.* did not discuss if the difference between static and dynamically rendered stimuli might have influenced the results of the experiment.

6.2 Methods

The aim of this experiment is to determine whether the ability of a participant to change the position of a soundscape relative to their view will lead to a different subjective evaluation of the soundscape. This experiment included many of the same assumptions as the experiments described in Chapters 4 and 5, including the following:

- Soundscapes reported as having a higher proportion of mechanical or natural sound sources would be reputed as having different subjective qualities, such as the presence or absence of environmental noise.
- The expectation is that stimuli with differing classification scores would elicit different states of affect which would reflect the physiological and subjective measures.

- Based on the reports of natural soundscapes being restorative or eliciting positive valence, a further expectation would be that soundscapes classified as being more natural would elicit higher valence.

- Another assumption was that ultra-short-term time domain heart rate measures are representative of changes in mood and emotion elicited by experiencing soundscapes.

The experiment followed a mixed design of between groups and repeated measures, where participants were randomly assigned to one of two groups. Each group of participants experienced the same set of soundscape stimuli which were presented via a video that included a 360 degree still image of the environment the recording was made in. This experiment is similar to that performed by Stevens *et al.* [206], though this experiment is designed to identify any differences in the subjective evaluation of soundscapes when experienced through static or dynamic binaural rendering. The independent variable that was different for each group was the method of binaural rendering used to present the soundscape stimuli. The first group experienced the soundscapes with static binaural rendering, where the orientation of the soundscape relative to the participant's look direction was fixed. The second group experienced the soundscapes with dynamic binaural rendering, where the position of the soundscape relative to the participant's head was controlled by the participant. This control of the orientation of the soundscape was achieved by using the computer mouse to click and drag across the video that was used to present the stimuli, in turn the 360 degree still image and Ambisonic sound field would be rotated relative to the view of the participant. The dependent variable was the subjective quality of the soundscape, which was measured using a questionnaire similar to those used in Chapters 4 and 5. This listening test was reviewed and given ethical approval by the University of York Department of Electronics Research Ethics Committee, and the documentation used in the application for ethical approval is included in Appendix A.

Table 6.1: Stimuli Selection Summary.

Location Type	Recording Number	Clip	NDSI Score	Location
Woodland	8	1	0.783	Rowntree Park Woods
Woodland	1	1	0.061	Knavesmire Wood
PedestrianZone	8	4	-0.667	Stonegate
PedestrianZone	6	2	-0.498	St. Helen's Square
Park	3	2	0.579	Yorkshire Museum Gardens
Park	1	4	-0.162	Rowntree Park
Busy Street	8	6	-0.364	St. Leonard's Place
Busy Street	5	1	-0.746	Micklegate Bar

6.2.1 Experimental Stimuli

The set of stimuli was selected from the EigenScape dataset, which was described in Section 4.2. The stimuli selection process was similar to that described in Section 5.2.2, with two notable exceptions. Firstly, the length of the soundscape stimuli was increased to one minute in length in order to allow participants to experience the soundscape for a longer period of time that the experiments described in Chapters 4 and 5. Secondly, the set of stimuli did not include sections of soundscape recordings used in prior experiments described in this thesis, though recordings from the same environments were used. The set of environments were also changed to include a set of environments that were focussed on outdoor urban, suburban and rural soundscapes around the city of York. The set of recordings were restricted to the city of York in order to consolidate the identity of to one general location while still including a range of different environment types. The stimuli set are summarised in Table 6.1. The meta data for these stimuli including access to the stimuli as hosted on YouTube is available in Appendix B in Figure B.1. As described in Section 4.2.2, the EigenScape dataset is organised into sets of recordings made in different types of environment, with each environment featuring a set of ten 10-minute recordings. Each recording is labelled with an identifier of the environment type and the recording number. Each of the stimuli summarised in Table 6.1 is indexed relative to the dataset by location type, the recording number from the dataset, and the clip number that references the position in the recording the clip was taken from. Table 6.1 also shows the NDSI score for each clip, which was calculated using the method described in Section 4.2.2. The NDSI score is a measure of the proportion of sound in a recording that is between 1kHz to 2kHz and 2kHz to 11kHz. The proportion of sound in the lower frequency band is theorized to be indicative of the presence of mechanical sounds, and the proportion of sounds in the upper frequency band is is theorized to be indicative of the presence of natural sounds [151]. A more positive NDSI score is therefore indicative of a higher proportion of natural sounds in a recording, and a more negative NDSI score is indicative of a higher proportion of mechanical sounds in a recording. Finally, each stimuli in Table 6.1 also includes the location of the environment the recording was made in, allowing the reader to reference the location from the online map of the EigenScape dataset [150].

Once selected, the set of soundscape stimuli were then peak normalised and recorded into two sets of files; one set was the first-order Ambisonic recordings and the other set was rendered into two channel binaural streams using the IEM plugin suite [155] as described in Section 5.2.2. During the recording of the EigenScape dataset, the authors also simultaneously recorded videos of the environment using a 360 degree camera that was attached to the microphone stand used in the recording of the dataset. The videos were recorded using the Samsung Gear 360 camera model *SM-C200*, a consumer grade 360 degree camera that records video at a resolution of 3840 x 1920 pixels at 30 frames per second [208]. The videos were recorded as two separate streams that were stitched together using the Samsung Gear 360 Action Director software, forming one 360 degree video stream when viewed on a flat screen using equirectangular projection [209]. The video for each location was sliced into one frame at the middle of the recording, and that one frame was manipulated into a one minute long video using the FFMPEG python package [210]. The identity of people captured in the video was then anonymised using the blurring tool in Davinci resolve, a video editing software package [211]. The set of eight videos were then copied into two sets of eight videos, one set for each of the two different audio conditions. Finally, each set of audio streams were bound to their respective video streams using the FFMPEG python package, and the appropriate meta data for each video was added to each video using the Google spatial media metadata injector [212]. The videos were uploaded to YouTube as unlisted videos.

6.2.2 Data Collection Instrument

The data collection instrument was a web-based survey which was created using the Qualtrics survey platform as described in Section 4.2.3. The survey question selection was a modified version of the questionnaire described by the Soundscape Indices Protocol [98], which itself is adapted from the Method A questionnaire described by the ISO 12913-2 technical specification [96]. Each survey page presented in an experiment interval was organised as follows:

- An embedded hyper-text markup language (HTML) element that presented a YouTube video player. This video player was configured to present the appropriate stimuli by utilising the Javascript programming interface built into Qualtrics.
- The self-assessment manikin (SAM) introduced in Section 4.2 was presented as a set of three 5-element items with pictographic labels.
- A set of eight five-point Likert items that were used to assess the perceived affective quality of the soundscapes.
- A five-point Likert item representing the perceived loudness of the environment.
- A five-element soundscape classification question.
- A text entry question where users were asked to identify the keynote of the soundscape.
- A seven-level drill-down question where users were asked to identify the keynote of the soundscape from a pre-defined taxonomy.

- A text entry question where users were asked to identify the sound-mark of the soundscape.
- A seven-level drill-down question where users were asked to identify the sound-mark of the soundscape from a pre-defined taxonomy.
- A five point Likert item surveying for the appropriateness of the combination of the visualisation of the environment and the soundscape.

The YouTube player was embedded into the survey using the html embed code provided by YouTube, configured with a width of 1120 pixels height of 630 pixels. Users were instructed that they were able to replay the video as many times as they wished, and that they were able to pause the video at any time. Further, they were able to view the video in full screen or in the native YouTube player by using the click-through links provided by the YouTube player.

The SAM was presented as a five-point Likert style item as in the previous experiments. The SAM was included as a means of assessing the state of affect (valence, arousal and dominance) of the participant due to the soundscape, and to allow for the comparison of the results of this experiment with the results of the previous experiments.

Following the SAM, a set of eight five-point Likert items were used as an instrument for assessing the perceived affective quality of the soundscapes. In this section of the questionnaire participants were asked to respond to the following question: *For each of the scales below, to what extent do you agree that the environment was...* The eight items were as follows: *Pleasant, Annoying, Vibrant, Monotonous, Calm, Chaotic, Eventful*, and *Uneventful* The five point scale for each item was labelled as follows: *Strongly Disagree, Somewhat Disagree, Neither Agree nor Disagree, Somewhat Agree*, and *Strongly Agree*. These eight items represent the different affective qualities of soundscape that were identified by Axelsson in [113] when applying the circumplex model of affect to soundscape as introduced in Section 3.9. These measures of the perceived affective qualities of soundscape were integrated into the ISO:12913 standard series [213, 214] which provides a formula for performing a two-dimensional orthonormal projection of the eight items into a two-dimensional coordinate on the circumplex of affect in the dimensions of pleasantness and eventfulness hereafter described as *ISOPleasant* and *ISOEventful* as used in [215, 216, 98]. These formulae are summarised in Equation 6.1 in Section 6.3.

The five point Likert item representing the perceived loudness of the soundscape was included to identify whether the two rendering methods had an effect on the perceived loudness of the soundscape, or if any one soundscape was perceived as louder than the others despite normalisation of the stimuli. The participants were asked: *How loud did you perceive the environment to be?* The scale of the question included the levels *Not at all*, *Slightly*, *Moderately*, *Very*, and *Extremely*.

Several elements were included that were used to assess the perceived structure of the soundscape. The first of these was a five element classification item that included the following classes: *Urban, Suburban, Rural, Wilderness*, and *Indoor*. This set of classes was adapted from the classification system described in [98]. The subsequent four items were used to assess the perceived keynotes and sound-marks of the soundscape. For each item two methods of description were provided, one in the form of a textbox and one in the form of a seven level drill-down with a structure that reflects the sound source taxonomy described in ISO:12913-2 [96] and discussed in Section 3.12 of this thesis. The drill-down questions took the form of several HTML drop down lists box elements that would successively populate with options as the user selected options from each list box, allowing users to successively click through the tree structure sound class taxonomy presented in Figure 3.5. Each level in the tree structure was represented as a drop down list box, and users were able to traverse the taxonomy structure by selecting options in each of the successive list boxes.

6.2.3 Experiment Design and Procedure

Upon landing on the experiment website, participants are presented with a participant information sheet that details the purpose of the experiment, the procedure of the experiment, the risks and benefits of the experiment and the contact details of the researcher. After reading the participant information sheet, participants are asked to provide in-

formed consent to participate in the experiment. Following this, the participants are asked to complete a short demographics questionnaire. In the demographics section of the study participants are surveyed for their age, ethnicity, occupational status and education level, as well as the type of environment that the participant grew up in and the type of environment the participant currently inhabited.
Following this, participants are surveyed for their current mood via the self-assessment manikin. This information could be used to control for the state of affect of participants prior to participating, if a difference in affect was reported between the groups.

Next an audio baseline task was performed to ensure participants were able to view the YouTube 360 videos with dynamic binaural audio. Given the online nature of the experiment the audio playback apparatus used by each participant was unknown, and the baseline task was required to check that the participants were able to view the stimuli as intended. The task included viewing a short YouTube 360 video with dynamic binaural audio, and then answering three binary questions about the the performance of the video. Participants were instruction to play the YouTube video and use the mouse pointer to move the camera view around the scene, and to listen to the audio playback. Using the mouse and cursor to click and drag on the video window rotates the 360° image and Ambisonic recording around the viewer, allowing the viewer to look around the scene. Participants were then asked to answer three true or false statements, and included the following:

- *The playback started, I could see and hear the content correctly.*
- *I could move the camera view around.*
- *I could hear the direction of sounds change as I moved the camera view.*

Each of the three questions were required to be answered correctly in order to proceed with the experiment, ensuring that all of the participants were able to view the stimuli as intended. Participants were excluded from the study if the equipment they used was unable to present the stimuli as intended.

Following this participants were presented with a version of the survey form that was annotated with explanatory text, introducing the participants to the concepts and context required to answer each part of the survey. The annotated survey was intended to mitigate the risk of participants misunderstanding the questions, and to ensure that all participants had the same understanding of the questions. Further, the annotated survey also provided another opportunity to remind the participants how to interact with the stimuli, using the mouse and cursor to rotate the view of the stimuli. When participants had finished reading the annotated survey, they moved onto the first of two practice trials. Each of the two practice trials used a different stimuli, one from a beach and one from a quiet street. The order that the practice trails were presented to the participants was randomised in order to minimize any systematic bias that might have been introduced by experiencing the stimuli in a specific order. Once the participant had completed the practice trials, they were presented with the eight main trials which were also presented in a random order. The randomization of the order of the practice and main trials was intended to mitigate any systematic bias that may have been introduced by the order of the trials. When performing the main trials, each participant was assigned to one of two groups, one with all statically rendered binaural stimuli, and one with dynamically rendered binaural stimuli. Both groups were able to rotate the view of the stimuli with their mouse pointers, but the orientate of the dynamic group would change and the orientation of the static group would not. Figures 6.1 and 6.2 present two excerpts of the user interface used in the listening test. Figure 6.1 shows the top of the survey user interface, including the YouTube video, followed by the valence dimension of the SAM. Note the icon in the top left of the YouTube video that indicates the interactive cursor that can be used to rotate the view of the video. Figure 6.2 shows the middle portion of the survey user interface, including the dominance dimension of the SAM, the eight Likert items describing perceived affective quality, the question pertaining to the loudness of the soundscape and the first of the classification questions.

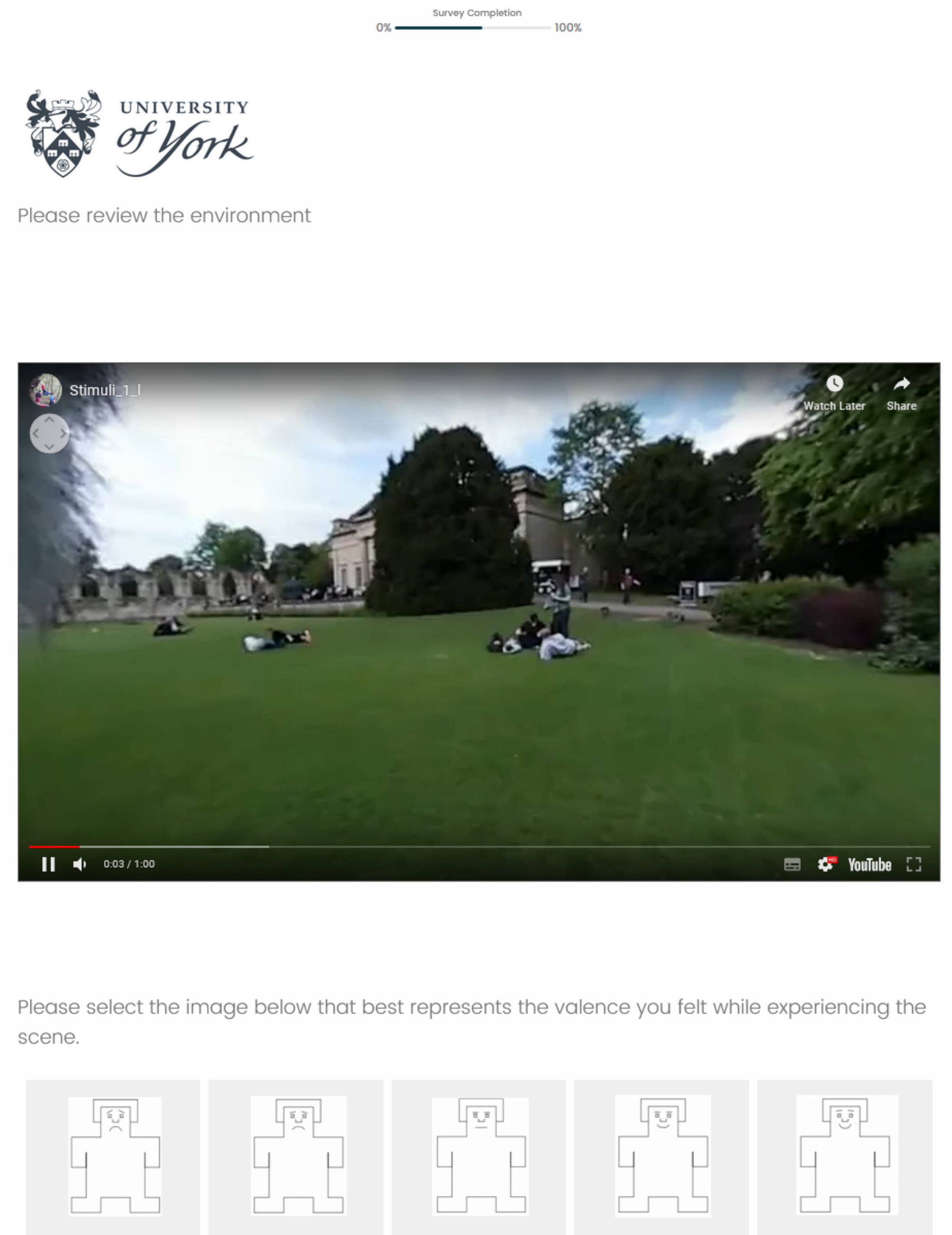


Figure 6.1: The top portion of the listening test web interface, including the YouTube video and the valence part of the SAM scale.

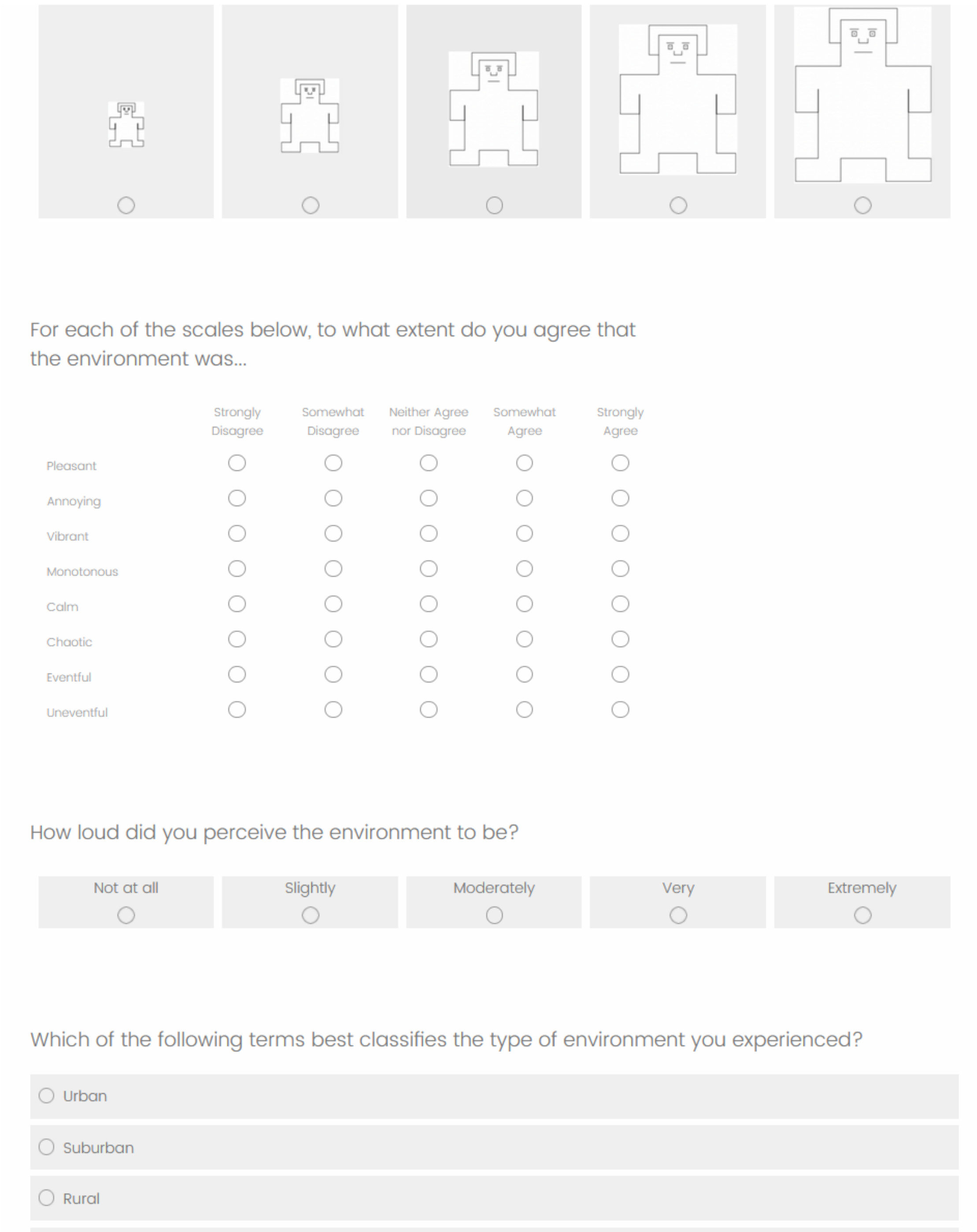
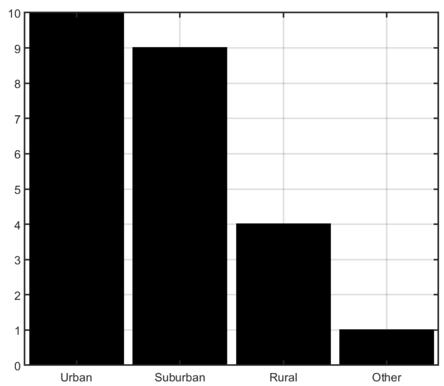


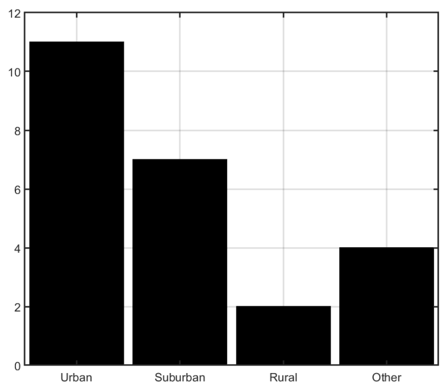
Figure 6.2: The middle portion of the listening test web interface, including the dominance portion of the SAM scale.

6.2.4 Participants

Participants were recruited from the student population of the University of York via email, and from members of the general public via social media. Participants were screened for any atypical hearing loss or hearing impairment prior to the experiment. Prior research has identified that demographic factors can have an influence on the perception of different soundscapes, particularly in the rating eventfulness or arousal [217, 218]. Twenty-seven potential participants viewed the participant information, three of which self-excluded from the experiment. A total of 24 participants were recruited for this experiment, of which 4 identified as female and 20 as male. From the participants who took part in the study 70% were educated to postgraduate level or higher, and 88% identified as being from white ethnicity. The ages of the participants were not evenly spread, with 60% of the participants aged between 18 and 34 years old, 25% aged between 35 and 44, and the remaining 15% were aged 45 and above. The majority of the participants at 70% identified as living in an urban or suburban environment, and 80% identified as having grown up in an urban or suburban environment. This data is presented in the form of two bar charts in Figures 6.3a and 6.3b.



(a) A bar chart showing the number of participants who identified as growing up in one of four different types of environment.



(b) A bar chart showing the number of participants who identified as currently residing in one of four different types of environment.

Figure 6.3: Two bar charts representing the number of participants who grew up and currently reside in different types of environment.

Figure 6.3a shows the frequency of participants that identify as having grown up in a particular type of environment, and Figure 6.3b shows the frequency of participants that identify as currently living in a particular type of environment. Figures 6.3a and 6.3b show that more participants identify as currently living in an urban environment than grew up in an urban environment, and fewer participants identify as currently living in a rural or suburban environment than identified as growing up in those environment types. Of the 24 participants, only 21 completed the full experiment, one of which failed to complete one experiment interval. The remaining 20 participants completed the experiment in full.

6.3 Results

The experimental data was first anonymised by removing any personal information from the data and only retaining the participant ID number, a randomly generated string of characters. The data was then checked for any missing values, and any incomplete responses were removed from the data set, resulting in 10 complete sets of responses for each stimuli, and 80 complete sets of responses per group. Following this the *ISOPleasant* and *ISOEventful* scores were calculated for each experiment interval as described in ISO:12913-3 [214]. These two scores were introduced in ISO:12913-3 The equations for calculating *ISOPleasant* and *ISOEventful* are shown in Equation 6.1.

$$\begin{aligned} P &= (p-a)+\cos(45°)(ch-ca)+cos(45°)(v-m) \\ E &= (e-u)+\cos(45°)(ch-ca)+cos(45°)(v-m) \end{aligned} \tag{6.1}$$

Where: a is annoying, ca is calm, ch is chaotic, e is eventful, m is monotonous, p is pleasant, u is uneventful, and v is vibrant. The data was scaled to a range between -1 to 1 by multiplying each *ISOPleasant* and *ISOEventful* coordinate by $(4+\sqrt{32})$ following the description in ISO:12913-3 [214].

The experiment design was a two group repeated measures design, with stimuli presented to each participant in a randomized order with each participant assessing each stimuli once. The type of the raw data was primarily ordinal, but neither interval, continuous, normally distributed nor independent, so non-parametric tests were used in the analysis presented below. An exception to this was the *ISOPleasant* and *ISOEventful* scores which are continuous. Statistical analysis was performed using MATLAB [165].

To determine if the stimulus type had an influence on affect, the Friedman test [138] was used to test if the distribution of the affect scores from the SAM were significantly different between the stimuli. The Friedman test is a non-parametric statistical test that can be used to determine the differences between treatments in a way that is similar to ANOVA (see Section 5.3), operating on ranked data. There was a significant difference between the stimuli for *valence* ($\chi^2(7) = 57.94$, $p = 3.881e^{-10}$) and *arousal* ($\chi^2(7) = 50.88$, $p = 9.681e^{-09}$) but not *dominance* ($\chi^2(7) = 4.6$, $p = 0.708$). As described in Section 5.3, *dominance* was found confusing as a measure in this form of study for participants, and was not considered for further analysis in Chapter 5 or this chapter. However, the results for *valence* and *arousal* show that the stimuli had an influence on self reported affect. To determine if the stimuli were scored as having different qualities of pleasantness and eventfulness, the same analysis was performed on the *ISOPleasant* and *ISOEventful* scores. There was a significant difference between the stimuli for *ISOPleasant* ($\chi^2(7) = 70.57$, $p = 1.135e^{-11}$) and *ISOEventful* ($\chi^2(7) = 68.09$, $p = 3.582e^{-12}$). This confirmed that the selection of stimuli was appropriately varied to elicit a range of responses from the participants. To determine if there were group effects on affect and the perceived affective quality scores of the stimuli, several tests were performed. Figure 6.4 shows the *ISOPleasant* and *ISOEventful* scores for all stimuli for both groups. Figure 6.4 presents the *ISOPleasant* and *ISOEventful* coordinates for every experiment interval, grouped by participant group. There are no clear clusters in the data, showing that there is no specific bias between either group in the distribution of the *ISOPleasant* and *ISOEventful* scores. However there is a clear negative relationship in the data, showing that the *ISOPleasant* and *ISOEventful* have a wider spread across the dimensions of *chaotic* & *calm* and a narrower spread across the dimension of *monotonous* & *vibrant*. This trend is similar to that described in [219], where Lionello *et al.* proposed a

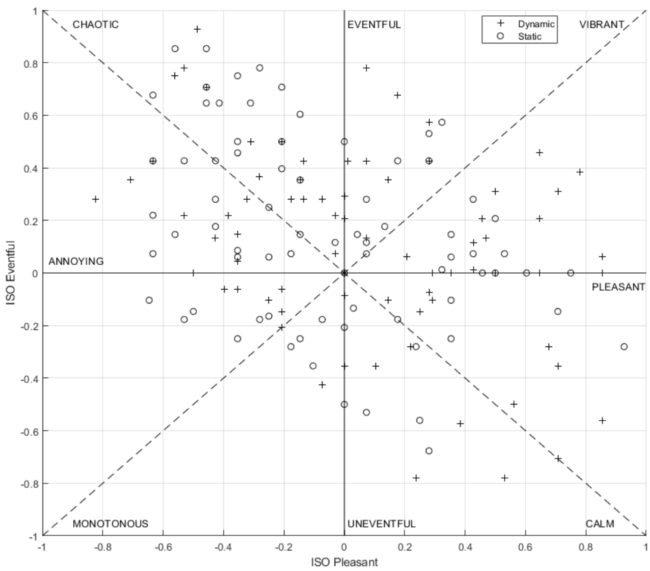
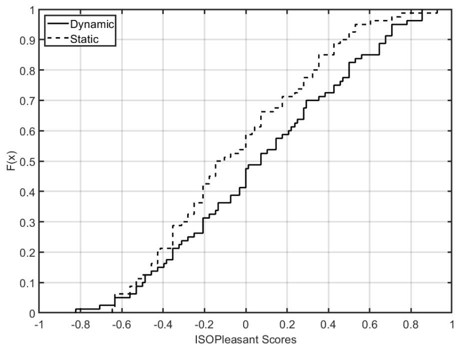


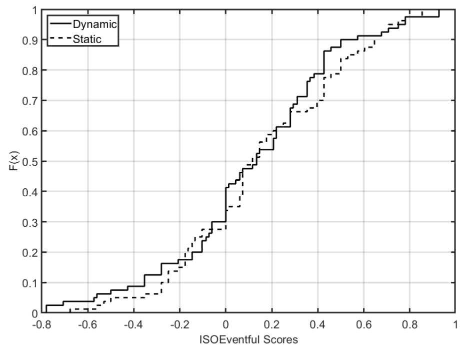
Figure 6.4: ISOPleasant and *ISOEventful* scores for all stimuli for both groups.

correction method to account for the bias present when comparing multiple Likert scales with different magnitudes of impression.

To determine if the *ISOPleasant* and *ISOEventful* scores between each group were drawn from the same distribution, the two-sample Kolmogorov-Smirnov test [138] was performed. The two-sample Kolmogorov-Smirnov test is a non-parametric statistical test of the equality of cumulative distribution functions of two independent variables, and is used to test the null hypothesis that two samples are drawn from the same distribution. Figure 6.5 shows the cumulative distribution functions of the *ISOPleasant* and *ISOEventful* scores for all experimental intervals for both groups of participants. Figure 6.5a presents the cumulative distribution function of the *ISOPleasant* scores for both groups of participants. The two distributions presented in Figure 6.5a are not identical, but the difference is not statistically significant ($k_{stat} = 0.162$ $p = 0.22$). Figure 6.5b presents the cumulative distribution function of the *ISOEventful* scores for both groups of participants. The two distributions presented in Figure 6.5b look more similar than



(a) The cumulative distribution function of *ISOPleasant* scores for both groups.

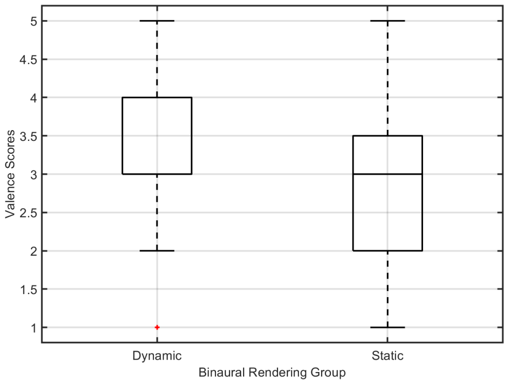


(b) The cumulative distribution function of *ISOEventful* scores for both groups.

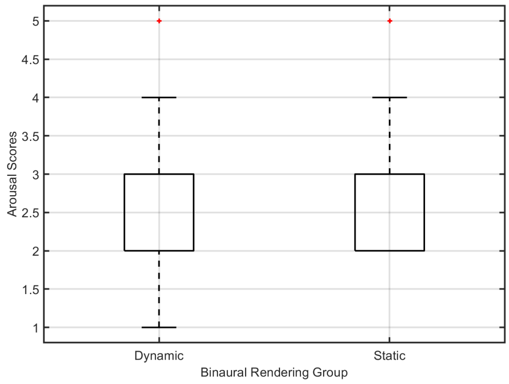
Figure 6.5: The cumulative distribution function for *ISOPleasant* and *ISOEventful* scores for all experimental intervals for both groups of participants.

those in Figure 6.5a, and the differences are also insignificant ($k_{stat} = 0.1125$ $p = 0.666$). The Kruskal-Wallis test [138] was used to test if the affect scores for each group were different. The Kruskal-Wallis test is a non-parametric version of the one-way ANOVA, and an extension of the Wilcoxon rank sum test to more than two groups. The test compares the medians of groups of data to determine if the samples come from the same population or different populations with the same distribution. Figure 6.6 present the distributions of *valence* & *arousal* scores for all experimental intervals for both groups of participants. Figure 6.6a presents a boxplot representing the distribution of the valence scores, and Figure 6.6b presents a boxplot representing the distribution of the arousal scores for all experiment intervals for both groups of participants. It is clear from both Figures 6.6a and 6.6b that the distribution of scores from each group are very similar for both valence and arousal. This is reflected in the results of the Kruskal-Wallis test which found no significant difference in *valence* scores ($\chi^2 = 3.714$ $p = 0.054$) or *arousal* scores ($\chi^2 = 3.008$ $p = 0.083$) at a confidence interval of $p = 0.05$.

Figure 6.7 presents a boxplot representing the distributions of *loud* scores over all test intervals per group. The boxplot in Figure 6.7 shows that the static group reported more *moderately loud* and fewer *slightly loud* scores than the dynamic group, and this is reflected in the results of the Kruskal-Wallis test. Performing the Kruskal-Wallis test on



(a) The boxplot of valence scores for both groups for all experiment intervals.



(b) The boxplot of arousal scores for both groups for all experiment intervals.

Figure 6.6: The distribution of self-reported valence and arousal scores for all experimental intervals for both groups of participants.

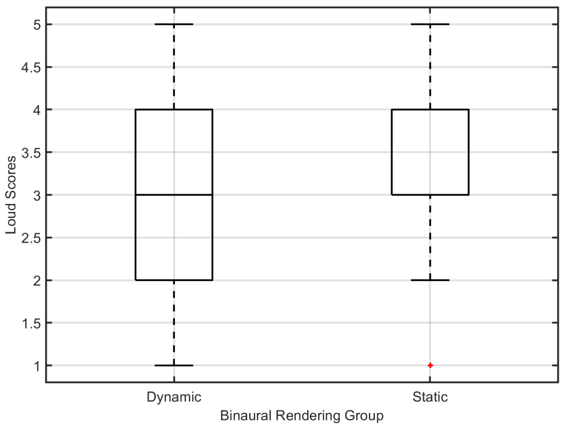
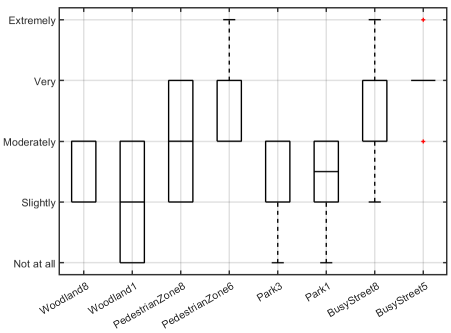
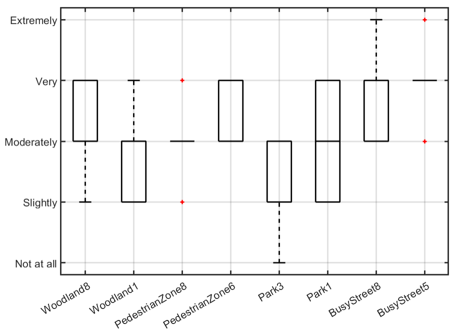


Figure 6.7: A boxplot of the distribution of *loud* scores for all experimental intervals for both groups of participants.



(a) The boxplot of loud scores for all stimuli for the dynamic group of participants.



(b) The boxplot of loud scores for all stimuli for the static group of participants.

Figure 6.8: The distribution of self-reported *loud* scores for each stimuli for both groups of participants.

reports of the *loud* Likert item resulted in identifying significant differences between the groups ($\chi^2 = 4.6371$ $p = 0.0313$), with the static group reporting higher *loud* scores than the dynamic group. Figure 6.8 presents boxplots of the distributions of *loud* scores over all stimuli for each participant group. The boxplots in Figure 6.8 show that members of the static group reported that the stimuli from the two woodland scenes and one of the park scenes were perceived as being *very*, *moderately* or *slightly* loud, whereas the dynamic group reported that these stimuli were *moderately*, *slightly*, or not loud at all. Further, a member of the dynamic group reported that the *BusyStreet8* stimuli was *slightly* loud.

6.4 Discussion

In this experiment 20 participants were randomly distributed into two groups, each group was exposed to the same set of soundscapes that was rendered using a different binaural renderer. Each participant was asked to evaluate the soundscapes using a set of Likert items, the SAM and classification criteria. The stimuli were presented to each participant in a randomised order. Prior to analysis the data was cleaned and balanced, and the eight Likert items were transformed into a geometric score following the formulas described in the ISO:12913-3 standard.

The data was first analysed to determine if the selection of stimuli elicited a variation of affective responses and perceived affective quality scores. There was a statistically significant effect of stimuli on both self-reported affect and perceived affective quality, suggesting the stimuli were effective in achieving a wide variety of subjective ratings. The data was then analysed to determine if the binaural renderer used to render the stimuli had an effect on the affective response of the participants, and this was achieved by performing group-wise analysis on the data. The results of the group-wise analysis showed no significant differences in either the self-reported affect or perceived affective quality scores between the two groups of participants. This would suggest that having the ability to render a soundscape with dynamic panning that is directed by the listeners mouse pointer does not have a significant effect on the affective response of the listener. Unfortunately, it was not possible to identify how many times participants interacted with the dynamic panning interface, and this would have been useful information to have as it would have allowed for a more detailed analysis of what spatial features in the soundscapes participants focussed on. This is an interesting result as it suggests that online soundscape evaluation experiments can be conducted using either static or dynamic binaural rendering without influencing the results of self-reported affect or perceived affective quality.

As highlighted in Section 5.3 there is a clear negative trend in the *ISOPleasant* and *ISOEventful* scores that is observable in Figure 6.4, . In their case study of urban soundscape data, Lionello *et al.* proposed a method for correcting the dilation and compression effects that can be observed in the comparison of several Likert scales data [219]. However, the size of the data set used in this experiment is too small to perform the same correction strategy proposed in that paper.

The analysis of the *loud* Likert item showed that the static group reported higher *loud* scores than the dynamic group. As the experiments were performed remotely it was not possible to assert if this was due to a feature of the binaural rendering procedure, the playback level or playback quality of the stimuli, or the playback system used by the participants. However, as the self-reported affect and perceived affective quality scores were not significantly different between the two groups of participants, it is unlikely that any differences in the loudness of the stimuli between the groups had any further effect on the perceived quality of the soundscapes. To understand why the two groups reported different *loud* scores a subsequent analysis was performed by recording the audio playback of each stimuli from YouTube and into Reaper, a digital audio workstation [220], and the loudness of each stimuli was measured using the LUFS metering tool provided as part of Insight 2 [221]. Insight 2 is a VST plugin developed by Izotope for use in digital audio workstations that provides a suite of metering tools for measuring the loudness of audio signals. It was found that the integrated loudness of the static group of stimuli was 4dB$_{LUFS}$ higher than the integrated loudness of the dynamic group of stimuli. The measure of dB$_{LUFS}$ was described in Section 5.2.2, and integrated dB$_{LUFS}$ average the loudness of the signal over the period of playback [179]. It is assumed this difference in loudness is caused by the differences between the two rendering methods.

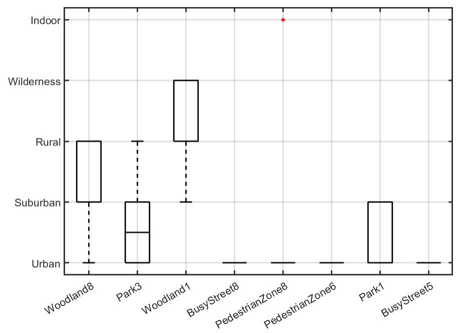
6.5 Limitations & Further Work

This experiment had several limitations that should be addressed in future work. The sample size of 20 participants used in the final analysis, 10 per group, was small and may not be representative of the wider population. Further, the demographic range of participants was limited, and further studies should aim to recruit participants from a wider range of backgrounds and ages in particular. Improving the variety of participants would help to improve the generalisability of the results, as demographic factors have been highlighted as influencing the perception of soundscapes in [218].

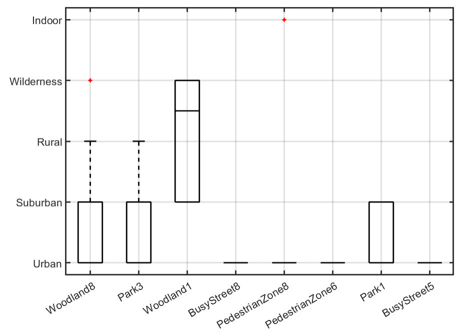
The choice to conduct this experiment remotely was, in part, initially due to the COVID-19 pandemic. Further, the data in this experiment were to be compared against another experiment that was to be conducted under laboratory conditions, using the same set of stimuli but performed using a head mounted display and utilising the tools described in Chapter 7. As this experiment was conducted remotely, it was not possible to control the playback system used by the participants. Further, the playback level and playback quality of the stimuli was not controlled. This lack of control will have increased the potential error in the range of results, and comparison under controlled conditions would be beneficial to understand the effect of playback system on the results of online soundscape evaluation experiments. However, as the results of the affect report between the groups of participants was similar, the effect of the playback system is likely to have been non-systematic and representative of the conditions of performing soundscape evaluation on a web-based platform. Future work should explore alternative platforms for dynamic binaural rendering of Ambisonics on the web such as the Omnitone library [222] or the HOAST library [223].

Although the selection of the stimuli was adequate to elicit a variety of affect quality scores, the quality of the stimuli reproduction may have had an influence on the results. The quality of the visual stimuli were low resolution which was a consequence of the recording system used to capture the visual stimuli. Also, the binaural renderer used by YouTube only supports up to first-order Ambisonics, resulting in lower quality spatial resolution in the binaural rendering of the stimuli compared to using higher order Ambisonic recordings. Because the IEM binaural renderer was used to render the static stimuli, the comparison of results between the groups of participants may have been influenced by this difference. Future work should aim to use the same binaural renderer to render both the static and dynamic stimuli, and to use higher order Ambisonic recordings to improve the spatial resolution of the stimuli.

Another limitation of this experiment was that although several questions related to the classification of the stimuli were included, this data was not included in the analysis section. This data was excluded as the experiment was designed to evaluate the effect of dynamic binaural rendering on the affective response of the listener, and because the classification of the stimuli was not the primary focus of the experiment. Figure 6.9 present two boxplots showing the distributions of *classification* scores for all stimuli for each group of participants. Figure 6.9a shows, and Figure 6.9b shows. The box plots of classification data in Figure 6.9 show that the distribution of the classification scores for



(a) The boxplot of classification scores for all stimuli for the dynamic group of participants.



(b) The boxplot of classification scores for all stimuli for the static group of participants.

Figure 6.9: The distribution of soundscape classification scores for all experimental intervals for both groups of participants.

the *Woodland8*, *Woodland1* and *PedestrianZone6* stimuli were slightly different between the two groups of participants. Future work should include a fully featured analysis of the classification data in this experiment, in order to determine if the classification of the stimuli, including its background keynotes and foreground sound marks, are perceived as being different due to the binaural rendering method that is used.

Finally, the experiment was conducted using a web-based platform that did not support head-tracked binaural rendering, and head mounted displays were not used. Utilizing head-tracked Ambisonic rendering or full immersion using a head mounted display may have improved the ecological validity of the stimuli and led to different results. Future work should explore these two options to determine if head-tracked dynamic binaural rendering leads to differences in perceived affective quality of soundscapes.

6.6 Conclusions

Twenty-four participants were recruited to evaluate a set of eight soundscapes, with each participant randomly assigned to one of two groups, and 20 participants completing the whole set of experiment intervals. The renderer used to decode the stimuli into binaural audio was different for each group. One group was presented with static binaural audio, audio that did not change with look direction, and the other group was presented with dynamic binaural audio that changed with look direction under control of the participant's mouse pointer. The results of the experiment confirmed that there was no difference in the perceived affective quality or the self-reported affect of the soundscapes between the two groups of participants. This result shows that the strategy of dynamic or static binaural rendering does not influence the perception of soundscapes in online soundscape evaluation experiments. However, the use of head-tracked dynamic binaural rendering of Ambisonic soundscape recordings may improve the immersion of the soundscape evaluation experience, and can lead to different results. The use of head-tracked dynamic binaural rendering for soundscape evaluation should be explored in future work.

Chapter 7

A Toolset for Soundscape Quality Evaluation in Virtual Reality

In Chapter 6 an experiment was presented that was designed to identify the importance of dynamic binaural rendering on the perceived affective quality of the soundscapes. The results of this experiment showed that in an online listening test, the ability to dynamically rotate the soundscape along with the camera view by using a computer mouse and cursor did not have a significant effect on the perceived affective quality of the soundscapes. However, due to the nature of the tools that were used to perform this experiment, it was not possible to test the effect of head-tracked binaural rendering on the perception of soundscapes. Neither was it possible to compare results that were gathered using several different modalities of visual presentation such as the use of a head-mounted display, a mobile phone or a desktop monitor. This chapter presents a toolset and environment that was developed for the purpose of conducting an experiment that would address these limitations. The toolset was developed for the purpose of performing soundscape evaluation and listening tests in virtual reality (VR), taking advantage of advances in mobile computing and game engine technology with the intention to improve the ecological validity of ex-situ soundscape evaluation. The toolset described in this chapter encapsulates future work, and an experiment is described at the end of the chapter that uses this framework for the subjective evaluation of soundscapes. A example usage of the toolset is described at the end of the chapter, providing a description of an experiment and an implementation of the toolset for the subjective and physiological evaluation of soundscapes.

7.1 Introduction

One of the challenges in the management of environmental noise is the communication of the estimated impact of noise on local communities, and an adjacent challenge is the collection of subjective data that reflects how people perceive the quality of an environment. Another challenge is gathering subjective data in controlled and repeatable conditions, or under circumstances where collecting subjective data in-situ is unfeasible. Further, improving the ecological validity of soundscape and multimodal environment reproduction may improve the accuracy and repeatability of the results of soundscape evaluation in a laboratory setting.

As described in Section 3.8, several methodologies have been developed for the evaluation of soundscapes, most of which have been developed for the purpose of in-situ soundscape evaluation. However, in-situ soundscape evaluation is not always possible or practical, and ex-situ soundscape evaluation including experiments performed at home or in a laboratory might be the only practical option. Virtual and augmented reality (VR) technologies can be used as a vehicle for the immersive presentation of soundscapes and landscapes. The quality and affordability of VR systems including spatial audio technologies have advanced significantly in the last decade, and researchers have begun exploring how these can be used in environmental noise and soundscape evaluation [224]. VR is also being utilized for acoustic evaluation by private sector organisations who are tendering for large public sector contracts. Arup have deployed virtual reality sound booths as part of the consultation for new works being proposed to one of the primary UK airports [225].

In Chapter 6 it was discussed that it is desirable to improve the ecological validity of soundscape evaluation by presenting soundscapes in a more immersive manner, improving the sense of presence, spatial immersion and emotional immersion experience by participants. The use of a head-mounted display (HMD) can have a significant impact on the perception of soundscapes, and the use of a HMD can also be used to present visual stimuli in a more immersive manner than a traditional desktop monitor. Jo and Jeon published an experiment that compared the results of a listening test performed with a computer monitor to the results of a listening test performed with a HMD [226]. The researchers reported that the sensitivity of the subjective assessment of soundscapes were increased when the soundscapes were presented using a HMD, however the authors also reported that the choice of visual stimuli presented in the experiments needed to be carefully considered when designing further experiments. If the use of a HMD increases the ecological validity of soundscape evaluation, causing participants to rate the perceived affective quality of soundscape more sensitively that in a traditional listening test, then it may be possible that the use of a HMD could improve the effect size of the effect of soundscape on the physiological responses of participants.

The influence of visual stimuli is an important consideration for researchers who are focussed on comparing the perception of soundscapes in different scenarios e.g. different acoustics noise barrier solutions. Further to the findings related to visual stimuli that were discussed by Jo and Jeon, Sanchez *et al.* found that the visual design of a noise barrier had a higher impact on the perceived quality of an environment than the acoustic effects of the barrier [227]. In this experiment the participants performed evaluations using a HMD and binaural rendering over headphones, and these results further support the importance of considering the design of visual stimuli used in soundscape evaluation experiments.

Researchers have begun to validate the ecological validity of immersive VR systems in the evaluation of soundscapes [228], finding several measures that can be used for this purpose. Luigi *et al.* reported that there were no significant differences in perceived affective quality of an environment evaluated in-situ and ex-situ evaluated using a HMD [229]. In this experiment soundscape stimuli were presented via a 5.1 surround sound loudspeaker system, as opposed to using an Ambisonic loudspeaker system or headphone based reproduction. This result suggests that participants may not be highly sensitive to the mode of soundscape reproduction when evaluating environments using a HMD. Hong *et al.* explored the effect of using static and dynamic binaural rendering of soundscape stimuli over headphones as well as the presentation of soundscapes over a first-order Ambisonic loudspeaker array while participants viewed the environments using a HMD [197]. The authors found that subjective results from soundscape evaluation utilizing several different spatial audio reproduction methods were not significantly different from in-situ evaluations, but differences in spatial quality were found between the different spatial audio reproduction methods.

The results described above appear to support the use of VR in soundscape evaluation, and researchers have begun to implement more general purpose VR systems from the purpose of environment evaluation. An example of a multi-platform system for the evaluation of soundscapes was developed by Puyana-Romero *et al.* [230], who designed the system to support several different modalities of usage. The authors found that there were no significant differences in subjective estimates of soundscapes between the modalities, and they further identified that a key challenge in maintaining the system was the disparity in the quality of spatial audio reproduction across the three modalities.

7.2 VR Application Development

The development of virtual reality experiments is primarily dictated by the choice of hardware and software that is used to implement the experiment, thanks to the support methods which support software development in VR. The choice of hardware and software platform is influenced by several key factors including the quality and cost of the hardware to be used, and the licensing scheme applied to the software tools used to develop the experiment. In the context of academic usage, most software is available free of charge, however the licensing scheme applied to the software may restrict the use of the software for commercial purposes. Applications that are intended for use with

HMDs are primarily developed with one of several game engines and a software development kit (SDK) that is supplied by the manufacturer of the target hardware such. A game engine is a software framework that is used to develop interactive applications, and typically includes support for the rendering of 3D graphics, spatial audio, physics simulation and user input. The SDK for a HMD will usually contain a set of precompiled software libraries that are required to interface with the hardware, along with an application programming interface (API) that can be used by the developer to integrate the software libraries with their application. The choice of game engine is therefore a key factor in the development of VR applications, as the game engine must support the SDK for the target hardware. One such game engine is Unity [231], which is a cross-platform game engine that supports the development of applications for a wide range of hardware platforms including desktop computers, mobile devices and HMDs. Another popular game engine is Unreal Engine [232], which is also a cross-platform game engine that supports the development of applications for a wide range of hardware platforms including desktop computers, mobile devices and HMDs. Both Unity and Unreal Engine support OpenXR, which is an open standard for the development of VR applications that was developed by the Khronos group [233]. The SDKs for many HMDs now support OpenXR, including the Valve Index, HTC Vive and Meta Quest 2, which means that applications developed using OpenXR can be deployed to a wide range of hardware platforms.

The range of available VR hardware platforms is constantly evolving, and the choice of hardware platform is a key factor in the development of VR applications. Some VR systems include the use of a large format projector to display images onto walls or a large domed surface, such as the Broomx MK360+ [234] and the Igloo systems [235]. These systems have the benefit that they can support the situation of several people at once and are combined with spatial audio reproduction systems to create a shared virtual environment for soundscape evaluation. This type of immersive system might be more appropriate than using multiple HMDs for the evaluation of soundscapes in a group setting, where open discussion and face to face interaction are desired. However, there are several drawbacks to using this type of system. One issue with using loudspeaker arrays is practicability of meeting the number of loudspeakers required to achieve the desired spatial resolution across a large frequency range [236]. Another issue is that there is a 'sweet-spot' where the sound field is most accurate, meaning the experience of each listener will not be the same [237]. Perhaps more ubiquitous than these large format display systems are the range of HMDs that are available. HMDs come in several sizes, categories and price brackets. Three popular HMDs are the Valve Index [238], HTC Vive pro [239] and Meta Quest 2 [240]. The Valve Index is a premium consumer headset that is designed to be used with a high performance gaming computer, and is therefore always tethered to that computer. The HTC Vive pro is a premium headset that is intended for the professional market, and is also designed to be used with a high performance gaming computer. Finally, the Meta Quest 2 is a relatively low cost consumer headset that is designed to be used without a computer, and is therefore a standalone device. However, the Meta Quest 2 can be tethered to a computer to enable the use of more computationally intensive applications.

Another important aspect of VR application development is the choice of audio system that will be used to reproduce audio content. To take advantage of spatial audio reproduction, the audio subsystem used with the game engine must be able to support the desired modality of spatial audio reproduction. For the reproduction of spatial audio over headphones, the game engine must support the use of binaural rendering of audio content. Both Unity and Unreal Engine natively support the reproduction of Ambisonic audio content, and the SDKs for all of the HMDs described above support the use of binaural rendering of Ambisonic audio content. For the multichannel reproduction of audio over loudspeakers, researchers have previously used external audio applications such as Max/MSP [241] to reproduce audio content [242].

Both Unity and Unreal have native support for first-order Ambisonics, and there are several third-party plugins that can be used to extend the support for Ambisonics in these game engines. Another method for extending the audio capabilities of a game engine is to use a third-party audio engine, typically described as audio middleware.

Two popular examples of audio middleware are FMOD [243] and Wwise [244]. Wwise is a middleware solution that is developed by Audiokinetic, and Fmod is a middleware solution that is developed by Firelight Technologies. Wwise has native support for the processing and playback of up to third-order Ambisonic audio, and has support for both native and third party binaural rendering plugins.

Both Wwise and Fmod as well as Unity and Unreal Engine are supported by Google Resonance Audio [245], an SDK that includes plugins for both binaural rendering and reverb generation. Meta provide the Oculus Audio SDK [246] for the development of spatial audio applications, including plugins for Unity, Unreal Engine, Fmod and Wwise. HTC provide the VIVE 3DSP Audio SDK [247] that supports up to third-order Ambisonic rendering in Unity and Unreal Engine. Finally, Valve provide the Steam Audio SDK [248] that supports Fmod, Unity and Unreal Engine. Each of these SDKs provides a wide variety of features including binaural rendering, reverb generation and occlusion effects.

7.3 Toolset

To study aspects of the use of VR in soundscape evaluation, a toolset was developed to support the creation of soundscape evaluation experiments. The toolset comprised of a set of software and hardware components that can be used to perform soundscape evaluation experiments. The headset that is used in this toolset is the Meta Quest 2 [240], which was chosen because it was relatively inexpensive, it is supported for development with both Unity and Wwise, it is a standalone unit that doesn't require a tether, and it continues to receive support and regular updates from Meta. Unity [231] was the game engine that was chosen for the toolset, because Unity has continued to integrate support for VR hardware include the OpenXR standard, the support for the Meta Quest 2 via the Oculus SDK has continued to receive support, and it has a licensing framework that is appropriate for academic use. Since the toolset was first being compiled, the license conditions for Unreal Engine have changed to make it more accessible for academic use, and the support for VR hardware has continued to improve. To support the use of spatial audio in the toolset, Wwise [244] was selected because it has native support for the use of Ambisonics and the rendering of single channel audio[4] into binaural audio via Google Resonance Audio [245]. Wwise also has a licensing framework that is appropriate for academic use. Finally, the toolset includes the VR Interaction Framework [249], a framework designed for Unity that integrates with several of the more popular HMD SDKs. This framework was chosen because it provides a suite of tools that ease the development of VR applications, it is inexpensive and it is purchasable via the Unity Asset Store.

An example of a user interface for an experiment that uses this toolset is presented in Figure 7.1. In an experiment that uses this toolset the participants are likely to experience a successive set of soundscapes which are paired with 360 still images that are presented via HMD. Figure 7.1 shows the view of a VR experiment environment as

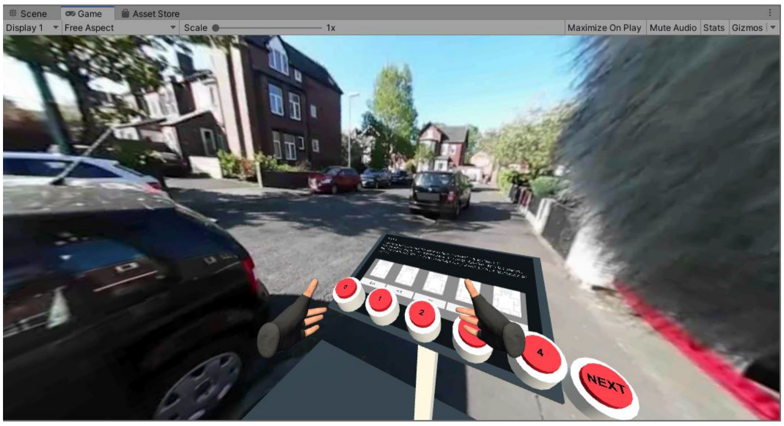


Figure 7.1: An implementation of a soundscape evaluation experiment as seen through the view of the software development environment.

seen from the Unity editor that is used as the software development environment for the experiment. This view would be stereoscopically rendered for a participant wearing a HMD, but is rendered in the Unity editor from a single camera perspective. The view in Figure 7.1 shows the visual stimuli of a 3D still image of a quiet street. The

[4]often referred to as object based audio

diegetic user interface presented to the participant is also visible, showing a section of the questionnaire presented on a text display and the buttons that participants can use to respond to the questionnaire. Finally the participant hand models are visible. Including a diegetic user interface that is presented as part of the game world is convenient as it enables the participant to remain immersed within the experiment situation and answer soundscape surveys in a way that is more similar to an in-situ soundscape assessment than a more traditional laboratory based listening test. However, the quality of text rendering and text legibility can be challenging in VR, and so a traditional pencil and clipboard form of presentation within the virtual world may be challenging to render successfully [250].

7.4 Implementation

The toolset was developed to support an experiment that can be used to identify if there are differences in the subjective experience of soundscapes, including physiological measures, between different modes of soundscape presentation including desktop, HMD and in-situ. The experiment follows a similar form to those described in Chapters 4, 5 and 6, in that participants experience a randomly ordered set of soundscapes and perform the same soundscape survey as in the experiment described in Chapter 6. However, the experiment is performed in VR using a HMD, and the same experiment can be performed on the desktop using the same stimuli and the same Unity software. If the same survey is performed in-situ and stimuli including Ambisonic recordings and 360 images, the three sets of soundscape evaluation data from each experiment can be cross-compared and validated against each other. An example Unity application for this experiment is included in the additional materials with this thesis as described in Appendix C.

To support the use of the toolset, a set of C# scripts for were developed that are intended for use in the creation of soundscape evaluation experiments. Figure 7.2 presents a block diagram of the software system that is used in the experiment described above. The block diagram in Figure 7.2 shows the context of the software components that are

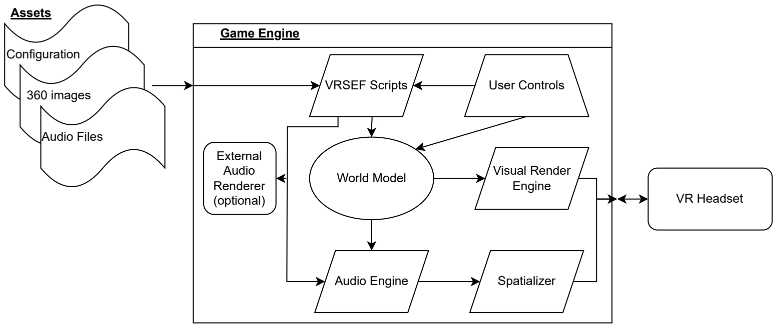


Figure 7.2: A high level block diagram overview of the components of the VR soundscape evaluation system.

used to create and deploy soundscape evaluation experiments. The game engine binary encapsulates various elements of the software system, including several scripted objects that are used to control the game world and the progression of the experiment. Several assets are included in the system that are external to the core application binary, represented on the left of the block diagram. The VR Headset hardware referenced on the right hand of the diagram interfaces with the binary, providing feedback and control information to the application. As represented in Figure 7.2, the game engine compiles a binary application and a set of resource files that are deployed to the VR headset, including the 360 images as textures, and the Ambisonic audio files as part of the Wwise soundbank for the project. Figure 7.3 presents a block diagram of the software objects that automate progress through the experiment. The block diagram in Figure 7.3 shows that the scripts developed for the toolset include interfaces between several singleton objects that also interface with the game engine asset loading system. The state machine directs several manager objects, setting the experiment stimuli and controlling the user interface elements presented to the participant. Where necessary, the game engine resource loading mechanism is used to read assets into the experiment. External input is used to trigger the state machine using the game engine call-back architecture. The scripts associated with the framework include several singleton objects that manage different parts of the system. A state machine organises the experiment procedure, and the configuration files in the assets are used to set up the state machine. As the participant

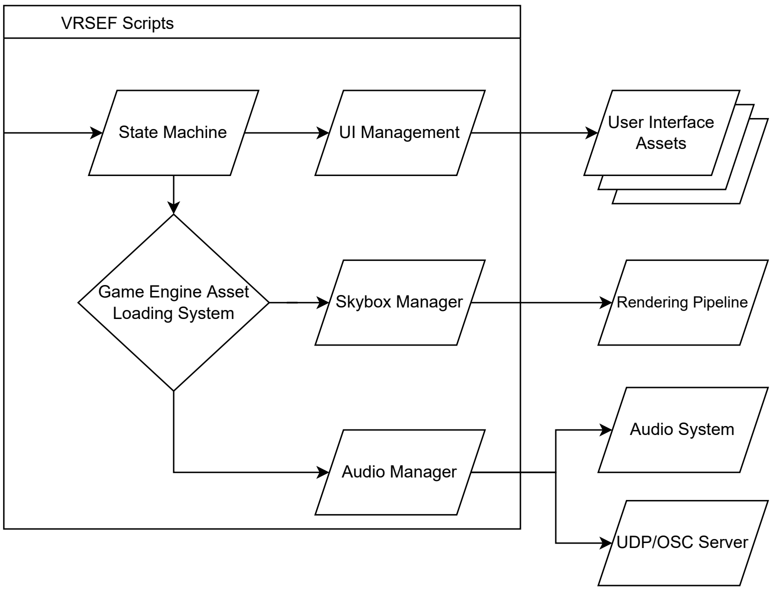


Figure 7.3: A block diagram of the connection between the singleton objects included in the scripts of the virtual reality soundscape evaluation framework.

is guided through the experiment procedure, the state machine interfaces with several singleton objects that control different aspects of the VR system. These include an audio manager that interfaces with the system's audio engine, and a user interface manager that presents visual information to the participant. Due to this modular architecture, the framework can be modified to support different audio systems and UI elements. This flexibility reflects the findings of previous research that has suggested loudspeaker rendering may be preferential in situations where verbal discussion is desirable, which is likely to be important when performing a group guided interview as part of a soundscape evaluation [224].

In the implementation of the experiment world at runtime, several key objects will need to be created and these will host the toolset scripts. Figure 7.4 shows the Unity editor when the game world is first instantiated. At the bottom of Figure 7.4 the scripts included in the toolset are visible, and in the left pane the Unity scene hierarchy is visible. Several base objects are presented including utility managers and game world assets that form the base of the game world. The key object as displayed in Figure 7.3

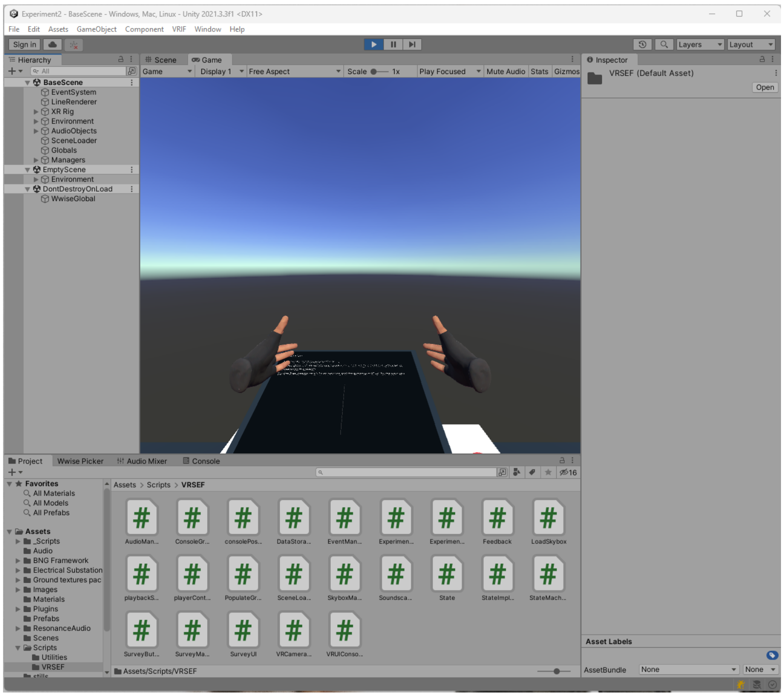


Figure 7.4: Example of the scripts and runtime objects that should be implemented as part of an experiment utilising the VR Soundscape Evaluation Framework for Unity.

is the experiment manager object that hosts the state and interacts with all of the other managers loaded into the experiment scene. The experiment manager is supported by several other managers that encapsulate different functionality required for the experiments to be performed. For example, the SceneLoader in the Hierarchy of Figure 7.4 is triggered by the experiment manager and loads and unloads scenes as the experiment progresses. Each experiment interval is a new scene that is loaded on top of the base scene. The Survey Manager encapsulates the user interface elements and progression for presenting and performing the same soundscape survey as shown in the experiment in Chapter 6, though without the drill down elements. The Environment Source object under the AudioObjects hosts the script responsible for audio playback. The Data Storage Manager is responsible for collecting and storing the resulting survey data in a way that can be recovered after the experiment. A JSON file is saved to the disk of the device used to perform the experiment, and in the case of an Oculus quest this data can be recovered via USB and searching the directory structure. The geometry objects that make up the game world in the base scene are kept under the Environment object in the Hierarchy in Figure 7.4.

The participant navigates through the experiment as directed via the test on the console that is presented to them at the start of the experiment. The participant uses the five option buttons to select the different options for each stage of the survey, and the next button is used to progress the experiment to the next stage of the survey. A view of the console as seen from the Unity editor is presented in Figure 7.5. Figure 7.5

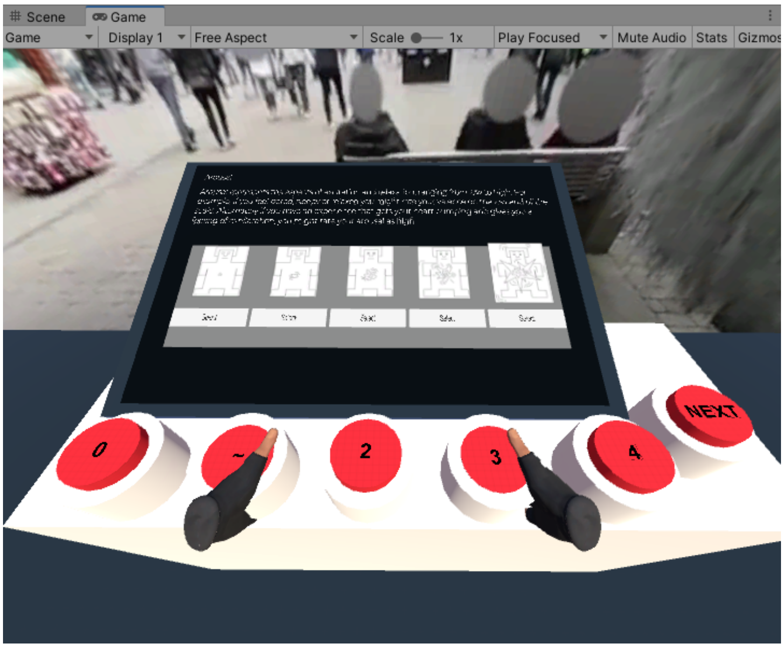


Figure 7.5: Example of an experiment interval scene having been loaded into the game world.

shows the collection of buttons and text display surfaces that make up the console that is used for performing the experiment. This presents a simple and legible text based interface, though an audio voice-over interface may be useful for further exploration of improved methodologies for soundscape evaluation in VR. The text in the view in Figure 7.5 may look poor in this particular image, but that is likely because of aliasing between the Unity editor and the screen resolution of the computer running the editor. When the experiment is performed, new scenes for each experiment interval are additively loaded on top of this base set of objects. An example of this is presented in Figure 7.6 which shows the same game world that has progressed in the test state, and now an experiment interval has been loaded. Each loaded soundscape scene has a script that is used to load

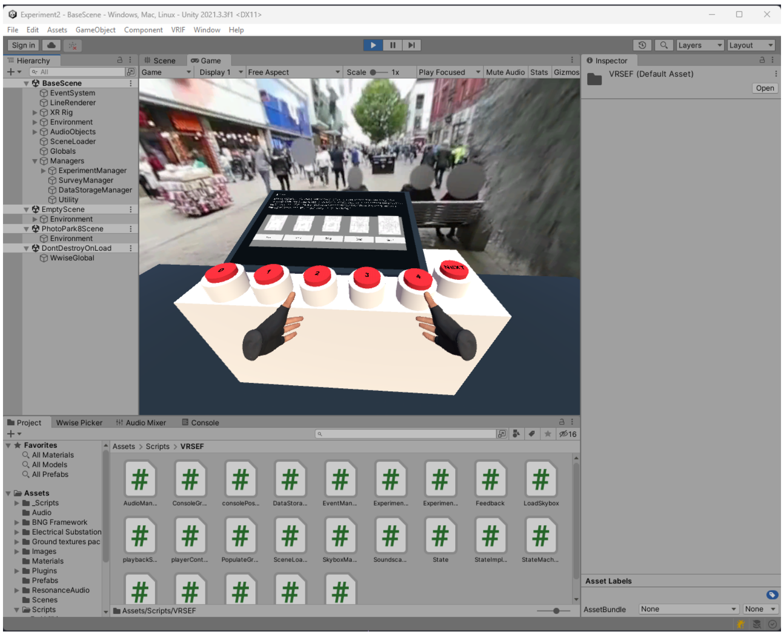


Figure 7.6: Example of an experiment interval scene having been loaded into the game world.

the 360 still image as visual stimuli, as is visible in Figure 7.6. However, the visual stimuli included in the environment of the additively loaded scene could also be things like loaded geometry and other graphical representations of the environment. Because of the modular nature of the project, all of the script based elements, managers and scenes are transferable to different implementations of the toolset to different environments and therefore soundscapes. This makes the toolset very useful for soundscape evaluation, as the practitioner can simply implement scenes into the Unity environment with their choice of visual and auditory stimuli, and have the experiment manager object load the scene at runtime as part of a randomised set of stimuli. Having this toolset and basis project available will greatly reduce the learning curve and complexity in implementing soundscape evaluation experiments in Unity.

7.5 Summary

Virtual reality systems are becoming increasingly accessible, and the use of virtual reality systems in soundscape evaluation could improve the ecological validity of soundscape assessment ex-situ. The development of soundscape evaluation experiments using head-mounted displays has been simplified thanks to the simplicity and accessibility of modern game engine technology. Soundscape evaluation in virtual reality has now begun to be implemented in research and industrial applications, and the intent of the toolset presented in this chapter is to simplify and minimise the effort required to develop soundscape evaluation experiments that are performed in virtual reality. However, another key finding in the literature related to soundscape evaluation in virtual reality is the choice of visual stimuli, and though it is beyond the scope of this chapter, choosing the correct visual stimuli is key to achieving accurate results in studies that use the toolset presented in this chapter. Although further work is required to validate the use of this toolset in soundscape evaluation across wider populations and scenarios, the results of literature summarised in this chapter are promising. In this chapter a toolset was described that is intended for the creation of soundscape evaluation experiments, simplifying the creation of soundscape evaluation experiment in Unity to the configuration of a collection of base objects and the creation of experiment interval environments as scenes. This toolset will be used in further research to investigate the importance of interactive and immersive technologies in the evaluation of soundscape, and serves as the starting point for any soundscape evaluation and psychoacoustics experiments that can be performed in VR.

Chapter 8

Conclusions

8.1 Summary

The work presented in this thesis has explored the use of physiological measures in the context of soundscape evaluation under laboratory or controlled ex-situ conditions, focussing on using low cost non-invasive sensors. This work was undertaken with the intention of identifying how the autonomic nervous system is influenced by soundscape, forming a basis for the use of soundscape and physiological measures in soundscape and well-being research. The foundations for acoustics and sound perception were described in Chapter 2, with a discussion of the physiological mechanisms of hearing and a description of the measurement, recording and reproduction of sound field contents. The intention of this chapter was to provide a basis for the discussion of the experimental work presented in the following chapters. Chapter 3 provided the context of environmental noise management and reviewed the current state of environmental noise management policy in the UK. This chapter provided context for the next chapter by highlighting the current state of environmental noise management and the limitations of this approach. Soundscape is then introduced in in this chapter as an alternative philosophy for sound management, paying attention to methods and measures used to assess soundscape quality. This chapter provides context for the form of experimental study undertaken in this thesis, with a discussion of the use of self-report measures in soundscape evaluation. Chapter 3 ends with a discussion of the use of physiological measures in soundscape research. Having established the context for this work, the following chapters then describe the experiments undertaken to investigate the use of physiological measures in soundscape evaluation. The first pilot experiment described in Chapter 4 investigated the use of photoplethysmography to measure heart rate in the context of soundscape evaluation. This chapter introduces an experimental design for the assessment of soundscapes using physiological measures, along with preliminary results. The key finding in this experiment was that the affective self-report of participants covaried with the nature of the stimuli, and soundscapes with predominantely natural sound sources were rated as more pleasant than those with predominantly mechanical sound sources. Although it was found that the experimental design was suitable for the assessment of the affective quality of soundscapes, further improvements to the experimental design were required to improve the reliability of the physiological measures. Building on this, a second experiment is described in Chapter 5 which investigated the use of electrocardiogram (ECG) based heart rate measurement in soundscape evaluation. This chapter describes improvements on the experimental design established in Chapter 4, include the analysis of heart rate variability measures, which are considered to be a more representative of affective state than heart rate alone. The key finding of this experiment was that the return rate of heart rate variability (rrHRV) was found to be significantly different between the stimuli, and further experiments that leverage larger sample sizes are likely to yield further significant results. Following this, Chapter 6 describes a third experiment which investigated the impact of dynamic binaural rendering on self-reported affect in soundscape evaluation. Finally, Chapter 7 describes a system designed for the evaluation of soundscapes using virtual reality (VR) and augmented reality (AR). This system is theorized to provide a more ecologically valid environment for the evaluation of soundscapes, and is intended to be used in future work.

8.2 Restatement of the Research Question

The work presented in this thesis aimed to test the following hypothesis: *The experience of soundscapes classified as featuring natural and mechanical sound sources evoke physiological and affective responses in listeners.* The experimental work presented in this thesis has aimed to identify physiological responses during the evaluation of recorded soundscapes, through the use of non-invasive low cost physiological sensors. In this thesis it has been shown that, under conditions that are typical for a controlled listening test, with participants seated in a listening room and interfacing with a computer, the auralisation of recorded soundscapes using binaural rendered techniques elicited states of affect in participants that covaried significantly with the nature of the stimuli. The experiment in Chapter 4 demonstrated that though heart rate and breathing rate may have been influenced by the nature of the stimuli, these changes were not able to be verified with statistical tests due to the small sample size used in the study. The subsequent experiment described in Chapter 5 demonstrated a statistically significant change in one heart rate variability metric, the return rate of heart rate variability (rrHRV). However, these changes were not consistent across participants and did not covary with the nature of the stimuli. In these experiments it was not possible to identify consistent and systematic changes in physiological measures that covaried with the state of affect that the participants reported under these experimental conditions. It was further demonstrated that if a binaural rendering technique is used in soundscape reproduction, the ability to dynamically change the orientation of the soundscape is unlikely to be a significant factor in the elicitation of affective responses to the soundscape. These findings provide evidence in support of the hypothesis that physiological and affective responses are evoked in listeners during the evaluation of soundscapes. Although the strength of the evidence is limited by the small sample size used in the experiments, the findings do, however, provide a basis for future work in the field of soundscape evaluation using physiological measures.

8.3 Contributions

In the thesis, the research that has been presented with a view to answering the hypothesis described above has resulted in the following contributions to the field:

- *The development of a novel experimental methodology for the evaluation of soundscapes using physiological measures.*

  The methodology described in this thesis improves on prior methodologies used for the investigation of physiological responses to soundscape stimuli by improving on both stimuli reproduction method and physiological measurement and analysis technologies that are used. The methodology developed through Chapters 4 and 5 incorporates the binaural rendering of naturally occurring soundscapes that were recorded in the Ambisonic format, the use of heart rate variability metrics as physiological measures of autonomic responses to soundscape stimuli, greater control over the timing of the experimental intervals when compared to similar studies identified in the literature, and advanced statistical analysis techniques that take advantage of multivariate statistical tools.

- *Knowledge pertaining to the application of low cost physiological measures in the evaluation of soundscapes.*

  The experimental work described in Chapters 4 and 5 of this thesis focus on the use of low cost readily available devices for the measurement of physiological responses to soundscape stimuli. These chapters describe two different sets of sensors, data collection tools, data analysis tools and supporting discussion that highlights benefits and challenges associated with using these tools in these experiments.

- *Evidence supporting the use of self-report measures in soundscape evaluation.*

  The results presented in Chapters 5 and 6 indicate that the use of self-report measures in the evaluation of soundscapes can yield consistent results across multiple experiments when comparing measures of affect and soundscape classification.

- *Evidence that the interactive nature of binaural rendering techniques used in the rendering of soundscape stimuli does not significantly influence the perceived affective quality of soundscapes in online listening tests.*

  The results presented in Chapter 6 indicate that in the context of an online listening test, the ability to interactively change the orientation of a soundscape does not significantly influence the perceived affective quality of the soundscape, hence indicating that static binaural rendering and dynamic binaural rendering may yield similar affective results in the context of soundscape evaluations.

- *The development of a toolset for the evaluation of soundscapes using virtual reality technologies and physiological measurements.*

  The toolset described in Chapter 7 provides a means of presenting soundscapes to participants in a manner that is theorised to elicit a higher degree of presence and spatial immersion than a traditional listening test that does not take advantage of immersive technologies.

8.4 Future Work

The work presented in this thesis has provided a basis for future work in the field, identifying a methodology that can be utilized for the assessment of soundscape quality using physiological measures. The choice of physiological measures explored in this thesis were somewhat limited, and an obvious development on top of the work presented in this thesis would be to increase the range of sensors that are explored. One example would be the use of electroencephalography (EEG) as recently discussed by Williams [251]. Further, a combination of sensors and sensor fusion techniques could further improve the quality of understanding in the effect of soundscapes on listeners. Li and Kang successfully identified statistically significant changes in several physiological indicators when participants experienced four different soundscapes [252]. However, further research is required to identify the extent to which these physiological responses can be used to infer the affective state of the participant. Confirming causality between physiological responses and affective state in response to soundscapes would cement the viability of using physiological measures in soundscape evaluation.

One key aspect of the experimental methodology presented in this thesis that could be greatly expanded upon was the mode of stimuli presentation. The presentation of stimuli and survey used in this thesis was limited to a form that could be described as a 'traditional' listening test. The experiment described in Chapter 6 showed that the minimum possible degree of interactivity in the presentation of stimuli did not significantly influence the perceived affective quality of the soundscape. However, it is the opinion of the author that there may be a minimum threshold of interactivity and spatial immersion that must be reached to elicit adequate engagement. It is has been established that physiological responses can be used to identify physiological behaviours in the contexts where participants are interactively engaged in activities within virtual worlds [253]. An extension of soundscape evaluation that could explore this would be the present soundscapes as part of a virtual environment to participants within a 3D game world. Further to this, the use of multisensory VR and AR technologies such as a head mounted display could provide an experience with a greater sense of sensory immersion and presence for participants. Exploring the application of VR and AR within the soundscape context could not only aid in the development of alternative methods of soundscape evaluation, but could also provide the environment required to improve the effect size of physiological measures in soundscape evaluation. Several recent studies have explored the use of VR in the context of soundscape evaluation [197, 254], and building on this base of work could provide a means of exploring the effect of soundscape on affective state in a more ecologically valid environment. Building upon the toolset described in Chapter 7, researchers can develop support for the importance of a high degree of presence and sensory immersion in lab based soundscape evaluation.

Finally, the work presented in this thesis has focussed on the use of recorded soundscape stimuli, eschewing the use of synthetic or artificially created stimuli in order to understand if physiological effects of real life soundscapes are observable in participants. Researchers could expand on this work by generating synthetic soundscapes that are designed to elicit specific responses in participants, and then investigate if physiological measures are able to identify these responses. However, the development and use of biofeedback techniques in the generation of soundscapes could provide an environment for better understanding the relationship between physiological responses and soundscapes.

8.5 Closing Remarks

This thesis has presented a body of work that forms the basis for exploring the use of physiological measures in soundscape evaluation. Being able to identify and understand why changes in human physiological behaviour occur, and infer information from those changes will help develop our understanding of how we experience the world. The work presented in this thesis has contributed to this goal, but further refinement of experiment methodologies and environment reproduction technologies will be crucial to this effort. A thorough review of the literature that is discussed in Chapters 2 and 3 has highlighted a wide breadth of knowledge that is required to perform research in this field, further showing how intrinsically multidisciplinary the required knowledge base is. It is the personal view of the author that achieving a complete understanding of the physiological and psychological processes involved in the perception of soundscapes will be impossible without greater multidisciplinary collaboration, and this is one area of effort that should be focussed on by researchers who are interested in this field of study. One of the primary outcomes of this body of research was an experimental methodology. The experiment presented in Chapter 4 serves as a starting point for this, and that is built upon with the experiment that is presented in Chapter 5. Developing a methodology through these two experiments has highlighted several practical challenges that will need to be overcome in further work. The work undertaken during the period of study that resulted in this thesis was greatly impeded by the COVID-19 pandemic, and under other circumstances the methodology would have been developed further through a subsequent experiment that is described in Chapter 7. However, the results of the experiment in Chapter 6 suggest that in the case of an experiment performed remotely via a web browser, that simple static binaural rendering may yield the same results from a soundscape survey as more complex binaural rendering schemes. It is the opinion of the author that, if the participant is not in a position to willingly suspend their disbelief at being situated in a virtual rendering of an environment, then there may be a minimum quality of spatial representation that is good enough to yield satisfactory experimental results. Utilising technologies like dynamic binaural rendering in combination with head mounted displays may provide an opportunity for experimental realism that may not be provided by a computer screen and a web browser alone. Taking advantage of the work presented in this thesis, researchers who are interested in the physiological evaluation of soundscapes are well positioned to build a better understanding of how we interact with our acoustic environment.

Appendices

Appendix A

Ethical Approval Documents

THE UNIVERSITY of York

Application Form for Physical Sciences Ethics Committee Approval

> ***Advice for applicants on completing the form***
>
> *Please ensure that the information provided is:*
> - *Accurate and concise*
> - *Clear and simple and easily understood by a lay person*
> - *Free of jargon, technical terms and abbreviations*
>
> *Further advice and information can be obtained from your departmental representative on the PSEC and at: http://www.york.ac.uk/admin/aso/ethics/cttee.htm*

Please return completed (typed) form to your departmental representative via email to:

elec-ethics@york.ac.uk

Title of project: Sounds Asleep - Towards better sleep quality with biometric noise metering and mediation

SECTION 1 DETAILS OF APPLICANTS

Details of principal investigator (name, appointment and qualifications)

Simon Durbridge
PhD Student
Audio Lab, Department of Electronics
BSc Sound, Light & Live Event Technology, MSc Audio Engineering

Names, appointments and qualifications of additional investigators *(student applicants should include their project supervisor(s) here)*

Prof. Damian Murphy (Primary Supervisor)
Professor
Audio Lab, Department of Electronics
BSc (Hons), MSc, DPhil

Dr Duncan Williams
Post-doctoral Researcher
Digital Creativity Labs, Department of Computer Science
PhD

Location(s) of project

Audio Lab Genesis 6 Innovation Way Heslington York

SECTION 2 FUNDERS

What is the funding source(s) for the project?

EPSRC – Doctoral Training Award

Please answer the following:

(i) Does the express and direct aim of the research or other activity raise ethical issues?

YES [x] NO [ ]

(ii) Is there any obvious or inevitable adaptation of research findings to ethically questionable aims?

YES [ ] NO [x]

(iii) Is the work being funded by organisations tainted by ethically questionable activities?

YES [ ] NO [x]

(iv) Are there any restrictions on academic freedoms – notably, to adapt and withdraw from ongoing research, and to publish findings?

YES [ ] NO [x]

If you answered **Yes** to any of the above, please give details below:

Participants will be asked to rate the pleasantness of soundscapes, and as such there might be unpleasant soundscapes in the experiment. However, all soundscapes used in the experiments are naturalistic and will be ecologically valid. i.e. the participants are likely to experience these soundscapes during day to day life. The risk of exposure to unpleasant and disturbing soundscapes is further mitigated by allowing participants to stop the experiment at any time without reason or penalty. Participant comfort will be monitored throughout the experiment.

SECTION 3 DETAILS OF PROJECT OR OTHER ACTIVITY

Aims (100 words max)

The aim of the project is to identify any relationships between human classification and rating of soundscapes, and human physiological parameter changes during and after listening to those soundscapes. This includes the use of graphical survey methods, and physiological sensing using non-invasive, commercially available, purpose built, worn physiological sensors; measuring physiological parameters such as heart rate, skin conductance and brain activity potential.

Background (250 words max)

The use of affective response survey methods such as the self-assessment manikin and sematic differential analysis is well established in the fields of psychology and in soundscape research. The results of soundscape rating experiments using semantic differential pairs often identify three principal dimensions, and these three dimensions are directly represented in the self-assessment manikin as valence (pleasantness), arousal(excitement) and dominance(control).

It is well identified in environmental noise epidemiological literature that long-term annoyance and sleep disturbance are linked to a wide range of negative health effects and diseases. There is a large body of strong evidence for physiologically measurable reactions of environmental noise exposure causing sleep disturbance.

While studies in the field of soundscape often rate the quality of soundscapes subjectively using the tools introduced above, there is a small pool of emerging literature that compares these subjective ratings to changes in physiological data. Finding relationships subjective evaluations, classification and physiological parameter variation could provide great utility to further soundscape research; providing a basis for soundscape studies using sleeping participants, study designs that allow participants to rate soundscapes without the use of surveys, and laying groundwork for the creation of community noise maps using physiological response data.

Brief outline of project/activity (250 words max)

The experiment will take place in the listening room in Genesis 6. The experiment will involve the presentation of five, five-minute long soundscape recordings. The soundscapes will be selected from the Eigenscape database and will be chosen to represent a range of locations such as woodland and urban. The soundscapes will be reproduced using an ambisonic loudspeaker array. Three minutes silence will be present between each soundscape presentation.

For the duration of the experiment participants externally measurable physiological parameters will be recorded using commercially available, purpose built, non-invasive sensor equipment. This physiological data will be recorded via a computer to an encrypted storage (hard) drive. After each soundscape presentation, participants will record a subjective rating of the soundscape that includes categorical and affective ratings.

Study design *(if relevant – e.g. randomised control trial; laboratory-based)*

The study is a laboratory-based randomised survey.

If the study involves participants, how many will be recruited?
Between 5 and 30 participants

If applicable, what is the statistical power of the study, i.e. what is the justification for the number of participants needed?
N/A

SECTION 4 RECRUITMENT OF PARTICIPANTS

How will the participants be recruited?

Participants will be recruited by email.

What are the inclusion/exclusion criteria?

Participants with known hearing impairment will be excluded in order to maintain data quality.
Participants with known heart ailments will be excluded in order to maintain data quality.
Participants with known neurological ailments will be excluded in order to maintain data quality.
Participants with damage to the skin where physiological measurements might take place will be excluded from the study, in order to minimise any risk of exposure.
Participants who are younger than 18 years of age will be excluded to simplify the consent process.
Participants who are older than 60 years of age will be excluded in order to maintain data quality.

Will participants be paid reimbursement of expenses? YES [ ] NO [X]

Will participants be paid? YES [ ] NO [X]

If yes, please obtain signed agreement

Will any of the participants be students? YES [X] NO [ ]

SECTION 5 DATA STORAGE AND TRANSMISSION

If the research will involve storing personal data, including sensitive data, on any of the following please indicate so and provide further details (answers only required if *personal* data is to be stored).

Manual files	**Yes**
University computers	**No**
Home or other personal computers	**No**
Laptop computers, tablets	**No**
Website	**No**

Please explain the measures in place to ensure data confidentiality, including whether encryption or other methods of anonymisation will be used.

Data will made confidential by not gathering the subjects name or other personal details that can be used to identify that person directly e.g. Sex, age range, confirmation of health and consent will be stored against a randomly assigned participant number. This survey data, and response data will be stored on an encrypted digital storage (hard)drive that will remain locked in the office of Simon Durbridge.

Please detail who will have access to the data generated by the study.

Simon Durbridge, Damian Murphy & Duncan Williams

Please detail who will have control of and act as custodian for, data generated by the study.

Simon Durbridge

Please explain where, and by whom, data will be analysed.

The data will be analysed by Simon Durbridge, at the Audio Lab on the University of York Campus.

Please give details of data storage arrangements, including where data will be stored, how long for, and in what form.

Data will be stored digitally on an encrypted portable storage device, and data will be accessed and temporarily stored in RAM for the process of analysis (as when RAM is cleared the local copy of the data is destroyed).

SECTION 6 CONSENT

Is written consent to be obtained?

YES	X	NO	

If yes, please attach a copy of the information for participants

If no, please justify

Will any of the participants be from one of the following vulnerable groups?

Children under 18	YES		NO	X
People with learning difficulties	YES		NO	X
People who are unconscious or severely ill	YES		NO	X
People with mental illness	YES		NO	X
NHS patients	YES		NO	X
Other vulnerable groups (if ‘yes’, please give details)	YES		NO	X

If so, what special arrangements have been made for getting consent?

N/A

SECTION 7 DETAILS OF INTERVENTIONS

Indicate whether the study involves procedures which:

Involve taking bodily samples	YES		NO	X
Are physically invasive	YES		NO	X
Are designed to be challenging/disturbing (physically or psychologically)	YES		NO	X

If so, please list those procedures to which participants will be exposed:
- N/A

List any potential hazards:
- N/A

List any discomfort or distress:
- N/A

What steps will be taken to safeguard

(i) the confidentiality of information

The identities of participants will not be collected, so participants will provide responses anonymously. All data will be stored on an encrypted hard drive, which will remain locked in the office of Simon Durbridge.

(ii) the specimens themselves?

Participation in the study is completely voluntary, and participants can choose to withdraw at any point without giving a reason. The study equipment shall be used such as described in all instruction manuals, employing a high regard for the safety and comfort of participants. The experimental conditions will be such that sound levels are at safe and comfortable levels.

What particular ethical problems or considerations are raised by the proposed study?

What do you anticipate will be the output from the study? ***Tick those that apply:***

Peer-reviewed publications	X
Non-peer-reviewed publications	X
Reports for sponsor	X
Confidential reports	X
Presentation at meetings	X
Press releases	
Student project	X

Is there a secrecy clause to the research?

YES		NO	X

SECTION 8 SIGNATURES

The information in this form is accurate to best of my knowledge and belief and I take full responsibility for it.

I agree to advise of any adverse or unexpected events that may occur during this project, to seek approval for any significant protocol amendments and to provide interim and final reports. I also agree to advise the Ethics Committee if the study is withdrawn or not completed.

Signature of Investigator(s): ………………………………………………………

………………………………………………………

Date: …12/6/2019………………………………………

Responsibilities of the Principal Researcher following approval • If changes to procedures are proposed, please notify the Ethics Committee • Report promptly any adverse events involving risk to participants

UNIVERSITY of York

Department of Electronics

CONSENT FORM

Name of Researchers: Simon Durbridge, Prof. Damian Murphy, Dr. Duncan Williams.

Title of Project: *Sounds Asleep – Towards biometric noise metering*
Experiment: 1A

Please tick all boxes

1. I confirm that I have read and understand the information sheet for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily. ☐

2. I have informed the principal investigator that I do not meet any of the exclusion criteria that are detailed in the information sheet. ☐

3. I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason. If I withdraw then I give consent for the data I have given up to that point to be retained. ☐

4. I understand that all data collected during the listening test will be anonymous and no identifiable personal information will ever be published. ☐

5. I understand that this consent form will be kept in a secure environment only available to Simon Durbridge, Prof. Damian Murphy and Dr. Duncan Williams. The information on this consent form will never be published. ☐

6. I give permission for the data obtained during the listening test to be retained for 5 years. ☐

7. I agree to take part in the above study. ☐

________________	________________	________________
Name of Participant	Date	Signature
Simon Durbridge	________________	________________
Name of Principal Investigator	Date	Signature

Confidential subject ID (to be filled in by the researcher): ☐

Department of Electronics

Experiment Outline for Participants

Name of Researchers: Simon Durbridge, Prof. Damian Murphy, Dr. Duncan Williams.

Title of Project: *Sounds Asleep – Towards biometric noise metering*
Experiment: 1A

Experiment Overview

Have you ever taken a walk through a park and heard the wind in the trees, or sat in quiet contemplation on a park bench as birds chirp around you? What makes the atmosphere of a café or a restaurant? The sonic landscape or *soundscape* is an integral part of the identity of the places we inhabit, and our perception of the soundscape might shape our emotional behaviour and identity.

The purpose of this experiment is to get a glimpse of how people classify and perceive different soundscapes, through a combination of survey fulfilment and biometric sensing.

Participant Exclusion Criteria

Due to the nature of this experiment, there are a range of exclusion criteria that you as the participant and the principal investigator must be are of.

You should be excluded from the study if you identify with any of the following:
- Any hearing impairment
- Any heart condition or ailment
- Sensitive or damaged (broken) skin around the wrist or fingers
- Any injuries around the skin or fingers
- Any know neurological condition or ailment
- Are younger than 18 years of age
- Are older than 60 years of age

You can choose not to disclose which exclusion criteria you meet to the principal investigator, but if you do meet any of these exclusion criteria please exclude yourself from the study.

Participant Activity

Classification

As a participant in this study you will be asked to listen to and then classify a series of soundscapes. Three classes or types of sound source are commonly used to identify sounds and qualities of soundscapes:

- Natural
- Human
- Mechanical

Further to this three-type classification scheme, a faceted taxonomy or structure of classes will be presented so that you can identify the primary sounds in the soundscape.

Self-Assessment Manikin

As a participant in this study you will be asked to rate your feelings after listen to a series of soundscapes. The rating will be performed using three scales which form the ‘Self-Assessment Manikin’. These three scales depicted in figure 1 relate to your rating of *Valence* (Sad <> Happy), *Arousal* (Calm <> Excited) and *Dominance* (Low Control <> High Control).

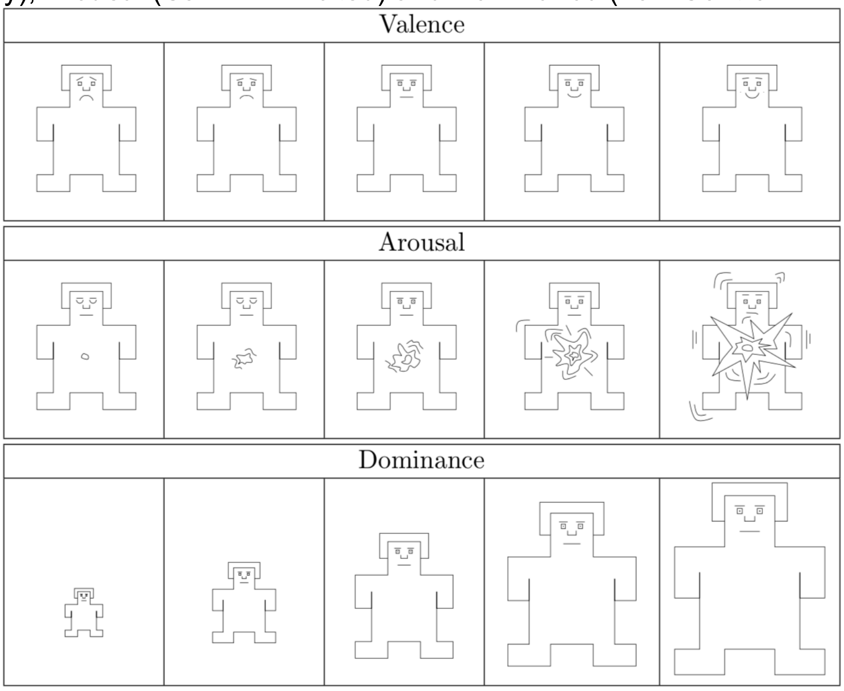


Figure 1The Self-Assessment Manikin [1]

Pleasantness, Calmness & Vibrancy

As part of your rating of the soundscape, you will be asked to rate the soundscape in terms of its pleasantness, calmness and vibrancy. Pleasantness will be rated on a scale of 0 to 10, from most unpleasant to most pleasant. The scales of calmness and vibrancy will be presented as two sliding scales from 0, 0 (bottom left of the following graph) to 10,10 (top right of the following graph). The following graph represents the plane of calmness and vibrancy:

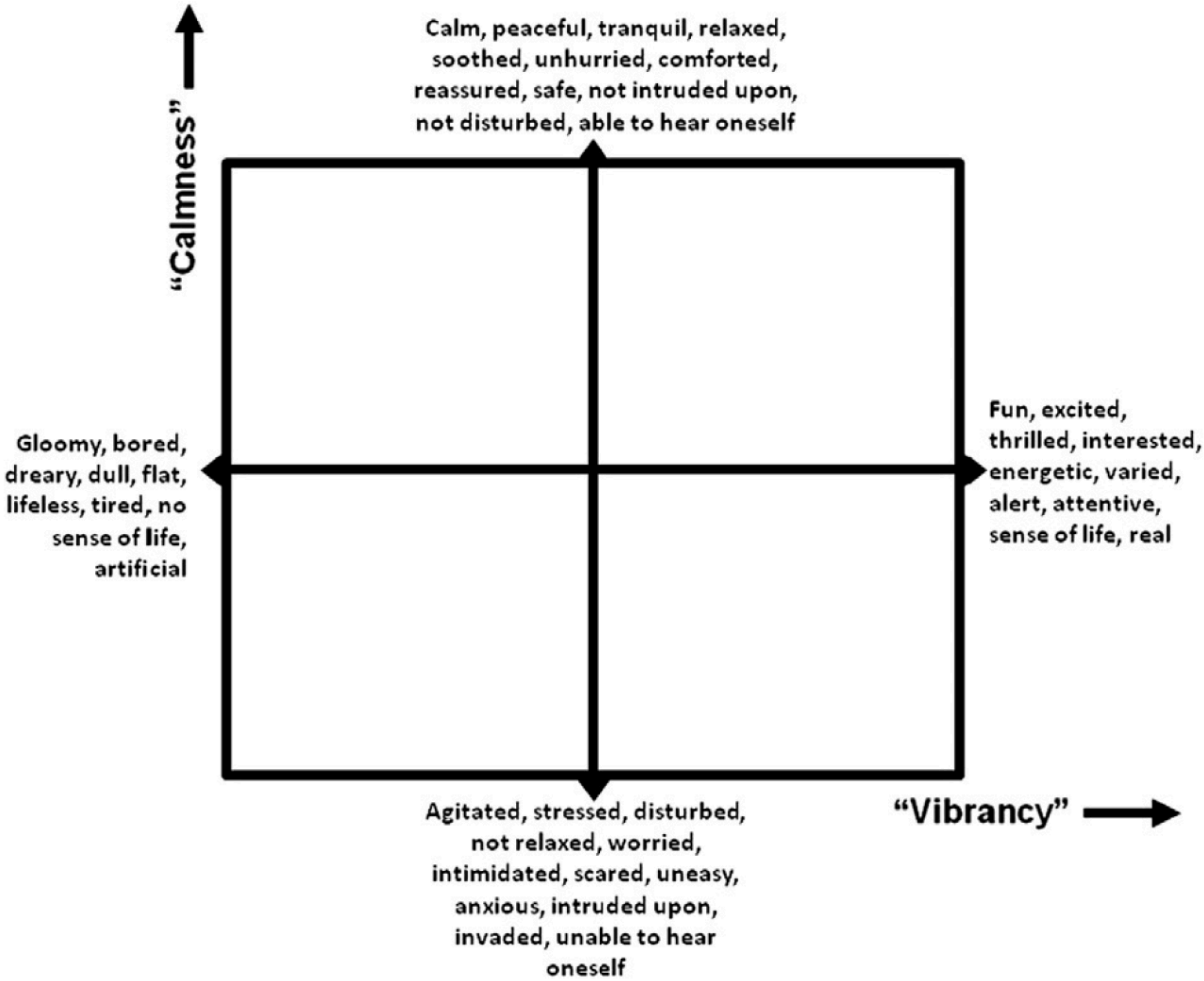


Figure 2Pleasantness Vibrancy Rating Graph [2]

Biometrics

To complement your emotional responses to the soundscapes, your biometrics will be recorded using a small non-invasive wearable device such as the Shimmer Sensing GSR+. Figure 2 shows the shimmer that you will wear during the experiment. The Shimmer Sensing GSR+ is a small, lightweight, battery powered device that senses biometric data such as heart rate, skin conductance and acceleration.

Figure 3Shimmer Biometric Sensing Device

Method

Upon reading this information sheet and signing the participation agreement, the experiment shall continue as such:

1. You shall be fitted with the biometric measurement device
2. You shall complete a short demographics survey
3. You be introduced to the experiment protocol by the principal investigator
4. You shall listen to one training soundscape, and you will complete one training survey
5. You shall rest until directed to begin the next round of the experiment; this is one round of the test procedure.
6. The acclimatising and training section is now complete, and you can continue to perform the cycle of listen>survey>rest until the experiment is complete.

The principal investigator will instruct you on the number and length of soundscapes in the experiment at the beginning of the experiment.

Participant Comfort

Your comfort as a participant is important, and in the event of any physical or emotional discomfort you can choose to terminate the experiment at any time. You will not be expected to give any reason for choosing to stop the experiment. If you have any concerns over discomfort that might occur during the experiment, please raise these with the principal investigator.

References (for further reading)

[1] F. Stevens, D. T. Murphy, and S. L. Smith, "the Self-Assessment Manikin and Heart Rate: Responses To Auralised Soundscapes," pp. 1–8, 2016.

[2] R. Cain, P. Jennings, and J. Poxon, "The development and application of the emotional dimensions of a soundscape," *Appl. Acoust.*, vol. 74, no. 2, pp. 232–239, 2013.

THE UNIVERSITY *of York*

Application Form for Physical Sciences Ethics Committee Approval

Advice for applicants on completing the form

Please ensure that the information provided is:
- *Accurate and concise*
- *Clear and simple and easily understood by a lay person*
- *Free of jargon, technical terms and abbreviations*

Further advice and information can be obtained from your departmental representative on the PSEC and at: http://www.york.ac.uk/admin/aso/ethics/cttee.htm

Please return completed (typed) form to your departmental representative via email to:

elec-ethics@york.ac.uk

Title of project: Towards the Evaluation of Soundscape Through Affective, Physiological and Psychoacoustic Measures 2A

SECTION 1 DETAILS OF APPLICANTS

Details of principal investigator (name, appointment and qualifications)

Simon Durbridge
PhD Student
Audio Lab, Department of Electronics
BSc Sound, Light & Live Event Technology, MSc Audio Engineering

Names, appointments and qualifications of additional investigators *(student applicants should include their project supervisor(s) here)*

Prof. Damian Murphy (Primary Supervisor)
Professor
Audio Lab, Department of Electronics
BSc (Hons), MSc, DPhil

Location(s) of project

Audio Lab
Genesis 6
Innovation Way
Heslington
York

SECTION 2 FUNDERS

What is the funding source(s) for the project?

EPSRC – Doctoral Training Award

Please answer the following:

(i) Does the express and direct aim of the research or other activity raise ethical issues?

YES	x	NO	

(ii) Is there any obvious or inevitable adaptation of research findings to ethically questionable aims?

YES		NO	x

(iii) Is the work being funded by organisations tainted by ethically questionable activities?

YES		NO	x

(iv) Are there any restrictions on academic freedoms – notably, to adapt and withdraw from ongoing research, and to publish findings?

YES		NO	x

If you answered **Yes** to any of the above, please give details below:

Participants will be asked to evaluate virtual representations of environments which are presented via head mounted display (VR). Participants will be surveyed and will be asked to use wearable physiological sensors, and physiological signals will be measured throughout the experiment.
The primary ethical issues raised in this experiment are those of the potential for the participants to experience simulation sickness from VR representation, and of personal privacy of the participants.

SECTION 3 DETAILS OF PROJECT OR OTHER ACTIVITY

Aims (100 words max)

The primary goal of this experiment is intended to identify if the representation of soundscapes in VR is sufficient to elicit changes in affect that are represented through physiological measures.
The secondary goal is to identify if self-report of subjective estimates of soundscape is itself a confound in identifying changes through physiological measures.
The tertiary goal is to identify if the representation of soundscapes through head mounted displays elicits similar subjective assessments as when those soundscapes are experienced through 360 video in a web browser.

Background (250 words max)

Soundscape studies attempt to identify the relationships between how people experience environments and the qualities of those environments, specifically from the perspective of the audible aspects of the environments. A key focus of soundscape research is to identify causal relationships between features of the soundscape. A challenge in this regard is to develop appropriate methodologies for generating and analysing data from which these relationships can be drawn.

The ground truth of the quality of the soundscape is a consequence of the perception of the person experiencing the soundscape, defining the necessity for listening tests that attempt to measure and interpret responses to soundscape stimuli. This presents significant issues related to experiment validity, repeatability and generalisability. Soundscape studies often employ surveys that include a mechanism for measuring affect, or the underlying experience of feeling, emotion, or mood. It could be that by performing self-report, participants are skewing their responses towards social expectations.

A growing body of soundscape research has begun to explore if physiological measurements can be used to identify participants responses to soundscape stimuli. An initial experiment first performed by Durbridge and Murphy in 2019 failed to identify significant physiological responses to soundscape stimuli in a small-scale study of responses to the binaural presentation of recorded soundscapes. This experiment is a continuation of that initial work. Relevant published works have often found conflicting results, identifying a high degree of uncertainty in the use of physiological measurement for affect classification.

Brief outline of project/activity (250 words max)

Brief outline of project/activity (250 words max)

The aim of this experiment is to identify the presence of different physiological and subjective responses to soundscape stimuli. The experiment will take place in Genesis 6. The experiment will involve the presentation of several soundscape stimuli supported by visual stimuli that will be presented via a VR head mounted display. The visual stimuli will include user-interface elements as well as graphical representations of environments, such as 360 still images. For the duration of the experiment, participants externally measurable physiological parameters will be recorded using commercially available, purpose built, non-invasive wearable sensors.

Study design *(if relevant – e.g. randomised control trial; laboratory-based)*

The study is a laboratory-based repeated measures listening test

If the study involves participants, how many will be recruited?

At least 60

If applicable, what is the statistical power of the study, i.e. what is the justification for the number of participants needed?

The presence of an effect and the effect size is contested in the literature and previous experiments from the lab, so estimation of an effect size is unlikely to be accurate. 30 participants per condition is generally considered a rule-of-thumb minimum due to assumptions of normality in data distributions. Experiments in the literature appear to follow this rule by including upwards 30 participants per condition, this experiment will have two conditions including self-report and no-self-report.

SECTION 4 RECRUITMENT OF PARTICIPANTS

How will the participants be recruited?

Participants will be recruited by email and organised through a private online calendar .

What are the inclusion/exclusion criteria?

Significant hearing impairment heart conditions pacemakers and other implanted electronics neurological conditions damaged or broken skin where wearable devices might be used Participants who are younger than 18 years of age Participants who are older than 60 years of age

Will participants be paid reimbursement of expenses?

YES		NO	X

Will participants be paid?

YES	X	NO	

If yes, please obtain signed agreement

Will any of the participants be students?

YES	X	NO	

SECTION 5 DATA STORAGE AND TRANSMISSION

If the research will involve storing personal data, including sensitive data, on any of the following please indicate so and provide further details (answers only required if *personal* data is to be stored).

Manual files	**No**
University computers	**Yes**
Home or other personal computers	**No**
Laptop computers, tablets	**No**
Website	**No**

Please explain the measures in place to ensure data confidentiality, including whether encryption or other methods of anonymisation will be used.

Participant identities (name) and contact information (email) will be collected as part of the informed consent procedure. Participant's identities will be kept separate from the experimental data. Informed consent forms will be stored on private university-owned storage (GDrive).

Please detail who will have access to the data generated by the study.

Simon Durbridge, Damian Murphy, Public

Please detail who will have control of and act as custodian for, data generated by the study.

Simon Durbridge

Please explain where, and by whom, data will be analysed.

The data will be analysed by Simon Durbridge, at the Audio Lab on the University of York Campus. The data may also be shared publicly as part of the publication procedure, subject to informed consent from the participants. As such anyone with access to the data will be able to analyse it.

Please give details of data storage arrangements, including where data will be stored, how long for, and in what form.

Data will be stored on the university private encrypted storage (GDrive) in the form of CSV files for 4 years or until the data custodian loses access to the university storage, whichever comes first.

SECTION 6 CONSENT

Is written consent to be obtained?

YES	X	NO	

If yes, please attach a copy of the information for participants

If no, please justify

Will any of the participants be from one of the following vulnerable groups?

Children under 18	YES		NO	X
People with learning difficulties	YES		NO	X
People who are unconscious or severely ill	YES		NO	X
People with mental illness	YES		NO	X
NHS patients	YES		NO	X
Other vulnerable groups (if 'yes', please give details)	YES		NO	X

If so, what special arrangements have been made for getting consent?

N/A

SECTION 7 DETAILS OF INTERVENTIONS

Indicate whether the study involves procedures which:

Involve taking bodily samples	YES		NO	X
Are physically invasive	YES		NO	X
Are designed to be challenging/disturbing (physically or psychologically)	YES		NO	X

If so, please list those procedures to which participants will be exposed:

- N/A

List any potential hazards:

- N/A

List any discomfort or distress:

- N/A

What steps will be taken to safeguard

(i) the confidentiality of information

The data generated in the experiment will be stored in an anonymised fashion.
Data that cannot be anonymised will be sorted privately and securely and will not be published.
All data will be stored on university-owned private secure storage (Gdrive).

(ii) the participants themselves?

Participation in the study is voluntary and participants can choose to withdraw at any point without giving a reason. The study equipment shall be used such as described in the manufacturer's instruction manuals, employing a high regard for the safety and comfort of participants. The experimental conditions will be such that sound levels are at safe and comfortable levels. The condition and comfort of participants will be observed throughout the course of the experiment.
Participants will be protected from slips, trips, falls, simulator sickness and exposure risk (Covid19) by following the risks identified in the associated risk assessment for performing VR based experiments in Genesis 6.
Participant identities will be protected and not exposed to the public.
Informed consent for participation and experimental (anonymised) data publication will be gathered prior to performing the experiment.

What particular ethical problems or considerations are raised by the proposed study?

Experiment data will be published alongside publication of the results of the study. To protect the identities of the participants all data that could be used to identify a participant will be anonymised or collected in a way that obfuscates the identification of participants.
Participants will be required to give informed consent for participation, data storage and publication before being allowed to participate in the experiment. An information sheet will be provided to participants, clearly informing participants of the experiment procedure and how the data will be used and shared.

What do you anticipate will be the output from the study? ***Tick those that apply:***

Peer-reviewed publications	X
Non-peer-reviewed publications	X
Reports for sponsor	X
Confidential reports	X
Presentation at meetings	X
Press releases	
Student project	X

Is there a secrecy clause to the research?

YES		NO	X

SECTION 8 SIGNATURES

The information in this form is accurate to the best of my knowledge and belief and I take full responsibility for it.

I agree to advise of any adverse or unexpected events that may occur during this project, to seek approval for any significant protocol amendments and to provide interim and final reports. I also agree to advise the Ethics Committee if the study is withdrawn or not completed.

Signature of Investigator(s): ……………………………………………………………

……………………………………………………………

Date: …23/9/2021………………………………………

Responsibilities of the Principal Researcher following approval

- If changes to procedures are proposed, please notify the Ethics Committee
- Report promptly any adverse events involving risk to participants

Participant Consent Form

University of York - Department of Electronic Engineering

Project Title: Towards the Evaluation of Soundscape Through Affective, Physiological and Psychoacoustic Evaluation

This experiment will be used to explore in influence of experiencing soundscapes in VR on peoples subjective evaluation of those soundscapes.

Please read the following questions and statements carefully. You are required to answer all sections and confirm that all of the statements are true, by ticking the checkboxes, before submitting the form.

Thank you for your time and participation.

1. Email *

   ____

2. I have read participant information sheet and I understand what what is involved in performing the experiment.

   Tick all that apply.

   ☐ True

3. I agree to take part in this project.

   Tick all that apply.

   ☐ True

4. I have been given the opportunity to ask any questions and had them answered to my satisfaction. I also understand that I will be free to ask further questions about the study after completing this consent form.

   Tick all that apply.

   ☐ True

5. I understand that my participation is voluntary and that I am free to withdraw from the study at any time without giving a reason.

   Tick all that apply.

   ☐ True

6. I understand that my participation in this project will be treated anonymously. I understand that my data will be stored securely and will be disposed-of in line with the universities data holding policies.

   Tick all that apply.

   ☐ True

7. I understand that data generated in this study may be published or made publicly available. I consent to the publication of the anonymised data, to be used for research purposes.

   Tick all that apply.

   ☐ True

8. I confirm that, to the best of my knowledge, I have no hearing impairments.

   Tick all that apply.

   ☐ True

9. I confirm that, to the best of my knowledge, I have no medical issues such as a heart condition or neurological impairment.

   Tick all that apply.

   ☐ Option 1

10. I confirm that, to the best of my knowledge, I have no implanted electronics or other electro-physiological systems that could interfere with the experiment apparatus.

Tick all that apply.

- [ ] Option 1

11. I confirm that, to the best of my knowledge, I do not have photosensitive epilepsy or any medical conditions where flashing images may trigger an adverse reaction, seizure, etc.

Tick all that apply.

- [ ] True

12. I am a trained/experienced listener

Tick all that apply.

- [ ] True
- [ ] False

13. I am over the age of 18 and under the age of 60

Tick all that apply.

- [ ] True

14. I understand that I will be remunerated with Amazon vouchers to the value of £20 upon completion of the test.

Tick all that apply.

- [ ] True

15. Participant's Full Name

____

Towards the Evaluation of Soundscape Through Affective, Physiological and Psychoacoustic Evaluation

UNIVERSITY of York

Department of Electronic Engineering

Participant Information Sheet

Researcher & Data Custodian: Simon Durbridge (sd1498@york.ac.uk)

Supervisor: Prof. Damian Murphy (damian.murphy@york.ac.uk)

Experiment: 2A

This project is being performed by Simon Durbridge (sd1498@york.ac.uk), who is a postgraduate research student at the University of York's AudioLab. This research is being supervised by Prof. Damian Murphy (damian.murphy@york.ac.uk). This project is funded by a UKRI doctoral training grant under the Engineering and Physical Sciences Research Council (EPSRC).

Before agreeing to take part, please read this information sheet carefully and let us know if anything is unclear or you would like further information.

Experiment Overview

Soundscape is a field of research that intends to make sense of the relationship between people and environments, with a special interest in the acoustic aspects of the environment. Soundscape studies often involve having participants experience different environments and assess how they feel within the given environment. This is usually achieved by asking participants to listen, look around and give feedback on their experience of the environment through a survey.

The main purpose of this study is to evaluate how participants respond to different environments when experienced through virtual reality (VR) technology, including a head mounted display and spatial audio. This experiment will include a combination of survey fulfilment and physiological sensing.

Participant Exclusion Criteria

Due to the nature of this experiment, there are a range of exclusion criteria that you as the participant and the principal investigator must be aware of.

You should be excluded from the study if you identify with any of the following:

- Any hearing impairment
- Any heart condition or ailment
- Any know neurological condition or ailment (including epilepsy)
- Any embedded electronics such as a pacemaker
- Sensitivity to flashing lights or adverse reactions to wearing a head mounted display
- Sensitive or damaged (broken) skin around the face, chest, wrist or fingers
- Are younger than 18 years of age
- Are older than 60 years of age

You can choose not to disclose which exclusion criteria you meet to the principal investigator, but if you do meet any of these exclusion criteria please exclude yourself from the study.

On the day

Participation in this study will be scheduled via a calendar and email.

There are a few constraints within 24 hours prior to your scheduled appointment:

- Please don't consume caffeine 2 hours prior to your appointment
- Please don't consume alcohol within 24 hours prior to your appointment
- Please don't consume unusual amounts of water prior to your appointment
- If you consume off-the-shelf medication such as ibuprofen or paracetamol prior to your appointment, please inform the researcher.

Please arrive 10 minutes prior to your scheduled appointment. When you arrive at Genesis 6 please head up the stairs and sit in the waiting area. Feel free to use the toilet facilities prior to your appointment.

Test procedure

One you have been greeted by the researcher hosting the session, you will be guided to the listening room. At first you will be introduced to the wearable sensors and the VR headset. You will have an opportunity to ask questions about the study. Once all details have been confirmed, you will be asked to follow a guide for fitting the experiment apparatus, with assistance from the researcher where necessary. The sensors are small wearable devices including a chest strap based heart rate sensor that you are required to fit in privacy. If you have any further questions regarding the physiological sensors, please contact the researcher (sd1498@york.ac.uk).

Once everything has been confirmed to be working correctly you will be asked to perform a few training exercises that will help you familiarise yourself with the test interface. This will also provide you with an opportunity to adjust the comfort of the test equipment. You will also have an opportunity to take a break from the head set before beginning the test proper.

The test will be conducted in a standing position and will take no more than 1 hour to complete. Once the test has been completed you will receive £15 in Amazon vouchers.

Potential Risks

Participation in this study is low risk. There are however associated risks with VR including flashing images that can trigger seizures in people with epilepsy. This is controlled for by the exclusion criteria, but if you have any questions or concerns related to sensitivity, please discuss these with the researcher(s) prior to the experiment.

Inattention to your physical surroundings while wearing the head mounted display can also lead to harm or injury. There can also be a risk of slips, trips and falls if the environment is not managed appropriately. The experiment area will be managed to minimise this risk, but please be aware that you might have to exercise caution in movement during the experiment.

There can be a risk of fatigue while undertaking the experiment due to wearing the headset, sensors and by using the hand controllers. To safeguard against fatigue there will be opportunities to take breaks during the experiment.

Simulator sickness can occur to some when wearing a head mounted display for VR purposes. To guard against this there will be opportunities to take breaks during the experiment. The experiment situation will also be static i.e. while you will be free to look around, you will be restricted to one location within the simulation. If you begin to feel unsteady, unwell or faint you should take the VR headset off immediately and raise the attention of the researcher hosting the experiment. You are under no obligation to complete the test and may stop at any time.

The level of sounds within the experiment will be managed to maximise the safety of your hearing. To mitigate the risk of over-exposure to unsafe sound levels the comfort of sound levels will be checked during the training intervals of the experiment. Wearing headphones presents a risk of isolation from sounds in the environment outside the experiment. To mitigate this risk, the researcher managing the experiment will notify you in the case that your attention needs to be raised e.g. a fire alarm is triggered.

You are once again reminded that your participation in this experiment is voluntary and you may pause the experiment to take a rest or stop the experiment entirely at any time.

Voluntary Participation

Participation in the experiment is voluntary. You will be required to fulfil an informed-consent participation form. You will be free to withdraw from the experiment at any time, without having to provide a reason. To withdraw from the experiment please raise the attention of the researcher hosting the experiment. When you withdraw any data associated with you will be deleted as soon as possible.

On what basis will you process my data?

Under the General Data Protection Regulation (GDPR), the University has to identify a legal basis for processing personal data and, where appropriate, an additional condition for processing special category data.

For further information and definitions of personal and special category data, please go to:

- https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/key-definitions/
- https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/special-category- data/

  Personal data is defined as data from which someone could be identified. For example, in this study we will be collecting your name and email address, which are needed in order to schedule the session and provide you with your amazon vouchers.

  In line with our charter which states that we advance learning and knowledge by teaching and research, the University processes personal data for research purposes under Article 6 (1) (e) of the GDPR: *Processing is necessary for the performance of a task carried out in the public interest*

Special category data is personal data which the GDPR says is more sensitive, and so needs more protection. For example, details of any hearing impairments. In this study, we will not be collecting any special category data. However, we do require you to confirm that, to the best of your knowledge, you do not have any hearing impairments if you wish to take part. This is due to the nature of the research.

Special category data is processed under Article 9 (2) (j):

- In line with ethical expectations and in order to comply with common law duty of confidentiality, we will seek your consent to participate where appropriate. This consent will not, however, be our legal basis for processing your data under the GDPR.

How will you use my data?

Data will be processed for the purposes outlined in this notice. Anonymised data will be analysed in the generation of research outcomes including the development of heuristics algorithms and falsification of hypotheses. Named data (covered under GDPR) will be used to schedule, communicate and arrange relative to the research activities in this research program.

Will you share my data with 3rd parties?

Anonymised data may be reused by the research team or other third parties for secondary research purposes. Data covered under GDPR will not be shared with any third parties.

How will you keep my data secure?

The University will put in place appropriate technical and organisational measures to protect your personal data and/or special category data. For the purposes of this project we will store data using secure University services provided by Google and the University Filestore.

Information will be treated confidentially and shared on a need-to-know basis only. The University is committed to the principle of data protection by design and default and will collect the minimum amount of data necessary for the project. The data that cannot be anonymised will be deleted at the earliest opportunity.

Will you transfer my data internationally?

Data will be held within the European Economic Area in full compliance with data protection legislation.

Only anonymised data may be available internationally.

Processing is necessary for archiving purposes in the public interest, or scientific and historical research purposes or statistical purposes

Research activities will only be undertaken where ethical approval has been obtained, where there is a clear public interest and where appropriate safeguards have been put in place to protect data.

Will I be identified in any outputs?

Any experimental data generated by this study will be anonymised, and participants will not be identified in any outputs.

How long will you keep my data?

Data will be retained in line with the University's Records Retention Schedule. The current guidance suggests that data is stored for no longer than is necessary. As such the data will be stored securely and disposed of at the first opportunity such as:

- When the data controller (Simon Durbridge) leaves the university
- There is no further need to retain contact information

What rights do I have in relation to my data?

Under the GDPR, you have a general right of access to your data, a right to rectification, erasure, restriction, objection or portability. You also have a right to withdrawal. Please note, not all rights apply where data is processed purely for research purposes. For further information see, https://www.york.ac.uk/records-management/generaldataprotectionregulation/individualrights/.

If you have any questions about this participant information sheet or concerns about how your data is being processed, please contact the Electronic Engineering Ethics Chair (elec-ethics@york.ac.uk). If you are further dissatisfied, please contact the University's Acting Data Protection Officer at dataprotection@york.ac.uk

Questions or concerns

If you have any questions or concerns about this experiment or the research project in general, please contact Simon Durbridge (sd1498@york.ac.uk) Prof. Damian Murphy (damian.murphy@york.ac.uk).

Right to complain

If you are unhappy with the way in which the University has handled your personal data, you have a right to complain to the Information Commissioner's Office. For information on reporting a concern to the Information Commissioner's Office, see www.ico.org.uk/concerns.

THE UNIVERSITY of York

Application Form for Physical Sciences Ethics Committee Approval

Advice for applicants on completing the form *Please ensure that the information provided is:* ● *Accurate and concise* ● *Clear and simple and easily understood by a lay person* ● *Free of jargon, technical terms and abbreviations* *Further advice and information can be obtained from your departmental representative on the PSEC and at: http://www.york.ac.uk/admin/aso/ethics/cttee.htm*

Please return completed (typed) form to your departmental representative via email to:

elec-ethics@york.ac.uk

Title of project: Towards the Evaluation of Soundscape Through Affective, Physiological and Psychoacoustic Measures 2B

SECTION 1 DETAILS OF APPLICANTS

Details of principal investigator (name, appointment and qualifications)

Simon Durbridge PhD Student Audio Lab, Department of Electronics BSc Sound, Light & Live Event Technology, MSc Audio Engineering

Names, appointments and qualifications of additional investigators *(student applicants should include their project supervisor(s) here)*

Prof. Damian Murphy (Primary Supervisor) Professor Audio Lab, Department of Electronics BSc (Hons), MSc, DPhil

Location(s) of project

Audio Lab
Genesis 6
Innovation Way
Heslington
York

SECTION 2 FUNDERS

What is the funding source(s) for the project?

EPSRC – Doctoral Training Award

Please answer the following:

(i) Does the express and direct aim of the research or other activity raise ethical issues?

YES	x		NO	

(ii) Is there any obvious or inevitable adaptation of research findings to ethically questionable aims?

YES			NO	x

(iii) Is the work being funded by organisations tainted by ethically questionable activities?

YES			NO	x

(iv) Are there any restrictions on academic freedoms – notably, to adapt and withdraw from ongoing research, and to publish findings?

YES			NO	x

If you answered **Yes** to any of the above, please give details below:

Participants will be asked to evaluate virtual representations of environments which are presented via 360 video. Participants will be surveyed on their subjective estimates and responses to the environments. The primary ethical issues raised in this experiment are the privacy and safety of the participants.

SECTION 3 DETAILS OF PROJECT OR OTHER ACTIVITY

Aims (100 words max)

The aim of the project is to identify any relationships between the perception of soundscapes and a person's emotions and physiological behaviour. This includes the use of the virtual representation of environments and the fulfilment of surveys related to experiencing those soundscapes. The primary aim of this experiment is to evaluate relationships between the aural features of the soundscapes and the subjective estimates of participants. The secondary aim of this experiment is to perform a parallel analysis of responses to soundscapes as in experiment, 2A in which participants will evaluate soundscapes in virtual reality. We will compare responses across experiment 2A and this experiment in order to identify if the improvements in ecological validity facilitated by virtual reality elicit different responses.

Background (250 words max)

Soundscape studies attempt to identify the relationships between how people experience environments and the qualities of those environments, specifically from the perspective of the audible aspects of the environments. A key focus of soundscape research is to identify causal relationships between features of the soundscape. A challenge in this regard is to develop appropriate methodologies for generating and analysing data from which these relationships can be drawn.

The ground truth of the quality of the soundscape is a consequence of the perception of the person experiencing the soundscape, defining the necessity for listening tests that attempt to measure and interpret responses to soundscape stimuli. This presents significant issues related to experiment validity, repeatability and generalisability. Soundscape studies often employ surveys that include a mechanism for measuring affect, or the underlying experience of feeling, emotion, or mood. However, there is not always a clear distinction between cultural influences and acoustics influences in these responses.

Brief outline of project/activity (250 words max)

The primary aim of this experiment is to identify if the presentation of soundscapes through 360 video elicits different subjective responses to experiencing the same soundscapes with static images and binaural rendering. This necessitates a repeated measures design with two groups of participants.
The secondary aim of this experiment is to identify if there are statistically significant relationships between participants' subjective estimates of the soundscape and the acoustical properties of the soundscape.
Participants will perform an online listening test hosted on the qualtrics platform. In the experiments participants will review a randomly ordered set of soundscape, presented as videos hosted on the Youtube platform. Some of these videos will be static and some will be 360. Participants will be surveyed for subjective and objective estimates of the soundscape, including dimensions of affect, immersion, soundscape classification and dimensions of pleasantness and vibrancy. The data from the experiment will be analysed using a generalised mixed effects linear model to compare the trends in these estimates against objective parameters associated with the experiment.

Study design *(if relevant – e.g. randomised control trial; laboratory-based)*

The study is a web-based online repeated measures listening test

If the study involves participants, how many will be recruited?

At least 60

If applicable, what is the statistical power of the study, i.e. what is the justification for the number of participants needed?

A pilot study found a greatest effect sizes of in the analysis of interest of 0.94 etta-squared N=15, but to pick up effects with smaller magnitudes reliably we would have at least 30 participants per condition.

SECTION 4 RECRUITMENT OF PARTICIPANTS

How will the participants be recruited?

Participants will be recruited by email, word of mouth and social media.

What are the inclusion/exclusion criteria?

Significant or diagnosed hearing impairment

Will participants be paid reimbursement of expenses?

YES		NO	X

Will participants be paid?

YES		NO	X

If yes, please obtain signed agreement

Will any of the participants be students?

YES	X	NO	

SECTION 5 DATA STORAGE AND TRANSMISSION

If the research will involve storing personal data, including sensitive data, on any of the following please indicate so and provide further details (answers only required if *personal* data is to be stored).

Manual files	**No**
University computers	**No**
Home or other personal computers	**Yes**
Laptop computers, tablets	**No**
Website	**Yes**

Please explain the measures in place to ensure data confidentiality, including whether encryption or other methods of anonymisation will be used.

Participant identities (name) and contact information (email) will be collected as part of the informed consent procedure. Participants's identities will be kept secure as part of the qualtrics storage until the data collection period has ended. Once data collection has ended and the data from the experiment is extracted, this extracted data will be anonymised and stored on the University secure cloud storage (GDrive) provided by Google.

Please detail who will have access to the data generated by the study.

Simon Durbridge, Damian Murphy, Public

Please detail who will have control of and act as custodian for, data generated by the study.

Simon Durbridge

Please explain where, and by whom, data will be analysed.

The data will be analysed by Simon Durbridge, at the Audio Lab on the University of York Campus. The data may also be published as part of the publication procedure, subject to informed consent from the participants. As such anyone with access to the data will be able to analyse it.

Please give details of data storage arrangements, including where data will be stored, how long for, and in what form.

Data covered under GDPR will be stored on the qualtrics platform until data collection is complete. When the data is extracted from qualtrics it will be anonymised and stored as comma-separated-value(CSV) files on the university private encrypted storage (GDrive). The anonymised data may be temporarily stored on the personal computer of the data custodian for the purposes of data analysis. The data will be stored for 4 years, until the data custodian loses access to the university storage or when the data is no longer needed, whichever comes first. If published, the anonymised experiment data will be published on Zenodo or similar data storage and sharing platform.

SECTION 6 CONSENT

YES	X	NO	

Is written consent to be obtained?

If yes, please attach a copy of the information for participants

If no, please justify

Will any of the participants be from one of the following vulnerable groups?

Children under 18	YES		NO	X
People with learning difficulties	YES		NO	X
People who are unconscious or severely ill	YES		NO	X
People with mental illness	YES		NO	X
NHS patients	YES		NO	X
Other vulnerable groups (if 'yes', please give details)	YES		NO	X

If so, what special arrangements have been made for getting consent?

N/A

SECTION 7 DETAILS OF INTERVENTIONS

Indicate whether the study involves procedures which:

Involve taking bodily samples	YES		NO	X
Are physically invasive	YES		NO	X
Are designed to be challenging/disturbing (physically or psychologically)	YES		NO	X

If so, please list those procedures to which participants will be exposed:
- N/A

List any potential hazards:
- N/A

List any discomfort or distress:
- N/A

What steps will be taken to safeguard

(i) the confidentiality of information

The data generated in the experiment will eventually be stored in an anonymised fashion. Data will be stored on the Qualtrics platform, and then on the University private cloud storage as anonymised data.

(ii) the specimens themselves?

Participation in the study is voluntary and participants can choose to withdraw at any point without giving a reason. The study equipment belongs to the participants and it is assumed the participants are familiar with how to use the equipment in a safe manner. Participants are reminded in the information sheet to monitor their own wellbeing through the experiment, self-managing appropriate sound levels.

What particular ethical problems or considerations are raised by the proposed study?

Experiment data will be published alongside publication of the results of the study. To protect the identities of the participants all data that could be used to identify a participant will be anonymised. Participants will be required to give informed consent for participation, data storage and publication of anonymised data before participating in the experiment. An information sheet will be provided to participants, clearly informing participants of the experiment procedure and how the data will be used and shared.

What do you anticipate will be the output from the study? ***Tick those that apply:***

Peer-reviewed publications	X
Non-peer-reviewed publications	X
Reports for sponsor	X
Confidential reports	X
Presentation at meetings	X
Press releases	
Student project	X

YES		NO	X

Is there a secrecy clause to the research?
No

<u>SECTION 8 SIGNATURES</u>

The information in this form is accurate to best of my knowledge and belief and I take full responsibility for it.

I agree to advise of any adverse or unexpected events that may occur during this project, to seek approval for any significant protocol amendments and to provide interim and final reports. I also agree to advise the Ethics Committee if the study is withdrawn or not completed.

Signature of Investigator(s): ....

....

Date: ...27/9/2021..

> ***Responsibilities of the Principal Researcher following approval***
> - If changes to procedures are proposed, please notify the Ethics Committee
> - Report promptly any adverse events involving risk to participants

Participant Consent Form

University of York - Department of Electronic Engineering

Project Title: Towards the Evaluation of Soundscape Through Affective, Physiological and Psychoacoustic Evaluation

This experiment will be used to explore in influence of experiencing soundscapes in VR on peoples subjective evaluation of those soundscapes.

Please read the following questions and statements carefully. You are required to answer all sections and confirm that all of the statements are true, by ticking the checkboxes, before submitting the form.

Thank you for your time and participation.

1. Email *

    ____

2. I have read and understood the participant information sheet about this project.

    Tick all that apply.

    ☐ True

3. I agree to take part in this experiment.

    Tick all that apply.

    ☐ True

4. I have been given the opportunity to ask any questions and had them answered to my satisfaction. I also understand that I will be free to ask further questions about the study after completing this con

    Tick all that apply.

    ☐ True

5. I understand that my participation is voluntary and that I am free to withdraw such participation at any time without giving a reason.

   Tick all that apply.

   - [ ] True

6. I understand that my participation in this project will be treated anonymously and I understand that my data will be stored securely.

   Tick all that apply.

   - [ ] True

7. I understand that data generated in this study may be published or made publicly available and I consent to the publication of the anonymised data.

   Tick all that apply.

   - [ ] True

8. I confirm that, to the best of my knowledge, I have no hearing impairments.

   Tick all that apply.

   - [ ] True

9. I am a trained/experienced listener

   Tick all that apply.

   - [ ] True
   - [ ] False

10. I am over the age of 18

    Tick all that apply.

    - [ ] True

11. Participant's Full Name

Towards the Evaluation of Soundscape Through Affective, Physiological and Psychoacoustic Evaluation

UNIVERSITY *of York*

Department of Electronic Engineering

Participant Information Sheet

Researcher & Data Custodian: Simon Durbridge (sd1498@york.ac.uk)

Supervisor: Prof. Damian Murphy (damian.murphy@york.ac.uk)

Experiment: 2B

This project is being performed by Simon Durbridge (sd1498@york.ac.uk), who is a postgraduate research student at the University of York's AudioLab. This research is being supervised by Prof. Damian Murphy (damian.murphy@york.ac.uk). This project is funded by a UKRI doctoral training grant under the Engineering and Physical Sciences Research Council (EPSRC).

Before agreeing to take part, please read this information sheet carefully and let us know if anything is unclear or if you would like further information.

Experiment Overview

Soundscape is a field of research that intends to make sense of the relationship between people and environments, with a special interest in the acoustic aspects of the environment. Soundscape studies often involve having participants experience different environments and assess how they feel within the given environment. This is usually achieved by asking participants to listen to and give feedback on their experience of the environment through a survey.

The main purpose of this study is to evaluate how participants respond to different environments when experienced through virtual reality (VR) technologies such as spatial audio. This experiment will include a combination of listening, 360 images and survey fulfilment. This experiment can be performed on your personal computer.

Participant Exclusion Criteria

The exclusion criteria of this study is a requirement to have no a-typical hearing loss or diagnosed hearing impairment. You are welcome to participate in this study if you consider yourself normal of hearing.

Items you will need

- A computer
  - Desktop or Laptop
  - Windows, Mac or Linux operating system
- A set of headphones that you can plug into the computer
  - Please make sure the headphones work correctly and that sound from the computer comes out of the headphones
- A web browser that is supported by Youtube 360
  - Google Chrome
  - Firefox
  - Microsoft Edge
  - Opera

How to participate

The study will be undertaken through your web browser using a combination of the qualtrics platform and Youtube. As you step through the experiment, you will be presented with soundscapes and 360 degree youtube videos. 360 videos allow you to use your mouse to 'look around' the scene in the video. You will be asked to experience these soundscapes and fill out a short questionnaire about each video.

Once you have completed reading this participant information sheet, you will be asked to read and fill out a form to declare your informed consent for participation. Once you have completed this form and have agreed to participate, you will be asked for some general demographic information.

Following this, you will be asked to complete a short sound calibration process. This will help the researchers get an idea of how the test is working in the context of your listening system (computer, headphones, ears etc).

Once this step is completed you will be taken to the training part of the experiment. Each interval in the experiment has 3 steps:

- Experience: Listen to the soundscape and view the image
- Respond: Fill out a very short questionnaire
- Rest: Wait a moment before you move onto the next interval

You are free to move through the experiment at your own pace. Please take your time to listen to the whole soundscape at least once before filling out the questionnaire. The soundscape may be presented with a 360 representation of the environment; you are encouraged to use the video viewer and your mouse to pan the view around the scene.

Once you have finished the experience step please fill out the questionnaire. The questionnaire will ask about the perceived (sound?) quality of the environment and how you feel about the environment. You are welcome to take one minute of rest once you have filled in the survey. When you are ready, move on to the next step in the experiment.

There will be 2 training soundscapes in which to get used to the format of a test interval. During the test interval, please make sure that your listening levels are comfortable. Once the training intervals are complete you will be directed to continue on to the test. The test should take no longer than 30 minutes to complete.

Potential Risks

Participation in this study is low risk. If you have any questions or concerns related to performing the experiment, please contact Simon Durbridge (sd1498@york.ac.uk).

It is necessary to control the level of sounds within the experiment. To mitigate the risk of over-exposure to unsafe sound levels the comfort of sound levels must be checked during the experiment. Please make sure that you adjust the levels of the sound to be clear, comfortable and not over-loud. If you think the levels are too high please adjust them accordingly during the training examples. Please avoid changing the levels during the test portion of the experiment, and only change them if you experience some discomfort. If you do this, please make a note as you will have the opportunity to provide feedback at the end of the study.

Wearing headphones presents a risk of isolation from sounds in the environment outside the experiment. Please exercise caution when performing the experiment. Please make sure you perform the experiment in a safe and quiet environment.

The experiment should not take more than 30 minutes to complete. However, if you start to feel fatigue you may pause the experiment to take a rest or stop the experiment entirely at any time.

Voluntary Participation

Participation in the experiment is voluntary. You will be required to fulfil an informed-consent participation form. You will be free to withdraw from the experiment at any time, without having to provide a reason. To withdraw from the experiment simply close the session in your browser. Incomplete sessions will be deleted when the data gathering period ends.

On what basis will you process my data?

Under the General Data Protection Regulation (GDPR), the University has to identify a legal basis for processing personal data and, where appropriate, an additional condition for processing special category data.

For further information and definitions of personal and special category data, please go to:

- https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/key-definitions/
- https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/special-category- data/

Personal data is defined as data from which someone could be identified. For example, in this study we will be collecting your name and email address, which are needed in order to notify you about the completion of the experiment and publication of the data and results, and for further correspondence and queries related to the research project.

In line with our charter which states that we advance learning and knowledge by teaching and research, the University processes personal data for research purposes under Article 6 (1) (e) of the GDPR: *Processing is necessary for the performance of a task carried out in the public interest*

Special category data is personal data which the GDPR says is more sensitive, and so needs more protection. For example, details of any hearing impairments. In this study, we will not be collecting any special category data. However, we do require you to confirm that, to the best of your knowledge, you do not have any hearing impairments if you wish to take part. This is due to the nature of the research.

Special category data is processed under Article 9 (2) (j):

- In line with ethical expectations and in order to comply with common law duty of confidentiality, we will seek your consent to participate where appropriate. This consent will not, however, be our legal basis for processing your data under the GDPR.

How will you use my data?

Data will be processed for the purposes outlined in this notice. Anonymised data will be analysed in the generation of research outcomes including the development of heuristics algorithms and falsification of hypotheses. Named data (covered under GDPR) will be used to communicate in relation to the research activities in this research program.

Will you share my data with 3rd parties?

Anonymised data may be reused by the research team or other third parties for secondary research purposes. Data covered under GDPR such as your identity or email address will not be shared with any third parties.

How will you keep my data secure?

The University will put in place appropriate technical and organisational measures to protect your personal data and/or special category data. For the purposes of this project we will store data using secure University services provided by Qualtrics, Google and the University Filestore.

Information will be treated confidentially and shared on a need-to-know basis only. The University is committed to the principle of data protection, and will collect the minimum amount of data necessary for the project.

Will you transfer my data internationally?

Data will be held within the European Economic Area in full compliance with data protection legislation.

Only anonymised data may be available internationally.

Processing is necessary for archiving purposes in the public interest, or scientific and historical research purposes or statistical purposes

Research activities will only be undertaken where ethical approval has been obtained, where there is a clear public interest and where appropriate safeguards have been put in place to protect data.

Will I be identified in any outputs?

Any experimental data generated by this study will be anonymised, and participants will not be identified in any outputs.

How long will you keep my data?

Data will be retained in line with the University's Records Retention Schedule. The current guidance suggests that data is stored for no longer than is necessary. As such the data will be stored securely and disposed of at the first opportunity such as:

- When the data controller (Simon Durbridge) leaves the university
- There is no further need to retain contact information

What rights do I have in relation to my data?

Under the GDPR, you have a general right of access to your data, a right to rectification, erasure, restriction, objection or portability. You also have a right to withdrawal. Please note, not all rights apply where data is processed purely for research purposes. For further information see,

https://www.york.ac.uk/records-management/generaldataprotectionregulation/individualsrights/.

If you have any questions about this participant information sheet or concerns about how your data is being processed, please contact the Electronic Engineering Ethics Chair (elec-ethics@york.ac.uk). If you are further dissatisfied, please contact the University's Acting Data Protection Officer at dataprotection@york.ac.uk

Questions or concerns

If you have any questions or concerns about this experiment or the research project in general, please contact Simon Durbridge (sd1498@york.ac.uk) Prof. Damian Murphy (damian.murphy@york.ac.uk).

Right to complain

If you are unhappy with the way in which the University has handled your personal data, you have a right to complain to the Information Commissioner's Office. For information on reporting a concern to the Information Commissioner's Office, see www.ico.org.uk/concerns.

Appendix B

Experiment 3 Stimuli

Sample	Locked Audio	Youtube	wav	mp4	still
1	0	https://youtu.be/q5zGX8wAwic	Park.3.wav	360_0128_Stitch_XHC.MP4	Park46.mp4
1	1	https://youtu.be/cAcVGL8moCc	Park.3.wav	360_0128_Stitch_XHC.MP4	Park46.mp4
2	0	https://youtu.be/hqPHtE4CfMM	BusyStreet.8.wav	360_0134_Stitch_XHC.MP4	BusyStreet52.mp4
2	1	https://youtu.be/hqPHtE4CfMM	BusyStreet.8.wav	360_0134_Stitch_XHC.MP4	BusyStreet52.mp4
3	0	https://youtu.be/eMIdEAsc6cw	PedestrianZone.6.wav	360_0112_Stitch_XHC.MP4	PedestrianZone31.mp
3	1	https://youtu.be/CtXxklHoPEY	PedestrianZone.6.wav	360_0112_Stitch_XHC.MP4	PedestrianZone31.mp
4	0	https://youtu.be/3fDCMYm02Fs	PedestrianZone.8.wav	360_0115_Stitch_XHC.MP4	PedestrianZone33.mp
4	1	https://youtu.be/cEnyAdJ0U1Q	PedestrianZone.8.wav	360_0115_Stitch_XHC.MP4	PedestrianZone33.mp
5	0	https://youtu.be/UoNXGS3CXPo	Park.1.wav	360_0109_Stitch_XHC.MP4	Park29.mp4
5	1	https://youtu.be/QGPIKPTbaIE	Park.1.wav	360_0109_Stitch_XHC.MP4	Park29.mp4
6	0	https://youtu.be/6MdaL8ChYqY	BusyStreet.5.wav	360_0120_Stitch_XHC.MP4	BustStreet38.mp4
6	1	https://youtu.be/ZuO3gEM2OWY	BusyStreet.5.wav	360_0120_Stitch_XHC.MP4	BustStreet38.mp4
7	0	https://youtu.be/F3NKQNVBaew	Woodland.1.wav	360_0086_Stitch_XHC.MP4	Woodland10.mp4
7	1	https://youtu.be/PgRyGrnM33A	Woodland.1.wav	360_0086_Stitch_XHC.MP4	Woodland10.mp4
8	0	https://youtu.be/Ak90Vl6PZBQ	Woodland.8.wav	360_0110_Stitch_XHC.MP4	Woodland30.mp4
8	1	https://youtu.be/UtbkqpVJuBE	Woodland.8.wav	360_0110_Stitch_XHC.MP4	Woodland30.mp4
Sample	**class**	**location**	**instance**	**time**	**date**
1	Park	Yorkshire Museum Gardens	3	15:52:00	18/05/2017
1	Park	Yorkshire Museum Gardens	3	15:52:00	18/05/2017
2	BusyStreet	St. Leonard's Place, York	8	11:44:00	23/05/2017
2	BusyStreet	St. Leonard's Place, York	8	11:44:00	23/05/2017
3	PedestrianZone	St. Helen's Square, York	6	15:28:00	11/05/2017
3	PedestrianZone	St. Helen's Square, York	6	15:28:00	11/05/2017
4	PedestrianZone	Stonegate, York	8	16:40:00	11/05/2017
4	PedestrianZone	Stonegate, York	8	16:40:00	11/05/2017
5	Park	Rowntree Park, York	1	14:00:00	11/05/2017
5	Park	Rowntree Park, York	1	14:00:00	11/05/2017
6	BusyStreet	Micklegate Bar, York	5	14:03:00	16/05/2017
6	BusyStreet	Micklegate Bar, York	5	14:03:00	16/05/2017
7	Woodland	Knavesmire Wood, York	1	11:41:00	08/05/2017
7	Woodland	Knavesmire Wood, York	1	11:41:00	08/05/2017
8	Woodland	Rowntree Park Woods	8	14:22:00	11/05/2017
8	Woodland	Rowntree Park Woods	8	14:22:00	11/05/2017

Figure B.1: Stimuli arrangement for experiment 3.

Appendix C

Accompanying Materials

Table C.1 gives a structured overview of the materials available with this thesis.

Folder	File	Type
Experiment_1_stimuli_test	BusyStreet.6.4	Experiment 1 Stimuli
Experiment_1_stimuli_test	QuietStreet.4.1	Experiment 1 Stimuli
Experiment_1_stimuli_test	QuietStreet.4.3	Experiment 1 Stimuli
Experiment_1_stimuli_test	QuietStreet.4.5	Experiment 1 Stimuli
Experiment_1_stimuli_test	Woodland.2.6	Experiment 1 Stimuli
Experiment_1_stimuli_test	Woodland.4.6	Experiment 1 Stimuli
Experiment_1_stimuli_test	Woodland.5.3	Experiment 1 Stimuli
Experiment_1_stimuli_test	Woodland.5.5	Experiment 1 Stimuli
Experiment_1_stimuli_train	Beach	Experiment 1 Stimuli
Experiment_1_stimuli_train	Shopping_Centre	Experiment 1 Stimuli
VR_Soundscape_experiment	VR_Soundscape_experiment.zip	Unity project for Soundscape Investigati
VR_Soundscape_experiment	AudioManager.cs	C# Utility Script for Unity
VR_Soundscape_experiment	ExperimentManager.cs	C# Utility Script for Unity
VR_Soundscape_experiment	SceneLoader.cs	C# Utility Script for Unity
VR_Soundscape_experiment	SurveyButtonInfo.cs	C# Utility Script for Unity
VR_Soundscape_experiment	ConsoleGrabbable.cs	C# Utility Script for Unity
VR_Soundscape_experiment	SkyboxManager.cs	C# Utility Script for Unity
VR_Soundscape_experiment	SurveyManager.cs	C# Utility Script for Unity
VR_Soundscape_experiment	consolePositionHandle.cs	C# Utility Script for Unity
VR_Soundscape_experiment	LoadSkybox.cs	C# Utility Script for Unity
VR_Soundscape_experiment	SoundscapeManagement.cs	C# Utility Script for Unity
VR_Soundscape_experiment	SurveyUI.cs	C# Utility Script for Unity
VR_Soundscape_experiment	DataStorageManager.cs	C# Utility Script for Unity
VR_Soundscape_experiment	playbackScript.cs	C# Utility Script for Unity
VR_Soundscape_experiment	State.cs	C# Utility Script for Unity
VR_Soundscape_experiment	VRCameraUtilities.cs	C# Utility Script for Unity
VR_Soundscape_experiment	EventManager.cs	C# Utility Script for Unity
VR_Soundscape_experiment	playerControlScript.cs	C# Utility Script for Unity
VR_Soundscape_experiment	StateImplementations.cs	C# Utility Script for Unity
VR_Soundscape_experiment	VRUIConsole.cs	C# Utility Script for Unity
VR_Soundscape_experiment	ExperimentDataPoint.cs	C# Utility Script for Unity
VR_Soundscape_experiment	PopulateGrid.cs	C# Utility Script for Unity
VR_Soundscape_experiment	StateMachine.cs	C# Utility Script for Unity

Table C.1: Table summarising the accompanying materials available with this thesis.

Bibliography

[1] H. L. F. Helmholtz, *On the Sensations of Tone as a Physiological Basis for the Theory of Music* (Cambridge Library Collection - Music), 4th ed., A. J. Ellis, Ed. Cambridge University Press, 2009. DOI: 10.1017/CBO9780511701801.

[2] SO/TC 43/SC 1/WG 54, "Acoustics. soundscape: Definition and conceptual framework," British Standards Institute, Standard, Sep. 2014.

[3] R. M. Schafer, *The Tuning of the World: Toward a Theory of Soundscape Design*, 1st. New York: Knopf, 1977.

[4] E. Peris, N. Blanes, J. Fons, M. Sainz de la Maza, and M. Jose Ramos, "Environmental noise in Europe - 2020," European Environment Agency, Copenhagen, Denmark, Tech. Rep. 22/2019, 2020, p. 104. DOI: 10.2800/686249.

[5] A. L. Brown and I. Van Kamp, "Who environmental noise guidelines for the european region: A systematic review of transport noise interventions and their impacts on health," *International Journal of Environmental Research and Public Health*, vol. 14, no. 8, 2017. DOI: 10.3390/ijerph14080873. [Online]. Available: %5Curl%7Bhttps://www.mdpi.com/1660-4601/14/8/873%7D.

[6] M. Lionello, F. Aletta, and J. Kang, "A systematic review of prediction models for the experience of urban soundscapes," *Applied Acoustics*, vol. 170, p. 107 479, 2020. DOI: https://doi.org/10.1016/j.apacoust.2020.107479. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0003682X20305831.

[7] Ö. Axelsson, "Soundscape revisited," *Journal of Urban Design*, vol. 25, no. 5, pp. 551–555, 2020. DOI: 10.1080/13574809.2020.1810006. [Online]. Available: %5Curl%7Bhttps://doi.org/10.1080/13574809.2020.1810006%7D.

[8] H. Ma and S. Shu, "An Experimental Study: The Restorative Effect of Soundscape Elements in a Simulated Open-Plan Office," *Acta Acustica united with Acustica*, vol. 104, pp. 106–115, 2018. DOI: 10.3813/AAA.919150.

[9] Durbridge, Simon and Murphy, Damian Thomas, "Assessment of soundscapes using self-report and physiological measures," *Acta Acust.*, vol. 7, p. 6, 2023. DOI: 10.1051/aacus/2022059. [Online]. Available: https://doi.org/10.1051/aacus/2022059.

[10] S. Durbridge and D. Murphy, "Soundscape Evaluation In The Virtual Reality: Tools for the creation of soundscape studies," in *Audio Engineering Society Conference: AES 2022 International Audio for Virtual and Augmented Reality Conference*, Aug. 2022. [Online]. Available: `%5Curl%7Bhttp://www.aes.org/e-lib/browse.cfm?elib=21843%7D`.

[11] S. Durbridge, "Efficient Acoustic Modelling of Large Spaces using Time Domain Methods," no. 5, pp. 1–8, 2017.

[12] D. Davis and C. Davis, *Sound System Engineering*, 2nd. Newton, MA: Focal Press, 1997. DOI: `10.4324/9780240818474`.

[13] J. Zhang, X. Xiao, D. Wang, Y. Yang, and J. Fan, "Source contribution analysis for exterior noise of a high-speed train: Experiments and simulations," *Shock and Vibration*, vol. 2018, 2018. DOI: `10.1155/2018/5319460`.

[14] *Condenser microphones*, Online; accessed 14. May 2023, May 2023.

[15] *Microphone polar patterns*, Online; accessed 14. May 2023, Apr. 2023.

[16] McGovern Medical School, *Ear anatomy schematics*, en, [Online; Accessed on 2023-08-30]. [Online]. Available: `%5Curl%7Bhttps://med.uth.edu/orl/online-ear-disease-photo-book/chapter-3-ear-anatomy/ear-anatomy-schematics/%7D`.

[17] F. A. Everest, *Master handbook of acoustics.* eng, Sixth edition. McGraw-Hill Education, 2015.

[18] "Tonotopic organization of human auditory cortex," *NeuroImage*, vol. 50, pp. 1202–1211, 3 Apr. 2010. DOI: `10.1016/j.neuroimage.2010.01.046`.

[19] J. G. Betts *et al.* OpenStax, 2022. [Online]. Available: `%5Curl%7Bhttps://openstax.org/books/anatomy-and-physiology-2e/pages/1-introduction%7D`.

[20] J. McGrath, P. Roy, and B. J. Perrin, "Stereocilia morphogenesis and maintenance through regulation of actin stability," *Seminars in Cell and Developmental Biology*, vol. 65, pp. 88–95, May 2017. DOI: `10.1016/j.semcdb.2016.08.017`.

[21] G. V. Carbajal and M. S. Malmierca, "The unique role of the non-lemniscal pathway on stimulus-specific adaptation (SSA) in the auditory system," *Proceedings of the International Symposium on Auditory and Audiological Research*, vol. 6, no. 2, pp. 95–106, 2018. [Online]. Available: `https://proceedings.isaar.eu/index.php/isaarproc/article/view/2017-12`.

[22] W. A. Yost, "Psychoacoustics: A brief historical overview," *Acoustics Today*, vol. 11, no. 3, pp. 46–53, 2015.

[23] M. S. Engel, A. Fiebig, C. Pfaffenbach, and J. Fels, "A review of the use of psychoacoustic indicators on soundscape studies," *Current Pollution Reports*, vol. 7, pp. 359–378, 2021. DOI: `10.1007/s40726-021-00197-1/Published`. [Online]. Available: `https://doi.org/10.1007/s40726-021-00197-1`.

[24] *ISO 226 Equal-Loudness-Level Contour Signal - File Exchange - MATLAB CentralFile Exchange - MATLAB Central*, [Online; accessed 15. May 2023], Mar. 2005. [Online]. Available: %5Curl%7Bhttps://uk.mathworks.com/matlabcentral/fileexchange/7028-iso-226-equal-loudness-level-contour-signal%7D.

[25] H. Fletcher and W. A. Munson, "Loudness, its definition, measurement and calculation," *Bell System Technical Journal*, vol. 12, no. 4, pp. 377–430, 1933.

[26] D. W. Robinson and R. S. Dadson, "A re-determination of the equal-loudness relations for pure tones," *British Journal of Applied Physics*, vol. 7, no. 5, p. 166, 1956.

[27] Y. Suzuki and H. Takeshima, "The new international organization for standardization (iso) standard for the equal-loudness contours: Comparison to earlier contours and procedures," *The Journal of the Acoustical Society of America*, vol. 119, no. 5, pp. 3292–3292, May 2006. DOI: 10.1121/1.4786219. [Online]. Available: %5Curl%7Bhttps://doi.org/10.1121/1.4786219%7D.

[28] H. Fastl and E. Zwicker, *Psychoacoustics: Facts and Models*, 2nd. Munich, 1999.

[29] B. R. Glasberg and B. C. J. Moore, "A Model of Loudness Applicable to Time-Varying Sounds," *Journal of Audio Engineering Society*, vol. 50, no. 5, pp. 331–342, 2002.

[30] W. Aures, "Procedure for calculating the sensory euphony of arbitrary sound signals," *Acustica*, vol. 59, no. 2, pp. 130–141, 1985.

[31] S. H. Swift and K. L. Gee, "Implementing sharpness using specific loudness calculated from the "procedure for the computation of loudness of steady sounds" engineering acoustics: Paper 5pnsa9," vol. 30, Acoustical Society of America, Jun. 2017. DOI: 10.1121/2.0000542.

[32] R. Sottek and K. Genuit, "Perception of roughness of time-variant sounds," *The Journal of the Acoustical Society of America*, vol. 133, p. 3598, May 2013. DOI: 10.1121/1.4806661.

[33] 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.

[34] G. Kearney, "Auditory scene synthesis using virtual acoustic recording and reproduction," Ph.D. dissertation, 2010.

[35] T. R. Letowski and S. T. Letowski, "Auditory spatial perception: Auditory localization," ARMY RESEARCH LAB ABERDEEN PROVING GROUND MD HUMAN RESEARCH and ENGINEERING ..., Tech. Rep., 2012.

[36] B. Moore, *An Introduction to the Psychology of Hearing: Sixth Edition*. Leiden, The Netherlands: Brill, 2013. [Online]. Available: https://brill.com/view/title/24210.

[37] W. C. Pirkle, *Designing audio effect plugins in C++: for AAX, AU, and VST3 with DSP theory*. Routledge, 2019.

[38] M. Zaunschirm, C. Schörkhuber, and R. Höldrich, “Binaural rendering of Ambisonic signals by head-related impulse response time alignment and a diffuseness constraint,” *The Journal of the Acoustical Society of America*, vol. 143, no. 6, pp. 3616–3627, 2018. DOI: 10.1121/1.5040489.

[39] C. Armstrong, L. Thresh, D. Murphy, and G. Kearney, “A perceptual evaluation of individual and non-individual hrtfs: A case study of the sadie ii database,” *Applied Sciences*, vol. 8, no. 11, 2018. DOI: 10.3390/app8112029. [Online]. Available: %5Curl%7Bhttps://www.mdpi.com/2076-3417/8/11/2029%7D.

[40] R. Y. Litovsky, H. S. Colburn, W. A. Yost, and S. J. Guzman, “The precedence effect,” *The Journal of the Acoustical Society of America*, vol. 106, no. 4, pp. 1633–1654, Oct. 1999. DOI: 10.1121/1.427914. eprint: https://pubs.aip.org/asa/jasa/article-pdf/106/4/1633/7519853/1633_1_online.pdf. [Online]. Available: https://doi.org/10.1121/1.427914.

[41] L. Gab, “15 Types of Headphones: All You Need to Know for Picking the Best Pair,” *Headphonesty*, May 2023. [Online]. Available: %5Curl%7Bhttps://www.headphonesty.com/2021/01/types-of-headphones%7D.

[42] B. N. Walker, R. M. Stanley, N. Iyer, B. D. Simpson, and D. S. Brungart, “Evaluation of bone-conduction headsets for use in multitalker communication environments,” in *Proceedings of the human factors and ergonomics society annual meeting*, SAGE Publications Sage CA: Los Angeles, CA, vol. 49, 2005, pp. 1615–1619.

[43] R. Burns, “Blumlein and the birth of stereo [biography],” *IEE Review*, vol. 45, no. 6, pp. 269–273, 1999. DOI: 10.1049/ir:19990607.

[44] A. D. Blumlein, “British patent specification 394,325 (improvements in and relating to sound-transmission, sound-recording and sound-reproducing systems),” *Journal of the Audio Engineering Society*, vol. 6, no. 2, pp. 91–130, 1958.

[45] h. a. m. clark h. a. m., g. f. dutton g. f., and p. b. vanderlyn p. b., “The ’stereosonic’ recording and reproducing system: A two-channel systems for domestic tape records,” *journal of the audio engineering society*, vol. 6, no. 2, pp. 102–117, Apr. 1958.

[46] Contributors to Wikimedia projects, *Blumlein pair - Wikipedia*, [Online; accessed 14. May 2023], Mar. 2023. [Online]. Available: %5Curl%7Bhttps://en.wikipedia.org/w/index.php?title=Blumlein_pair&oldid=1142630518%7D.

[47] L. Stelmach, W. J. Tam, D. Meegan, and A. Vincent, “Stereo image quality: Effects of mixed spatio-temporal resolution,” *IEEE Transactions on circuits and systems for video technology*, vol. 10, no. 2, pp. 188–193, 2000.

[48] *Loudness Concepts & Panning Laws*, [Online; accessed 14. May 2023], May 2023. [Online]. Available: %5Curl%7Bhttps://www.cs.cmu.edu/~music/icm-online/readings/panlaws%7D.

[49] W. M. Hartmann and A. Wittenberg, "On the externalization of sound images," *The Journal of the Acoustical Society of America*, vol. 99, no. 6, pp. 3678–3688, Jun. 1996. DOI: `10.1121/1.414965`. eprint: `https://pubs.aip.org/asa/jasa/article-pdf/99/6/3678/8102171/3678\_1\_online.pdf`. [Online]. Available: `%5Curl%7Bhttps://doi.org/10.1121/1.414965%7D`.

[50] [Online; accessed 14. May 2023], May 2023. [Online]. Available: `%5Curl%7Bhttps://www.neumann.com/en-en/products/microphones/ku-100%7D`.

[51] H. Krer and G. Flege, "Artificial head for the electroacoustic recording of three dimensional sound field," *Georg Neuman & Co, Audio Export, Tech. Rep*, 1975.

[52] B. Wiggins, *An investigation into the real-time manipulation and control of three-dimensional sound fields*. University of Derby (United Kingdom), 2004.

[53] M. Gerzon, N. R. Dev, and M. Gerzon, *Sound reproduction systems*, [Online; accessed 14. May 2023], Feb. 1981. [Online]. Available: `%5Curl%7Bhttps://patents.google.com/patent/WO1981002502A1/en?oq=WO1981002502A1%7D`.

[54] m. a. gerzon michael a. and g. j. barton geoffrey j., "Ambisonic decoders for hdtv," *journal of the audio engineering society*, Mar. 1992.

[55] R. T. Seeley, "Spherical harmonics," *The American Mathematical Monthly*, vol. 73, no. 4P2, pp. 115–121, 1966.

[56] j. daniel jerome and s. moreau sebastien, "Further study of sound field coding with higher order ambisonics," *journal of the audio engineering society*, May 2004.

[57] K. Hamasaki and K. Hiyama, "Reproducing spatial impression with multichannel audio," in *Audio Engineering Society Conference: 24th International Conference: Multichannel Audio, The New Reality*, Audio Engineering Society, 2003.

[58] T. Carpentier, "Normalization schemes in ambisonic: Does it matter?" In *Audio Engineering Society Convention 142*, May 2017. [Online]. Available: `http://www.aes.org/e-lib/browse.cfm?elib=18645`.

[59] C. Nachbar, F. Zotter, E. Deleflie, and A. Sontacchi, "Ambix-a suggested ambisonics format," in *Ambisonics Symposium*, vol. 2011, 2011.

[60] F. Zotter and M. Frank, "All-round ambisonic panning and decoding," *Journal of the audio engineering society*, vol. 60, no. 10, pp. 807–820, 2012.

[61] A. Politis, S. Tervo, and V. Pulkki, "Compass: Coding and multidirectional parameterization of ambisonic sound scenes," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 6802–6806. DOI: `10.1109/ICASSP.2018.8462608`.

[62] M. Noisternig, A. Sontacchi, T. Musil, and R. Holdrich, "A 3d ambisonic based binaural sound reproduction system," in *Audio Engineering Society Conference: 24th International Conference: Multichannel Audio, The New Reality*, Audio Engineering Society, 2003.

[63] A. Politis, L. McCormack, and V. Pulkki, “Enhancement of ambisonic binaural reproduction using directional audio coding with optimal adaptive mixing,” in *2017 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2017, pp. 379–383. DOI: 10.1109/WASPAA.2017.8170059.

[64] T. McKenzie, D. T. Murphy, and G. Kearney, “Interaural level difference optimization of binaural ambisonic rendering,” *Applied Sciences*, vol. 9, no. 6, p. 1226, 2019.

[65] C. Armstrong, L. Thresh, D. Murphy, and G. Kearney, “A perceptual evaluation of individual and non-individual hrtfs: A case study of the sadie ii database,” *Applied Sciences*, vol. 8, no. 11, p. 2029, 2018.

[66] *GRASSound&Vibration*, [Online; accessed 14. May 2023], Apr. 2015. [Online]. Available: %5Curl%7Bhttp://kemar.us%7D.

[67] *SADIE Spatial Audio For Domestic Interactive Entertainment*, [Online; accessed 14. May 2023], Jul. 2018. [Online]. Available: %5Curl%7Bhttps://www.york.ac.uk/sadie-project/news.html%7B%5C#%7DGoogleVR%7D.

[68] C. Armstrong, “Improvements in the measurement and optimisation of head related transfer functions for binaural ambisonics,” Ph.D. dissertation, University of York, 2019.

[69] T. Rudzki, D. Murphy, and G. Kearney, “Xr-based hrtf measurements,” in *Audio Engineering Society Conference: AES 2022 International Audio for Virtual and Augmented Reality Conference*, Audio Engineering Society, 2022.

[70] C. Murray, *NT-SF1 Ambisonic Microphone RØDE*, [Online; accessed 15. May 2023], May 2023. [Online]. Available: %5Curl%7Bhttps://rode.com/en/microphones/360-ambisonic/nt-sf1%7D.

[71] P. G. Craven and M. A. Gerzon, *Coincident microphone simulation covering three dimensional space and yielding various directional outputs*, US Patent 4,042,779, Aug. 1977.

[72] D. G. Malham, “Spatial hearing mechanisms and sound reproduction,” *University of York*, 1998.

[73] H. L. Helmholtz, *On the Sensations of Tone as a Physiological Basis for the Theory of Music*. Cambridge University Press, 2009.

[74] European Union, “Directive 2002/49/EC relating to the assessment and management of environmental noise (END).,” *Official Journal of the European Communities*, no. L189, pp. 12–25, 2002. DOI: 10.2779/171432. [Online]. Available: http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32002L0049%7B%5C&%7Dfrom=EN.

[75] *Fundamentals of acoustics / Lawrence E. Kinsler ... [et al.]*. eng, 4th ed. New York: Wiley, 2000.

[76] WHO Regional Office for Europe, “Burden of disease from Environmental Noise,” p. 126, 2011.

[77] European Commission, *Science for Environment Policy FUTURE BRIEF : What are the health costs of environmental pollution?* Luxemborg: European Commission, 2018. DOI: 10.2779/88198. [Online]. Available: http://ec.europa.eu/environment/integration/research/newsalert/pdf/health%7B%5C_%7Dcosts%7B%5C_%7Denvironmental%7B%5C_%7Dpollution%7B%5C_%7DFB21%7B%5C_%7Den.pdf.

[78] European Commission's Directorate General for the Environment, *Environmental noise: Noise makes its mark in the long-term*, 2017. [Online]. Available: https://www.youtube.com/watch?v=FAg%7B%5C_%7DZ905E%7B%5C_%7D4 (visited on 04/09/2019).

[79] website, "Coronary heart disease," *Nhs*, Apr. 2023. [Online]. Available: %5Curl%7Bhttps://www.nhs.uk/conditions/coronary-heart-disease%7D.

[80] J. M. Barrigón Morillas, D. Montes González, and G. Rey Gozalo, "A review of the measurement procedure of the iso 1996 standard. relationship with the european noise directive," *Science of The Total Environment*, vol. 565, pp. 595–606, 2016. DOI: https://doi.org/10.1016/j.scitotenv.2016.04.207. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S004896971630955X.

[81] International Standards Organisation, "Acoustics - Description, measurement and assessment of environmental noise - Part 2: Determination of sound pressure levels - ISO 1996-2:2017[E]," International Standards Organisation, Switzerland, Tech. Rep., 2017.

[82] R. L. S. Pierre Jr, D. J. Maguire, and C. S. Automotive, "The impact of a-weighting sound pressure level measurements during the evaluation of noise exposure," in *Conference NOISE-CON*, 2004, pp. 12–14.

[83] M. van den Berg *et al.*, "Position paper on eu noise indicators : A report produced for the european commission," European Commission, 2000.

[84] U. Government, "The Environmental Noise (England) Regulations," no. 2238, p. 20, 2006. [Online]. Available: http://www.legislation.gov.uk/uksi/2006/2238/pdfs/uksi%7B%5C_%7D20062238%7B%5C_%7Den.pdf.

[85] M. Burdett, *Extrium England Noise and Air Quality Viewer*, [Online; accessed 14. May 2023], Mar. 2020. [Online]. Available: %5Curl%7Bhttp://www.extrium.co.uk/noiseviewer.html%7D.

[86] M. Southworth, "The sonic environment of cities," *Environment and Behavior*, vol. 1, no. 1, pp. 49–70, 1969. DOI: 10.1177/001391656900100104. eprint: https://doi.org/10.1177/001391656900100104. [Online]. Available: https://doi.org/10.1177/001391656900100104.

[87] R. Murray Schafer, *Book Of Noise*, Vancouver, 1967.

[88] R. M. Schafer, *The soundscape : our sonic environment and the tuning of the world / R. Murray Schafer.* eng. Rochester, Vt.: Destiny Books, 1994.

[89] R. M. Schafer and D. New, *R. Murray Schafer: Listen*, 2009. [Online]. Available: https://www.youtube.com/watch?time%7B%5C_%7Dcontinue=5%7B%5C&%7Dv=r0lxuXHWfHw (visited on 04/12/2019).

[90] K. Wrightson and E. Ward, "An Introduction to Acoustic Ecology," *Journal of Acoustic Ecology*, pp. 10–13, 1978.

[91] J. Kang, K. Chourmouziadou, K. Sakantamis, B. Wang, and Y. Hao, "COST Action: TD0804 - Soundscape of European Cities and Landscapes," *Inter-Noise 2013*, 2013.

[92] *Oxford english dictionary online [electronic resource]*, eng, Oxford.

[93] Ö. Axelsson, "How to measure soundscape quality," in *Proceedings of Euronoise*, Maastricht, 2015. [Online]. Available: https://www.conforg.fr/euronoise2015/proceedings/data/articles/000067.pdf.

[94] S. Bech, "Methods for subjective evaluation of spatial characteristics of sound," in *Audio Engineering Society Conference: 16th International Conference: Spatial Sound Reproduction*, Mar. 1999. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=8010.

[95] S. Bech and N. Zacharov, *Perceptual Audio Evaluation - Theory, Method and Application.* Wiley, 2007. [Online]. Available: https://books.google.co.uk/books?id=1WGPJai1gX8C.

[96] P. D. Iso, "PD ISO / TS 12913 - 2 : 2018 BSI Standards Publication Acoustics — Soundscape," 2018.

[97] M. S. Engel, A. Fiebig, C. Pfaffenbach, and J. Fels, "A review of socio-acoustic surveys for soundscape studies," *Current Pollution Reports*, vol. 4, pp. 220–239, 3 Sep. 2018. DOI: 10.1007/s40726-018-0094-8.

[98] A. Mitchell *et al.*, "The soundscape indices (ssid) protocol: A method for urban soundscape surveys- questionnaires with acoustical and contextual information," *Applied Sciences (Switzerland)*, vol. 10, 7 Apr. 2020. DOI: 10.3390/app10072397.

[99] S. Harriet and D. T. Murphy, "Auralisation of an urban soundscape," *Acta Acustica united with Acustica*, vol. 101, no. 4, pp. 798–810, 2015. DOI: 10.3813/AAA.918874.

[100] T. Oberman, K. Jambrošić, M. Horvat, and B. B. O. Šćitaroci, "Using virtual soundwalk approach for assessing sound art soundscape interventions in public spaces," *Appl. Sci.*, vol. 10, no. 6, 2020. DOI: 10.3390/app10062102.

[101] C. Guastavino, B. F. G. Katz, J.-d. Polack, D. J. Levitin, and D. Dubois, "Ecological Validity of Soundscape Reproduction," no. May 2014, pp. 1–19, 2004.

[102] C. Guastavino and B. F. G. Katz, "Perceptual evaluation of multi-dimensional spatial audio reproduction," *The Journal of the Acoustical Society of America*, vol. 116, pp. 1105–1115, 2 Aug. 2004. DOI: 10.1121/1.1763973.

[103] U. Bronfenbrenner, “Toward an experimental ecology of human development.,” *Am. Psychol.*, vol. 32, no. 7, pp. 513–531, 1977. DOI: 10.1037/0003-066x.32.7.513.

[104] M. A. Schmuckler, “What is Ecological Validity? A Dimensional Analysis,” *Infancy*, vol. 2, no. 4, pp. 419–436, 2001. DOI: 10.1207/S15327078IN0204_02.

[105] J. J. Gibson, “The ecological approach to the visual perception of pictures,” *Leonardo*, vol. 11, no. 3, pp. 227–235, 1978.

[106] W. Davies *et al.*, “The Positive Soundscape Project: A synthesis of results from many disciplines,” *38th International Congress and Exposition on Noise Control Engineering 2009, INTER-NOISE 2009*, vol. 1, no. December 2013, 2009.

[107] *Apa dictionary of psychology*, [Online; Accessed 1-5-2022], 2022. [Online]. Available: %5Curl%7Bhttps://dictionary.apa.org/affect%7D.

[108] L. F. Barrett and E. Bliss-Moreau, “Affect as a psychological primitive,” *Advances in Experimental Social Psychology*, vol. 41, pp. 167–218, 2009. DOI: 10.1016/S0065-2601(08)00404-8.

[109] C. E. Osgood, *The measurement of meaning [by] Charles E. Osgood, George J. Suci [and] Percy H. Tannenbaum.* eng. Urbana: University of Illinois Press, 1957.

[110] J. A. Russell, “A circumplex model of affect,” *Journal of Personality and Social Psychology*, vol. 39, no. 6, pp. 1161–1178, 1980. DOI: 10.1037/h0077714.

[111] F. Stevens, “Strategies for Environmental Sound Measurement , Modelling , and Evaluation,” PhD Thesis, University of York, 2018.

[112] J. Kang and M. Zhang, “Semantic differential analysis of the soundscape in urban open public spaces,” *Building and Environment*, vol. 45, pp. 150–157, 2009.

[113] Ö. Axelsson, M. E. Nilsson, and B. Berglund, “A principal components model of soundscape perception,” *The Journal of the Acoustical Society of America*, vol. 128, no. 5, pp. 2836–2846, 2010. DOI: 10.1121/1.3493436. [Online]. Available: http://asa.scitation.org/doi/10.1121/1.3493436.

[114] J. Fan, M. Thorogood, and P. Pasquier, “Automatic soundscape affect recognition using a dimensional approach,” *AES: Journal of the Audio Engineering Society*, vol. 64, pp. 646–653, 9 Sep. 2016. DOI: 10.17743/jaes.2016.0044.

[115] C. Tarlao, D. Steele, and C. Guastavino, “Assessing the ecological validity of soundscape reproduction in different laboratory settings,” *PLOS ONE*, vol. 17, no. 6, pp. 1–26, Jun. 2022. DOI: 10.1371/journal.pone.0270401. [Online]. Available: https://doi.org/10.1371/journal.pone.0270401.

[116] N. L. Ping Zhang, *The importance of affective quality*, [Online; accessed 15. May 2023], May 2005. [Online]. Available: %5Curl%7Bhttps://dl.acm.org/doi/fullHtml/10.1145/1081992.1081997?casa_token=buEpxHIO5V0AAAAA:UpTFoLuP0go1Sztl5r6ncBYIWnmee6QXdOyEH8PDnp2O4QAtiFsQPXvQHA484ZwjZA3xGYc4EZs%7D.

[117] P. Kogan *et al.*, "Application of the swedish soundscape-quality protocol in one european and three latin-american cities," 2016. [Online]. Available: https://www.researchgate.net/publication/327079940.

[118] M. M. Bradley and P. J. 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. DOI: 10.1016/0005-7916(94)90063-9. arXiv: 0005-7916(93)E0016-Z.

[119] I. Bakker, T. van der Voordt, P. Vink, and J. de Boon, "Pleasure, Arousal, Dominance: Mehrabian and Russell revisited," *Current Psychology*, vol. 33, no. 3, pp. 405–421, 2014. DOI: 10.1007/s12144-014-9219-4.

[120] J. Fan, M. Thorogood, and P. Pasquier, "Emo-soundscapes: A dataset for soundscape emotion recognition," *2017 7th International Conference on Affective Computing and Intelligent Interaction, ACII 2017*, pp. 196–201, 2018. DOI: 10.1109/ACII.2017.8273600.

[121] G. R. Watts and R. J. 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. arXiv: ISSN0003-682X. [Online]. Available: http://eprints.whiterose.ac.uk/97667/.

[122] A. M. Maehr, G. R. Watts, J. Hanratty, and D. Talmi, "Emotional response to images of wind turbines: A psychophysiological study of their visual impact on the landscape," *Landscape and Urban Planning*, vol. 142, pp. 71–79, 2015. DOI: 10.1016/j.landurbplan.2015.05.011.

[123] F. Stevens, D. T. Murphy, and S. L. Smith, "Soundscape Preference Rating Using Sematic Differential Pairs and the Self-Assessment Manikin," in *Proceedings SMC 2016*, Hamburg, 2016.

[124] A. Dzedzickis, A. Kaklauskas, and V. Bucinskas, "Human emotion recognition: Review of sensors and methods," *Sensors*, vol. 20, no. 3, 2020. DOI: 10.3390/s20030592. [Online]. Available: %5Curl%7Bhttps://doi.org/10.3390/s20030592%7D.

[125] A. Schell and M. Dawson, "Psychophysiology," in *International Encyclopedia of the Social & Behavioral Sciences*, N. J. Smelser and P. B. Baltes, Eds., Oxford: Pergamon, 2001, pp. 12 448–12 452. DOI: https://doi.org/10.1016/B0-08-043076-7/03424-0.

[126] F. Shaffer, R. McCraty, and C. L. Zerr, "A healthy heart is not a metronome: an integrative review of the heart's anatomy and heart rate variability," *Front. Psychol.*, vol. 5, no. 9, pp. 1–19, 2014. DOI: 10.3389/fpsyg.2014.01040.

[127] M. Erfanian, A. Mitchell, and J. Kang, "The neurophysiology and physiology of soundscape : A review of the empirical literature The 6th European Conference on Psychology & the Behavioral Sciences (ECP2019) The neurophysiology and physiology of soundscape : A review of the empirical literature," no. 2, pp. 1–2, 2019.

[128] 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. DOI: 10.1016/j.ijpsycho.2004.02.002.

[129] A. Irwin, D. A. Hall, A. Peters, and C. J. Plack, "Listening to urban soundscapes: Physiological validity of perceptual dimensions," *Psychophysiology*, vol. 48, no. 2, pp. 258–268, 2011. DOI: 10.1111/j.1469-8986.2010.01051.x.

[130] K. Hume and M. Ahtamad, "Physiological responses to and subjective estimates of soundscape elements," *Applied Acoustics*, vol. 74, no. 2, pp. 275–281, 2013. DOI: 10.1016/j.apacoust.2011.10.009. [Online]. Available: http://dx.doi.org/10.1016/j.apacoust.2011.10.009.

[131] J. J. Alvarsson, S. Wiens, and M. E. 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. DOI: 10.3390/ijerph7031036.

[132] D. Hodges, *Music in the Human Experience: An Introduction to Music Psychology*. Taylor & Francis, 2019. [Online]. Available: https://books.google.co.uk/books?id=lx0lEAAAQBAJ.

[133] K. Genuit and A. Fiebig, "Psychoacoustics and its benefit for the soundscape approach," *ACTA ACUSTICA UNITED WITH ACUSTICA*, vol. 92, pp. 1–1, 2006.

[134] C. Trudeau and C. Guastavino, "Classifying soundscapes using a multifaceted taxonomy," pp. 2487–2492, 2018. [Online]. Available: http://www.euronoise2018.eu/docs/papers/411%7B%5C_%7DEuronoise2018.pdf.

[135] O. Bones, T. Cox, and W. Davies, "An evidence based soundscape taxonomy," in *International Congress of Sound and Vibration*, 2017.

[136] O. Bones, T. J. Cox, and W. J. Davies, "Sound categories: Category formation and evidence-based taxonomies," *Frontiers in Psychology*, vol. 9, 2018. DOI: 10.3389/fpsyg.2018.01277. [Online]. Available: https://www.frontiersin.org/article/10.3389/fpsyg.2018.01277.

[137] T. H. Park, J. H. Lee, J. You, M.-J. Yoo, and J. Turner, "Towards soundscape information retrieval (sir)," Sep. 2014. [Online]. Available: https://www.freesound.org/.

[138] A. Field and G. Hole, *How to design and report experiments*. Sage, 2002.

[139] F. Stevens, S. L. Smith, and D. Forest, "Soundscape Categorisation and the Self-Assessment Manikin," pp. 481–488, 2017.

[140] F. Stevens, D. T. Murphy, and S. L. Smith, "The Self-Assessment Manikin and Heart Rate: Responses To Auralised Soundscapes," pp. 1–8, 2016. [Online]. Available: %5Curl%7Bhttps://api.semanticscholar.org/CorpusID:195258140%7D.

[141] P. Lehrer *et al.*, "Heart rate variability biofeedback improves emotional and physical health and performance: A systematic review and meta analysis," *Applied psychophysiology and biofeedback*, vol. 45, pp. 109–129, 2020.

[142] L. Shu *et al.*, "Wearable emotion recognition using heart rate data from a smart bracelet," *Sensors*, vol. 20, no. 3, p. 718, 2020.

[143] M. Mather and J. F. Thayer, "How heart rate variability affects emotion regulation brain networks," *Current Opinion in Behavioral Sciences*, vol. 19, pp. 98–104, 2018, Emotion-cognition interactions. DOI: `https://doi.org/10.1016/j.cobeha.2017.12.017`. [Online]. Available: `%5Curl%7Bhttps://www.sciencedirect.com/science/article/pii/S2352154617300621%7D`.

[144] O. Medvedev, D. Shepherd, and M. J. Hautus, "The restorative potential of soundscapes: A physiological investigation," *Applied Acoustics*, vol. 96, pp. 20–26, 2015. DOI: `10.1016/j.apacoust.2015.03.004`. [Online]. Available: `%5Curl%7Bhttp://dx.doi.org/10.1016/j.apacoust.2015.03.004%7D`.

[145] S. R. Payne, "The production of a perceived restorativeness soundscape scale," *Applied Acoustics*, vol. 74, no. 2, pp. 255–263, 2013. DOI: `10.1016/j.apacoust.2011.11.005`. [Online]. Available: `http://dx.doi.org/10.1016/j.apacoust.2011.11.005`.

[146] X.-C. Hong, S. Cheng, J. Liu, E. Dang, J.-B. Wang, and Y. Cheng, "The physiological restorative role of soundscape in different forest structures," *Forests*, vol. 13, no. 11, p. 1920, 2022.

[147] M. Green and D. Murphy, "EigenScape: A Database of Spatial Acoustic Scene Recordings," *Applied Sciences*, vol. 7, no. 11, p. 1204, 2017. DOI: `10.3390/app7111204`.

[148] E. Bates, M. Gorzel, L. Ferguson, H. O'Dwyer, and F. M. Boland, "Comparing ambisonic microphones – part 1," in *Audio Engineering Society Conference: 2016 AES International Conference on Sound Field Control*, Jul. 2016. [Online]. Available: `%5Curl%7Bhttp://www.aes.org/e-lib/browse.cfm?elib=18317%7D`.

[149] E. Bates, S. Dooney, M. Gorzel, H. O'Dwyer, L. Ferguson, and F. M. Boland, "Comparing ambisonic microphones — part 2," in *Audio Engineering Society Convention 142*, May 2017. [Online]. Available: `%5Curl%7Bhttp://www.aes.org/e-lib/browse.cfm?elib=18607%7D`.

[150] M. Green and D. Murphy, *Eigenscape dataset*, Dataset, Oct. 2017. DOI: `10.5281/zenodo.2628463`. [Online]. Available: `https://doi.org/10.5281/zenodo.2628463`.

[151] E. P. Kasten, S. H. Gage, J. Fox, and W. Joo, "The remote environmental assessment laboratory's acoustic library: An archive for studying soundscape ecology," *Ecological Informatics*, vol. 12, pp. 50–67, 2012. DOI: `10.1016/j.ecoinf.2012.08.001`.

[152] P. Devos, "Soundecology indicators applied to urban soundscapes," *Inter-noise 2016*, no. 4, pp. 177–184, 2016.

[153] mh acoustics, "em32 Eigenmike microphone array release notes v15.0," mh acoustics, Tech. Rep., 2013, p. 7.

[154] P. Guyot, *Acoustic Indices Github Repository*, comp. software, 2019. [Online]. Available: `%7Bhttps://github.com/patriceguyot/Acoustic_Indices%7D` (visited on 07/09/2020).

[155] I. of Electronic Music & Acoustics, *IEM Plug-in Suite*, Software, Graz, Austria, 2019. [Online]. Available: `https://plugins.iem.at/` (visited on 03/13/2020).

[156] C. Schörkhuber, M. Zaunschirm, and R. Höldrich, "Binaural Rendering of Ambisonic Signals via Magnitude Least Squares," *Fortschritte der Akustik – DAGA 2018*, no. 3, pp. 339–342, 2018.

[157] *Qualtrics XM - Experience Management Software*, [Online; accessed 17. Apr. 2023], Apr. 2023. [Online]. Available: `https://www.qualtrics.com`.

[158] *: The Embed Audio element - HTML: HyperText Markup Language MDN*, [Online; accessed 15. May 2023], Apr. 2023. [Online]. Available: `%5Curl%7Bhttps://developer.mozilla.org/en-US/docs/Web/HTML/Element/audio%7D`.

[159] S. Sensing, *Shimmer3 gsr+ unit - shimmer wearable sensor technology*, [Online; Accessed 10-7-2022], 2018. [Online]. Available: `%5Curl%7Bhttps://shimmersensing.com/product/shimmer3-gsr-unit/%7D`.

[160] S. Sensing, "Gsr+ user guide revision 1.13," 2018.

[161] P. van Gent, H. Farah, N. Nes, and B. Arem, "Heart rate analysis for human factors: Development and validation of an open source toolkit for noisy naturalistic heart rate data," Jun. 2018.

[162] J. Park, H. S. Seok, S.-S. Kim, and H. Shin, "Photoplethysmogram analysis and applications: An integrative review," *Frontiers in Physiology*, vol. 12, 2022. DOI: `10.3389/fphys.2021.808451`. [Online]. Available: `%5Curl%7Bhttps://www.frontiersin.org/articles/10.3389/fphys.2021.808451%7D`.

[163] A. Reisner, P. A. Shaltis, D. McCombie, H. H. Asada, D. S. Warner, and M. A. Warner, "Utility of the Photoplethysmogram in Circulatory Monitoring," *Anesthesiology*, vol. 108, no. 5, pp. 950–958, May 2008. DOI: `10.1097/ALN.0b013e31816c89e1`. eprint: `https://pubs.asahq.org/anesthesiology/article-pdf/108/5/950/367170/0000542-200805000-00024.pdf`. [Online]. Available: `https://doi.org/10.1097/ALN.0b013e31816c89e1`.

[164] *Shimmer MATLAB Instrument Driver - File Exchange - MATLAB CentralFile Exchange - MATLAB Central*, [Online; accessed 17. Apr. 2023], Jun. 2022. [Online]. Available: `https://uk.mathworks.com/matlabcentral/fileexchange/43712-shimmer-matlab-instrument-driver`.

[165] *MATLAB version 9.10.0.1739362 (R2021a) Update 5*, Software, Natick, Massachusetts: The Mathworks, Inc., 2021.

[166] P. van Gent, H. Farah, N. van Nes, and B. van Arem, "Heartpy: A novel heart rate algorithm for the analysis of noisy signals," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 66, pp. 368–378, 2019. DOI: `https://doi.org/10.1016/j.trf.2019.09.015`. [Online]. Available: `%5Curl%7Bhttps://www.sciencedirect.com/science/article/pii/S1369847818306740%7D`.

[167] A. Horvers, N. Tombeng, T. Bosse, A. W. Lazonder, and I. Molenaar, "Detecting emotions through electrodermal activity in learning contexts: A systematic review," *Sensors*, vol. 21, 23 Dec. 2021. DOI: `10.3390/s21237869`.

[168] M. Sharma, S. Kacker, and M. Sharma, "A brief introduction and review on galvanic skin response," *International Journal of Medical Research Professionals*, vol. 2, 6 Dec. 2016. DOI: `10.21276/ijmrp.2016.2.6.003`.

[169] A. P. Field, *How to design and report experiments.* London: Sage Publications Ltd, 2003.

[170] S. Laborde, E. Mosley, and J. F. Thayer, "Heart rate variability and cardiac vagal tone in psychophysiological research - Recommendations for experiment planning, data analysis, and data reporting," *Front. Psychol.*, vol. 8, no. FEB, pp. 1–18, 2017. DOI: `10.3389/fpsyg.2017.00213`.

[171] M. Egger, M. Ley, and S. Hanke, "Emotion Recognition from Physiological Signal Analysis: A Review," *Electron. Notes Theor. Comput. Sci.*, vol. 343, pp. 35–55, 2019. DOI: `10.1016/j.entcs.2019.04.009`. [Online]. Available: `https://doi.org/10.1016/j.entcs.2019.04.009`.

[172] J. F. Thayer and R. D. Lane, "A model of neurovisceral integration in emotion regulation and dysregulation," *J. Affect. Disord.*, vol. 61, no. 3, pp. 201–216, 2000. DOI: `10.1016/S0165-0327(00)00338-4`.

[173] S. Porges, "Orienting in a defensive world: Mammalian modifications of our evolutionary heritage. a polyvagal theory," *Psychophysiology*, vol. 32, pp. 301–318, 4 1995. DOI: `10.1111/j.1469-8986.1995.tb01213.x`. [Online]. Available: `https://doi.org/10.1111/j.1469-8986.1995.tb01213.x`.

[174] S. W. Porges, "The polyvagal perspective," *Biol. Psychol.*, vol. 74, no. 2, pp. 116–143, 2007. DOI: `10.1016/j.biopsycho.2006.06.009`.

[175] F. Shaffer and J. P. Ginsberg, "An Overview of Heart Rate Variability Metrics and Norms," *Front. Public Heal.*, vol. 5, no. 9, pp. 1–17, 2017. DOI: `10.3389/fpubh.2017.00258`.

[176] G. G. Berntson *et al.*, "Heart rate variability: Origins methods, and interpretive caveats," *Psychophysiology*, vol. 34, pp. 623–648, 6 1997. DOI: `10.1111/j.1469-8986.1997.tb02140.x`.

[177] G. G. BERNTSON and J. R. STOWELL, "Ecg artifacts and heart period variability: Don't miss a beat!" *Psychophysiology*, vol. 35, no. 1, pp. 127–132, 1998. DOI: `10.1111/1469-8986.3510127`.

[178] L. Hejjel and L. Kellenyi, “The corner frequencies of the ecg amplifier for heart rate variability analysis,” *Physiological Measurement*, vol. 26, no. 1, p. 39, Dec. 2004. DOI: 10.1088/0967-3334/26/1/004. [Online]. Available: https://dx.doi.org/10.1088/0967-3334/26/1/004.

[179] International Telecommunication Union Radiocommunication Bureau, “Algorithms to measure audio programme loudness and true-peak audio level,” Geneva, Tech. Rep., 2017. [Online]. Available: http://www.itu.int/ITU-R/go/patents/en.

[180] Polar Research and Technology, “Polar h10 heart rate sensor system,” Polar Electro Oy, White Paper, Nov. 2019. [Online]. Available: https://www.polar.com/sites/default/files/static/science/white-papers/polar-h10-heart-rate-sensor-white-paper.pdf.

[181] *ABC of clinical electrocardiography [electronic resource] / edited by Francis Morris, William J. Brady, John Camm.* (ABC series (Malden, Mass.)), eng, 2nd ed. Malden, Mass. ; Oxford: Blackwell, 2008.

[182] R. Gilgen-Ammann, T. Schweizer, and T. Wyss, “Rr interval signal quality of a heart rate monitor and an ecg holter at rest and during exercise,” *European Journal of Applied Physiology*, vol. 119, no. 3, pp. 1525–1532, 2019. DOI: 10.1007/s00421-019-04142-5. [Online]. Available: https://doi.org/10.1007/s00421-019-04142-5.

[183] M. Schaffarczyk, B. Rogers, R. Reer, and T. Gronwald, “Validity of the polar h10 sensor for heart rate variability analysis during resting state and incremental exercise in recreational men and women,” *Sensors*, vol. 22, 17 Sep. 2022. DOI: 10.3390/s22176536.

[184] Polar Electro Oy, “Polar H10,” Polar Electro Oy, Kempele, Finland, White Paper, 2018. [Online]. Available: %5Curl%7Bhttps://www.polar.com/uk-en/products/accessories/polar_h10_heart_rate_sensor%7D (visited on 07/23/2020).

[185] G. D. Clifford *et al.*, “Af classification from a short single lead ecg recording: The physionet/computing in cardiology challenge 2017,” in *2017 Computing in Cardiology (CinC)*, IEEE, 2017, pp. 1–4.

[186] K. Martinmäki and H. Kinnunen, “R-R interval measurement and heart rate variability in polar products,” Polar Electro Oy, Kempele, Finland, White Paper, May 2011.

[187] *Google Pixel 4 XL - Full phone specifications*, [Online; accessed 15. May 2023], May 2023. [Online]. Available: %5Curl%7Bhttps://www.gsmarena.com/google_pixel_4_xl-9895.php%7D.

[188] J. Happonen, *Polar sensor logger*, Software, 2019. [Online]. Available: %5Curl%7Bhttps://play.google.com/store/apps/details?id=com.j_ware.polarsensorlogger%7D (visited on 03/13/2020).

[189] M. Vollmer, "HRVTool - an open-source matlab toolbox for analyzing heart rate variability," in *2019 Computing in Cardiology Conference (CinC)*, Computing in Cardiology, Dec. 2019. DOI: `10.22489/cinc.2019.032`. [Online]. Available: `https://doi.org/10.22489%2Fcinc.2019.032`.

[190] M. Vollmer, "A robust, simple and reliable measure of heart rate variability using relative RR intervals," in *2015 Computing in Cardiology Conference (CinC)*, IEEE, 2015, pp. 609–612. DOI: `10.1109/CIC.2015.7410984`.

[191] L. Salahuddin, J. Cho, M. G. Jeong, and D. Kim, "Ultra short term analysis of heart rate variability for monitoring mental stress in mobile settings," *Annu. Int. Conf. IEEE Eng. Med. Biol. - Proc.*, pp. 4656–4659, 2007. DOI: `10.1109/IEMBS.2007.4353378`.

[192] M. Malik *et al.*, "Heart rate variability: Standards of measurement, physiological interpretation, and clinical use," *European Heart Journal*, vol. 17, no. 3, pp. 354–381, Mar. 1996. DOI: `10.1093/oxfordjournals.eurheartj.a014868`. eprint: `https://academic.oup.com/eurheartj/article-pdf/17/3/354/1312587/17-3-354.pdf`. [Online]. Available: `https://doi.org/10.1093/oxfordjournals.eurheartj.a014868`.

[193] Ipek, *Normality test package*, Software, MATLAB Central File Exchange, 2021. [Online]. Available: `https://www.mathworks.com/matlabcentral/fileexchange/60147-normality-test-package` (visited on 03/13/2020).

[194] R. W. Rumberger, "Hierarchical linear models: Applications and data analysis methods," eng, *Economics of education review*, vol. 16, no. 3, p. 348, 1997.

[195] A. Nielsen, *Practical time series analysis: Prediction with statistics and machine learning.* O'Reilly Media, 2019.

[196] D. Ciuk, A. K. Troy, and M. C. Jones, "Measuring Emotion: Self-Reports vs. Physiological Indicators," *SSRN Electronic Journal*, pp. 1–29, Apr. 2015. DOI: `10.2139/ssrn.2595359`.

[197] J. Y. Hong *et al.*, "Quality assessment of acoustic environment reproduction methods for cinematic virtual reality in soundscape applications," *Build. Environ.*, vol. 149, no. May 2018, pp. 1–14, 2019. DOI: `10.1016/j.buildenv.2018.12.004`. [Online]. Available: `https://doi.org/10.1016/j.buildenv.2018.12.004`.

[198] J. Y. Jeon, J. Y. Hong, C. Lavandier, J. Lafon, Ö. Axelsson, and M. Hurtig, "A cross-national comparison in assessment of urban park soundscapes in france, korea, and sweden through laboratory experiments," *Applied Acoustics*, vol. 133, pp. 107–117, 2018. DOI: `https://doi.org/10.1016/j.apacoust.2017.12.016`. [Online]. Available: `%5Curl%7Bhttps://doi.org/10.1016/j.apacoust.2017.12.016%7D`.

[199] D. S. Quintana and J. A. J. Heathers, "Considerations in the assessment of heart rate variability in biobehavioral research," *Frontiers in Psychology*, vol. 5, p. 805, 2014. DOI: `10.3389/fpsyg.2014.00805`. [Online]. Available: `https://www.frontiersin.org/article/10.3389/fpsyg.2014.00805`.

[200] J. Williams, S. Shepstone, and D. Murphy, "Understanding immersion in the context of films with spatial audio," in *Audio Engineering Society Conference: AES 2022 International Audio for Virtual and Augmented Reality Conference*, Aug. 2022. [Online]. Available: `%5Curl%7Bhttp://www.aes.org/e-lib/browse.cfm?elib=21878%7D`.

[201] M. Slater and S. Wilbur, "A framework for immersive virtual environments (five): Speculations on the role of presence in virtual environments," *Presence: Teleoperators & Virtual Environments*, vol. 6, pp. 603–616, 1997.

[202] C. Zhang, A. Perkis, and S. Arndt, "Spatial immersion versus emotional immersion, which is more immersive?" In *2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX)*, 2017, pp. 1–6. DOI: `10.1109/QoMEX.2017.7965655`.

[203] A. Tellegen and G. Atkinson, "Openness to absorbing and self-altering experiences (" absorption"), a trait related to hypnotic susceptibility.," *Journal of abnormal psychology*, vol. 83, no. 3, p. 268, 1974.

[204] *Home - YouTube VR*, [Online; accessed 8. May 2023], May 2023. [Online]. Available: `%5Curl%7Bhttps://vr.youtube.com%7D`.

[205] *YouTube 360 VR Ambisonics Teardown!* [Online; accessed 8. May 2023], Jan. 2017. [Online]. Available: `%5Curl%7Bhttps://www.brucewiggins.co.uk/?p=700%7D`.

[206] F. Stevens, D. T. Murphy, and S. L. Smith, "Ecological validity of stereo UHJ soundscape reproduction," in *Audio Engineering Society Proceedings of the 142nd Convention*, 2017, pp. 1–13.

[207] F. Stevens and S. L. Smith, "Soundscape Auralisation and Visualisation : a Cross-Modal Approach To Soundscape Evaluation," no. 1509136, pp. 133–140, 2018.

[208] *What is Samsung Gear 360 camera? Samsung UK*, [Online; accessed 8. May 2023], Dec. 2020. [Online]. Available: `%5Curl%7Bhttps://www.samsung.com/uk/support/mobile-devices/what-is-samsung-gear-360-camera%7D`.

[209] C. Brown, "Bringing pixels front and center in VR video," *Google*, Mar. 2017. [Online]. Available: `%5Curl%7Bhttps://blog.google/products/google-ar-vr/bringing-pixels-front-and-center-vr-video%7D`.

[210] kkroening, *ffmpeg-python*, [Online; accessed 8. May 2023], May 2023. [Online]. Available: `%5Curl%7Bhttps://github.com/kkroening/ffmpeg-python%7D`.

[211] B. Design, *DaVinci Resolve 18 Blackmagic Design*, [Online; accessed 8. May 2023], May 2023. [Online]. Available: `%5Curl%7Bhttps://www.blackmagicdesign.com/uk/products/davinciresolve%7D`.

[212] Google, *Spatial-Media*, [Online; accessed 8. May 2023], May 2023. [Online]. Available: `%5Curl%7Bhttps://github.com/google/spatial-media%7D`.

[213] EH/1/1, "Acoustics — Methods for calculating loudness," British Standards Institute, Standard 9, 2018.

[214] BS ISO 12913-3:2019, "Acoustics – Soundscape – Part 3: Data analysis," The British Standards Institution, London, UK, Technical Specification, Dec. 2019.

[215] M. Lionello, F. Aletta, A. Mitchell, and J. Kang, "Introducing a method for intervals correction on multiple likert scales: A case study on an urban soundscape data collection instrument," *Frontiers in Psychology*, vol. 11, 2021. DOI: 10.3389/fpsyg.2020.602831. [Online]. Available: %5Curl%7Bhttps://www.frontiersin.org/articles/10.3389/fpsyg.2020.602831%7D.

[216] A. Mitchell *et al.*, "Investigating urban soundscapes of the covid-19 lockdown: A predictive soundscape modeling approach," *The Journal of the Acoustical Society of America*, vol. 150, no. 6, pp. 4474–4488, 2021. DOI: 10.1121/10.0008928. eprint: https://doi.org/10.1121/10.000892. [Online]. Available: https://doi.org/10.1121/10.000892.

[217] J. Y. Jeon, J. Y. Hong, C. Lavandier, J. Lafon, O. Axelsson, and M. Hurtig, "A cross-national comparison in assessment of urban park soundscapes in france, korea, and sweden through laboratory experiments," *Applied Acoustics*, vol. 133, pp. 107–117, 2018. DOI: 10.1016/j.apacoust.2017.12.016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0003682X1730912X.

[218] M. Erfanian, A. Mitchell, F. Aletta, and J. Kang, "Psychological well-being and demographic factors can mediate soundscape pleasantness and eventfulness: A large sample study," *J. Environ. Psychol.*, vol. 77, no. February 2020, p. 101 660, 2021. DOI: 10.1016/j.jenvp.2021.101660. [Online]. Available: https://doi.org/10.1016/j.jenvp.2021.101660.

[219] M. Lionello, F. Aletta, A. Mitchell, and J. Kang, "Introducing a Method for Intervals Correction on Multiple Likert Scales: A Case Study on an Urban Soundscape Data Collection Instrument," *Frontiers in Psychology*, vol. 11, no. 1, 2021. DOI: 10.3389/fpsyg.2020.602831.

[220] *REAPER Audio Production Without Limits*, [Online; accessed 13. May 2023], May 2023. [Online]. Available: %5Curl%7Bhttps://www.reaper.fm%7D.

[221] *Insight 2 Intelligent Metering Plug-in iZotope*, [Online; accessed 8. May 2023], May 2023. [Online]. Available: %5Curl%7Bhttps://www.izotope.com/en/products/insight.html%7D.

[222] *OMNITONE*, [Online; accessed 8. May 2023], Jan. 2019. [Online]. Available: %5Curl%7Bhttps://googlechrome.github.io/omnitone%7D.

[223] T. Deppisch, N. Meyer-Kahlen, B. Hofer, T. Latka, and T. Zernicki, "Hoast: A higher-order ambisonics streaming platform," in *Audio Engineering Society Convention 148*, May 2020. [Online]. Available: %5Curl%7Bhttp://www.aes.org/e-lib/browse.cfm?elib=20828%7D.

[224] C. Rajguru, M. Obrist, and G. Memoli, "Spatial Soundscapes and Virtual Worlds: Challenges and Opportunities," *Front. Psychol.*, vol. 11, 2020. DOI: 10.3389/fpsyg.2020.569056.

[225] H. Harris and Arup, *Virtual reality soundbooths - communicating changes in aircraft noise for a public consultation*, [Online; Accessed 22-07-2022], 2022. [Online]. Available: `%5Curl%7Bhttps://www.arup.com/projects/virtual-reality-soundbooths%7D`.

[226] H. I. Jo and J. Y. Jeon, "Perception of urban soundscape and landscape using different visual environment reproduction methods in virtual reality," *Applied Acoustics*, vol. 186, p. 108 498, 2022. DOI: `https://doi.org/10.1016/j.apacoust.2021.108498`. [Online]. Available: `%5Curl%7Bhttps://www.sciencedirect.com/science/article/pii/S0003682X21005922%7D`.

[227] G. M. Echevarria Sanchez, T. Van Renterghem, K. Sun, B. De Coensel, and D. Botteldooren, "Using virtual reality for assessing the role of noise in the audio-visual design of an urban public space," *Landscape and Urban Planning*, vol. 167, pp. 98–107, 2017. DOI: `https://doi.org/10.1016/j.landurbplan.2017.05.018`. [Online]. Available: `%5Curl%7Bhttps://www.sciencedirect.com/science/article/pii/S0169204617301226%7D`.

[228] C. Xu, T. Oberman, F. Aletta, H. Tong, and J. Kang, "Ecological Validity of Immersive Virtual Reality (IVR) Techniques for the Perception of Urban Sound Environments," *Acoustics*, vol. 3, no. 1, pp. 11–24, 2020. DOI: `10.3390/acoustics3010003`.

[229] M. Luigi, M. Massimiliano, P. Aniello, R. Gennaro, and P. R. Virginia, "On the validity of immersive virtual reality as tool for multisensory evaluation of urban spaces," *Energy Procedia*, vol. 78, pp. 471–476, 2015, 6th International Building Physics Conference, IBPC 2015. DOI: `https://doi.org/10.1016/j.egypro.2015.11.703`. [Online]. Available: `%5Curl%7Bhttps://www.sciencedirect.com/science/article/pii/S1876610215024352%7D`.

[230] V. Puyana-Romero, L. S. Lopez-Segura, L. Maffei, R. Hernandez-Molina, and M. Masullo, "Interactive soundscapes: 360-video based immersive virtual reality in a tool for the participatory acoustic environment evaluation of urban areas," *Acta Acust. united with Acust.*, vol. 103, no. 4, pp. 574–588, 2017. DOI: `10.3813/AAA.919086`.

[231] U. Technologies, *Unity Game Engine*, [Online; Accessed 1-5-2022], 2021. [Online]. Available: `%5Curl%7Bhttps://unity.com/%7D`.

[232] *Unreal Engine 5*, [Online; Accessed 13. May 2023], May 2023. [Online]. Available: `%5Curl%7Bhttps://www.unrealengine.com/en-US/unreal-engine-5%7D`.

[233] *OpenXR - High-performance access to AR and VR —collectively known as XR— platforms and devices*, [Online; accessed 13. May 2023], Dec. 2016. [Online]. Available: `%5Curl%7Bhttps://www.khronos.org/openxr%7D`.

[234] *The World's First Immersive Projector*, [Online; accessed 13. May 2023], May 2023. [Online]. Available: `%5Curl%7Bhttps://broomx.com/devices/mk360%7D`.

[235] *The Igloo CAVE Shared Immersive Spaces Igloo Vision*, [Online; accessed 13. May 2023], May 2023. [Online]. Available: %5Curl%7Bhttps://www.igloovision.com/products/technology/igloo-caves%7D.

[236] J. S. Bamford and J. Vanderkooy, "Ambisonic Sound for Us," in *Audio Engineering Society Convention 99*, Oct. 1995. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=7628.

[237] M. Frank, "How to make ambisonics sound good," in *International Conference 7th Forum Acusticum Krakow 2014*, European Acoustics Association (EAA), Sep. 2014.

[238] *SteamVR - Valve Developer Community*, [Online; accessed 13. May 2023], Apr. 2023. [Online]. Available: %5Curl%7Bhttps://developer.valvesoftware.com/wiki/SteamVR%7D.

[239] *VIVE Developers Get started developing for VIVE*, [Online; accessed 13. May 2023], May 2023. [Online]. Available: %5Curl%7Bhttps://developer.vive.com/eu%7D.

[240] Meta, *Meta Quest 2*, [Online; Accessed 1-5-2022], 2020. [Online]. Available: %5Curl%7Bhttps://store.facebook.com/gb/quest/products/quest-2%7D.

[241] *What is Max? Cycling '74*, [Online; accessed 13. May 2023], May 2023. [Online]. Available: %5Curl%7Bhttps://cycling74.com/products/max%7D.

[242] t. rudzki tomasz, c. earnshaw chris, d. murphy damian, and g. kearney gavin, "Salte pt. 2: On the design of the salte audio rendering engine for spatial audio listening tests in vr," *journal of the audio engineering society*, Oct. 2019.

[243] Firelight Technologies Pty Ltd, *FMOD*, [Online; Accessed 1-5-2022], 2022. [Online]. Available: %5Curl%7Bhttps://fmod.com/%7D.

[244] Audiokinetic Inc, *Wwise*, [Online; Accessed 1-5-2022], 2022. [Online]. Available: %5Curl%7Bhttps://www.audiokinetic.com/%7D.

[245] *Resonance Audio - Google*, [Online; accessed 13. May 2023], Sep. 2022. [Online]. Available: %5Curl%7Bhttps://resonance-audio.github.io/resonance-audio%7D.

[246] *Unity Audio: Unity Oculus Developers*, [Online; accessed 13. May 2023], May 2023. [Online]. Available: %5Curl%7Bhttps://developer.oculus.com/documentation/unity/unity-audio%7D.

[247] *VIVE 3DSP Audio SDK — VIVE 3DSP Audio SDK v 1.2.8 documentation*, [Online; accessed 13. May 2023], Dec. 2021. [Online]. Available: %5Curl%7Bhttps://hub.vive.com/storage/3dsp/vive_3dsp_audio_sdk.html%7D.

[248] *Steam Audio*, [Online; accessed 13. May 2023], Mar. 2023. [Online]. Available: %5Curl%7Bhttps://valvesoftware.github.io/steam-audio%7D.

[249] *Home*, [Online; accessed 13. May 2023], May 2023. [Online]. Available: https://wiki.beardedninjagames.com.

[250] R. Rzayev, P. Ugnivenko, S. Graf, V. Schwind, and N. Henze, “Reading in vr: The effect of text presentation type and location,” in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, Yokohama, Japan: Association for Computing Machinery, 2021. DOI: 10.1145/3411764.3445606. [Online]. Available: https://doi.org/10.1145/3411764.3445606.

[251] D. Williams, “What our bodies tell us about noise,” *Frontiers for Young Minds*, vol. 10, p. 689 100, Aug. 2022. DOI: 10.3389/frym.2022.689100.

[252] Z. Li and J. Kang, “Sensitivity analysis of changes in human physiological indicators observed in soundscapes,” *Landscape and Urban Planning*, vol. 190, p. 103 593, 2019. DOI: https://doi.org/10.1016/j.landurbplan.2019.103593. [Online]. Available: %5Curl%7Bhttps://www.sciencedirect.com/science/article/pii/S0169204619303792%7D.

[253] A. Koshy *et al.*, “An observation of common physiological parameters during esports activity,” *International Journal of Esports*, vol. 1, no. 1, 2020.

[254] J. Marín-Morales, J. L. Higuera-Trujillo, J. Guixeres, C. Llinares, M. Alcañiz, and G. Valenza, “Heart rate variability analysis for the assessment of immersive emotional arousal using virtual reality: Comparing real and virtual scenarios,” *PLoS One*, vol. 16, no. 7 July, pp. 1–16, 2021. DOI: 10.1371/journal.pone.0254098.