UNIVERSITY of York

School of Physics, Engineering, and Technology

MSc Project Report

2022/2023

Student Name: Richard Cully

Project Title: The Virtual Drum Room (VDR): Simulating Room Acoustics through Real-Time Auralisation

Supervisors: Dr Frank Stevens

Dr Aglaia Foteinou

School of Physics, Engineering, and Technology
University of York
Heslington
York
YO10 5DD

Abstract

Reverberation, whether natural or artificial, is a highly sought-after audio effect that many musicians find desirable. In particular, the unique reverberation characteristics of different rooms can greatly enhance the sound of instruments, making it especially important for drummers due to the acoustic nature of the instrument. Currently, drummers face challenges in performing in varied acoustic environments without physically being present in those spaces, as existing artificial reverberation effects, such as convolution reverberation plug-ins, do not provide real-time and spatially informative experiences. This project aims to address this limitation by developing a system that can simulate different acoustic environments for drummers using real-time auralisation techniques. The focus will be on designing and implementing an interactive auralisation system specifically tailored for drums.

This thesis comprehensively reviews relevant academic literature, encompassing areas such as room acoustics, virtual acoustics, and interactive music performance systems. The literature review shed light on suitable methodologies for capturing source material and conducting Spatial Room Impulse Response (SRIR) measurements, which form integral parts of the auralisation chain upon which this system is built. The literature review emphasised the significance of incorporating directivity in order to achieve a realistic and immersive auralisation experience. This can be effectively captured through the use of multi-channel auralisation techniques.

The literature review informed the design and implementation of the Virtual Drum Room (VDR), which comprised a 4-channel auralisation system. A drum kit was placed in an anechoic chamber which was captured using the 4 microphones, processed within Reaper using the X-MCFX convolver and played back to the drummers binaurally via headphones. The system used measured SRIRs which were captured in two spaces at the University of York. In order to evaluate the success and performance of the system, a perceptual listening test was conducted. The evaluation sought to investigate how different microphone configurations were perceived and the preferences for them. The results indicate a unanimously positive user experience, supporting its motivation as a recreational tool. Additionally, the results indicate that the use of spatialised auralisations is noticeable and preferred by the participants, further supporting the motivation behind the system.

Contents

List of Figures

Acknowledgements

I would like to express my heartfelt gratitude to several individuals who have played pivotal roles in the success of my research project. Foremost, I extend my profound appreciation to my dedicated supervisor, Dr. Frank Stevens, for his unwavering guidance, support, and initial encouragement that paved the way for this research project. I also extend a special thank you to my second supervisor, Dr. Aglaia Foteinou, for her invaluable insights and consistent support throughout the project.

I am extremely grateful to Andrew Chadwick for his generous assistance in assisting with the practical aspects and logistics of the project. Lastly, my thanks go to Sinuo Feng for her contributions to the SRIR measurements and her aid with the camera work.

Chapter 1

Introduction

Room acoustics play a crucial role in shaping the sound of musical instruments during performances. While all instruments are affected by the acoustic properties of a room, drums are particularly susceptible due to their unique characteristics. The drums' inherent loudness enables them to actively engage with the surrounding environment, setting them apart from quieter instruments. Their percussive nature produces forceful, resonant sounds that effortlessly fill a room. Consequently, the room's acoustics become essential in determining how the drum sound is projected and dispersed. The room's reflections, resonances, and reverberations greatly influence the perceived sound of drums, impacting clarity, punch, and tonal balance. The complex nature of a drum kit, spanning a wide performance range and encompassing diverse frequencies, intensifies the significance of room acoustics. From the low-end bass drum frequencies to the high-end frequencies of the hi-hats, each frequency behaves uniquely within the room, affecting its travel and decay. In addition, the drums' impulsive nature leads to time-varying sound alterations in a reverberant environment, unlike instruments with soft and sustained sounds that are more affected by timbre changes caused by room reverberation.

Research has demonstrated that musicians adjust their playing in response to the acoustic parameters of the performance space [1]. Adaptions to their performance are based on how they perceive their own sound which is influenced by the aural feedback of the acoustic space in which they are performing [2]. Traditionally, the only way for a drummer to fully experience and benefit from the acoustic characteristics of a particular room is to physically perform in that space. However, this approach is impractical and restrictive for drummers due to the logistical challenges involved.

Currently, drummers can simulate the response of a room using audio effects such as convolution reverberation plug-ins which use impulse responses recorded in specific environments that simulate the acoustic sound of a specific location. These plug-ins can be used to create a high-quality reverberation effect for your drum sound. However, these methods tend to lack spatial information and more importantly, as post-production tools, can't be used in real-time for musical performance.

1.1 Aims & Objectives

Based on the background and context mentioned above, this project aims to investigate applying virtual acoustics to the drums. By exploring real-time auralisation techniques that incorporate spatial information, the Virtual Drum Room (VDR) developed in this project aims to provide a technologically advanced alternative to what is already available. The VDR will allow drummers to experience the sonic characteristics of different acoustic environments without the need for physical presence in each space. The motivation behind the VDR is primarily as a recreational tool, however, this doesn't mean that the VDR does not have other applications such as virtual reality or virtual acoustic recording. The intention was not to create a system that provided a hyper-realistic simulation, but rather an experience that provided moderate realism which incorporates spatial and immersive audio techniques to provide a real-time auralisation system for drummers that is believed to be the first of its kind.

In order to achieve the aim of the project above, the following objectives are defined:

1. Explore real-time auralisation methodologies and technologies for musical applications and specifically investigate source capture and impulse response techniques for the purpose of real-time auralisation for musical performance.

2. Design and build the VDR informed by the knowledge gained from the literature review.

3. Evaluate the VDR to determine its success and performance via perceptual listening tests.

4. Analyse the results obtained from the evaluation.

This project will make use of Spatial Room Impulse Responses (SRIRs) to provide a spatial and immersive simulation. In order to create the VDR, the methodologies of creating a real-time auralisation system had to be investigated and therefore is an aim and objective of this project. This required investigating microphone capture and SRIR techniques, which are informed by the literature review and a methodology was devised.

The VDR would be evaluated through a listening test which involved getting participants (drummers) to experience the VDR and provide feedback on their experience. The aim of the evaluation was not to obtain an objective analysis of the performance of the VDR but rather to obtain a subjective analysis of the VDR regarding the participant's view on its performance and their preferences. The aims and objectives are discussed further in Section 4.1.

1.2 Thesis Structure

The thesis is structured accordingly:

- Chapter 2 - Literature Review: This chapter provides a detailed review of the relevant literature to inform the design and implementation of the VDR.
- Chapter 3 - Design & Implementation: This chapter provides a detailed account of the entire design and implementation of the VDR, giving details of the software, hardware and methodologies used.
- Chapter 4 - Evaluation: This penultimate chapter provides an account of the experimental protocol of the listening test that was used to evaluate the VDR. The results from the listening test are presented, followed by a discussion of the results. The chapter ends with a discussion of some limitations associated with the VDR and its evaluation.
- Chapter 5 Conclusion: This final chapter provides a summary of the entire thesis followed by a restatement of the aims and objectives of the project. The time management of the

project is discussed along with a discussion about future work. The chapter ends with a personal reflection on the whole project.

- Appendices: A collection of material to support the documentation provided in this thesis. The material includes Matlab files, impulse response, a Reaper project and other miscellaneous files.

Chapter 2

Literature Review

To effectively develop the system proposed in the project, it is imperative to carefully select appropriate techniques. This necessitates conducting a comprehensive review of the relevant literature, which is presented throughout this chapter, to understand the techniques and methods that will be most appropriate for the development of the VDR. As stated in the previous chapter, it is believed this system is the first of its kind so therefore there are no set ways on how to go about this. This is why this literature review is so vital as it will extrapolate and transfer methods and techniques from other research to inform this project and obtain the best result.

As well as providing a detailed literature review, this chapter also serves to inform on the background theory of the techniques of auralisation and its use in musical performance. Section 2.1 provides an overview of auralisation and the steps involved. Section 2.2 delves into the specific incorporation of auralisation for interactive musical performance, while also addressing the associated challenges. This section explores how auralisation techniques can be tailored to suit the unique requirements and constraints of interactive musical scenarios. In Section 2.3, two spotlight papers of particular relevance to this project are examined. These papers have played a significant role in informing and shaping the current stage of the project, contributing valuable insights and perspectives.

2.1 Simulating Room Acoustics

To develop a real-time auralisation system for drummers it is necessary to define and understand auralisation and the process behind how it is achieved as it underpins the whole methodology behind this project. This section begins with defining auralisation and the auralisation chain followed by further discussion of each step of the chain and the implementation of existing methods and techniques.

2.1.1 Auralisation

Auralisation is a technique used to simulate or recreate the auditory experience of a sound source in a specific environment. It is similar to visualisation, which creates visual representations, but auralisation focuses on sound. Auralisation allows people to hear what something would sound like before it is experienced. This is often performed using digital signal processing techniques such as convolution (see Sec 2.1.2.3).

2.1.2 Auralisation Chain

The auralisation chain can be considered as the following steps:

- Room Impulse Response (RIR) Measurement
- Capturing the Source Material
- Convolution
- Rendering and Reproduction

This chain is illustrated in Fig. 2.1. Each step is important to consider based on the desired outcome as it can significantly affect the result.

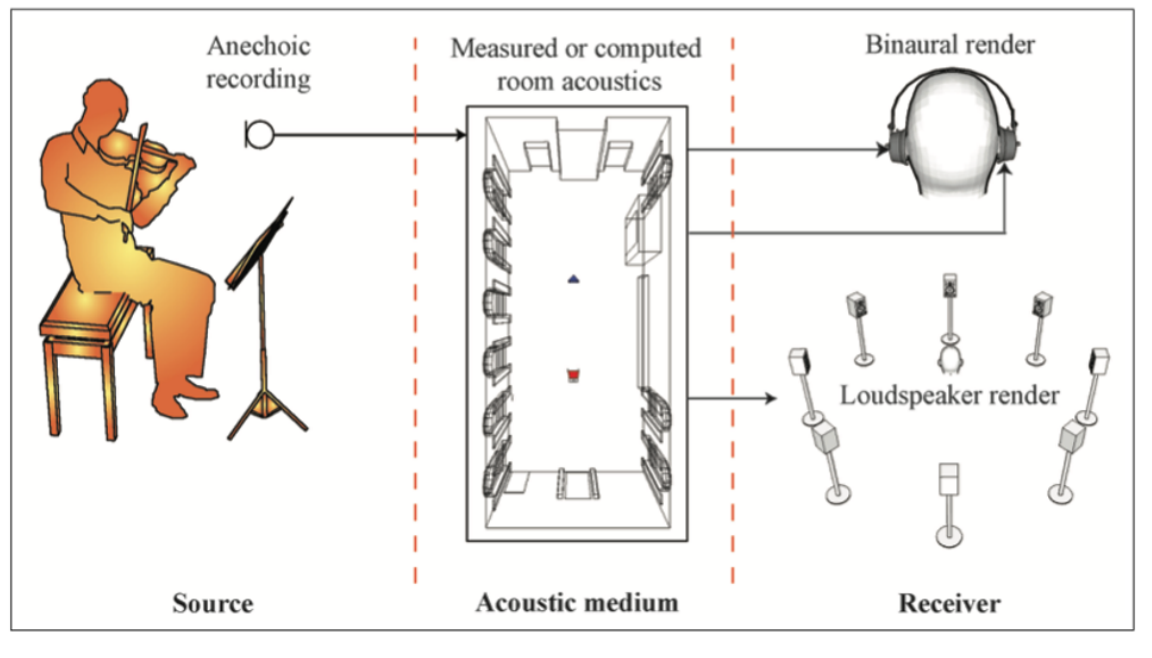


FIGURE 2.1: *"The auralisation concept: Source audio taken in an anechoic environment is filtered with measured or computed room acoustic responses, and the results rendered for headphone or loudspeaker listening."* [3]

2.1.2.1 Room Impulse Response Measurements

A Room Impulse Response (RIR) is the "acoustic fingerprint" of a room, representing how it responds to an excitation signal for a specific source and receiver combination. It can be obtained by either measuring it directly within the space or by simulation using computer software. To measure the SRIR, an audio excitation signal is played in the room and the response is recorded. The excitation signal used must have enough energy across the frequency spectrum to produce a reliable signal-to-noise ratio. Two main methods are available: direct measurements with impulsive excitation signals, or indirect measurements with wideband signals like noise or narrowband signals like sine sweeps or time-stretched pulses. An example of an RIR measurement using a sine sweep is shown in Fig. 2.2. Here a loudspeaker source on the left of the foreground and a microphone receiver on the right of the background can be seen. The receiver position represents the listening position and the source position represents the source material position of the virtual acoustic environment.

In the past, conventional methods relied on loud, sudden stimuli like gunshots, canons, or electrical sparks to capture room impulse responses. This approach eliminated the need for

FIGURE 2.2: An RIR measurement taken at the University of York Sports Centre. Image taken from OpenAIR [4]

subsequent adjustments since the recorded gunshots directly represented the room's impulse response in the time domain. However, these methods have become less popular primarily due to the non-uniform frequency response of such stimuli across the frequency range, making it challenging to ensure an adequate energy level of the signal.

Random or pseudo-random broadband noise signals emitted by a loudspeaker can be utilised for measuring RIRs. Once the noise source is turned off, the decay curve of the room's response can be measured. This approach allows for the extraction of time-based room acoustic parameters. However, it should be noted that energy-based parameters like Clarity (C80) cannot be accurately evaluated using this method. This limitation arises because the method does not generate an impulse response necessary for calculating the energy content across the frequency spectrum.

The exponential swept sine (ESS) technique enables the measurement of a system's impulse response while removing any distortion present in the system. Farina (2000) [5] developed this method and has proven to be reliable for various applications. To perform the measurement, a logarithmic sine sweep with a constant amplitude is synthesised, which increases exponentially in frequency per unit time. The sweep moves through the lower frequencies slowly but accelerates as it reaches higher frequencies, leading to a 6dB attenuation per octave in the spectrum. This sine sweep is then output into the room via loudspeaker, recorded, and deconvolved with a time-inversed filter of the original input. The inverse filter must account for the amplitude envelope of the input sweep, which means that an amplitude envelope needs to be applied to obtain an impulse response with a flat spectrum. The result is an impulse response with an initial delay that is equal to the length of the input signal [5], [6]. Any harmonic distortion present in the loudspeaker or recording equipment is now visible in the time domain and appears as a series of lower amplitude impulse responses before the main RIR. These can be easily excluded by editing out the initial time delay. The ESS technique has been widely used for in-situ room impulse response measurements in concert halls and other performance spaces [7], [5], [8], [9].

At present, resources like the Open Acoustic Impulse Response Library (OpenAIR) [4] offer access to a wide range of over 50 Room Impulse Responses (RIRs) that can be utilised for auralisation or acoustic analysis. However, it is important to note that these resources, along with similar ones, may not be suitable for projects requiring RIRs measured with specific methodologies. Consequently, for projects like the one at hand, resources such as OpenAIR are not suitable due to the need for RIRs obtained using precise and tailored methodologies.

RIR Source

Early methods of auralisation relied on omnidirectional point sources as the measurement source for the RIR measurement and monophonic anechoic recordings as the source material. However, these methods produce auralisations with little to no spatial or directional information for the listener and have been replaced by newer techniques that aim to capture and retain the directional properties of the music or speech being auralised. To achieve a successful and natural-sounding auralisation, it is essential to consider and replicate the directivity of the source material throughout the auralisation process both by how the source material is captured and how the RIRs are measured. It is advantageous to use a sound source in the RIR measurement stage that possesses radiation characteristics similar to the instrument being auralised [10].

Using an omnidirectional source for auralisations (such as a gunshot or electrical spark) can lead to erroneous reproductions in terms of measures of C80 and RT60. Low frequencies from a loudspeaker are almost omnidirectional, however, higher frequencies are much more directional by nature. This results in the room being less excited at higher frequencies and so the receiver will capture less reverberant energy at these higher frequencies. Wang & Vigeant (2004)'s [11] subjective testing revealed that highly directional sources were distinguished from those using an omnidirectional source when convolved with anechoic recordings of an instrument or voice.

Kessler (2005) [12] has broadly outlined a method of approximating source directivity by taking multiple spatial impulse response measurements at the same position, but at different loudspeaker rotations. An approximated directional characteristic can then be derived by simply exciting the room in different directions from the source position, capturing the RIRs in each source direction and combining the impulse responses in a given ratio for auralisation.

Kearney (2010) [3] has developed a technique where multiple RIRs are combined in appropriate proportions to approximate the directivity characteristics of the source sound. This method improved the subjective evaluation of virtual acoustic recordings. This research showed that including some form of source directivity information in the signal chain, even if simplified, such as averaged across perceptually relevant frequency bands, leads to higher naturalness ratings for auralisations from listeners.

The choice of the sound source for the RIR measurement and its placement in a room may vary depending on the individual instrument being auralised. To incorporate directivity information into the auralisation process, the RIR measurement stage can output the measurement source signal using a loudspeaker that replicates the source's directivity, such as the instrument or voice to be included in the auralised sound field. However, it has been found by Albrecht et al. (2005) [13] that the directivity characteristics of a natural musical instrument do not have to be completely identical to the original response, provided that the end listener is not familiar with every single aspect of the particular instrument's directivity. A close approximation is generally sufficient to create a plausible reproduction.

2.1.2.2 Capturing the Source Material

A vital aspect to consider is the content preparation, i.e. how the source audio is captured. Recording an acoustic event represents the first, and perhaps the most important step in any audio production chain and the same is true for recording for virtual acoustics. The sound source, in this case, the drums, must be accurately captured in real-time whilst avoiding unwarranted colouration of the signal from microphone placement. Monophonic source material can be used in auralisation but for complex sound sources (such as orchestras, music ensembles and percussion) a point source is inadequate [14]. Various approaches have been developed to address this issue including directional filtering of an omnidirectional source ([15]) and the use of multi-channel recordings ([16]).

Multi-channel auralisation involves representing a sound source as a point source but with the emission of distinct anechoic signals in different directions [17]. To achieve this, anechoic recordings must be made simultaneously using multiple microphones, and during auralisation, all of these recorded signals are utilised. In this technique, the directivities of instruments are modelled more accurately compared to techniques such as directional filtering. However, employing a large number of signals for rendering can become burdensome. For instance, using 20 channels per instrument and 50 instrument positions would result in 1000 signals being processed during auralisation. Nonetheless, as proposed by Vigeant et al. (2008) [18] and Otondo & Rindel (2005) [17], such a high number of channels is likely unnecessary. Alternatively, the computational load could potentially be reduced by combining nearby sound sources before applying them in the auralisation process.

Using close microphone techniques to capture the sound source allows for a higher level of direct sound compared to reverberant sound, enabling increased loudspeaker output levels while maintaining stability. However, this approach sacrifices the directivity of the sound source and may introduce colouration to the output signal ([19]).

Lokki & Pätynen (2009) [20] have recorded anechoic orchestral performances where each instrument is recorded individually, maintaining its radiation characteristics by use of a multi-channel technique. This work is particularly relevant for this project as they include percussion instruments in the orchestral performances, which were modelled as point sources. Lokki and

Pätynen go on to discuss the complex nature of the directivity of percussion instruments. Take, for instance, a tam-tam that exhibits a radiation pattern that changes over time. Certain modes of vibration are excited or dampened more rapidly than others, leading to a shifting directivity as the instrument vibrates freely following a strike. Furthermore, cymbals are typically spread apart and rotated after a forceful strike to prolong the audible decay for the listeners. These properties make it impractical to model directivity using time-invariant directional filters; instead, a multi-channel technique should be employed. The same principle holds true for timpani. Since they are often encountered in pairs or larger groups, their directivity is influenced by one another, making it unreasonable to model the directivity using filters. Despite Tam Tam and Timpani not being used in this project, they have similarities for the individual instruments of the drum kit that will be used here. Other authors who have used multi-channel recording techniques to preserve the source radiation characteristics of a single instrument, singer, or group of instrumentalists include Vigeant et al. (2006) [21], Vigeant et al. (2011) [22], Otondo et al. (2001) [15]).

Here there are cases for recording the direct sound of the instrument and further from the instrument. Both have advantages and disadvantages. However, a consideration is that because the simulation for this project is not entirely virtual, i.e. the drums are physically present in the room for the performer, perhaps generating the direct sound of the drums is not as important as that will already exist, so perhaps the anechoic method could work as it would generate the virtual reverberant sound rather than the direct sound that maybe is unnecessary.

Recent implementations of virtual auditory environments for performance have measured RIRs using directional sources. Rindel et al. (2004) [23] used a large number of microphones arranged around the instrumentalist and used the resulting directivity pattern in the sound source in a modelled auralisation, as shown in Fig. 2.7. Subjective testing showed that a directional sound source was preferred by listeners.

Anechoic Source Material

The majority of the studies previously mentioned utilize source material recorded in an anechoic environment. This is common practice as the use of the anechoic material avoids any unwanted reverberant properties in the signal which would ultimately impact the auralisations authenticity.

Iaird et al. (2011) [24], Amengual et al. (2016) [25] and Brereton (2014) [19] have investigated using non-anechoic source material as anechoic spaces can create inherent problems. These problems include limited accessibility as well as adversely affecting certain musicians' performance due to an absence of auditory feedback, such as singers, as discussed by Brereton (2014) [19]. To avoid unwanted room noise in non-anechoic source capture, close miking techniques are often used which increase the ratio of direct-to-reverberant. However, close-miking can lead to a reduction in the directivity of the sound source and can cause colouration of the output signal. Gorzel et al. (2010) [26] have shown that there are observable tonal distortion effects present in auralisations utilising a close, directional microphone technique.

2.1.2.3 Convolution

In the context of audio processing, convolution refers to a mathematical operation that combines two audio signals to produce a third signal. It is commonly used for various audio effects, such as reverberation, echo, and spatialisation. The process of convolution involves taking an audio signal, often referred to as the input signal or the source, and convolving it with an impulse response. The impulse response represents the characteristics of a particular audio effect or system. When the input signal is convolved with the impulse response, the resulting output signal reflects the effect of the impulse response on the original audio. Calculating the convolution of the RIR with a sound source can be done through either finite impulse response filters (FIR) in the time domain, or through Fourier transformations (FFT) in the frequency domain, as mentioned in [27], p138.

2.1.2.4 Rendering & Reproduction

To provide an auditory experience that simulates a real space or environment, the use of spatial audio reproduction becomes necessary. Spatial audio refers to technologies and techniques used to create and reproduce sound in a way that simulates the perception of sound in a three-dimensional space. It aims to recreate a realistic auditory experience by capturing and reproducing the spatial attributes of sound, including the direction, distance, and position of sound sources relative to the listener. When listening through headphones, this can be accomplished by generating Binaural Room Impulse Responses (BRIRs), which depict how the source interacts with the measured or simulated acoustic environment while considering the head and ear characteristics. For loudspeaker-based reproduction, Spatial Room Impulse Responses (SRIRs) will be measured to enable the reconstruction of sound pressure and velocity at specific listening points within the room using loudspeakers [3]. This is commonly done using Ambisonics which is a spatial audio technique used for capturing, encoding, and reproducing sound in a three-dimensional space. It allows for the creation of immersive audio experiences by representing sound as a spherical sound field rather than individual audio channels. The spatial sound field is represented using four coincident microphone signals, which are subsequently decoded to simulate the pressure and velocity components of the sound field at the central listening position [28], [29]. Ambisonics can also be used to produce reproductions for headphone-based auralisations by the use of decoder plug-ins and applying Head Related Transfer Functions (HRTFs). HRTFs can be fundamental to spatial audio processing, and simulate the way sound is filtered by the human head and ears before reaching the eardrums.

Spatial sound rendering involves a two-stage procedure: encoding audio signals to and decoding these signals for playback via loudspeakers or headphones. Some of the numerous spatial audio rendering techniques include Wavefield Synthesis (WFS), Ambisonics, Vector Based Amplitude Panning (VBAP) and Spatial Impulse Response Rendering (SIRR). Kearney conducted a comparison of various techniques for delivering audio to multiple listeners and determined that VBAP was effective for localizing stationary sources, while Ambisonics performed best in rendering moving sound sources [30], [3].

Ambisonics serves as both a rendering and reproduction method. It involves capturing the spatial sound field using four coincident microphone signals (A-format) and then translating it into B-format, which consists of four signals representing the omnidirectional sound field (W channel), as well as the front-back (X), left-right (Y), and up-down (Z) directions. A diagram of this is shown in Fig. 2.3. The four-channel B-format representation can be decoded for playback through headphones or various loudspeaker arrangements. Loudspeaker-based auralisation is not examined in this report due to the inherent problems of using it with acoustic drums (see Section 2.2.4.3 for more on this point).

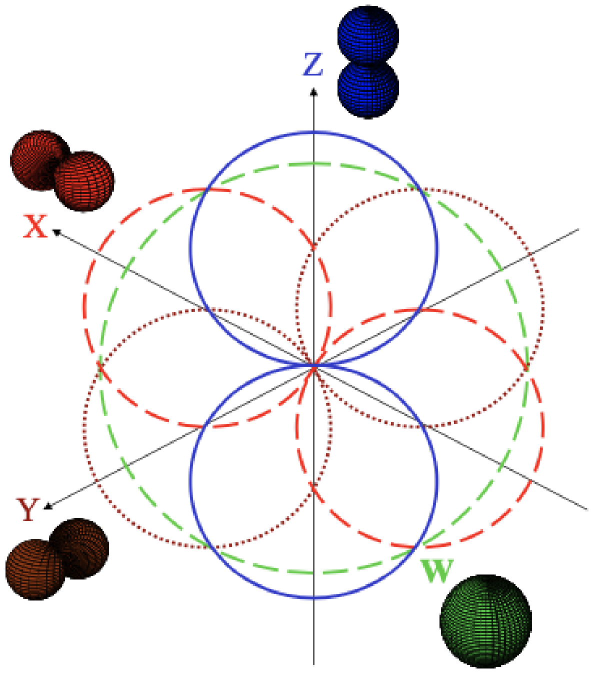


FIGURE 2.3: "*Ambisonic B-Format soundfield representation. X, Y, and Z represent the velocity components in the x, y, and z directions respectively and W represents the omnidirectional pressure component.)*" [3]

Reproduction Methods

Headphone playback is a practical option for mobile applications, offering convenience and portability. However, achieving convincing binaural playback necessitates the careful generation of ear signals. The perception of spatial sound can be achieved by convolving a source signal with corresponding head-related impulse responses (HRIR) or binaural room impulse responses (BRIRs). Utilising the loudspeaker feeds as source signals enables the recreation of ear signals from any loudspeaker arrangement within a room. To enhance localisation and plausibility, head-tracking is employed, allowing the system to track the listener's head movements [31]. Incorporating head movements, though, requires the sophisticated interpolation of HRIRs/BRIRs [32]. Alternatively, using Ambisonics for headphone playback offers a straightforward approach to incorporate head movements through rotation matrices [33, 34]. This approach provides the added benefit of incorporating head-tracking, which can render the acoustic sound scene static to the listener and improve the immersion of the virtual acoustics experience.

The advancement of binaural decoding for headphones has been driven by notable contributions from researchers in the field. Building upon Jot's work in the 1990s, Sun, Ben-Hur, and Brinkmann have made significant progress in addressing fundamental issues [35–37]. The TAC and MagLS decoders, developed by Zaunschirm and Schörkhuber, have emerged as crucial advancements in this domain, effectively tackling challenges associated with HRTF delays and optimizing HRTF phases at high frequencies to avoid spectral artefacts [38, 39]. Through the application of interaural covariance correction, the MagLS/TAC decoders consistently reproduce diffuse fields, incorporating the formalism introduced by Vilkamo et al. [40].

While obtaining personalised head-related impulse responses (HRIRs) can be laborious for most individuals, online databases such as the SADIE II DATABASE [41] provide a collection of freely available HRIRs for download. Although HRIRs are typically specific to each person, having a wider selection allows users to customize their listening experience. With advancements in technology, obtaining HRIRs for a larger number of individuals may become more accessible and efficient in the future.

Colouration

In the context of audio reproduction, colouration refers to any alteration or distortion of the original sound. It involves changes in the frequency response or other characteristics of the audio signal, resulting in a modified perception of the sound. This can occur from the microphones, loudspeakers and headphones used throughout the auralisation chain. These modifications can introduce tonal shifts, imbalances, or other undesirable effects that deviate from the true or natural representation of the sound. Colouration is often used to describe any non-linear or frequency-dependent changes that impact the overall quality or accuracy of the reproduced audio.

Equalization of Convolution Chain

It is important to mitigate the colouration that they can cause by using inverse filters. An inverse filter, also known as a deconvolution filter, is a signal processing technique used to reverse or compensate for the effects of a previous filter applied to a signal. It aims to undo the alterations or distortions introduced by the original filter, thereby restoring the signal to its original form or compensating for specific frequency response characteristics. The inverse filter is designed based on the known frequency response or impulse response of the original filter. By applying the inverse filter to the filtered signal, the goal is to cancel out the frequency-dependent modifications and restore the signal's original spectral content or other desired characteristics.

2.1.3 Summary

In this subsection, auralisation and the auralisation chain has been defined as well as a discussion of each step of the chain looking at implementations of existing methods and techniques. The academic literature reviewed in this section highlights the importance of directivity modelling to increase the realism of the auralisation and has also highlighted the benefits of using multi-channel auralisation to achieve this.

2.2 Simulating Room Acoustics for Musical Performance

In this section "real-time" auralisation for interactive room aocustic simulation for musical performance will be investigated, which is vital as it is the tehcnology that underpins this project. There are certain challenges and adaptations regarding the auralisation chain that surround real-time auralisations which will need to be considered in this project and will be discussed in this section.

2.2.1 Virtual Acoustic Environments

Techniques used in room acoustic simulation are closely related to those used in the production of virtual acoustic environments, of which auralisation is a subset. In literature, the terms auralisation and virtual acoustics are often used interchangeably, however Savioja et al. (1999) [14] use the term "Virtual acoustics" specifically to cover the modelling of three main aspects of acoustic communication: 1) source, 2) transmission medium (room) and 3) receiver (listener). They suggest the term "auralisation" is a subset of virtual acoustics, namely the modelling and reproduction of sound fields ([14], p675). Hence, auralisation involves stages 2 and 3 of VAE implementation. To create a live interactive simulation of the room acoustics for musicians, auralisation methods will be used, but some adjustments are necessary at various stages in the auralisation process.

2.2.2 Interactive Room Acoustic Simulations

Many studies discussed above utilize auralisation in a "non-real-time" manner, where the audio material is processed in advance of being presented to the listener. However, there is growing interest in the use of "real-time" auralisation within interactive virtual acoustic environments, for both the listener and performer to investigate subjective responses to virtual acoustic displays. Savioja et al. (199) [14] have incorporated interaction into a virtual acoustic environment (VAE) that permits a listener to change their listening position within a simulated concert hall. Other forms of interactivity involve enabling the listener to physically move around within a room auralisation, as seen in studies by Lokki et al. (2007) ([42] and Kajastila (2007) n[43]).

In recent times, there has been a growing interest in the development of real-time room acoustic simulations. Researchers such as Buchholz & Favrot (2010) [44] and Favrot (2010) [45] have successfully implemented loudspeaker-based room acoustics simulations, primarily for the purpose of advancing auditory perception research. In some cases, these simulations have incorporated interactive features that allow performers to hear themselves as if they were in an actual performance space. For instance, Ueno & Tachibana (2011) [46] have created a real-time acoustic environment with six channels specifically designed for instrumentalists, building upon their previous work in Ueno & Tachibana (2003) [47]. Similarly, Kalkandjiev & Weinzierl (2013a) [7] and Kalkandjiev & Weinzierl (2013b) [48] have focused on simulating stage acoustics for solo cellists. Furthermore, Woszczyk et al. (2010) [49] and Woszczyk & Martens (2008) [50] have developed virtual stage acoustics to enhance the existing acoustics of concert hall venues, thereby assisting instrumentalists during performances or recreating the acoustic characteristics of historical venues.

2.2.3 Interactive Room Acoustic Simulations for Performance

A VAE requires three components that must be modelled, namely the source, the medium and the receiver, whereas an interactive virtual acoustic for real-time performance always includes a real source and receiver. Therefore, only the room acoustic characteristics need to be modelled, whether that be from measured or synthetic RIRs. Because of this, these real-time performance systems can be thought of as a subset of virtual acoustics, Brereton (2014) [19] suggests referring to this subset as “Real-time Room Acoustic Simulation”.

A Real-time Room Acoustic Simulation (RRAS) for musical performance requires interactivity of sound, meaning the listener is able to make a sound and hear back the response of the simulated room in real time. It is important to acknowledge that even with the use of advanced technology, there will always be a certain amount of delay inherent in the processing and sound rendering, which means that the system cannot operate strictly in real time. In this context, the term ”real-time” is employed to differentiate from ”off-line” processing, and includes systems that work in ”pseudo-real-time”. This means that the system operates with minimal latency times, which are either not perceptible or tolerated by the users of the system. More on this in Section 2.2.4.1.

Real-time room acoustics simulations for performance differ from non-real-time methods, in that the performer’s sound is used as the sound source and must be convolved with the SRIR of the simulated space in real-time or as close to it as possible. To achieve this, the SRIR must be modelled or measured from the position of the performer in the space, rather than from a listening position in the audience as is typically done. In Brereton (2014)’s [19] work, this meant that the source and receiver positions should be co-located as the source was the singing voice.

With interactive simulations for performance, the source is live and present within the room the performer is situated and so the direct sound is already present. Therefore with RRASs, only the reverberant sound needs to be modelled for the performer. This will dictate how the musical source is captured and modelled for SRIR measurement.

As previously mentioned, it would be ideal for the radiation patterns of the measurement source to match those of the actual source during real-time auralisation. An example of this is the study by Martens and Woszczyk where they ”modelled the source directivity during the RIR measurement in historic concert halls by using a group of omnidirectional loudspeakers in order to simulate the complex directional radiation of the pianoforte. During their subsequent performance experiment in the VAE, a spaced microphone array was used to capture the sound of the pianoforte which was then convolved with the measured RIR.” ([51]).

Recent implementations of virtual auditory environments for performance have measured SRIRs using directional sources. Rindel et al. (2004) [23] used a large number of microphones arranged around the instrumentalist and used the resulting directivity pattern in the sound source in a modelled auralisation, as shown in Fig. 2.7. Subjective testing showed that a directional sound source was preferred by listeners.

A growing body of work has encompassed the use of real-time acoustic simulation for a number of applications. Some researchers have used it to gather subjective responses from musicians for different room acoustics e.g. [2], [50], [52] [53] and [54]. A few research groups use measured RIRs to replicate a musician’s location on stage or in the hall, with the aim of presenting an interactive acoustic environment or room acoustic simulation for the performer. These often lead to the construction of a "Virtual Performance Studio" (VPS) which enables a musician to perform in a virtual acoustic simulation of a real performance venue e.g. [19], [25] and [24]. These particular VPSs are constructed with a sphere of speakers placed around the musician. The instrument is captured and auralised in real-time with real or modelled SRIRs and then fed back through the loudspeaker array. An example of the set-up can be seen in Fig. 2.4 and 2.5.

Amengual et al. (2016) [25] used real-time auralisation technology to reduce the logistical and procedural complexity for studying the influence of room acoustics on musicians as well as aiming to provide a framework for the study of live music performance in different acoustic conditions. Iaird et al. (2011) [24] used the technology as an opportunity for musicians to get accustomed to the performance space and adapt their technique during rehearsals instead of doing it on stage in front of an audience. They mention the several research avenues coming from this, including exploring stage fright among musicians and potential applications in concert hall stage design.

While this project may not strictly fall under the category of a VPS due to the fact that it simulates a variety of performance environments that may not always conform to conventional standards, the fundamental concept remains consistent.

2.2.4 Considerations

Creating a real-time room acoustic simulation with source interactivity for performance requires careful consideration at every stage of the auralisation chain. Various challenges are encountered during the implementation process, which are discussed in the following sections. For this project, the particular focus is on: 1) investigating the source location and directivity of the SRIR measurement and 2) source recording methods for the drums. So far, the interest and previous research around RRASs have been highlighted, however, it appears that there is little research into these systems for drums. Therefore, the decisions that will be taken on how to record the drums as well as the location and positioning of the source for the SRIR measurements need to be understood and supported based on an extrapolation of methods that have been outlined in similar research.

FIGURE 2.4: A photograph of Amengual et al. (2016)'s [25] virtual acoustic simulation rig with trumpet player performing.

FIGURE 2.5: "*Photograph of the Virtual Singing Studio as set up in the acoustically treated listening room (KEMAR denotes position singer would take)*" [19]

2.2.4.1 Real-time Convolution and Latency

The convolution of the source sound with the simulated room impulse response (SRIR) can be performed either in the time or frequency domain for "offline" processing. Frequency domain processing offers quicker results, but both methods can be computationally demanding, necessitating an alternative approach for real-time applications. In real-time convolution, the input signal is divided into frames and processed sequentially, ensuring low latency to maintain the perception of timely early reflections and reverberant sound. The processed results are output in a sequential manner [55].

In contrast to the "off-line" auralisation discussed previously, real-time convolution requires the source sound to be convolved with the RIR of the simulated room in real-time, achieved through hardware or software applications like a VST plug-in within a Digital Audio Workstation. It is crucial to minimize latency to avoid perceptible delays in early reflections and reverberation. According to studies by Miller et al. (2003) [56], the minimum perceptible end-to-end latency for a virtual audio system during head movements, assuming a source velocity of 180°/s, is approximately 70 ms. They suggest that this threshold applies to various types of source-listener motion, even when the source remains fixed while the listener moves. Wenzel (1998) [57] proposes a method termed Total System Latency (TSL) to measure end-to-end latency, anticipating thresholds of 92 ms, 69 ms, and 59 ms for slow, moderate, and fast-moving sources, respectively. Chafe et al. (2004) [58] demonstrated that musicians successfully synchronized their clapping with a partner's delayed sound up to time delays of 77 ms. However, delays exceeding 14 ms resulted in a deceleration of tempo, while short delays up to 11.5 ms aided in stabilizing the tempo.

Drummers, due to the inherent haptic nature of their instrument, exhibit a heightened sensitivity to latency. With their physical connection to the drums, while playing, they possess a natural inclination towards rhythm and time, accompanied by notable time-keeping abilities. Consequently, drummers are more likely to be acutely aware of latency issues, rendering them more sensitive to such delays compared to singers, for instance.

2.2.4.2 RIR editing

When employing a real-time room auralisation technique, whether using synthetic or measured RIRs, it is necessary to edit the RIR by eliminating the direct sound component as this will be present in the room from the source (the drums in this case). During this editing process, it is crucial to maintain the Initial Time Delay Gap (ITDG), which refers to the time interval between the direct sound and the first reflection. Additionally, if the listening room where the RIR is presented includes a floor, it is important to remove the first-floor reflection as well ([59], [46]).

Ueno & Tachibana (2003) [47] and Ueno et al. (2010) [2] introduced a room auralisation system for musicians within a six-sided anechoic room. However, they acknowledged the challenge of simulating the first-floor reflection, especially since the distance between the musician and the floor in a real performance venue could be as close as 1.5 meters (or even closer if seated), which is much shorter than the distance to the loudspeakers positioned beneath the mesh floor in the anechoic chamber.

2.2.4.3 Rendering and Reproduction

The aforementioned studies have generally been equal in weighting regarding the reproduction class used (i.e. over headphones or loudspeakers). Both classes show advantages and disadvantages. However, the studies that incorporate real-time convolution seem to favour loudspeaker reproduction class compared with headphones. Perhaps this is due to the fact that headphones can reduce the direct sound of the instrument and therefore affect how much the performer is able to hear themselves. This is particularly true with singers as Brereton (2014) [19] explains "*Most singers report that wearing headphones alters the balance of bone-conducted and airborne aural feedback, which can be detrimental to their singing performance.*" Brereton also drew on research from Libeaux et al. (2007) [60] where they reported that they preferred the virtual acoustic environment to be presented over an array of loudspeakers rather than binaural rendering over headphones. Eley et al. (2021) [61] compared singers' experience using headphones and loudspeakers in a virtual acoustic environment of Notre Dame. The general preference for the musicians was for the multichannel loudspeaker system rather than binaural headphones although one singer had mixed feelings and noted that the acoustic feedback from the headphones sounded more realistic but they found it more difficult to hear themselves.

In the context of individual listening, headphone playback is preferred as it allows for disregarding the acoustics of the listening environment. Nevertheless, achieving high-quality 3D audio reproduction through headphones presents challenges. The use of headphones introduces timbral colouration in the resulting auralisation, particularly noticeable at higher frequencies. This colouration occurs due to variations in the headphone response caused by minor positional adjustments each time the headset is worn [62], [63].

2.2.4.4 Musician Movement

Musicians naturally exhibit movement during performances, serving as a means of expression and personal comfort. Depending on the instrument type, musicians either synchronise their movements with the instrument, as seen with trumpeters or violinists or move independently from it, as observed in drummers or pianists. Real-time auralisation encounters various challenges associated with musician movement, which vary depending on the instrument being played. The auralisation is simulated precisely for the receiver position used in the SRIR measurement. For drummers, a significant issue arises from the fixed positioning of the receiver in SRIR measurements, specifically aligned with the average height of a seated drummer's head from the ground. However, this average height fails to accurately represent the head height of individual drummers, as some may have lower or higher positioning.

Of greater importance is the dynamic nature of a drummer's head and body movements during performance. To address head rotations, the incorporation of head tracking techniques becomes essential. When an Ambisonic microphone serves as the receiver, compensation for rotation around this point becomes feasible, and incorporating head tracking technology allows for headphone reproduction that maintains a static virtual acoustic environment (refer to Section 2.1.2.4). This implementation creates an auditory illusion for the listener, wherein the simulated sound field remains static and unaffected by the drummer's head rotations.

However, it is crucial to note that head tracking solely accounts for rotational movements around the receiver's axis. If the drummer moves their head away from the receiver's initial position, the system no longer accurately simulates the sound of the room from the drummer's new head position. Unfortunately, effectively addressing this issue necessitates a substantial number of SRIR measurement points, which presents a considerable challenge. As a result, when the drummer moves excessively far from the receiver position, the realism of the simulation may be diminished. Regrettably, this occurrence is likely to transpire, thus constituting a current limitation that necessitates acceptance and adaptation.

2.2.5 Summary

In this subsection, the literature associated with real-time auralisation has been reviewed and the practical challenges associated with interactive room acoustic simulation for musical performance have been examined. Techniques to mitigate some of these challenges have been identified. However, while there are many cases of successful implementation of real-time auralisation, there is little study on its application for the drums. Therefore, specific techniques for SRIR measurement and source material capture have not yet been fully identified.

2.3 Spotlight Papers

In this section, two academic papers will be reviewed that have been of particular significance in this literature review. Otondo & Rindel (2005) [17] provides inspiration for methodologies for the SRIR measurement and source material capture that will be adapted for the development of the system proposed in this project. Brereton (2014) [19] provides valuable insights into other research discussed throughout this literature review as well as providing a detailed account of a VPS that has highlighted many of the challenges associated with real-time auralisation as previously discussed.

2.3.1 Otondo & Rindel (2005)

The directional properties of sound sources have been identified as a crucial factor to consider in auralisation, as evidenced by previous studies [64]. The directivity pattern of musical instruments, known for its complexity, gives rise to a distinct acoustic response in a room that varies across the range of pitches played ([17], [65], [15]). This is illustrated in Fig. 2.6.

In conventional auralisations, the directivity of musical instruments is typically represented as a constant value across their entire range of pitches. This fixed directivity is derived from average measurements [66], and a single recording is used for sound radiation. However, a study focusing on room auralisations has revealed that this representation is insufficient in capturing the variations in radiation patterns when different tones are played on different musical instruments [10].

Taking this as motivation, Otondo & Rindel created objectives for their study that were to enhance the fidelity of musical instrument representation in auralisations by considering the dynamic changes in instrument radiation. Their aim was to provide a more realistic and immersive auditory experience by designing a recording and reproduction method that effectively incorporates the spatial and tonal contributions of a sound source as perceived by a listener in a room.

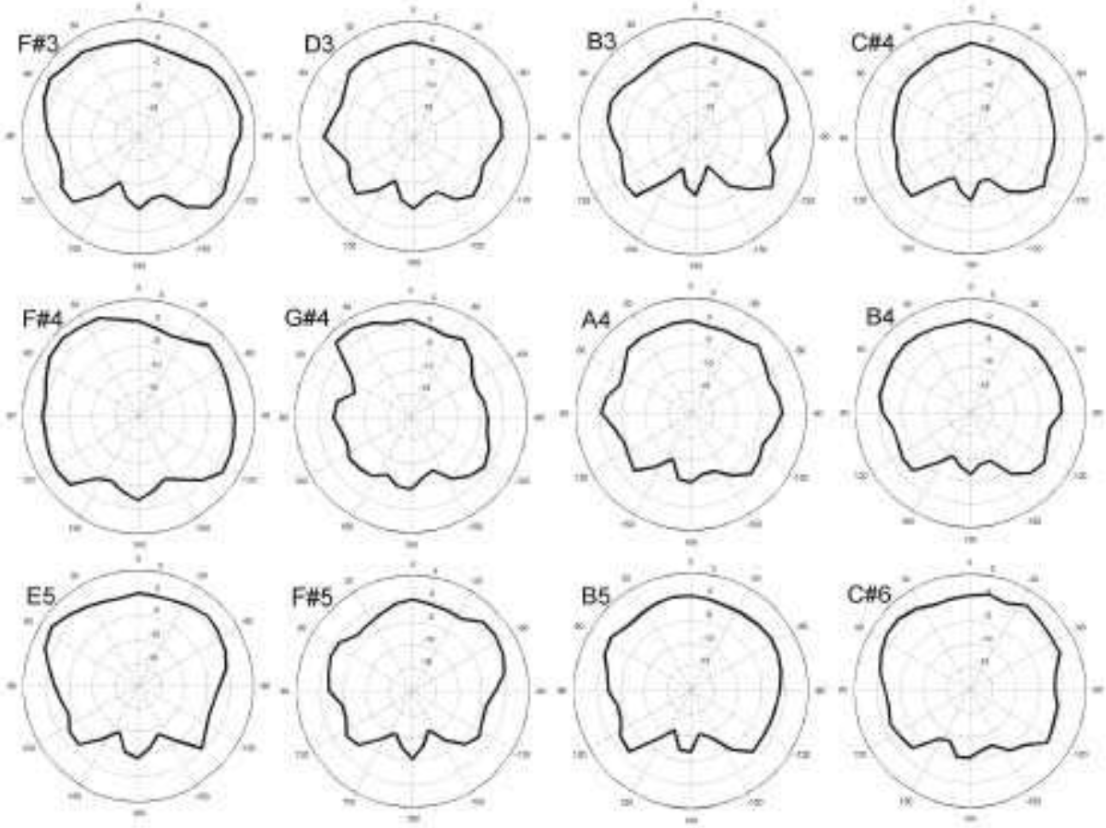


FIGURE 2.6: "*Evolution of the directivity of a Bb clarinet in the horizontal plane for different tones in three musical octaves measured in the 1kHz octave band. The instrument is facing upwards in the plots.*" From [17]

In traditional room auralisations, a musical instrument's source representation typically involves a fixed directional characteristic and a monophonic recording. Otondo & Rindel discuss that a more accurate depiction of the instrument's spatial sonic qualities can be achieved by incorporating multiple samples of the sound field generated by the source, which can then be utilised in the reproduction process. They say that one approach to capturing the instrument's radiated sound in various directions is to conduct simultaneous anechoic recordings using microphones positioned around and above the source. The recorded sound in different tracks will contain spatial information about the source, including any asymmetries in the instrument, performer movements, and variations in radiation across different tones (now known as multi-channel auralisation).

Otondo & Rindel measured the directivity of different musical instruments in a performance situation and compared it to the averaged directivity that is traditionally used for auralisation purposes. The two directivity representations were compared through means of auralisation in order to evaluate the perceptual importance of changes in radiation in a room. The outcomes of their listening tests revealed discernible variations in the sound produced when employing different representations of directivity, which were noticeable to the listeners. These findings indicated that the conventional fixed directivity representation in room acoustic software would result in a diminished sense of presence in the representation of sound sources and potentially lead to a lack of realism in the reproduced sound. They found that the directivity measurements of musical instruments in a performance situation have proved that significant changes occur in the directivity in the horizontal and vertical planes.

Otondo & Rindel proposed a method of capturing sound sources that vary in directivity in time in auralisations. The method incorporates the use of multi-channel recordings of the instrument which are then played by a virtual source in the auralisation, according to the original positions of the recordings. An example of this can be seen in Fig. 2.7 in which a four-channel anechoic recording of an instrument is used in a room auralisation with a compound source consisting of four virtual sources. Each virtual source depicted in the diagram exhibits omnidirectional behaviour within a quarter-sphere range, emitting sound waves towards the 0, 90, 180, and 270-degree directions on the horizontal plane. When combined, these four virtual sources form a composite source that emits sound uniquely in each of the four directions. This distinctive radiation pattern influenced by variations in level, movements, asymmetries, and orientation of the original source is captured by the individual microphones. The synthesised RIRs were captured for an audience listening position.

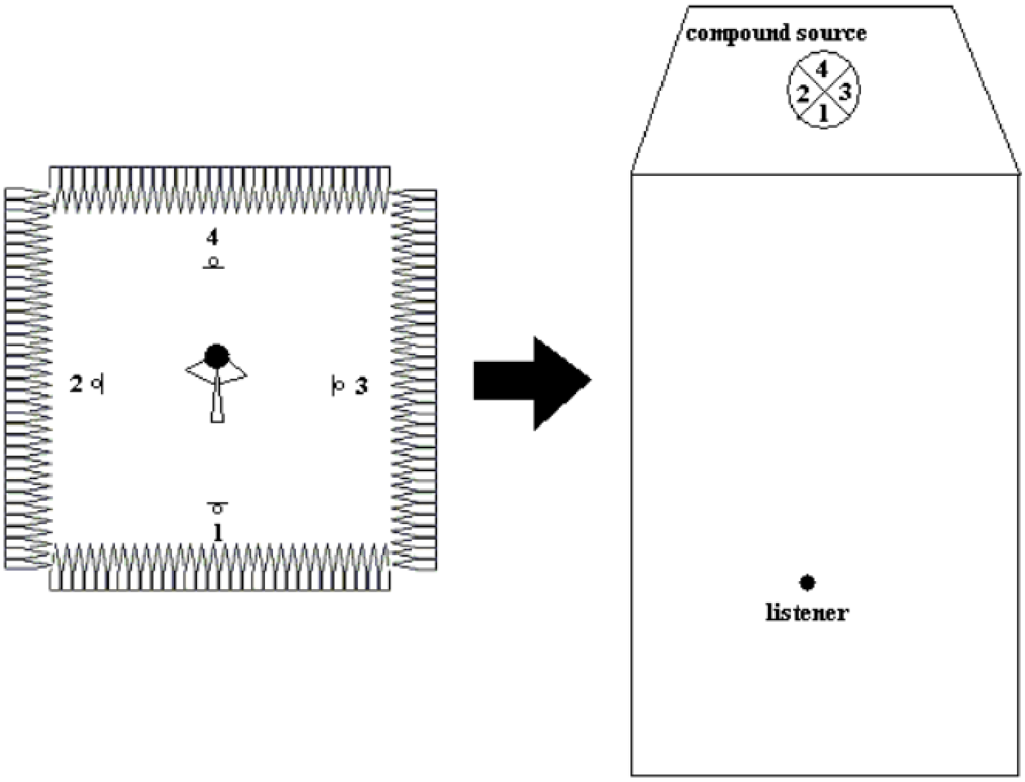


FIGURE 2.7: *"Horizontal view of a four-channel anechoic recording of a musical instrument and the reproduction method in a room auralisation with a compound source representation using four virtual sources.)"* From [17]

To validate the method, a listening experiment was carried out that compared auralisations using different microphone configurations of the setup in order to test which configuration would give the best sound representation of the source. One configuration for a clarinet using a 13-channel recording set-up can be seen in Fig. 2.8. The auralisations for the experiments involved the use of four distinct directivity representations. One representation utilized the averaged directivity of the clarinet, while the other three representations employed the proposed method with varying numbers of sources. These representations were utilized for comparisons in the experiments.

The assessment of the system has demonstrated a noticeable enhancement in the audibility and naturalness of timbre in auralisations when employing the suggested method. However, the outcomes concerning spaciousness lacked consistency. Listeners encountered challenges in perceiving spaciousness and seemed to associate more with reverberation, similar to findings

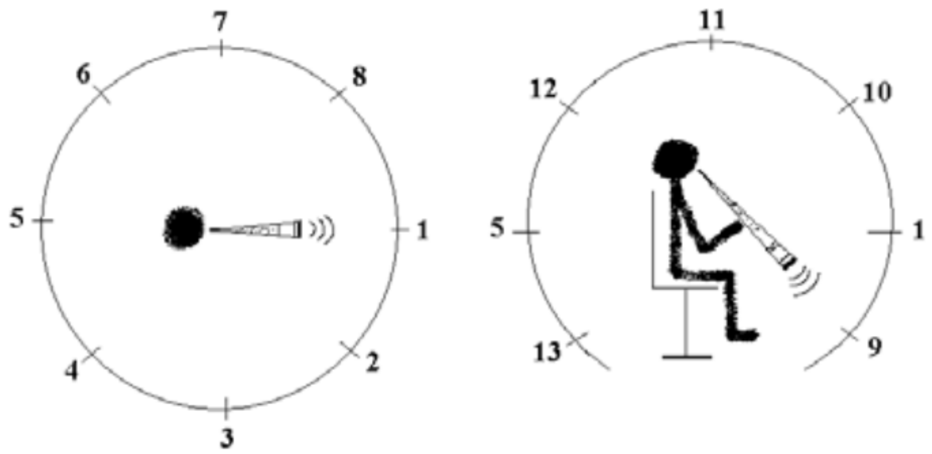


FIGURE 2.8: "*Setup for the multi-channel recordings and directivity measurements with 13 microphones. The left part of the figure shows the setup in the horizontal plane and the right part shows it in the vertical plane. Microphones 1 and 5 appear in both planes*" From [17]

from previous studies on auralisations. As the subjects compared various configurations within the system, they observed an enhancement in the naturalness of timbre quality as the number of sources employed in the system increased. Considering the results of the listening experiments used to validate the proposed method, it was concluded that this approach significantly improves the quality of reproduced sound. It faithfully represents the sound of the source as perceived by listeners, surpassing traditional auralisation methods that rely on single monophonic recordings and a fixed representation of source directivity.

Ontondo & Rindel mention that the efficacy of this method relies on utilizing the original sounds emitted by instruments rather than a theoretical approach to their radiation. It can be inferred that tonal directional instruments are better suited for this method compared to instruments with wideband signals, transients, and a directivity closer to that of an omnidirectional source.

2.3.2 Brereton (2014)

Brereton (2014) [19] primary objective of her research was to examine how singing performances are perceived in actual acoustic environments and room acoustic simulations, with a specific emphasis on the perceived similarities as assessed by listeners. To investigate these performances, an interactive real-time simulation of a performance space with realistic room acoustics was created and evaluated, known as the Virtual Singing Studio (VSS), as shown in 2.5. Professional singers were involved in the evaluation process to determine if the simulation was capable of eliciting changes in their singing performance, similar to what would be expected in different real-life performance spaces with varying acoustic conditions.

The signal chain involved in the implementation of the VSS is illustrated in Fig. 2.9. The singer's voice was captured by a head-worn DPA4066 microphone and was converted into digital format via the RME Fireface800 external soundcard. The voice signal was then convolved over four channels in the Reaper Digital Audio Workstation with the measured SRIR. The reverberant sound was then decoded for Ambisonic presentation over a 3-dimensional array of loudspeakers. Brereton used a loudspeaker reproduction method as opposed to headphones because "*Most singers report that wearing headphones alters the balance of bone-conducted and airborne aural feedback, which can be detrimental to their singing performance.*" Brereton also drew on research from Libeaux et al. (2007) [60] where they reported that they preferred the virtual acoustic environment to be presented over an array of loudspeakers rather than binaural rendering over headphones.

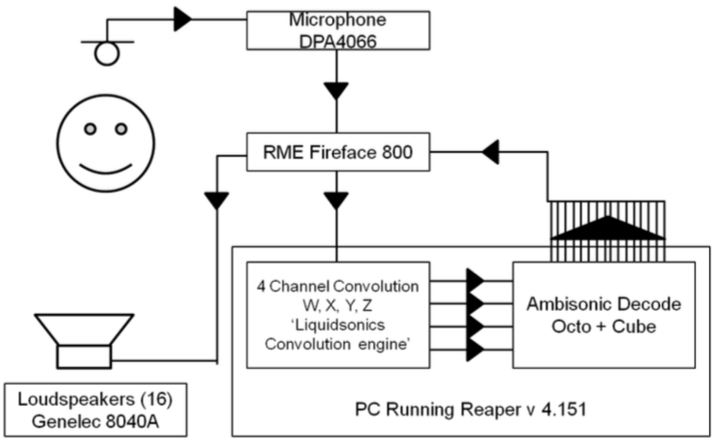


FIGURE 2.9: "*Graphical representation of the processing chain involved in the VSS*" From [19]

The B-format room impulse responses were measured in the real performance space at a sampling rate of 96 kHz using the ESS methods developed by Farina (2000) [5] (as described in Section 2.1.2.1). Brereton incorporated directivity into the SRIR measurements by using a Genelec 8040 loudspeaker as the source. The SRIRs were captured to specifically emulate the performer's experience i.e. the source and receiver co-located as seen in Fig. 2.10.

FIGURE 2.10: "*Position of Soundfield Microphone placed above the Genelec 8040 loudspeaker used to measure performer position SRIRs in the real performance space*" From [19]

The close mic technique of the voice provided the advantage of a high ratio of direct-to-reverberant sound which reduced the effects of acoustic feedback as the system was reproduced over loudspeakers. Different miking techniques were investigated, with an overhead microphone placement being a strong contender, however, Brereton reported that this method resulted in singers perceiving "*the simulated reverberant field to be sharper in pitch than expected.*" [19] - an obvious hindrance for singers but perhaps not for drummers. However, Brereton reported certain drawbacks associated with the use of a head-mounted microphone which is important to consider for this project. The head-mounted microphone resulted in the loss of directivity information in the signal. Therefore the voice is in effect "spatially sampled" and treated as a point source within the auralisation and hence loses its directivity pattern. This can cause a mismatch between the directivity pattern of the simulated room reflections and reverberated sound and that of the singing voice. Brereton suggested that the directivity of the Genelec 8040a loudspeaker used is similar to that of the singing voice, which would minimize the negative impact of this mismatch. Brereton goes on to discuss future improvements to the VSS involving developing the real-time implementation of methods proposed by Vigeant et al. (2007) [16] and Kearney (2010) [3] to capture and replicate the directivity of the singing voice in the simulation.

2.4 Conclusion

After conducting a comprehensive literature review, valuable insights have been gained that will guide the design and implementation of the VDR presented in the following chapter. To facilitate the development of the VDR, a thorough review of the literature was undertaken, presented in this chapter, which will inform the design that the VDR will be built around. This literature review has investigated simulating room acoustics through auralisation and outlined the definition and process of this technology. It has specifically examined interactive room acoustic simulation for musical performance through real-time auralisation and discussed the technologies and considerations associated. Of particular interest was investigating the methodologies for capturing the source material (the drums) and measuring the associated SRIRs. The literature review highlighted the importance of capturing and replicating the directivity of an instrument in order to create an immersive and realistic virtual acoustic experience. The literature review has also highlighted that such methods in the context of auralisation for drums/percussion are significantly lacking, particularly for real-time systems of which none were discovered. Despite Otondo & Rindel remarking that their methodologies for capturing and replicating the directivity of an instrument (in their case a clarinet) are not suited to instruments with a directivity closer to that of an omnidirectional source (such as the drums), their methodologies (presented in Section 2.3.1) are deemed to be the most appropriate for the aims of the system proposed in this project. Therefore their work will inspire the methodologies for source capture and SRIR measurement outlined in the next chapter.

Chapter 3

Design & Implementation

This chapter will provide a detailed account of the design and implementation of the Virtual Drum Room (VDR). The chapter is split into two main sections, Section 3.2 and Section 3.3, which both explain the implementation of the initial testing phase and the final build of the VDR respectively. The two phases drew from the same design methodology of the VDR which was built around the auralisation chain discussed in Chapter 2. The main difference in the implementation of the VDR for the initial testing and final build was the number of microphones employed. The initial testing phase comprised a 2-channel system and the final build comprised a 4-channel system. In terms of the software, hardware and set-up in the anechoic chamber, much was kept the same from the initial testing phase into the final build. Any discrepancies are noted.

3.1 Design

The methodology used for the VDR draws inspiration from the work of Otondo & Rindel (2005) [17], as discussed in Section 2.3.1. Although they stated that their proposed method was likely not suitable for instruments with a directivity closer to that of an omnidirectional source, of which the drum kit is, it is believed to be the most suitable methodology for this project. A significant factor in creating a system that includes spatial information and a level of immersion requires directivity to be employed in the auralisation methodology. This method allows for an inclusive and non-biased representation and employment of the directivity of the drum kit. Despite there being few significant disparities in the methodology used for this project and the methodology used by Otondo & Rindel such as the listening position, the nature of the instrument and the type of SRIRs obtained (real vs. modelled), their methodology is still considered the most appropriate for this project and predicted to work just as effectively despite the disparities.

Although methodologies for this project have not significantly been drawn from Brereton (2014) [19], the reduced directivity from the close miking techniques further supports the notion of using an alternative method here. Additionally, the drum kit is a complex and multi-voice instrument and its directivity can not be reproduced by the implementation of a loudspeaker, such as Brereton did. Therefore multichannel auralisation is necessary to capture the wideband signal of the drum kit.

Source Capture

The drum kit will be situated in an anechoic chamber to capture the drum sounds for this project as it allows for experimentation with a wider selection of microphone configurations. A non-anechoic environment would warrant close miking configurations which would limit the directivity of the auralisations and complicate the SRIR source measurements, as discussed in Section 2.1.2.2.

The multi-channel auralisation technique for the final build will ultimately involve a setup of four microphones. It would have been preferable to use configurations consisting of 4, 6, and 8-channel microphone arrangements in order to investigate the effect of the number of microphones used and their impact on the VDRs performance and listener preference. However, because the size and logistical problems with the anechoic chamber to be used in this project, it means that only a maximum of 4 microphones will be used.

SRIR Measurement

The recording of SRIRs will be performed through actual measurements conducted in real rooms, rather than relying solely on modelling software. This is inspired by the intention to capture and reproduce the authentic acoustics of real environments as well as experiment with methodologies of SRIR measurements in real places specifically. Nevertheless, it does not exclude the possibility of incorporating modelled responses in future iterations. Such an addition could prove advantageous for this project, as it would provide a less time-consuming method for remodelling sound and receiver positions.

Considering the constraints of alternative methods, ESS has been selected as the preferred approach for this project. This choice is grounded not only in its benefits for editing SRIRs but also in its control over directivity modelling. Fixed positions can be established, allowing for precise measurements and enhanced control. This methodology further improves the accuracy of reproducing measurements across various rooms in future research.

SRIR Source & Receiver Positioning

To obtain the SRIRs for the final build, the receiver will be placed at the drummer's head position when sitting at the drum kit. The microphone used will be an Ambisonic microphone in order to allow for head tracking and the possibility of applying personalised HRTFs for different drummers in the reproduction stage. The source loudspeakers will be precisely located equivalent to where the microphones were initially placed to capture the drums, corresponding to the specific configuration used. They will be placed at the same height as the microphones. The loudspeakers will be positioned facing away from the receiver so the resultant IR will project the sound into the room, much like the radiation pattern of a drum kit.

The receiver position in Otondo & Rindel's research was situated for an audience listening position whereas in this project, the listening position is situated for a performing listening position. Therefore, for this project, the receiver will be situated inside the loudspeaker array, while Otondo & Rindel's work positioned it outside the array. Initially, concerns arose regarding the potential colouration of the recorded SRIRs due to this receiver placement, similar to the effect of listening to a speaker from directly behind, potentially resulting in a muffled SRIR. However, this concern is deemed irrelevant in the context of the VDR since the drums themselves are physically present in the room and can be directly heard by the drummer during a performance. In contrast, a traditional auralisation process necessitates replicating the direct sound. Nonetheless, the SRIRs still require editing, as discussed in Section 2.2.4.2, to remove the direct sound and solely replicate the reverberant sound field.

Rendering & Reproduction

For this project, it is logical to prioritise headphone reproduction over loudspeakers. This choice stems from the inherent loudness of drums as an instrument, which could potentially overpower the sound produced by loudspeakers. Moreover, drummers commonly wear headphones or some form of hearing protection during performances, making the concerns raised by Brereton (2014) and Eley et al. (2021) [61] less significant in this context. Another advantage of headphone usage is the avoidance of re-auralisation issues, where the sound reproduced by the loudspeakers is captured by microphones that pick up the direct sound from the musical instrument. This concern becomes particularly problematic if methods similar to those proposed by Otondo & Rindel (2005) are employed, especially when the instrument is not closely miked, resulting in a reduced direct sound-to-room sound ratio. Using loudspeakers would heighten the potential for re-auralisation, adding to the complexity and challenges of the project.

3.2 Initial Testing

Part of building the VDR required a development stage that sought to give a proof of principle for the system as a whole which consisted of a scaled-down set-up. This stage of the development, labelled as "Initial Testing" also was useful as it highlighted that some of the methods and technologies required adjustment or change to optimise the system for the final proposed build. This initial testing stage encompassed a two-channel set-up and convolution performed with alternative SRIRs to see whether the system would work as hoped. Although not the full set-up, this pilot scheme provided valuable insight. The pilot scheme as well as the adjustments made from its results will be discussed in this chapter.

The initial testing set-up consisted of two microphones placed to the left and right of the drum kit. At this stage of the development and building of the system, specific SRIRs had not yet been captured and instead, other SRIRs were used which had directionality in their measurements (more on this later). These SRIRs were captured in 3rd order and so the plug-ins used in Reaper had to be for 3rd order ambisonics.

3.2.1 Software

Initially, MAX/MSP was proposed as the software to perform the convolution. However, it quickly became apparent that Reaper was a preferable choice. Little research was conducted during the literature review about exactly which convolution algorithms would be used for this system. It was known that real-time convolution was possible and required some different techniques.

To perform real-time convolution, a plug-in and a plug-in host needed to be selected. The plug-in used would require an algorithm that didn't incur latency. The plug-in chosen was "X-MCFX" by Angelo Farina [67] as it uses unequal partition overlap-add, with the first partition in the time domain rather than FFT. This allows for zero latency compared with other methods such as the equal partition overlap-add method which incurs latency. Reaper was chosen as the plug-in host for this plug-in as MAX/MSP can be a computationally plug-in host and Reaper is computationally efficient at low buffer sizes. Additionally, previous research work that uses real-time convolution uses Reaper so it was known that Reaper worked efficiently and effectively [19].

As well as X-MCFX, three other plug-ins were used in the Reaper project. Two of the plug-ins came from the "ambiX v0.3.0 – Ambisonic plug-in suite" [68] which were the Ambix rotator, which would be used for head tracking (however head tracking was not used in the initial testing stage, therefore, more on that in later sections), and the Ambix binaural decoder which was used to decode the 16-channel 3rd order ambisonic convolutions for binaural headphone listening using the "ring24-3h0v" preset. The third plug-in used was ReaEQ from the "ReaPlugs VST FX Suite" [69] which was used to apply a low shelf filter of the Right and Left channel tracks, seen in Fig. 3.3, as it was noticed there was a low background hum around 200 Hz.

Figure 3.1 presents a signal flow chart illustrating the Reaper project's signal path. While this flow chart depicts the signal path for a single microphone, it's crucial to note that during the initial testing phase, there are two microphones in operation. Therefore, to accommodate additional microphones, you must replicate the components within the blue dashed boundary and route them into the *VST: ambix_rotator_o3*, as depicted in the figure. It's important to emphasise that for each newly added microphone, a dedicated SRIR will be employed. Additionally, keep in mind that the VST plug-ins are active on distinct channels within the Reaper project.

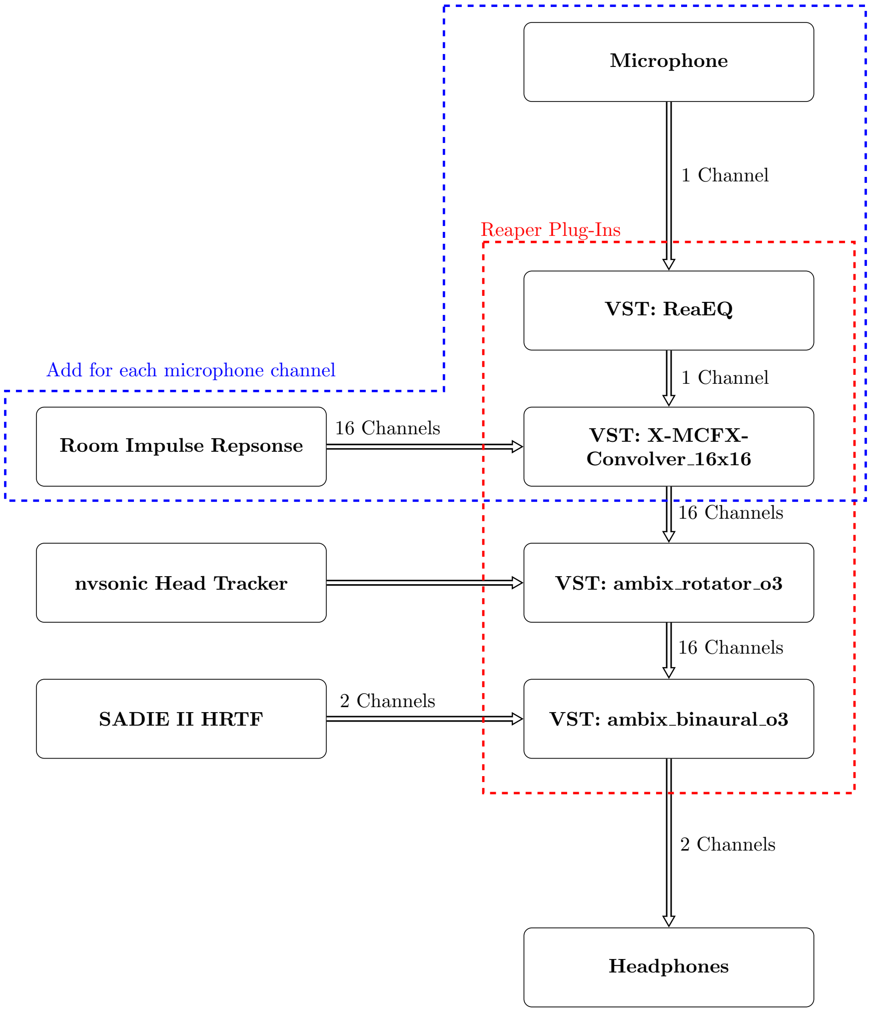


FIGURE 3.1: A flow chart of the signal flow within the Reaper project. This is representative of one microphone however in the initial testing stage there are two microphones and in the final build, there are four microphones.

Fig. 3.2 shows an image of the Reaper project that was used in the initial testing stage. Fig. 3.1 illustrates the signal flow of the Reaper project to better understand what is happening.

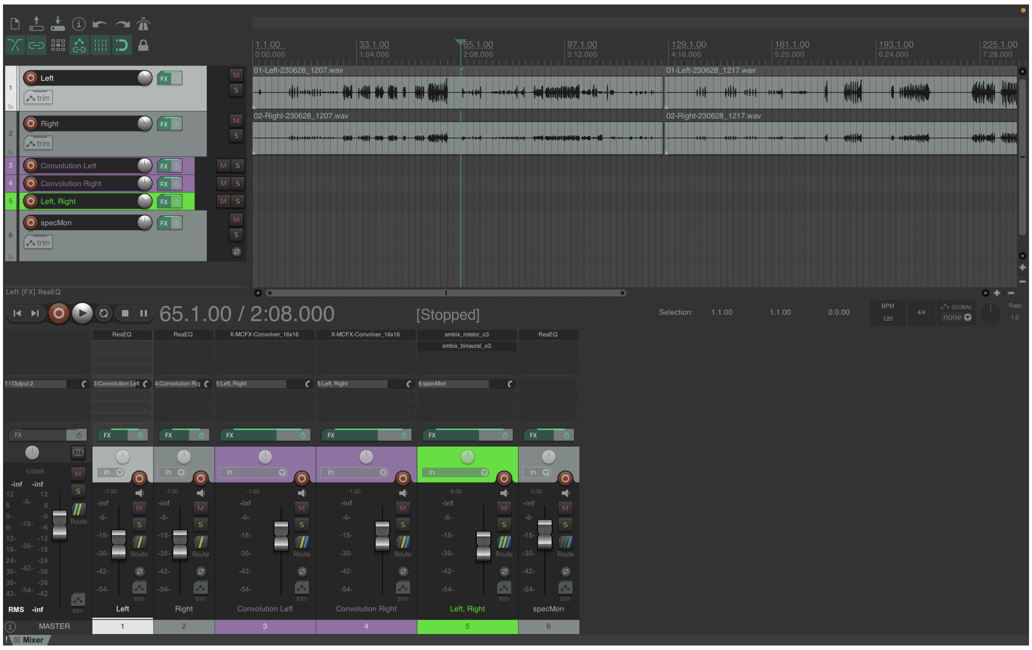


FIGURE 3.2: The Reaper session used in the initial testing stage.

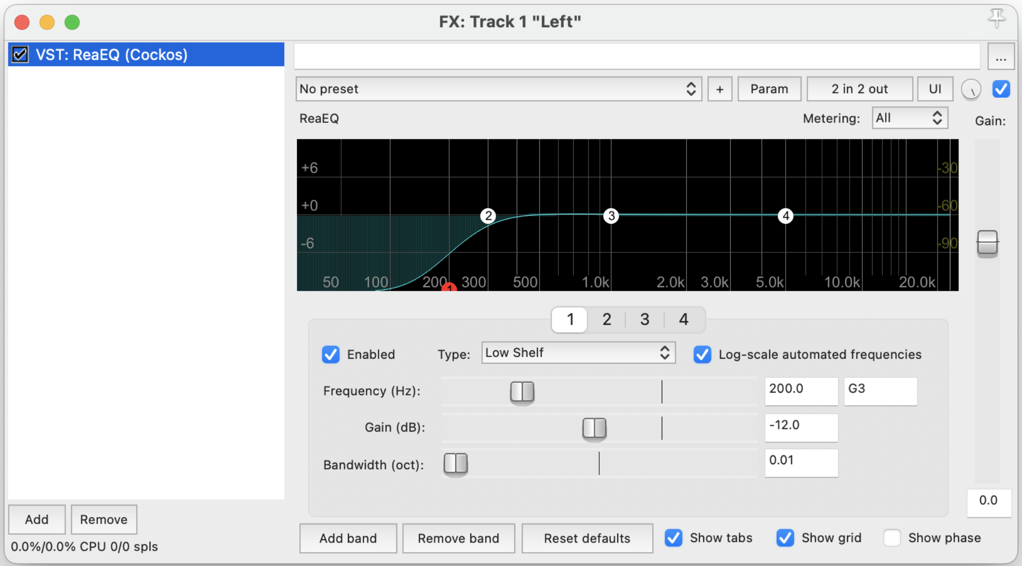


FIGURE 3.3: RealEQ plug-in used to apply a low shelf filter at 200 Hz for both left and right channels (left channel shown here).

3.2.2 SRIRs

During the initial testing phase, SRIRs provided by the Audio Lab at the University of York were utilised. These SRIRs were originally acquired through measurements conducted at the

Guildhall in York city centre as part of a previous Audio Lab project. The reason for their selection stemmed from the fact that project-specific SRIRs had not yet been measured. The choice of these Guildhall SRIRs was strategic as they incorporated directional source positioning relative to the receiver, enabling an evaluation of perceived directionality during the initial testing stage. To facilitate this assessment, the source and receiver were positioned in close proximity, and measurements were taken for four orientations: north, south, east, and west. For testing purposes, the east and west orientations were chosen, with the east direction assigned to the right channel and the west direction assigned to the left channel. While the precise alignment of the source and receiver was not tailored specifically to the proposed system, it served well in assessing the VDRs directional capabilities.

The SRIRs were recorded in 3rd order ambisonics so the the plug-ins used had to be 3rd order, as discussed above. 3rd order will be used throughout the remainder of the project as this initial testing stage proved that it works and there is no good reason to switch to a different order. Additionally, 3rd order generally sounds better compared to 1st or 2nd order [3].

3.2.3 Anechoic Chamber

The drum kit was set up in the anechoic chamber located at the University of York. The floor of the chamber measures roughly 3.5x3.5m. All VDR testing and use was carried out within this anechoic chamber. The drum kit used, shown in Fig. 3.4, was a collection of drums from a Mapex Horizon Birch drum kit and consisted of a 14" snare drum and a 16" bass drum. The cymbals used were a pair of Dream Dark Matter 14" hi-hat cymbals and a Dream Contact 20" ride cymbals. This drum kit was chosen due to the limited size of the anechoic chamber, and the complexity introduced by increasing the number of elements.

The drums were initially placed in the centre of the chamber, with the snare drum acting as the centre of the drum kit, however, it was moved forward as the floor of the chamber consists of a wire frame which bounces and the drum placement of the drum in the centre was creating too much bounce. Therefore the head of the drummer was now at the centre of the room.

(A) Front Right View

(B) Back View

FIGURE 3.4: Mapex Horizon Birch Drum Kit and Dream cymbals used in the initial testing stage. Two photos showing the drum kit from different angles.

In the chamber, there are four mounting points to place the four microphones that would be used in the VDR (despite only using two for the initial testing phase). These mounting points were positioned at the cardinal points at 116 cm from the centre of the chamber. When the drum kit was initially placed in the centre of the room the microphone cross-section went through the centre of the drum kit. However, when the drum kit was moved forward the cross-section of the two microphones went through the drummer's head. This was beneficial as it fitted in with the framework proposed. However, there are two problems with this. One is the drum kit moved forward means there is no space for the microphone which will be placed in front of the kit for the microphones to be equally spaced. Therefore the drum kit will need to be moved back. A mounting rig would be built to mitigate the bouncing. Secondly, the cross sections of the microphones need to be going through the centre of the drum kit to accurately and equally pick up the drum sound. To accommodate for the simulated head positioning movement, the SRIR measurement will have to be adapted from the proposed method. More on this later.

The drum kit was placed upon a 120cm x 150cm x 18 mm MDF board as the wires mesh floor of the chamber could not support a drum kit. The board was covered in a drum rug, as seen in Fig. 3.5, to prevent the drum kit from slipping around on the MDF board and also to create more accurate floor reflections. Drum rugs are commonly used to place drums upon no matter the room so despite not being an identical material underneath to the room in which the SRIRs were measured, the drum rug provided a similar effect.

Fig. 3.6 shows a floor plan schematic of the layout of the drum kit and microphones in the anechoic chamber for the initial testing phase. As previously stated, the listening position (drummer's head) was positioned at the centre of the anechoic chamber (marked as the red dot). The drum kit was positioned at 0.48cm in front of the centre position and the snare drum was used as the measurement reference position. Fig. 3.7 shows an illustrated top view of the drum kit showing all its components. The figure serves to show how the drum kit would appear in the floor plan schematics. The red point represents the drummer's head position or the listening position and the blue point represents the centre of the drum kit which will correlate to the point where the distance of the drum kit to the centre point is measured in Fig. 3.6.

FIGURE 3.5: Drum rug used to cover the MDF board reducing slip and providing more accurate floor reflections.

3.2.4 Hardware

To capture the drums two AKG 414s were placed 116cm from the centre of the drum kit and set roughly at the drummer's ear height (in this case 140cm). The microphones were set up to capture in an omnidirectional pattern, 0 dB and a 40 Hz low pass filter, as shown in Fig. 3.9. The omnidirectional pattern was chosen as it creates less colouration of the signal and as the recordings were anechoic there was no reflection from the walls so only the side. Fig. 3.8 shows the set-up in the anechoic chamber during the initial testing phase and the microphone positions can be seen (as annotated). The left channel microphone at 270° and the right channel microphone at 90°. The microphones were fed into a Fireface UC interface (shown in Fig. 3.10) and then into a Mac Book Pro running Reaper, as previously discussed.

The resulting convolution was played via a Behringer stereo headphone amplifier (shown in Fig. 3.11a) and then over a pair of Sennheiser HD650 open-back headphones (shown in Fig. 3.11b). Open-back headphones were used as they allowed the direct sound of the drums to come into

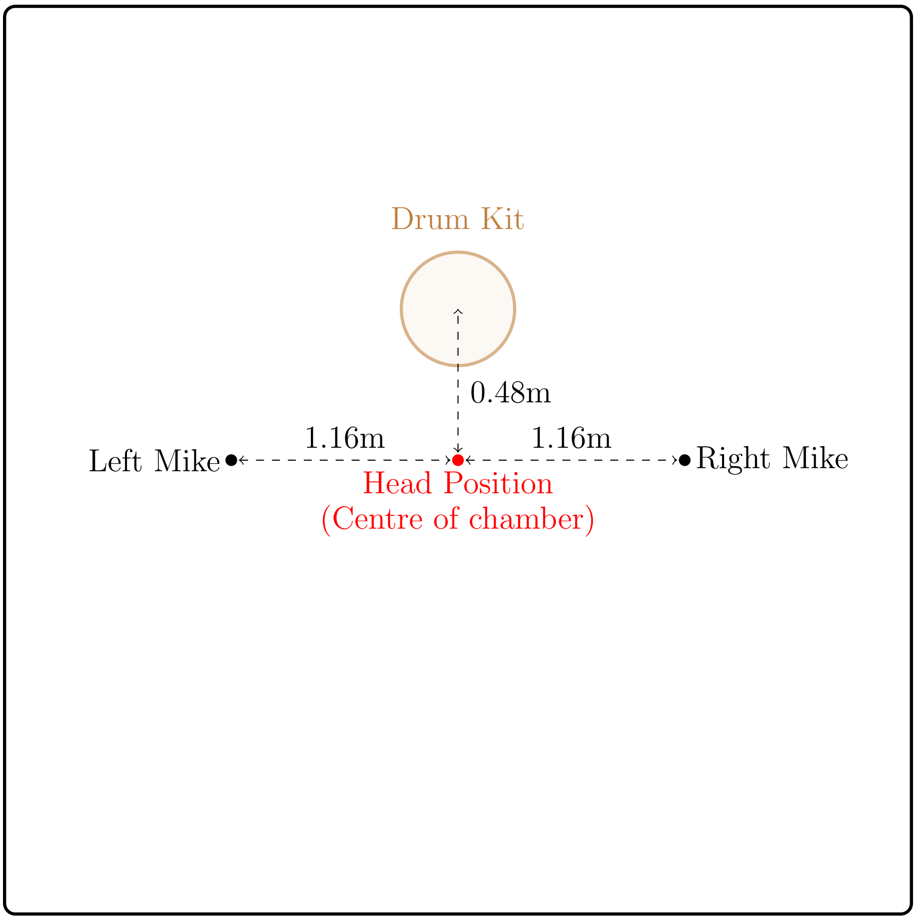


FIGURE 3.6: A floor plan schematic of the set-up in the anechoic chamber for the initial testing phase. Note that the centre position and the microphone positions correlate to anchor points in the floor structure of the anechoic chamber. The centre of the brown circle representing the drum kit acts as the centre of the snare drum.

the drummer's ears naturally, further enhancing the naturalness and realism of the experience. However, as the headphones were open-back there was no hearing protection from the drums. The VDR was trialled with and without hearing protection worn under the headphones and both still allowed for an effective experience, however not wearing hearing protection meant the reverberant sound was not muffled and more realistic. However, as the headphones were open back, when wearing hearing protection, both the direct and reverberant sound was muffled so no bias was created.

The Behringer stereo headphone amplifier was last in the signal chain before the headphones. Therefore it set the overall final volume level that the auralisation was played back to the listener. The level was set at two notches down from the maximum. The level of the auralisation playback was not set at a level that aimed to be identical to Hendrix Hall, as firstly not enough

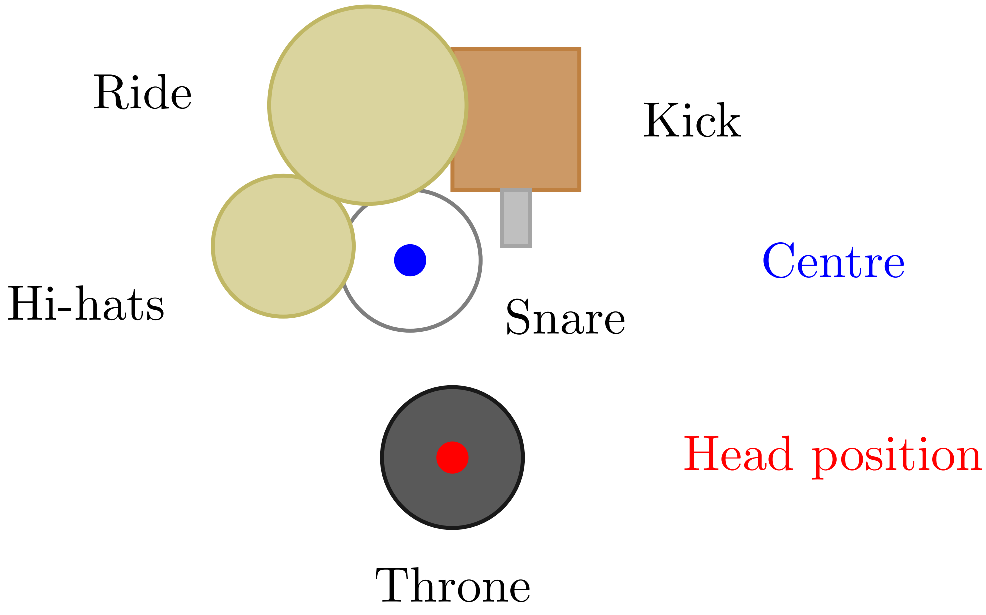


FIGURE 3.7: An illustrated top view of the drum kit showing all its components as well the centre point and head position locations.

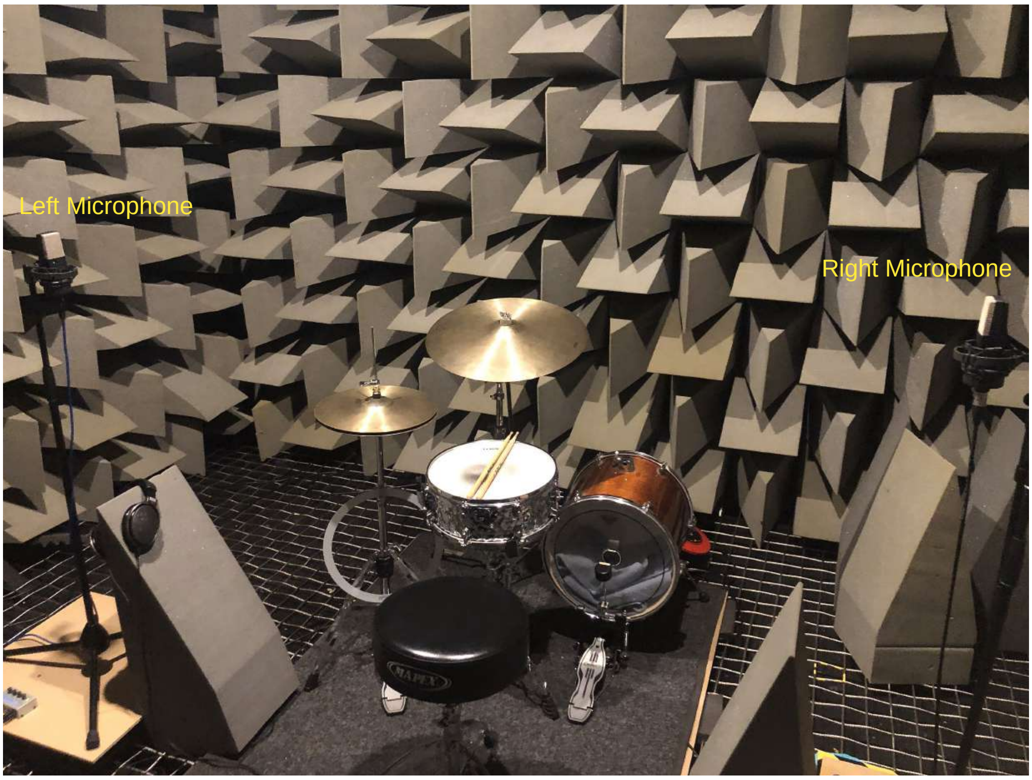
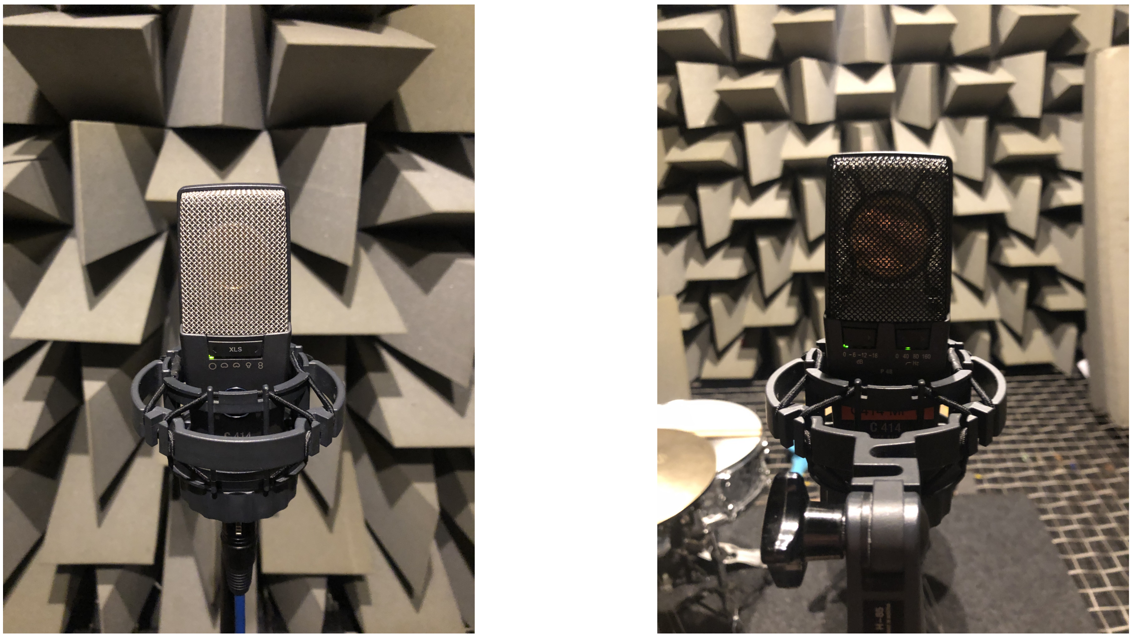


FIGURE 3.8: Anechoic chamber set up in initial testing phase with left and right microphones annotated.

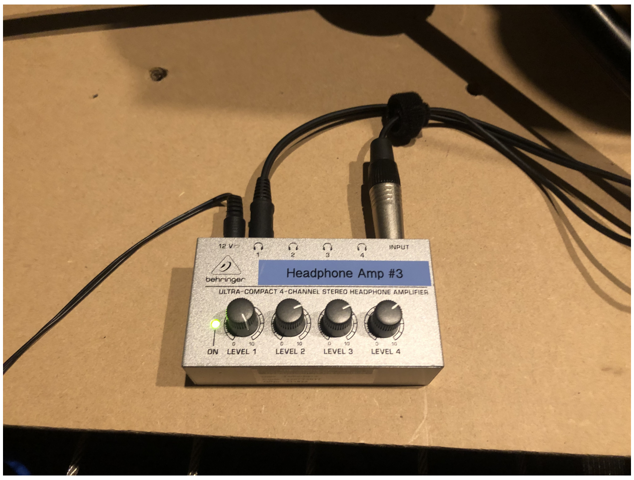


(A) Front view (B) Rear view

FIGURE 3.9: AKG 414 microphone configuration during initial testing.

FIGURE 3.10: Fireface UC interface.

measurements were taken in Hendrix Hall to attempt to replicate this and secondly, it is very difficult to achieve. Therefore the volume level was chosen subjectively to sound pleasant but not stray too far from a realistic/believable sound. The reverberation wanted to be loud enough for the listener to be aware of the sound but not too loud that it felt unrealistic. Because of the natural loudness of the drums, the headphone level was set relatively high.



(A) Behringer Stereo headphone amplifier.

(B) Sennheiser HD650 open-back headphones.

FIGURE 3.11: Images of headphone amp and headphones used in the initial testing phase and subsequently the final build.

3.2.5 RIR Editing

The SRIRs were edited to truncate and remove the direct sound from the convolution as the direct sound was present in the anechoic chamber from the drums themselves so only the reverberant sound needed to be simulated. This was performed by a modified version of two Matlab functions *directTruncate* & *removeDirect* which was developed by Patrick Cairns, a PhD student in the Audio Lab at the University of York. *directTruncate* is used to truncate an IR from the beginning of an audio file to the start of the direct sound on a specified channel with a specified pre-ring time and a linear fade-in. *removeDirect* works by using the source and receiver heights, the distance between the source and the receiver, and the speed of sound to calculate the time difference between direct sound arrival and the shortest path reflection (here considered to be the first ground reflection). The function finds the direct sound (as the maximum sample value in the IR) and estimates the point of the first reflection based on the air propagation time difference calculated. The beginning of the IR is then truncated to the estimated time of the first reflection, and a linear fade-in of the specified length is added.

Fig. 3.12 shows a graphical representation of the truncation process. The upper graph shows the original IR with a red dashed line which represents the maximum loudness. The lower graph then shows the resultant truncated IR.

Fig. 3.13 shows the result graphically where the upper graph shows the original IR (or in this case the truncated IR as this was done prior) with the direct sound at the red dashed line and the calculated first reflection at blue dashed line. The lower graph then shows the IR with the direct sound removed.

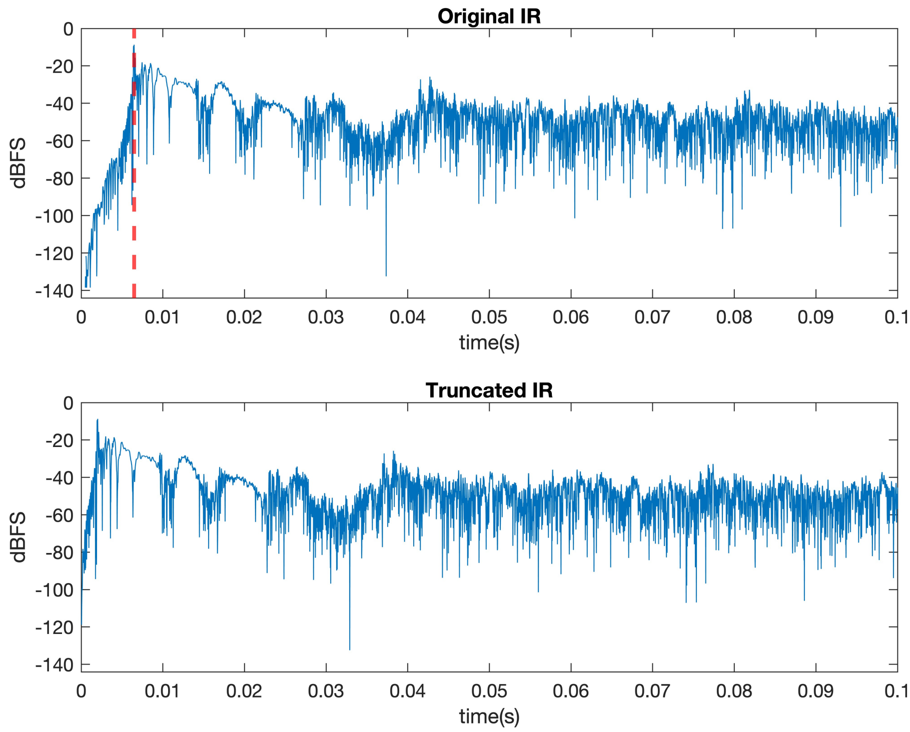


FIGURE 3.12: A graph to show a comparison of the original IR and the truncated IR. IR was taken in the Guild Hall with the source facing west.

It was noticed that there wasn't a significant difference in the sound between using edited and non-edited didn't SRIRs. This wasn't thoroughly tested but a noteworthy observation.

3.2.6 Objective Analysis

Overall, the results of the initial testing provided a proof of principle for the VDR and showed that it not only works but works well. There is no perceivable latency and the quality of the convolutions is both high and believable. Even with only two microphones, no head tracking

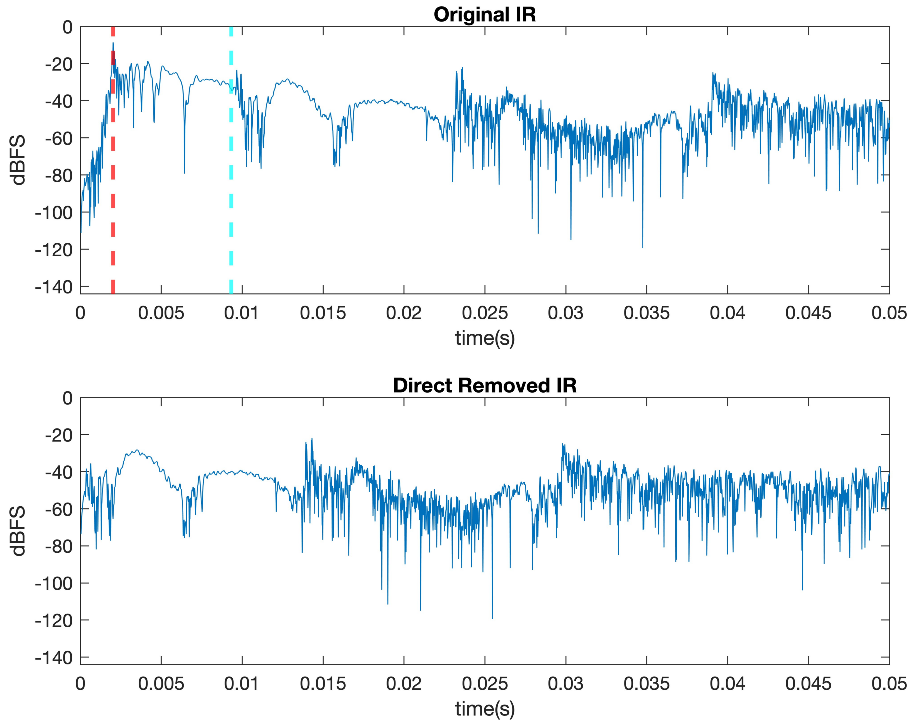


FIGURE 3.13: A graph to show a comparison of the truncated IR and the IR with the direct sound removed. IR was taken in the Guild Hall with the source facing west.

and SRIRs that are not tailored specifically to the system, the results were very positive and provided an enjoyable experience when playing the drums.

During the initial testing stage, some recordings were taken for further analysis as well as a demonstration to the reader. Testing whether there was any directionality to the VDR was difficult while playing as the drums were present in the room so it was difficult to discern whether the directionality was coming from the drums themselves or the convolutions. However, from listening back to the recordings, it can be heard that the sounds are coming from one particular direction.

It can be seen on the dB readings and also heard that the levels are not equal for the direct sound of the drums. The left channel is peaking around -6 dB and the right channel is peaking around -18 dB. This is likely due to the drum kit not being fully centred between the two microphones and the two microphones having slightly different calibrations. However, what the centre of the drum kit should be is not fully decided upon. It could be the snare drum as this is mostly the loudest and most featured part of the drum kit. But even then the hi-hats and the ride cymbal are left of the snare drum so it makes sense for the left channel to be louder.

Additionally, the bass drum, which is to the right of the snare drum, is quieter on average than the cymbals. Therefore the right channel being quieter isn't completely unexpected. The purpose of this microphone set-up was the capture the directivity of the drum kit and tailor it to the radiation characteristics of the instrument, so perhaps this shows that it is working as hoped and simulating the propagation of the sound of the drum kit into the room.

3.3 Final Build

After the initial testing phase, it was then on to the final build of the VDR, which would be evaluated. These two stages did not happen in succession as it required some time between to make design adaptions, measure the SRIRs and design the implementation and evaluation methodology. This section will provide a detailed account of the final build of the VDR.

Much of the software, hardware and other elements such as the drum kit and the anechoic chamber used in the final build are identical to that of the initial testing phase. Therefore some aspects may not be mentioned as thoroughly in this section as in the initial testing phase.

3.3.1 Software

The software used in the final build was identical to the software used in the initial testing phase discussed in Section 3.2.1. The only addition was the extra channels in the Reaper project for the four microphones used compared to the two used in the initial testing phase. The signal flow shown in Fig. 3.1 in Section 3.2.1 also holds true for the final build.

3.3.2 Hardware

The hardware, and the methodologies associated, that were used in the final build were almost identical to the hardware used in the initial testing phase discussed in Section 3.2.4. One alteration was that four AKG 414 microphones were used rather than two. Additionally, different microphone stands were used for the microphones as seen in Fig. 3.15. The other alteration was using a Zoom F8 interface, seen in Fig. 3.14, instead of the Fireface UC interface because it allowed for a four-channel input rather than two.

FIGURE 3.14: An image of the Zoom F8 multitrack field recorder used in the final build of the VDR.

3.3.3 SRIR Measurement

As discussed in Section 3.2.3, it was discovered that some work was required to adapt the SRIRmeasurement process to accommodate for the fact that the centre of the drum kit and the player's head are significantly far apart from each other. The centre of the drum kit is required to be in the cross-section of all four of the microphones so they capture the directivity and radiation acoustics without bias. However, this shifts the listening position (drummer's head) back towards the back microphone. If the SRIRmeasurements were continued as normal, the receiver position would be measuring a listening position at the cross-section of the microphones which is not where the drummer's head is located. Therefore, the receiver needs to be shifted back to the drummer's head when playing the drums relative to the four microphone positions.

FIGURE 3.15: Microphone stands used in the final build.

FIGURE 3.16: Images showcasing PL/001 as a location for SRIR measurements.

Location

Two rooms were chosen to take room impulse response measurements. These were Hendrix Hall and PL/001, both located at the University of York. Hendrix Hall is a large hall which has retractable tiered seating and is used for a variety of events and sports. PL/001 is a medium-sized lecture theatre with tiered seating. Fig. 3.16 and Fig. 3.17 show some images of PL/001 and Hendrix Hall respectively. Fig. 3.18 and Fig. 3.19 show schematics of the floor plans of Hendrix Hall and PL001 respectively, which include the dimensions of the rooms as well as the positioning of the centre point of theSRIR measurements.

Methodology

Fig. 3.20 shows a schematic of the layout for the SRIR measurements that were taken for this project. In the schematic, the blue dot represents the centre point which is the cross-section of the four loudspeaker positions and the red dot represents the receiver position which represents the listening position (drummer's head). There are four source positions labelled SF, SL, SR and SB - source front, source left, source right and source back respectively.

To measure the SRIRs, the centre point, receiver positions and four source positions were marked out on the floor of the room, as shown in Fig. 3.21. The receiver was placed in position with the centre of the microphone above the marked point. In turn, the source was placed in all four positions facing 180 ° away from the receiver. The source was also placed so the centre of the speaker was directly above the marked point. Only one source was used in the

FIGURE 3.17: Images showcasing Hendrix Hall as a location for SRIR measurements.

measurements and therefore the source was moved to each of the marked source positions starting at SF (source front) and moving clockwise around through to SL (source left). The receiver was placed at the head height of a drummer when sat down on a drum throne, which in this case was 140 cm, measured to the centre of the microphone. The source was also positioned at a height of 140 cm, with the midpoint between the woofer and the tweeter as the marker. This was to match the positioning of the microphones used to capture the drum sound which were also at a height of 140 cm.

Fig. 3.22 and Fig. 3.23 show photographs of the measurement process in Hendrix Hall and PL001 respectively. In both figures, the photographs show the source in each of the four positions.

Equipment

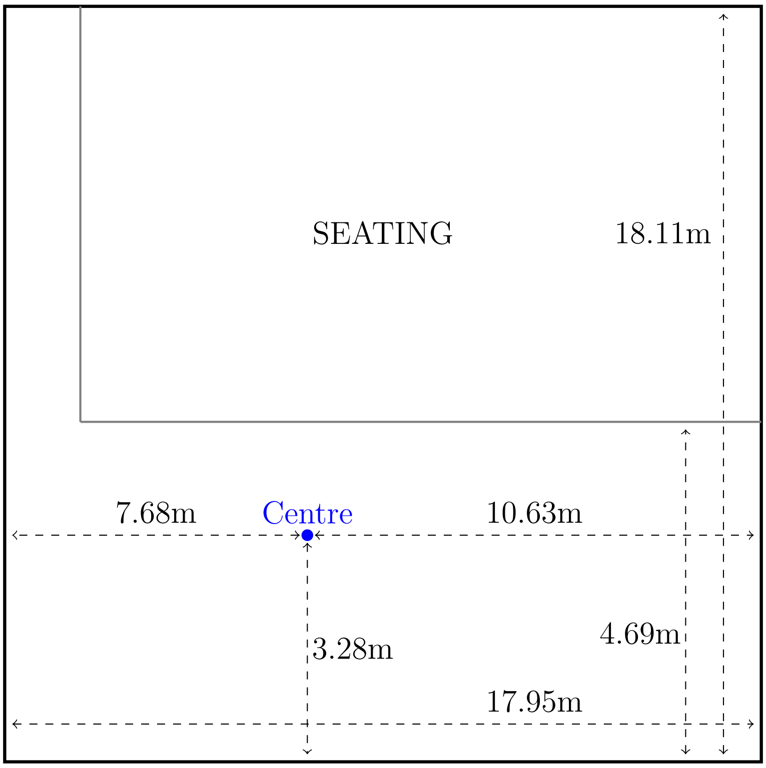


FIGURE 3.18: A floor plan schematic of Hendrix Hall.

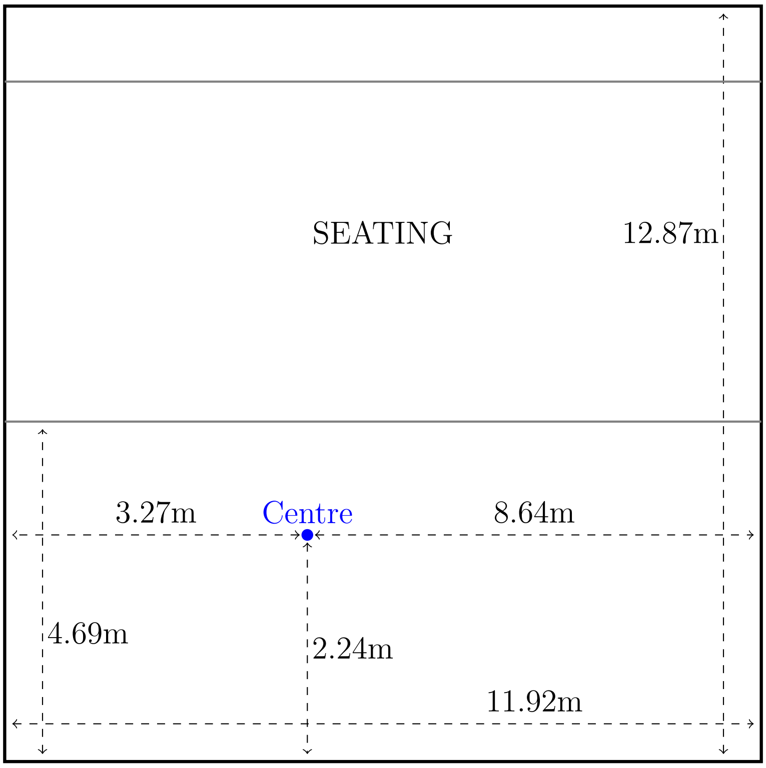


FIGURE 3.19: A floor plan schematic of PL001.

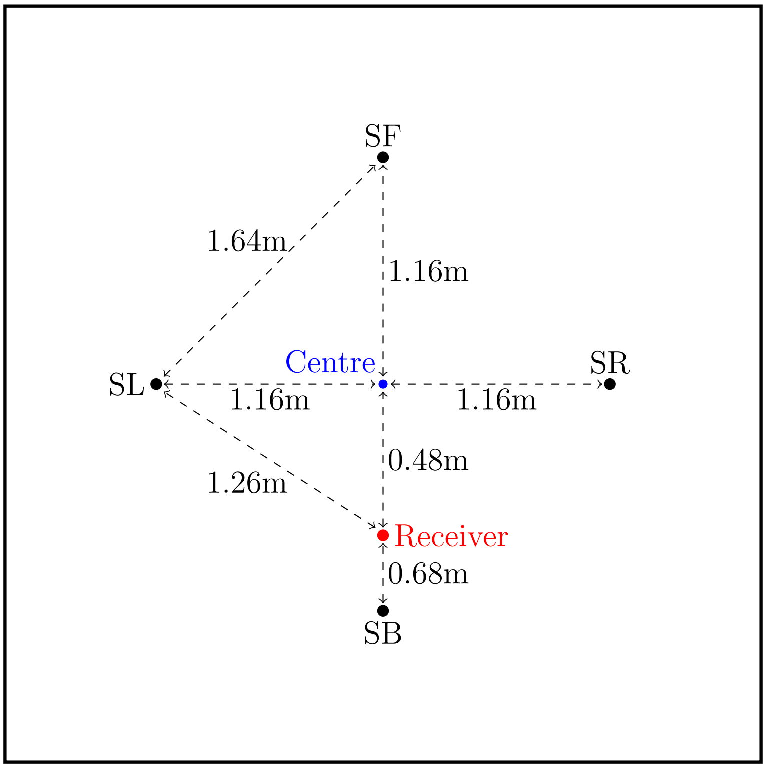
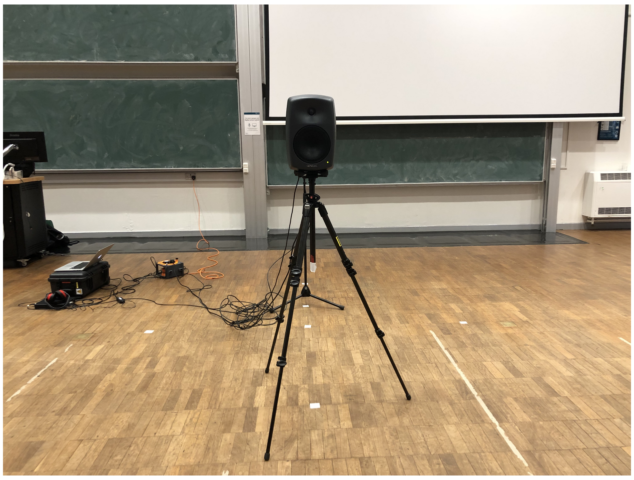


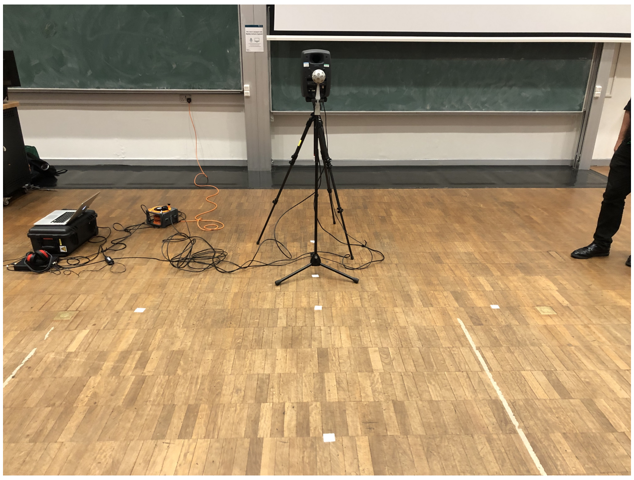
Figure 3.20: A floor plan schematic of the SRIR measurement in Hendrix Hall and PL001. SF - source front, SR - source right, SB - source back, SL - source left

Figure 3.21: Floor markings of source, receiver & centre positions forSRIRmeasurements in PL001.



(A) Source at SF position.

(B) Source at SR position.

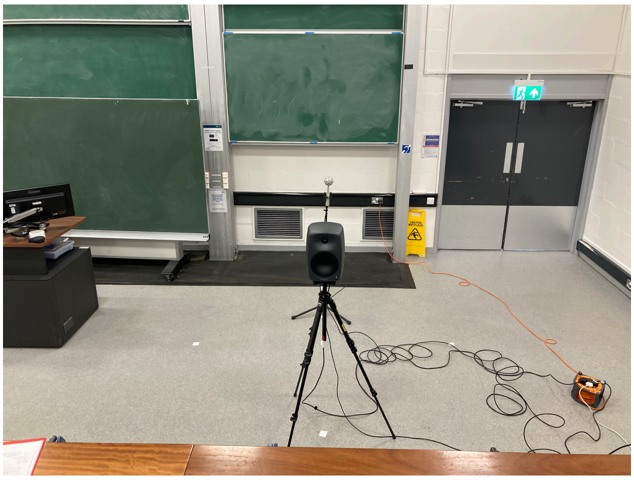


(C) Source at SB position.

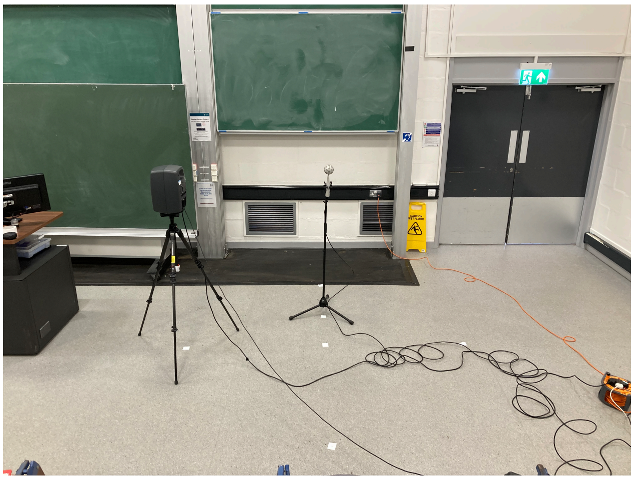
(D) Source at SL position.

FIGURE 3.22: Photographs of the four source and receiver positions for the SRIR measurements in Hendrix Hall.

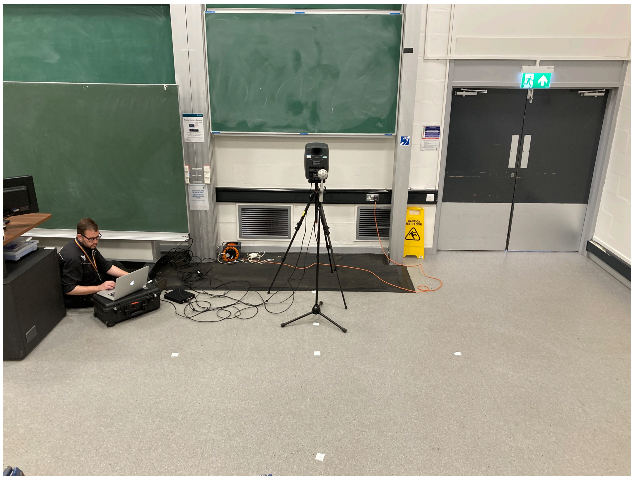
The receiver used was an Eigenmike Fig. 3.24b which allows for recording in 3rd order ambisonics and the source used was a Genelec 8040A loudspeaker Fig. 3.24a. Using Mac Book Pro, a 30-second sine sweep from 20 Hz to 20 kHz at a sampling rate of 48 kHz was generated using MATLAB (code found in Appendix A). This was copied into a Reaper project (found in Appendix A) and played through the Genelec Speaker while simultaneously recording using the Eigenmike. The Eigenmike's gain was required to increase to + 30 dB (the maximum value) in order to get a recording that was loud enough.



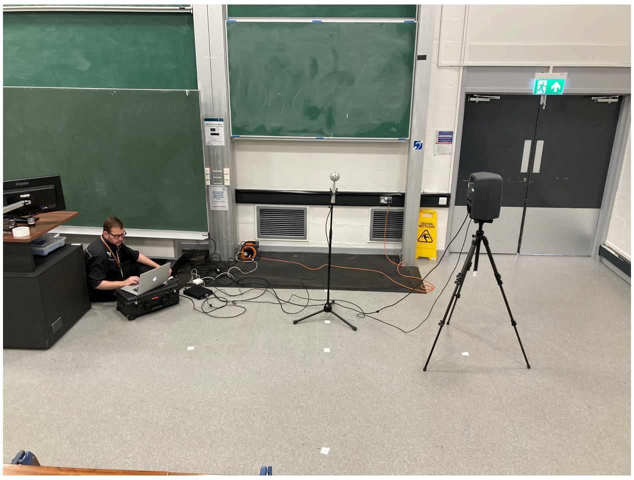
(A) Source at SF position.



(B) Source at SR position.



(C) Source at SB position.



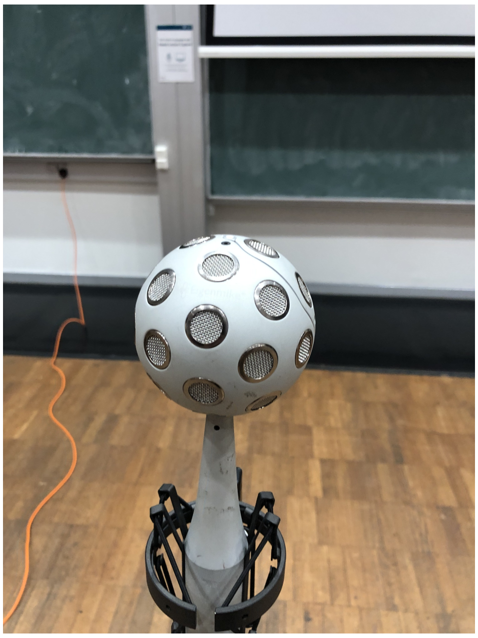
(D) Source at SL position.

FIGURE 3.23: Photographs of the four source and receiver positions for the SRIR measurements in PL001.

3.3.4 SRIR Editing

After deconvolving the recorded sine sweeps to generate the SRIRs, they were then truncated and the direct sound removed using the *directTruncate.m* & *directRemove.m* functions as in the initial testing stage in section 3.2. Fig. 3.12 shows a graphical representation of this process. The upper graph shows the original IR with a red dashed line representing the maximum loudness. The lower graph then shows the resultant truncated IR. Fig. 3.13 shows the result graphically where the upper graph shows the original IR (or in this case the truncated IR as this was done prior) with the direct sound at the red dashed line and the calculated first reflection at blue dashed line. The lower graph then shows the IR with the direct sound removed.

(A) Source used in SRIR measurement- Genelec 8040A



(B) Receiver used in SRIR measurement - Eigenmike

FIGURE 3.24: Loudspeaker source and microphone receiver used in SRIR measurement.

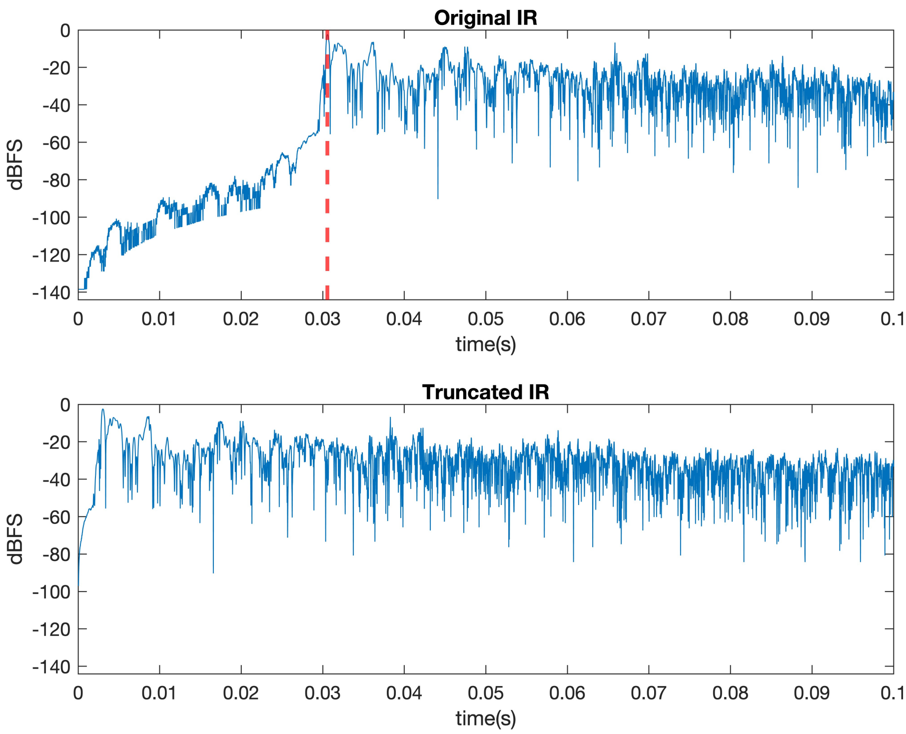


FIGURE 3.25: A graph to show a comparison of the original IR and the truncated IR. IR was taken in Hendrix Hall with the source in the back position.

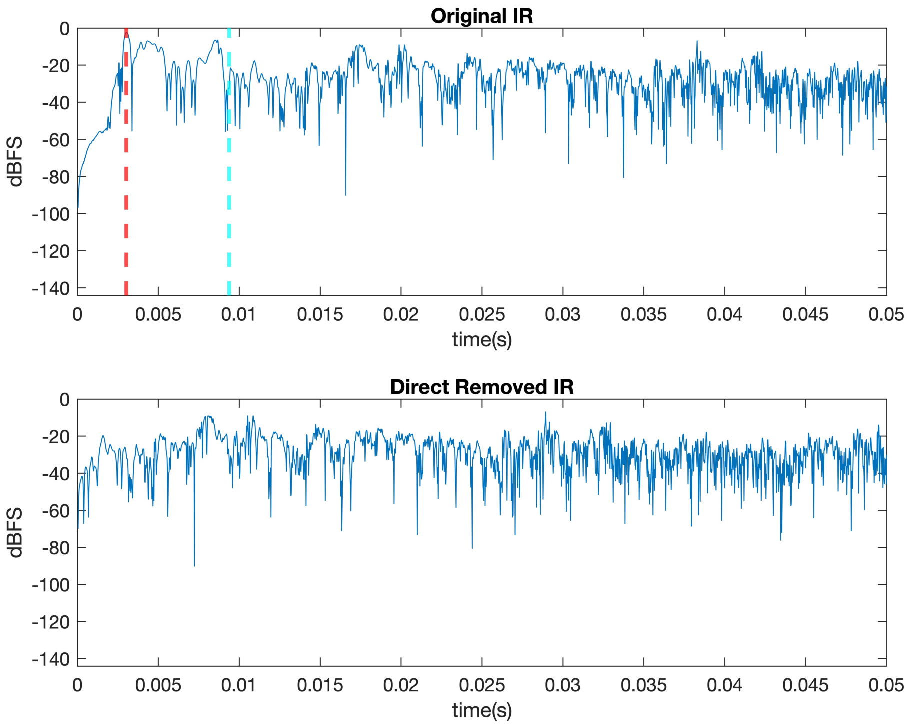


FIGURE 3.26: A graph to show a comparison of the truncated IR and the IR with the direct sound removed. IR was taken in Hendrix Hall with the source in the back position.

Unfortunately, in both Hendrix Hall and PL001, there was significant background noise caused by air circulation systems and computer banks which could not be turned off. In the sine sweep recordings (found in Appendix A) this can be heard throughout as a low hum. Because the Eigenmike has a particularly low noise floor, it had to be boosted to +30 dB to get a suitable recorded signal that was loud enough. However, this led to the Eigenmike capturing more of this background noise which ultimately affected the SRIRs. When listening to the SRIRs individually, any anomalies may not be readily discernible. However, when these SRIRs are applied to convolve with the drum sounds, a subtle but perceptible low wooshing noise becomes apparent. This auditory artifact becomes particularly pronounced after the direct drum sound has naturally decayed, leaving only the reverberant sound. The most conspicuous manifestation of this wooshing noise occurs with the snare drum, as it tends to activate frequencies associated with the background noise. This wooshing sound persists for the duration of the SRIR, which typically lasts around 3 seconds, and then abruptly ceases. It's worth noting that when the drums are played with rapid, successive hits, this effect is less noticeable. Conversely, when individual hits are spaced apart, exceeding the reverberation time, the wooshing becomes more prominent in the auditory experience.

To reduce some of the unwanted noise, the IRs were processed through the Matlab script *fadeIR.m* (found in Appendix A) which was provided by Andrew Chadwick from the Audio Lab at the University of York. The processing was done prior to the truncation and direct sound removal. The script works by splitting a signal into an arbitrary number of bands and envelopes each according to its rate of decay such that the noise floor is smoothly reduced to -Inf dB. The script was successful in reducing the noticeable cut of sound when the SRIR finishes. This can be heard when comparing the auralisations using the faded and non-faded SRIRs.

In the event of repeating this process, it would be advantageous to consider recording in an environment devoid of such background noise or employ a microphone with a lower noise floor. The Eigenmike was chosen primarily due to its status as the sole available 3rd order microphone at the time. The decision to utilise 3rd order ambisonics stemmed from the use of 3rd order SRIRs during the initial testing phase, which prompted the setup of the Reaper project to be configured for 3rd order as well. This choice was made to streamline setup procedures and mitigate potential complications in the later stages of the project.

FIGURE 3.27: Images of the NV Sonic head tracker attached to a pair of headphones and worn by a model head.

3.3.5 Head Tracking

To incorporate head tracking into the VDR, the NV Sonic head tracker [70] was used which was developed by Tomasz Rudzki, a PhD student in the Audio Lab at the University of York. The device consisted of an Arduino Pro Micro board with an MPU-9250 orientation sensor attached. Documentation for how to build this device and all other information can be found in Appendix A. The device was attached to the top of the headphones via a velcro strip and connected to the computer running the Reaper project. Fig. 3.27 shows images of the NV Sonic head tracker attached to a pair of headphones and worn by a model head. The data from the head tracker is sent out via an OSC plug-in, shown in Fig. 3.28, which then communicates to the Ambix rotator plug-in within Reaper. The OSC plug-in allows for a reset of the device so the origin coordinates can be set for each user.

3.3.6 Microphone Configurations

Of the four microphones used in the final build, there were four configurations that would be used and investigated in the listening test. The four configurations were as follows:

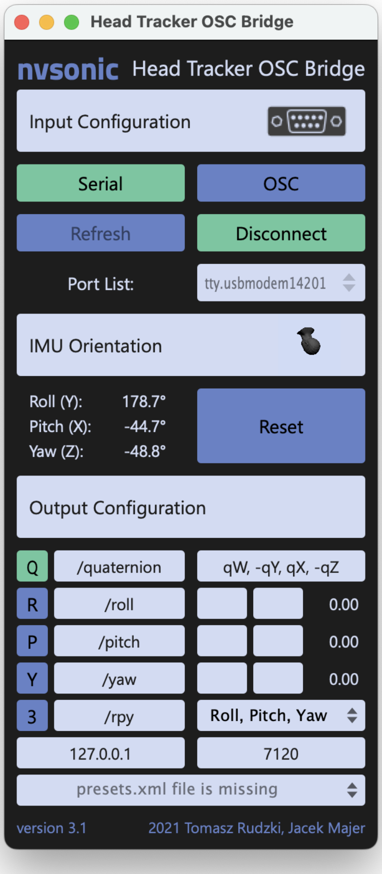


Figure 3.28: NV Sonic head tracker OSC bridge.

- Front & Back (FB) - The front and back microphone is active.
- Left & Back (LR) - The left and right microphone is active.
- Quad (All four) - All four microphones are active.

As mentioned in Section 3.1, the maximum number of microphones that could be feasibly used in the chamber was four. Investigating the VDRs performance and listener preference based on the number of microphones used was desired but with only four microphones, the configurations available wouldn't provide enough data to draw a valid and definitive conclusion. However, this was not to say that a 4-channel configuration and a 2-channel configuration could not be tested and compared. Additionally, as there were only a small number of possible configurations, it was decided to investigate the differences between a Left & Right channel configuration and a Front

& Back configuration. A Quad mono condition was also added to the VDR which consisted of all four microphones but with the master track of the Reaper project set to mono so no spatial information was present. This was for the purposes of the listening test and discussed further in Section 4.2.

3.3.7 Anechoic Chamber

The SRIR measurements are directly proportional to the microphone layout used to capture the drums. Fig. 3.29 shows a schematic of the microphone and drum kit positioning within the anechoic chamber for the final build. Comparing Fig. 3.29 to Fig. 3.20, it can be seen that the source positions and microphone positions are co-located as well as the drummer's head position and the receiver position. The drum kit is located in the centre of the anechoic chamber with the snare drum acting as the centre of the drum kit, as discussed in 3.2.6. The drum kit used in the final build was identical to the drum kit used in the initial testing phase and was set up in an identical format.

As in the initial testing phase, the microphones are positioned 116cm away from the centre position of the anechoic chamber on the left and right sides. For the final build, two extra microphones were added (as mentioned in Section 3.3.2) in front and behind the drum kit, again at 116cm from the centre position. As before all the microphones used were AKG 414s. The microphone set-up can be seen in Fig. 3.31.

For the final build, the drum kit was placed upon a 210cm x 210cm x 18 mm MDF platform that was made up of two equal-sized pieces of MDF. The platform was attached to an aluminium frame which was mounted to five mounting points within the floor of the anechoic chamber. The platform is shown in Fig. 3.30. The platform was rigid and secure and prevented the drum kit from bouncing up and down whilst playing as it did during the initial testing phase. The platform was large enough to have the four microphone stands placed upon it. As in the initial testing phase, a drum rug was placed down on the platform for the drum kit to stand on (the same drum drug was used).

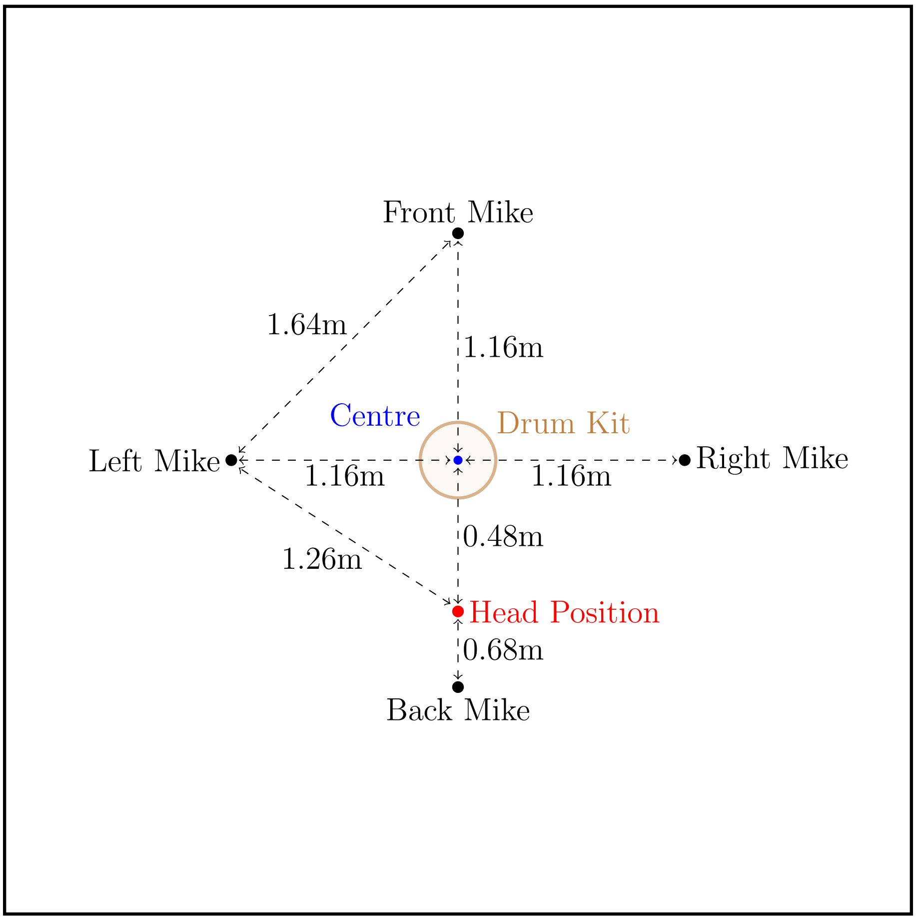


FIGURE 3.29: A floor plan schematic of the set-up in the anechoic chamber for the final build of the VDR. Note that the centre position and the microphone positions correlate to anchor points in the floor structure of the anechoic chamber. The centre of the brown circle representing the drum kit acts as the centre of the snare drum.

Regrettably, the chamber fell short of the desired anechoic quality due to the presence of various equipment and materials within. Elements such as microphone stands, hardware, a music stand, and the partially covered MDF floor – as the drum rug didn't provide complete coverage – contributed to this limitation. While an entirely anechoic environment would have been optimal, this deviation wasn't considered a substantial concern. The alterations were relatively minor, and their potential impact on the auralisations seems unlikely, given the drum's inherent loudness, which can effectively mask such nuances. Moreover, the chamber already exhibited a highly damped acoustic environment. Additionally, a faint low-frequency hum originating from a nearby building at approximately 30 Hz was being picked up by the microphones. However, proactive measures were taken to address this issue. Equalisation adjustments were introduced to the input channels, coupled with a 40 Hz filter applied directly to the 414 microphones

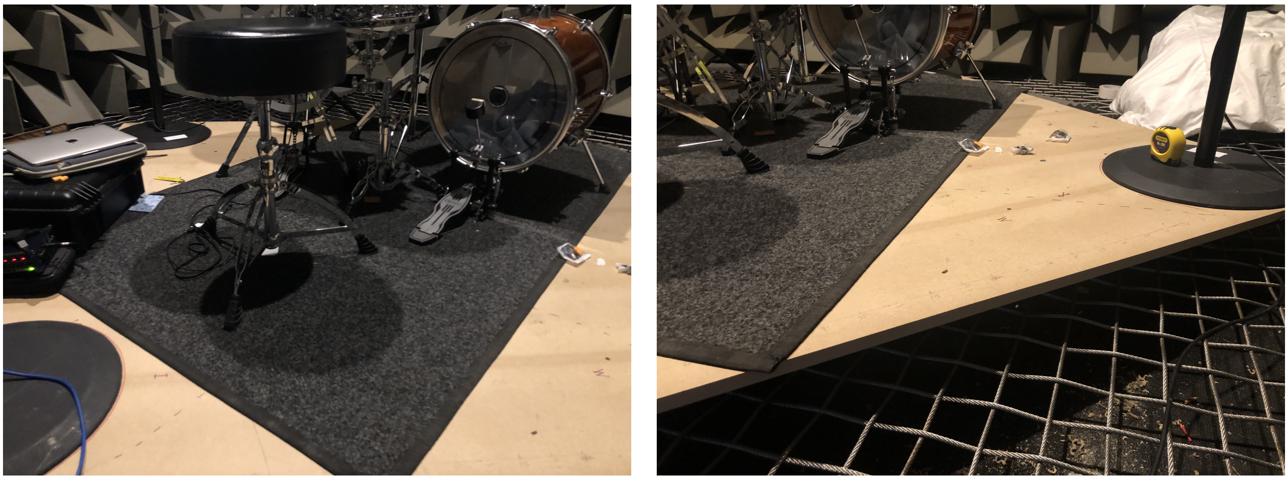


FIGURE 3.30: Images of the MDF platform the drum kit was situated on in the anechoic chamber.

themselves. These interventions succeeded in mitigating the interference and minimising its influence on the captured audio.

Some images of the final set-up of the VDR are shown in Fig. 3.31.

3.3.8 Objective Analysis

The final build is a further testament to the success of the design and implementation of the VDR. The final build of the VDR performs well and the sound is again of high quality and a believable nature. The VDR is very enjoyable to use and simulates playing in the rooms effectively. The addition of head tracking adds a new depth of realism and immersion to the experience. The addition of two more microphones is noticeable and provides a larger sense of space to the sound of the auralisations. The difference in sound between initial testing and the final build is significant but is not as distinct as expected. With the initial testing, the directivity of the sound was more noticeable but with the final build, it was less so. The reasons for this are not fully understood. The quality of the SRIRs measured in Hendrix Hall and PL/001 is not as high quality compared to the SRIRs measured in the Guildhall, as discussed in Section 3.3.4, however, this is a minor discrepancy and is believed not to take too much away from the overall quality of the experience in the final build.

FIGURE 3.31: Set up of the final build of the VDR.

3.4 Reaper Project

As a supplement to the thesis, the Reaper project referred to in this chapter has been provided in the digital appendix (Appendix A). Along with this, a user guide has also been provided which will provide instructions on how to navigate the Reaper project for the purpose intended. The user guide can be found in the *Reaper Project* folder in the digital appendices and is titled *Reaper User Guide*.

Text and pictures only go so far as to document the VDR. Therefore it was thought it would be beneficial to allow the reader to hear how the VDR sounds for themselves. Obviously, the reader will not be able to use the VDR as intended, i.e. play the drums and experience the VDR in the flesh, however listening to the sound of the VDR will still be valuable for a reader to get a better understanding of the VDR.

Recordings of the drums were taken during both the initial testing phase and the final build, which are found in the Reaper project. If these are played back through the input microphone channels (grey-coloured tracks), the recordings will be sent through the signal flow and the user will able to hear the drum sound auralised. When a drummer is using the VDR, the direct sound comes from the drum kit itself and the auralised reverberant sound is played back to the drummer. In this case, the drum kit is not present and therefore the recordings are sent directly to the master channel so the user can hear as close as possible to what the drummer would hear.

The user is advised to listen to the recorded drum tracks in all the different configurations and associated SRIRs to their desire.

Along with the Reaper project, a short video which of the VDR being used is provided in Appendix A.

3.5 Conclusion

This section has documented in detail the design and implementation of the VDR. Based on the knowledge gained through the literature review, this would inform the design of the system which would then be implemented through the initial testing phase (Section 3.2) and the final build (Section 3.3). The initial testing phase sought to build a scaled-down version of the VDR which ultimately showed proof of principle and determined the design worked effectively. The initial testing phase was also used to ascertain any alterations or additions needed in the design. The final build stage then sought to build the finalised design of the VDR ready for evaluation. There were some aspects that were not used in the initial testing phase such as head-tracking and the project-specific SRIRs. Therefore, any alterations that were required had to be investigated during the final build. Both the initial testing phase and the final build were successful in implementing the design of the VDR and producing a quality experience.

Chapter 4

Evaluation

In this chapter, the methodologies utilised to evaluate the Virtual Drum Room (VDR) will be presented, along with the results, a discussion of the results and limitations of the VDR and evaluation methodologies.

For the purpose of this evaluation, only the Hendrix Hall SRIRs will be used in the listening test. Incorporating the SRIRs from PL/001 would have increased the duration of the listening test significantly. The Hendrix Hall SRIRs were chosen over the PL/001 SRIRs as the reverberation time was longer and the sound was deemed more interesting. These were believed important factors for a novel system such as this, as the simulated reverberation wanted to be noticeable and different from what drummers may be used to.

4.1 Aims & Objectives

Section 1.1 discussed and provided the aims and objectives for the project as a whole. For the evaluation, there are some more specific aims and objectives which are presented below.

While a comprehensive perceptual listening test is beyond the scope of this project, an informal test can still be conducted. Instead of acquiring datasets and extensively analysing the data, this project will rely on verbal feedback from the drummers, which will be documented. Due to the nature of the aims and objectives of this study, subjective analysis is the primary type of analysis that is presented and discussed in this chapter. Unlike the work of Otondo discussed in Chapter 2, this project does not aim to investigate and compare the microphone configurations objectively, i.e. measuring directivity patterns, as this would not be relevant to what the study is fundamentally trying to discover. Additionally, it would not be feasible within the time frame of the project. The study is primarily about perception and human preference so a subjective analysis isn't necessary and appropriate. As this listening test is in essence a trial run for the VDR and an indication of how to conduct further testing, an objective analysis could be involved in future work.

The evaluation process aims to investigate three essential factors. The three factors under investigation in the evaluation are as follows:

1. Factor 1: The first aspect centres around evaluating the proposed methodology's ability to accurately capture and simulate the directivity of the drums within a virtual acoustic environment. This will involve investigating how the spatial information is perceived/preferred, and investigating how the different microphone configurations are perceived/preferred. The goal is to determine whether using spatialised auralisations is noticed and preferred over non-spatialised auralisations as is the motivation for the VDR.

2. Factor 2: The second aspect involves assessing whether drummers find using the VDR enjoyable as a recreational tool and what their preferences are for the VDRs configurations.

3. Factor 3: The third factor is to assess the effectiveness of the evaluation methodology i.e. the experimental protocol. The preliminary listening test will be useful to determine whether the experimental protocol used is suitable for effectively evaluating the VDR. Therefore, the results from the evaluation will also provide comments and insight into future testing procedures.

These three factors will be key in determining the success of the VDR so will be referred to throughout this chapter.

4.2 Experimental Protocol

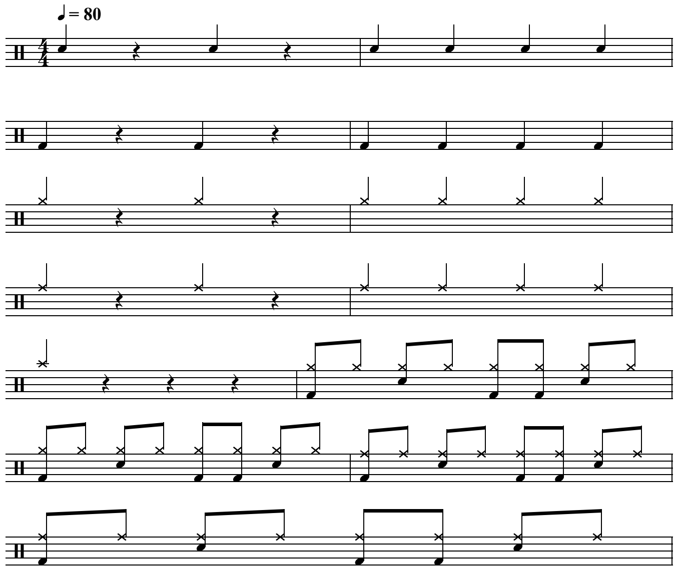
This section will explain in detail the experimental protocol for the listening test, which encompasses assessing the three factors presented above. The listening test comes in two parts, Part A and Part B, so the experiment protocol will be explained as such. Factor 3 is assessed in Part B and Factor 1 and 2 are assessed across both Part A and Part B.

To assess the VDR, an informal listening test was carried out encompassing an assessment of both the factors stated above. The first factor will be assessed through an informal listening test whereby the participants are asked to play extracts of drum notation with the different system configurations applied, as outlined in Section 3.3.6. The participant was not aware of the number of configurations tested, nor the order in which the participant will experience them. This is in order to assess whether the participants can notice a difference in the configurations without bias. The second factor is assessed by asking the participants to play whatever they wish on the drums and allowing them to switch between the microphone configurations freely. This will allow them to experience the VDR within their own playing style and comment more freely on their experience.

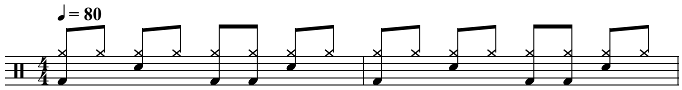
Both the fast-track ethics approval form as well as the participant consent form for the listening test can be found in Appendix A.

4.2.1 Part A

Part A of the listening test aims to investigate the performance of the VDR regarding its ability to accurately capture the drum kit's directionality and effectively reproduce it in a virtual acoustic environment. It would assess how different microphone configurations affect this. This part of the listening test would get the participants to compare the different microphone configurations with a control (in this case a mono auralisation with no spatial information) in order to assess whether the spatialisation was noticeable and effective. This would allow for the assessment of Factor 1.



(A) Exercise 1



(B) Exercise 2

FIGURE 4.1: Drum notation of exercise 1 & 2 for the evaluation of the VDR.

The participants were first asked to warm up, as is common before a performance. The participant was then instructed to play through the drum score exercises, as seen in Fig. 4.1, in the order that is seen in Fig. 4.2. For each participant, the order of the configurations was randomised to prevent bias. The conditions labelled "*Conv. Reverb - Microphone Configuration 1, 2 and 3*" are just placeholders.

As shown in Fig. 4.2, the participants first played through 4.1b once anechoically (VDR turned off). This allowed them to become accustomed to the acoustic environment and to act in a neutral tone. Following this, they were asked to play through 4.1a once; this time they performed with a convolution reverb (VDR turned on) and all four microphones were active. However, here the master tracks were switched to mono so no spatial information was present. This acted as a reference point for the participants to compare the three configurations (discussed in Section 3.3.6) to. Following the anechoic and mono conditions, the participants were instructed that there would be three conditions that they would play in, where they would perform through

4.1a. Each condition proceeded by playing through 4.1b with the mono condition in order to remind them of the comparison condition.

	Time --->						
Condition	Anechoic	Conv. Reverb - Mono condition	Conv. Reverb - Microphone Configuration 1	Mono condition	Conv. Reverb - Microphone Configuration 2	Mono condition	Conv. Reverb - Microphone Configuration 3
Exercise	Exercise 2	Exercise 1	Exercise 1	Exercise 2	Exercise 1	Exercise 2	Exercise 1

FIGURE 4.2: A timeline of the factor 1 part of the listening test. Microphone configurations are labelled 1-4 as the order is different for each participant as it was randomised.

After each condition, the participants were then asked to assess three distinct sound attributes in the comparisons. These were:

- The perceived spaciousness of sound in virtual the room [71]
- The perceived naturalness of the drum's timbre
- The perceived realism of the simulation

These specific sound characteristics were selected based on previous findings from similar listening tests involving auralisations. The reason for choosing these attributes was that they were easily understandable for the test subjects during the evaluation processes in [10] and [17].

By inquiring about the perceived spaciousness of sound, the aim was to measure how well the subjects could detect changes in the source's directivity in space. On the other hand, asking about the naturalness of timbre was intended to evaluate to what extent the representation of the source's radiation influences the quality of the reproduced sound of the instrument.

Once this section of the listening had come to an end the participants were asked whether they had a preference for any of the four conditions they performed in (three microphone conditions and the mono condition) as well as whether they thought each of the three configurations were all different, all the same, or somewhere in between compared to the mono condition.

4.2.2 Part B

Part B of the listening test aims to investigate the participants' preferences for the different configurations they were presented with as well as their overall opinion and feeling on using the VDR. This would allow for assessment of Factor 2, as well as insights into future development of the VDR.

The participants were instructed that they had 10 minutes to play freely and switch between the microphone configurations as they wished. At this point, they were informed about the three configurations and the mono condition. They were encouraged to try out all configurations during this section. During and after the 10 minutes, the participants were encouraged to be vocal about their thoughts and feelings about the VDR, which were written down by the listening test administrator. They were also asked again whether they had a preference for one configuration and whether they had any suggestions for future improvements to the VDR.

This section allowed for more feedback from the participants regarding their experience using the VDR which was more than just numbers and ratings. With such a sample size, achieving accurate data for ratings and numerical data is very hard without a representative sample. Allowing the participants to be more open and vocal about their thoughts allowed further insight into the VDR and suggestions for future improvements.

4.3 Results

This section will present the results of the listening test. As with the experimental protocol discussed above, the results will be presented as Part A and Part B.

4.3.1 Participants

Four participants undertook the listening test and were all student or staff members within the Audio Lab at the University of York. They were all considered to be expert listeners and were familiar with the technology used in the VDR. Coincidentally, two of the participants had trialled the VDR in the initial testing phase. This proved beneficial as they both commented on the performance of the VDR with the SRIRs from the Guildhall and Hendrix Hall. This is discussed further in Section 4.3 and 4.4.

All participants had been playing drums for a significant amount of time, ranging from 10 to 28 years. All had played very regularly in past years but in recent years they hadn't been playing as regularly. They mixed from self-taught to formally taught by teachers or schools. One participant played electronic drums predominantly. Fig. 4.3 shows an image of one of the participants playing the drums while using the VDR.

FIGURE 4.3: An image of one of the participants playing the drums while using the VDR.

4.3.2 Part A

Recall that at this point, the participants were unaware of which of the three configurations they played in the listening test were which microphone configuration.

Perceived Spaciousness, Naturalness and Realism

Fig. 4.4 shows the average of the ratings given by the four participants for the perceived naturalness, spaciousness and realism. In Appendix A the individual ratings from the participants are also provided. Fig. 4.4 shows that on average, the perceived naturalness, spaciousness and realism for all three of the microphone configurations were rated higher compared to the mono condition rather than less or the same. However, as seen in the individual ratings in Appendix A, in a few instances, the perceived naturalness, spaciousness and realism were rated lower than the mono condition. DESCRIBE

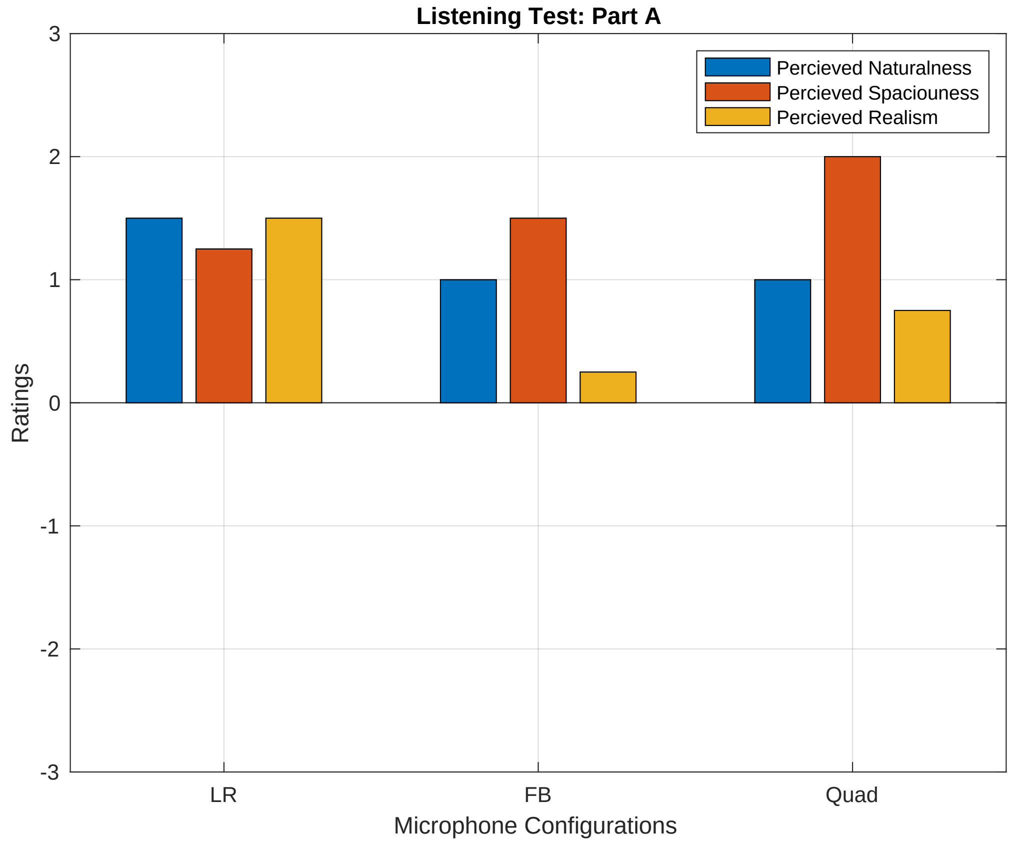


FIGURE 4.4: A bar chart to show the average scores of the three aspects being assessed in Part A of the listening test.

For the perceived naturalness, Left & Right performed the best with an average rating of 1.5 with Front & Back and Quad rating at 1. For the perceived spaciousness, Quad performed the best with an average rating of 2 with Front & Back at a rating of 1.5 and Left & Right at a rating of 1.25. For the perceived realism, Left & Right performed the best with an average rating of 1.5 with Front & Back at a rating of 0.25 and Quad at a rating of 0.75.

Fig. 4.4 also shows that on average the Left & Right configuration performed the best having the highest perceived naturalness and realism ratings compared to the other two configurations. The Quad configuration achieved the highest score for perceived spaciousness as well as the highest average score rating at +2. The Front & Back configuration did not achieve the highest score for any of the three attributes. It did score higher than the Left & Right configuration for the perceived spaciousness, however, for the other two attributes it had the lowest scores on average.

Preference of Microphone Configurations

When asked at the end of Part A, all four participants unanimously preferred the Quad microphone configuration, with the exception of one participant who stated they preferred both the Quad and the Left & Right configuration equally over the mono condition and that they preferred the mono condition over the Front & Back configuration. One of the participants who preferred the Quad configuration did provide a caveat with their response that they found it difficult to remember all the configurations.

Number of Configurations

When asked whether they thought the three configurations were different, the same or a mixture compared to the mono condition, 50% of the participants stated that thought all of the conditions were different and the other 50% stated that some were the same as the mono condition and some were different. One participant specified specifically that they thought that the Front & Back configuration was the same as the mono condition.

4.3.3 Part B

In this section, the participants were allowed free control of the VDR in order to play around and experiment with the configurations as they wished, whilst being informed of what microphones the configurations consisted of. All four of the participants provided a vast amount of feedback that was provided in Part B of the listening test. Whilst extremely beneficial, it is not feasible to go over all the feedback obtained in this part of the listening test. Therefore, a few highlights will be selected from each participant, however, the full document with all the notes taken during the listening test for all four participants will be provided in Appendix A. The results will be presented with general comments that were made by the participants and then will be broken down into comments made by each participant.

General Comments

Here is a collection of comments that were made by either a majority or all of the participants during part B of the listening test:

- Across the board, all participants stated that after learning what the three microphone configurations in the listening were, they still had a preference for the Quad configuration. It was commented that the Quad configuration was more reverberant than the mono condition and that the Quad configuration was the most natural and spatialised. One participant stated that the Quad configuration felt nicer to play but that it wasn't necessarily more realistic compared to the other configurations.

- All the participants very much enjoyed the experience they had using the VDR and would very much like to use it again for their own recreation if they were to have access to it. One participant even commented that they had thought of building a similar system at home without spatialisation (due to a lack of technology). They also all stated that if they had access to it in the future they would like to see it have a selection of rooms and spaces to play in, i.e. more SRIRs. Therefore, regarding Factor 2, the VDR can be considered a success.

- All the participants stated that the ride cymbal felt "washed out" and the reverberation was not as noticeable on this part of the drum kit compared with the others.

- Two of the participants who undertook the study had also used the VDR in the initial testing phase when the Guildhall SRIRs were used. They both stated in the listening test that they preferred the sound of the Guildhall SRIRs compared to the Hendrix Hall SRIRs. They mentioned how objectively the quality seemed superior but also they subjectively preferred the sound of the reverberation in the Guildhall.

Individual Comments

Following is a selection of the feedback provided by the four participants.

Participant 1

- They felt that the shape of the anechoic chamber and the position of the drum kit within the anechoic chamber differ visually from the shape and the simulated position of the drum kit within Hendrix Hall. Because of this, the left and right channels threw them off as the audio they were hearing didn't match the visuals they were seeing. They said it would be interesting to investigate the influence of Virtual Reality on this effect.
- They felt there was a strong left-leaning bias with more sound appearing to be coming from the left.
- They commented that the VDR created nice nuances, such as rim shots, and brought them to life. They expressed how the VDR effectively highlight how the acoustics can significantly change the sound of the drums and result in the increased enjoyment of playing them.

Participant 2

- They expressed how the reverberation is much more convincing in terms of the perceived realism when it turns with head movement, i.e. the head tracking is very effective. They were particularly complimentary of how well the head tracking worked and how much it added to the experience.

- They mentioned how the HRTFs weren't accurate enough which resulted in everything sounding like it was coming from behind them. They expressed how they would like to trial the VDR with personalised HRTFs.
- They mentioned how the close proximity of the back wall was very noticeable, which added to the realism (*Note that at this point the participants weren't aware of the simulated position of the drums within Hendrix Hall.*
- They expressed how their performance was massively affected by the levels of reverberation. They didn't comment specifically but it was implied that having reverberation positively influenced their performance.

Participant 2 also expressed their thoughts on the testing methodology:

- They felt that reading music requires cognitive load so their attention is focused on that rather than thinking about the sound.
- They mentioned that reading the sheet music while playing meant their head was in a fixed position and therefore wasn't experiencing the head tracking as much.
- They expressed how they found the rating system hard in Part A of the listening test. They wanted to give a rating of 1.5 at times, which wasn't possible. They also mentioned how they were reluctant to rate too highly in case a following configuration warranted a higher rating.

Participant 3

- They expressed how the directionality didn't seem to correspond with the microphone configuration, e.g. under the Front & Back configuration the sound appeared to be coming from the left and right (when their head was facing toward to front mike).
- They felt that the VDR was slightly limited by not having 6DOF (6 degrees of freedom) and that they would be keen to try the system with it incorporated.

- They mentioned how they would like to have an image of the room that they were virtually playing in as it would have let them relate what they were hearing to the room.
- They expressed how they noticed the walls they could hear in the simulation would change with different configurations and that they, therefore, felt that the configurations were affected by the specific room the SRIRs were measured in.
- Participant 3 also mentioned some issues with the head tracking, specifically that the sound would change disproportionately to their head movement as well as the rotation suddenly and randomly changing throughout playing.

Participant 4

- The participant felt that their timing while playing was improved with reverberation compared to performing in a studio which often has a dry sound.
- They expressed how they had thought of setting up a similar system at home but without the anechoic chamber and spatialisation.
- They mentioned that the snare and the hi-hat sounded very good in all the configurations but that the bass drum felt fluttery and sounded digital at times. Despite this (including the comments about the ride cymbal made above), they mentioned that when playing the whole kit, the sounds blended well and the things they didn't like about individual parts of the drum kit were suppressed.
- They expressed how the spatialisation sounded "natural" and "right" and how that although the spatialisation wasn't super prominent, they appreciated how it wasn't too obvious.

4.3.4 Summary

Four participants undertook the listening test and were all experienced drummers, having played for a significant amount of time, and were all considered expert listeners due to their profession.

Fig. 4.4 shows the average scores from Part A of the listening test where the three attributes of perceived naturalness, spaciousness and realism were compared to the mono condition. From the results, the Left & Right configurations performed the best achieving the highest average ratings for perceived naturalness and realism. Quad performed next best, achieving the highest average rating for the perceived realism which was also the highest average rating score. The Front & Back configuration performed the least well not having the highest average rating for any of the three attributes.

All four participants expressed a preference for the Quad configuration with the exception of one participant who shared a preference for the Quad and the Left & Right configurations. 50% of the participants believed that all three configurations were different compared to the mono condition and the other 50% participants believed that some of the configurations were the same as the mono condition and some were different.

The participants unanimously commented that they very much enjoyed using the VDR and would use it again if it were accessible to them. They all said after learning about the configurations that the Quad configuration was their preferred configuration, commenting that it felt like a "bigger" sound.

The individual comments were then recorded and presented in Section 4.3.3. The comments were extremely beneficial and would provide valuable insight into future iterations of the VDR.

4.4 Discussion

Due to the small sample size of participants undertaking the listening test, not too much weight can be given to the results presented above. While they can be discussed in the context of this project, it is important to remember this limitation.

In Part A of the listening, the Quad configuration was unanimously preferred with the exception of one participant who preferred the Quad and the Left & Right configurations equally. This is a notable result as at that point in the listening test, the participants were not aware of what configuration they were playing in, disregarding any sort of bias to a "more means better" mentality, in regards to the number of microphones. As discussed above, due to an insufficient amount of conditions, assessing whether an increase in the number of microphones used is preferred amongst the participants was not feasible. Although the results presented above can not provide much weight to indicate a trend, the fact that the Quad configuration was unanimously preferred could indicate the possibility of the notion that more microphones are preferred. Perhaps this is to be expected based on similar findings by Otondo (2004) [10]. Participants made comments such as "The Quad configuration sounded much bigger". To fully investigate this notion in the context of the VDR, further study is required. While this thesis does not delve into an exhaustive discussion of the subject, it is reasonable to anticipate that participants' preference for the Quad and Left & Right configurations over the Front & Back configuration stems from a combination of anatomical and psychoacoustic factors. Specifically, humans possess a heightened capacity for sound localisation in the frontal, left, and right directions. Therefore the Quad and Left & Right configurations provide a more spatially dynamic sound for the listeners which the results suggest is preferred.

One participant commented that they preferred the mono condition over the Front & Back configuration, however, besides this, no one had a preference for the mono condition. Additionally, all participants were not under the impression that all the configurations were the same as the mono condition. 50% of the participants believed that all three configurations were unique, with the other half believing some were unique and the others were the same as the mono condition. The overall preference for the Quad configuration and the lack of preference for the mono condition suggests that the spatialised auralisations are not only noticeable compared to non-spatialised auralisation (mono) but also preferred, promoting the motivation and effectiveness of the VDR. Additionally, the fact that the different configurations were mostly distinguishable, promotes the significance of different microphone configurations. It is important to note however, the participants were informed during Part A of the listening test about the mono condition as the control and the VDR was using SRIRs. This could have invoked a bias where the participants thought spatialised auralisations would sound better than mono (non-spatialised) auralisations. During Part B of the listening test, the participants were informed of the different microphone configurations and the Quad configuration still came out unanimously on top as the preferred configuration. However, once knowing the configurations a "bigger means better" mentality could have had an influence.

The results presented in Fig. 4.4 further support the preference for spatialised auralisations over non-spatialised auralisations. All three microphone configurations were on average rated more highly for all three attributes compared to the mono track. From initial inspection, this could be inferred as the three configurations perform better than the mono condition regarding the three attributes, but further inspection is required. From inspecting the individual ratings, it can be seen that the three configurations were not always given a positive number rating. There were six instances where a participant gave a rating of 0, meaning the attribute in question was perceived equally compared to the mono condition. There were three instances where a participant gave a negative number rating, however, two of these instances were given by one participant so it could be considered an anomaly. Moreover, this participant had problems with the head tracker cutting out during their listening test, which could the explain said anomaly. Nevertheless, individual ratings show that the configurations do not always rate higher than the mono condition across the three attributes.

Despite how increasing the number of microphones used affects the VDRs performance was not under investigation here, it is interesting to note that the results shown in Fig. 4.4 do not align with the results found by Otondo, which are highlighted in Section 2.3.1. Otondo observed that employing a higher number of microphones to capture the sound source leads to improved naturalness and timbre quality. It would have been reasonable to hypothesise that a similar effect will be observed in this project. However, the parameters were different for both studies and multiple factors could have influenced this such as the instrument used and more importantly the fact that the study by Otondo incorporated many more microphones and thus more data. This study had very little data so not much weight can be given nor are the results reliable.

The combination of the results presented in Fig. 4.4 and the overall preference for the Quad configuration suggest that the three attribute ratings don't necessarily correlate to the most preferred configuration. The Left & Right configuration was rated the highest for two out of three attributes (perceived naturalness and realism) but was not the preference of the participants. The Quad configuration was rated the highest for perceived spaciousness, so perhaps spaciousness is related to preference but there is little data to support this notion and further study would be required. The Quad configuration having the highest rating for perceived spaciousness is perhaps to be expected - the greater the number of mics, the more of the directivity field is captured, and the more SRIRs are used in simulating the space, leading to a feeling of more space. The Front & Back configuration was not rated the highest for any of the three attributes and, although no participant stated that it was their least preferred configuration, some participants made comments about how it sounded narrower compared to the other configurations, despite the perceived spaciousness ratings suggesting differently. With the exception of one participant who gave an equal rating to all three attributes across the three configurations, the ratings for the three attributes were not proportional to each other across the three configurations. This could suggest that the three attributes aren't necessarily connected, although this wasn't predicted. Future studies could investigate what factors affect the perception of such attributes and which attributes are linked with enjoyment and pleasantness.

When looking at the results in Fig. 4.4, the average ratings are pejoratively 1.5 or less, with the exception of one rating which is 2. This shows that despite the average ratings all being greater than 0, they are fairly low ratings (considering the highest rating is 3). This could mean that although the three configurations were rated higher than the mono condition for the three attributes, it was not by a large margin and therefore not too much weight should be given. However, a caveat to this notion is a point that one participant specifically mentioned in Part B of their listening test - the rating system was problematic as they wanted to give higher scores but were aware that a better configuration may be to come so rated conservatively. If this was felt by all participants then it could explain why the average ratings are greater than 0 but relatively low. This was a consideration when designing the listening test as without an upper bound being presented the participants may not know when to give the highest rating.

A factor that could have influenced the results presented in Fig. 4.4 is that the three attributes could have been open to interpretation. The definitions of the attributes as well as the scoring system were loosely defined to the participants prior to the listening test beginning, so although they had an indication of the terminology, their choices of ratings could have been subjective. There were no audio examples to provide examples of the definitions of the attributes - perhaps an addition to future studies. Moreover, judging and comparing different attributes of audio is naturally complex and will involve many factors that create difficulty in allowing the participants to answer effectively themselves and consistently across a group. In any such listening test, these factors must be considered when drawing conclusions from the results.

One thing that stood out from the feedback received in Part B of the listening test, was the comments about the positioning of the SRIR measurements in Hendrix Hall. It appears that the position where the SRIRs were measured in Hendrix Hall affects the configurations presented to the participants. The position of the SRIR measurements was chosen not because of any compelling reason. The room was not completely free to position anywhere as there was a seating rig in the way and a lectern. As it had been a performance space, the measurements were positioned where a drum kit could likely be placed, i.e. off-centre (in this case it was stage left as the lectern was stage right). The choice of positioning itself is not discredited, however, perhaps some of the implications of the choice were not thought about. One comment was made about how there was a mismatch between the position of the drums in the anechoic chamber (centre of the room) and the position the drums were simulated in within Hendrix Hall. This was not really something that could have been worked around regarding the positioning of the drum kit in the anechoic chamber, but positioning in Hendrix Hall could have. The participant commented on the mismatch in the visuals of the room they were in with the audio of the room they had simulated. This could have thrown them off when rating the comparisons in Part A but also could have taken away from the enjoyment of the experience so would need to be considered in future iterations. They expressed how using Virtual Reality (VR) could help mitigate this effect, however, VR was beyond the scope of the project as well as not a desire for the VDR. Participants also commented on how the positioning of the SRIR measurements also influenced the different microphone configurations. The closer proximity to the left wall compared to the right made the auralisation very left-biased. In fact, the left channel was usually the loudest channel which could have affected the results as the Front & Back configuration was the only configuration without the left channel. The positioning of the SRIR measurements was also very close to the back wall which could have affected the results depending on people's preference for sound. Changing the configurations affected the sound more so and in different ways than originally thought. The flutter echo effect that was present in the room seemed to be predominantly coming from the left and right wall, therefore when the left and right channels were not active, this effect was reduced which could have affected the results. Ultimately the different configurations clearly are noticeable and have an effect on the attributes measured and the personal preference of the VDR. However, it seems that the results are dependent on variables such as SRIR measurement positioning. Perhaps if the SRIRs were taken in a different position, the results would be different. This is an important factor to consider in future iterations and means that conclusions from these results can only be made in the context of the exact positioning and location of the SRIRs taken and not generally.

Participant 2 commented on the evaluation methodology of the listening test. They discussed how the concentration on reading and playing the sheet music provided was distracting from thinking about and being aware of the changes in the sound of the different configurations. Although only mentioned by one participant, it likely affected all four. This could have caused participants not to have been paying attention to the sound and given made-up ratings within the listening test. The concentration of reading the sheet music also caused most of the participants to not move their heads whilst playing, which Participant 2 also commented on, causing them not to fully experience the head tracking over a wider range. This is also likely to have influenced the results.

4.5 Limitations

When assessing the VDR in terms of whether the number of microphones improves the experience, the spatialisation is noticeable and whether the VDR is enjoyable to use for recreational purposes, the results presented in Section 4.3 appear to indicate that the VDR was overall a success. However, as with any system in the early stages of research and development, there are limitations to consider. The limitations discussed in this section are both ones that were known prior to building and evaluating the VDR and ones that were discovered whilst building and evaluating the VDR.

The listening test used to evaluate the VDR involved judgements made by human listeners interacting with technology. Naturally, listeners are susceptible to a range of limitations that can affect the accuracy and reliability of the results. Equally, the results can be affected by the integrity and design of the technology used to perform the test. It is important to consider limitations that can arise from perceptual listening tests to ensure the reliability and generalisability of the results.

Although unlikely to have caused significant disruption to the results obtained from the listening test, the unfortunate effect of the background noise present in the original sine sweep recordings created the tail end of the reverberation from the auralisation to sound unnatural. As discussed previously, this mainly was unnoticeable when playing multiple successive hits on the drums, i.e. when playing a groove, however, this reduction in naturalness could have impacted the results and is essential to take into consideration, particularly with the perceived naturalness ratings. The fact that both participants who used the VDR with the Hendrix Hall SRIRs and the Guildhall SRIRs stated they preferred the sound of the Guildhall SRIRs, as Section 4.4, could further support this.

Perhaps one of the more significant limitations to consider regarding the evaluation of the VDR was the small number of participants who undertook the listening test. Only four participants undertook the listening test, which produced a very limited set of results. There were only four participants due to factors causing time restraints and limited access to drummers. A small sample size can lead to unreliable or invalid results due to limited representation and chance variability. Limited representation, can lead to biased results that do not accurately reflect the perceptions of the larger population. Chance variability is where random variation in the responses of just a few participants can have a significant impact on the overall results. Although not too much weight can be given to the results presented in Section 4.3, the results can still give some indication of the success of the VDR and provide insight into future work and adaptions to the methodology of both the design of the VDR and the evaluation of the VDR.

Perhaps one of the more significant limitations to consider regarding the design of the VDR was the number of microphones used to capture the drum kit and therefore the number of SRIRs taken. Initially, the proposal was for the VDR to have the capability of having an 8-channel and a 6-channel configuration on top of the configurations that were used in the final build of the VDR. This would allow for more vigorous testing of the effect of how an increase in microphones and therefore a better spatially sampled drum kit would affect the experience of performing using the VDR, i.e. "does more equal better". However, in order for the 2, 4, 6 and 8-channel configurations to have 180, 90, 60, and 45 degrees separation respectively for the associated microphones, there would need to be a total of 12 microphones, and therefore 12 SRIRs. To record 12 SRIRs was not problematic, however having 12 microphones within the anechoic chamber was. As mentioned previously, the drum kit was placed on an MDF platform within the anechoic chamber. Because of the dimensions and the orientation of the platform, it would not have been large enough to accommodate all 12 microphones. The mesh floor of the anechoic chamber was not sturdy enough to support microphones. Therefore only 4 microphones were used. This resulted in a limited evaluation of the performance of the VDR regarding the number of microphones capturing the drum kit. If the 2, 4, 6 and 8-channel configurations were able to be tested this would provide four cases of configurations with differing numbers of microphones. This would have allowed for more robust testing on the effect of adding more microphones. However, only 4 microphones were used in the final build. There were only two cases of configurations with differing numbers of microphones. This did not provide enough cases to effectively test this variable. However, although this variable wasn't able to be tested as thoroughly as previously thought, other factors were possible to evaluate and yield beneficial results.

A limitation regarding the final VDR was the HRTFs were not personalised for the participants. Although each participant would have experienced the same HRTFs, each participant would have had a different experience. Some participants' ears may be more closely to matched the HRTFs sued than others incurring a bias. One participant commented on how the HRTFs were not personalised and not working well for them. Although perhaps not dramatically, this will have affected the results in some way. The participant who pointed out the problems with HRTFs was an expert listener and familiar with HRTFs, perhaps a less expert listener drummer would be unaware of the issues. Using personalised HRTFs would have been impractical for this study as the measure of personalised HRTFs would have been beyond the scope of the project in terms of the time frame and technology available.

As mentioned in Section 4.4, the head tracker was troublesome in that the connection between it and the computer was sporadic and kept having to be reconnected. This was particularly the case for Participant 3. This created a significant limitation for the study as the issues will have likely affected the study's results, although as mentioned the points at which the head tracker stopped transmitting data are unknown so it is unknown exactly where the results have been affected.

As mentioned previously, the head tracker was prone to cutting out while the participants were undertaking the listening test. The problem was mitigated as much as possible by reconnecting the head tracker frequently during the listening test, however, it is still important to consider the effects on the results and the possibility of skewing them. The head tracker was particularly troublesome with one participant. As seen in Section 4.3.3, the participant commented on this in Part B of the listening test where they remarked that the sound would change disproportionately with their head movement as well as the rotation suddenly changing throughout playing despite their head not. Unfortunately, it was not known when these problems exactly occurred so it is difficult to conclude whether they would have affected their perceptions and thus ratings. As a result of this, it was thought that it would have been wise for another pair of headphones to be available for the listening test administrator to wear so they could hear what the participant was hearing. This is perhaps a consideration for future listening tests.

A factor that could have affected the results was the volume level at which the auralisation was played back via headphones to the participants. As mentioned in Chapter 3, the level of the auralisation playback was not set at a level that aimed to be identical to Hendrix Hall. However, because the level was set subjectively by the listening test administrator, this could have caused a bias in the sound levels and different participants could have responded differently to the listening test based on their preferences. In fact, one participant did ask if the level of the headphone playback could be increased but they were instructed this was not possible in order to keep the listening test unbiased. Another participant wore earplugs during their listening test, and although they stated that the earplugs used had a flat frequency response and wouldn't cut out any high-end, it was not known exactly how they would affect what the participant was hearing. Ideally, the level of the headphone playback would have been set to exactly replicate Hendrix Halls, as this would decrease any chance of bias. However, as stated before, this is extremely difficult to achieve. Because the level was set subjectively, a bias was incurred and could have affected the results. How the results would have been affected isn't exactly known, but it is an important factor to consider.

Subsequent to the listening tests, a realisation emerged concerning the height at which the microphones were situated, revealing that this placement might not have been optimal for the intended VDR performance. The microphones were positioned at a fixed elevation of

140cm, while the drum kit maintained an average height of 100cm, accounting for the various components of the instrument. Consequently, a significant disparity existed in the horizontal planes of the drum kit and the microphones. It's important to note that the height at which a microphone is positioned, influences the frequency balance of the captured sound. Specifically, higher microphone placement tends to accentuate high-frequency components, while lower positioning accentuates low-frequency elements. Given that the microphones were elevated above the drum kit, the auralisation playback was affected by this phenomenon. The rationale behind situating the microphones at this height pertained to achieving alignment between the source and receiver heights in the SRIR measurements, where the listener's head height served as the basis. This was intended to enhance the authenticity of the auralisation; however, it appears that the underlying principle was misinterpreted. In retrospect, it might be more prudent for the microphone's height position to dictate the source's placement, rather than the inverse. Revisiting the approach, it becomes evident that situating the microphones at a lower height could be more conducive to capturing the drum sound with a balanced frequency response. This adjustment could be incorporated into the SRIRs, with the source positions also lowered accordingly while maintaining the receiver's position at 140cm. When viewed through the lens of the SRIRs, the challenge becomes evident: the auralistion envisions the drum sound emanating from sources aligned with the listener's head height, projecting into the room at the same elevation. Yet, the drum components themselves do not occupy this height. Consequently, the final auralisation construction fails to faithfully replicate the drum sound's true spatial characteristics within the room and does not optimise the sound capture process. A more comprehensive exploration is warranted to gain a deeper understanding of this intricate interplay and to arrive at a well-informed resolution.

4.6 Conclusion

While conducted with a limited sample size, the listening test offers valuable insights into the VDRs performance. Although the data may not carry significant weight due to the small sample, the results remain key in achieving the objectives of the evaluation of the VDR, namely factors 1, 2 and 3. These objectives were achieved in this chapter and can deemed a success. Not only did the results from the listening test provide extremely beneficial results in terms of how the microphone configurations altered the experience of using the VDR, but they also have been extremely productive for insight into the perceptual aspects related to human auditory experiences as well as improvements and alterations into the VDR and the test paradigm.

The results indicate that the technology and methodology used were successful in creating a virtual acoustic environment for drummers that provided spatial information to the player. The results also indicate that the use of spatial information in convolution reverberation is preferred over the alternative (non-spatialised). The unanimously preferred configuration was Quad and all three configurations were on average rated more highly regarding the perceived naturalness, spaciousness and realism compared to the mono condition (non-spatialised).

Chapter 5

Conclusion

This thesis has presented a project aiming to develop and create a spatially informative real-time auralisation system for the drums. The proposed system is named the Virtual Drum Room (VDR) and was successfully built and evaluated throughout the project. This chapter concludes the project by providing a summary of the thesis, a restatement of the aims and objectives introduced in chapter 1, a discussion of the time management of the project and a discussion of future work.

5.1 Summary

Chapter 1 introduced the project by providing contextualisation to the VDR. Room acoustics and their influence on the drums were discussed as well as people's desire to create and experience the resulting reverberation. The current ways to experience or create reverberation for drums and their limitations were outlined followed by the aims and objectives for the project which were to develop a system that can simulate acoustic environments virtually by exploring real-time auralisation techniques that incorporate spatial information. The structure of the thesis was given with a summary for each chapter.

A comprehensive literature review is given in Chapter 2 which is used to inform the methodology for the VDRs development and build. The chapter begins by discussing how room acoustics are simulated through auralisation. Auralisation as an overall concept is explained as well as its process. It is broken down into an auralisation chain which consists of impulse response measurement, capturing of the source, convolution and rendering & reproduction. Following this, auralisation for musical performance is contextualised and examples of previous studies that have used these techniques are presented and discussed. Considerations for interactive room acoustic simulation for performance are also discussed such as latency and SRIR editing. Finally, two papers that are particularly relevant and provide inspiration to this project (Otondo & Rindel [17] and Brereton [19]) are spotlighted and discussed. The literature review highlights that techniques and methodologies for real-time auralisation for music performance are well established, however, specific cases for the drum kit are rare if not nonexistent. Despite this, previous research presented in this chapter is sufficient to provide a methodology that will be used for the VDR built in this project. The literature review highlights the importance of directivity when creating an interactive virtual acoustic environment for music performance and the benefit of multi-channel auralisation to capture and replicate this.

Chapter 3 provides comprehensive documentation on the design and implementation of the VDR covering the initial testing phase and the final build. The initial testing phase aimed to give a proof of principle for the VDR as well as an opportunity to identify any alterations that would be required of the design for the final build. Fortunately, there were very few alterations that were necessary. The only one of note was attempting to find a more sturdy of positioning the drum kit on the wire mesh floor of the anechoic chamber. Fortunately, this was solved by constructing an MDF platform. The initial testing phase consisted of building a scaled-down version of the VDR, which consisted of two microphones rather than the four used in the final build. The impulse responses for the final build had not been measured at this stage so a set of impulse responses measured at the Guildhall were used. These SRIRs were not measured specifically for this VDR so there were discrepancies in the design, however, these did not appear to cause significant issues. The final build had the addition of two other microphones, creating a four-channel auralisation system. There was also the addition of head-tracking technology. Both the initial testing phase and the final build were built around the same design of a drum kit recorded in an anechoic chamber using a multi-channel auralisation approach. The captured sound was convolved with SRIRs using the X-MCFX convolver within Reaper. The resulting auralisation was played back via headphones. For the final build, specific SRIRs were measured in Hendrix Hall and PL/001 which were tailored specifically for the VDRs microphone arrangement. For the final build the VDR comprised of three microphone configurations that were to be evaluated, which were Front & Back, Left & Right and Quad.

Chapter 4 presents the methodologies used to evaluate the VDR and provides the results of the evaluation as well as a discussion of the results and limitations of the evaluation and the system as a whole. The evaluation of the VDR was conducted via a listening test in which four drummers partook. The aims and objectives of the evaluation were to gain a subjective analysis of the system rather than an objective analysis. Participants' experiences and preferences were recorded and analysed as well as their opinions on how the different microphone configurations affected the perceived spaciousness, naturalness and realism of the VDR. The chapter provides a detailed account of the experimental protocol of the listening test. The results from the listening test are provided and then discussed. The results indicate that the VDR is well enjoyed amongst drummers and that they would desire to use it as a recreational tool if it were accessible to them. The participants preferred the Quad configuration overall and on average rated all the configurations higher in terms of the three attributes of perceived naturalness, spaciousness and realism compared to the mono condition which contained no spatial information. This suggested that the use of spatial audio to simulate room acoustics was preferred and supports the motivation behind this project. Sections 4.4 and 4.5 provided a discussion of the results and points were raised regarding the reliability and representation of the results as well as limitations, both anticipated and unanticipated, of the VDRs performance and the evaluation methodology.

5.2 Restatement of Aims & Objectives

This section recaps the aims and objectives of the project and compares them with the achievements. The initial aim was to create a spatially informative real-time auralisation system for the drums. Four objectives were created which are presented in Chapter 1. Here are their recaps and post-project evaluations:

1. Explore real-time auralisation methodologies and technologies for musical applications and specifically investigate source capture and impulse response techniques for the purpose of real-time auralisation for musical performance.

    - After The objective of exploring real-time auralisation methodologies and technologies for musical applications, with a specific focus on investigating source capture and impulse response techniques for real-time auralisation in musical performance, has been successfully met through this comprehensive literature review (presented in Chapter 2. The review has provided valuable insights that served as a guiding foundation for the subsequent chapters of the project. It not only defined and elucidated the concept of simulating room acoustics through auralisation but also delved into the intricacies of interactive room acoustic simulation for musical performance. Of particular significance was the investigation into methodologies for capturing source material, specifically drums, and measuring the corresponding SRIRs. This examination emphasised the critical importance of capturing and replicating the directivity of musical instruments to achieve an immersive and realistic virtual acoustic experience. While the literature review revealed a dearth of methodologies tailored to auralisation for drums and percussion, it identified the work of Otondo and Rindel [17] as the most relevant for this project's objectives. Consequently, their methodologies served as a key source of inspiration in the design of source capture and SRIR measurement techniques.

2. Design and build the VDR informed by the knowledge gained from the literature review.

    - The objective of designing and building the VDR, guided by the knowledge acquired from the comprehensive literature review, has been effectively achieved. Chapter 3 meticulously outlines the process of developing the VDR, with careful consideration of insights gleaned from the literature review. The initial testing phase, a crucial part of the design process, successfully established the proof of concept and ensured the effectiveness of the chosen design. It also provided an opportunity to identify and address any necessary alterations or additions. Subsequently, the final build stage translated the refined design into a functional system ready for evaluation. Notably, some features, such as head-tracking and project-specific SRIRs, were reserved for implementation during the final build. Both phases—initial testing and final build—have been successful in their respective objectives, culminating in the realisation of the designed system and the delivery of a high-quality user experience.

3. Evaluate the VDR to determine its success and performance via perceptual listening tests.

    - Chapter 4 provides a detailed account of the experimental protocol that was used to perform a perceptual listening test. The perceptual listening test was designed to determine the performance and the success of the VDR through subjective methods. The participants trialled the VDR and rated certain attributes against a control. The participants were also asked for their individual verbal feedback on their experience of using the VDR. Objective 3 was successfully achieved with the perceptual listening test that was designed and carried out. The combination of qualitative and quantitative data from the listening test facilitated a thorough assessment which allowed the success and performance of the VDR to be determined within the confines of the limited sample size of participants that was used.

4. Analyse the results obtained from the evaluation.

    - The analysis of the results is presented in Chapter 4. The results were successfully analysed resulting the objective 4 being successfully achieved. The results are limited by the sample size used so not much weight can be given to them, however, they still provide valuable insights into how the VDR performed and was perceived by drummers, as well as highlighting areas of improvement for future works. The results indicate that the VDR was an overall success. All participants very much enjoyed using the VDR and desired to use it again if the opportunity arose. The participants preferred the spatialised configurations over the non-spatialised configuration (mono), rating them all higher for the three attributes on average. This supports the notion that such a system as this which incorporates spatialised auralisations will be enjoyed as well as unique to users compared to non-spatialised auralisation.

5.3 Time Management

Below, two Gantt charts are presented: Figure 5.2 illustrates the initially proposed project timeline, while Figure 5.1 showcases the realised project schedule. These charts commence with the design and implementation phase and extend all the way to the final thesis submission. Upon comparing the two charts, noticeable disparities emerge. The initial project timeline was devised not only to establish project structure but also to serve as a tool for monitoring the achievement of time-related objectives. In general, the project adhered well to the proposed timeline; however, a few exceptions did arise

Upon initial examination, the two project plans may appear somewhat distinct; however, they share a common foundational structure encompassing development, construction, evaluation, and documentation phases. Formulating a proposed timeline presents a challenge, as accurately predicting the duration of specific tasks and accounting for unforeseen external factors that can potentially disrupt the schedule can be inherently complex.

In the initial project plan, the first five weeks (S10 to V4) were allocated for VDRs development, although the specific breakdown of these weeks underwent significant modifications. The development stage in the initial plan was intentionally left somewhat vague, given that the methodology had not been fully outlined and was expected to evolve. The actual implementation of the development stage can be observed in Figure 5.1. To illustrate some of the changes, for instance, the timing of impulse response measurements shifted from V2 to V3 due to room availability constraints. The listening test remained scheduled for week V6, but the experimental protocol design was extended over two weeks instead of one. Furthermore, the final build construction coincided with the listening test week because the time required for each phase was less than initially anticipated. The analysis of results also took less time, spanning only week V8 instead of the originally speculated weeks V7 and V8, as indicated in Figure 5.2. This adjustment was partly influenced by a one-week holiday, marked in week V7 in Figure 5.1. While Figure 5.2 implied that note-taking would be an ongoing process throughout the project, this was not the case. Instead, notes were primarily recorded after the experiment was conducted, with the write-up process commencing earlier than initially anticipated, as evident in Figure 5.1. The write-up phase persisted from post-experiment stages through to the final hand-in.

Comparing these two Gantt charts has shed light on areas where project optimisation is possible. While it was intentional to maintain a certain level of vagueness in the development stage, in retrospect, a more detailed understanding of the development phase would have been beneficial. This hindsight is based on the observation that there were several weeks during this stage when practical activities were halted due to dependencies on the availability of specific resources, such as the anechoic chamber or rooms for SRIR measurements. Although these weeks were productively utilised for tasks like writing up notes, better advanced planning could have averted these downtime periods.

In summary, the project's time management can be deemed successful. Despite variations between the two project plans, the fundamental sequence of development, construction, evaluation, and documentation persevered. These modifications encompassed both anticipated and unanticipated situations. Nevertheless, in the face of unforeseen challenges, adjustments to the project plan were promptly executed to maintain efficiency. During periods of abundant time, as exemplified in the development phase, it was carefully utilised to prevent an accumulation of tasks toward the end of the project.

5.4 Future Work

The nature of this project was to design and develop the VDR, which can be thought of as a prototype. Naturally, despite the aims and objectives being accomplished at a high level, this will mean there are factors that have either come to light throughout the project or factors that were not possible to consider or take into account due to the time and resource limitations of the project. This section will discuss some of these factors that would be acted upon for future renditions and iterations of the VDR and its evaluation if it were to be repeated or continued in the future.

As mentioned throughout this thesis, the number of microphones used in the VDR was limited by logistical constraints revolving around the anechoic chamber. It would be desirable to incorporate more microphones into a future VDR not only to hear how the sound differs but to enable or more thorough evaluation of how it affects perceptions and experiences. It would be interesting to see whether a positive correlation between perceived naturalness and the number of microphones was present just as Otondo [17] observed. If resources and logistics weren't a concern, adding a 6 and 8-channel configuration in the horizontal plane would likely be the priority as this would be the simplest to incorporate. Configurations with more than 8-channels would likely need some sort of rig to be built around the drum kit rather than the microphones being on individual stands. Measuring the SRIRs wouldn't be as difficult as they are done individually so spaces aren't a concern, although the time factor would be increased but considering a sine sweep measurement takes about one minute, this wouldn't be too significant. Incorporating microphones and resulting SRIRs in the vertical plane, i.e. positions with non-zero elevation angles, would also be something of interest however this would require more time and thought as the microphones and speakers would have to be angled and mounted in awkward positions - some sort of rig to mount the equipment would be necessary.

The results in Chapter 3 support the notion that the VDR is enjoyed by its users as a recreational tool and that the users would want to use it again if they access it. Perhaps the largest limitation of the VDR being used as a recreational tool is the use of the anechoic chamber. Anechoic chambers are rare and are mostly owned by research institutes such as universities. This makes the idea of the current VDR being used as a recreational tool completely impractical. It would therefore be interesting to redesign the VDR for non-echoic use, which could involve the VDR being situated in a studio that would be semi-echoic. The effects of using a non-anechoic are not known but it could create re-auralisation issues, as discussed in Chapter 2. It would be beneficial to investigate whether this was an issue and how significant it was as it would determine the likelihood of people being able to use a similar system in the future. If time permitted, the VDR would have been rebuilt in a semi-echoic room (a studio), however, this was not possible for this project.

As this project also sought to assess the effectiveness of the evaluation methodology of the VDR, future versions would introduce new experimental protocols and redesign the listening test. Incorporating more microphones, for example, would require a redesigned listening test.

Future renditions would seek to streamline the evaluation process by revising the discussion and limitations sections of this thesis.

Other things to incorporate into future works also include:

- Take directivity measurements of the drum kit to compare how well effectively the difference configurations replicate this and determine whether this VDR is appropriately designed for this. This would allow for a more objective analysis of the VDR rather than a subjective analysis.
- Adding personalised HRTFs for the users. This would require them to have a set of HRTFs ready as measuring HRTFs is a significant process.
- The VDR could be incorporated into recordings of drums as some participants mentioned how the VDR positively affected their playing.
- Measuring SRIRs in more spaces. This could include indoor and outdoor spaces.
- As mentioned in Chapter 4, it was questioned whether the height of the microphones and sources in the SRIR measurements was the optimum choice for the VDR. In future works, it would be intriguing to alter the heights and determine whether a difference in the sound and experience is noticeable.

5.5 Personal Reflection

I have found immense satisfaction in every facet of this individual research project. It afforded me the opportunity to seamlessly integrate the various elements of the master's course that I am most passionate about, including signal processing, virtual acoustics, spatial audio, and psychoacoustics. The synergy of these fields with my deep-rooted love for drumming elevated the entire experience.

The VDR not only met but exceeded my initial expectations in terms of its performance and sound quality. It left me yearning to engage with it for hours on end, and I was delighted to witness the enthusiasm it garnered from others as well. I genuinely wish I could have extended this project further, delving deeper into its development and conducting more comprehensive evaluations.

In terms of project management, I am pleased to report that everything unfolded smoothly and as planned. There was no last-minute rush, and I effectively managed all aspects of the work. The only segment that felt somewhat rushed was the listening test; however, it was an informal pilot test and did not require a rigorous approach.

One disappointment in the project was the presence of background noise in the measured SRIRs, which affected the quality compared to the Guildhall SRIRs. Ideally, I would have repeated these measurements in a different space or with a microphone having a lower noise floor to mitigate this issue.

Another setback was the occasional interruption of the head tracker, which could have potentially interfered with one participant's results. For future iterations, there is a significant scope for improvement. This includes taking SRIR measurements in various locations and employing multiple microphones to capture drum sounds and SRIRs more comprehensively.

I believe that drummers could find value in using this VDR themselves, as evidenced by the positive results from our listening tests. The only drawback lies in the impracticality of relying on an anechoic chamber. Exploring the VDRs performance in non-anechoic conditions, such as a semi-echoic studio, could provide valuable insights. If successful, it would make the VDR more accessible for drummers, expanding its usability beyond the confines of a fully anechoic environment.

In sum, I view this project as a resounding success. It has been a source of enjoyment for me, both in its theoretical and practical dimensions, enabling me to fortify my expertise in signal processing and virtual acoustics while yielding promising results.

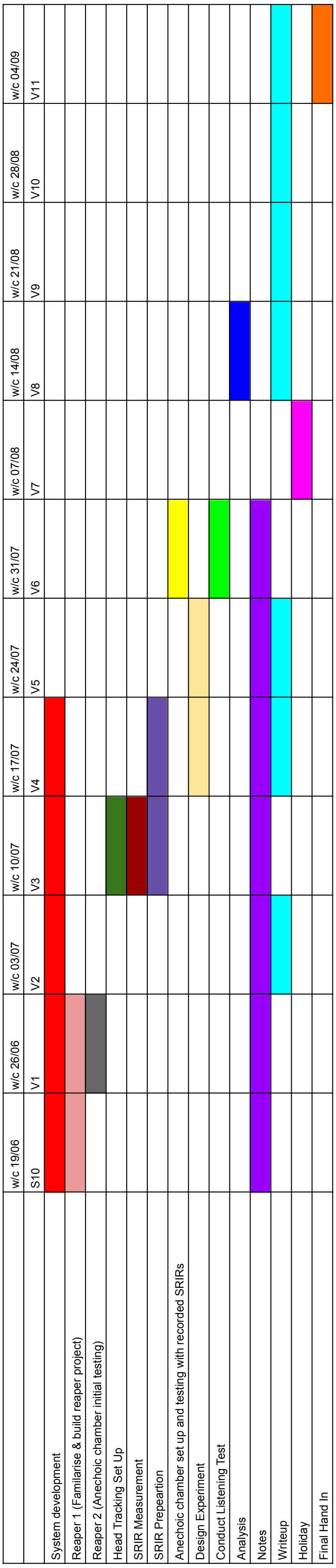


Figure 5.1: Gantt chart for the actual timeline of the project.

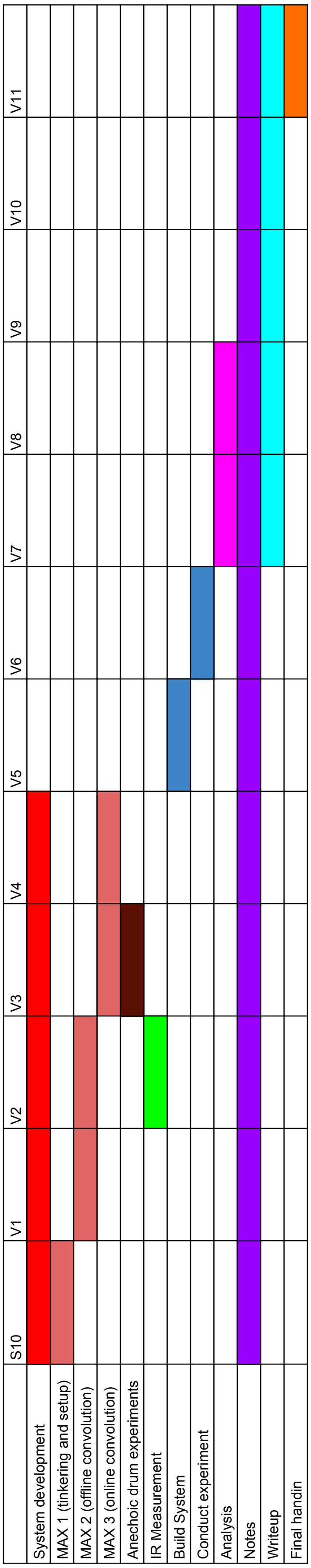


Figure 5.2: Gantt chart for the proposed timeline of the project.

Appendix A

Digital Appendices

This appendix is a collection of digital files which can be located in the accompanying zip folder titled "*Digital Appendices*" which is provided with this report. Included in this appendix are a collection of sub-folders that contain the various Matlab functions and scripts, the Reaper project, listening test associated files, the SRIRs and a collection of images. A video is also provided within the digital appendices named "*VDR Demo.mp4* as mentioned in Section 3.4.

Found in the "Digital Appendices" folder are:

The "Matlab Files" subfolder containing:

- *directTruncate.m*
- *truncateScript.m*
- *generatesweep.m*
- *deconvolveScript.m*
- *deconvolve.m*
- *removeDirect.m*
- *removeDirectScript.m*
- *fadeIR.m*

- *RichieIRs.m*

The “Reaper Project” sub-folder containing:

- *Drumming in Spaces.RPP*
- *Reaper User Guide.pdf*

The “Listening Test” sub-folder containing:

- *Literature Review.pdf*
- *Fast Track Ethics Form.pdf*
- *Listening Test Part B Feedback.pdf*
- *Listening test consent form.pdf*
- *Part A Individual Scores Part A.pdf*

The “Impulse Responses” sub-folder. The sub-folder is too large to list all the files within. This sub-folder contains all the impulse response recordings taken as part of the project right from the raw recordings of the Eigenmike right through the processed IRs used in the final build.

The “Images” sub-folder contains a number of photos taken throughout the SRIR measurement stage in PL/001 and Hendrix Hall.

References

[1] A. Marshall, "Acoustical determinants for the architectural design of concert halls," *Architectural Science Review*, vol. 11, no. 3, pp. 81–87, 1968.

[2] K. Ueno, K. Kato, and K. Kawai, "Effect of room acoustics on musicians' performance. part i: Experimental investigation with a conceptual model," *Acta Acustica united with Acustica*, vol. 96, no. 3, pp. 505–515, 2010.

[3] G. Kearney, "Auditory scene synthesis using virtual acoustic recording and reproduction," 2010.

[4] "Openair library." [Online]. Available: https://www.openair.hosted.york.ac.uk

[5] A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Audio engineering society convention 108.* Audio Engineering Society, 2000.

[6] P. Fausti and A. Farina, "Acoustic measurements in opera houses: comparison between different techniques and equipment," *Journal of Sound and Vibration*, vol. 232, no. 1, pp. 213–229, 2000.

[7] Z. S. Kalkandjiev and S. Weinzierl, "The influence of room acoustics on solo music performance: An empirical case study," *Acta Acustica united with Acustica*, vol. 99, no. 3, pp. 433–441, 2013.

[8] J. Merimaa, T. Peltonen, and T. Lokki, "Concert hall impulse responses-pori, finland," *Report and data available at: http://www. acoustics. hut. fi/projects/poririrs*, 2005.

[9] A. Farina, P. Martignon, A. Azzali, and A. Capra, “Listening tests performed inside a virtual room acoustic simulator,” *I seminario Música Ciênca e Tecnologia” Acústica Musical”, São Paolo do Brasil*, pp. 3–5, 2004.

[10] F. Otondo and J. H. Rindel, “The influence of the directivity of musical instruments in a room,” *Acta acustica united with Acustica*, vol. 90, no. 6, pp. 1178–1184, 2004.

[11] L. M. Wang and M. C. Vigeant, “Objective and subjective evaluation of the use of directional sound sources in auralizations,” 2004.

[12] R. Kessler, “An optimized method for capturing multidimensional “acoustic fingerprints”,” in *Audio Engineering Society Convention 118*. Audio Engineering Society, 2005.

[13] B. Albrecht, D. de Vries, R. Jacques, and F. Melchior, “An approach for multichannel recording and reproduction of sound source directivity,” in *Audio Engineering Society Convention 119*. Audio Engineering Society, 2005.

[14] L. Savioja, J. Huopaniemi, T. Lokki, and R. Väänänen, “Creating interactive virtual acoustic environments,” *Journal of the Audio Engineering Society*, vol. 47, no. 9, pp. 675–705, 1999.

[15] F. Otondo, J. H. Rindel, and C. L. Christensen, “Directional patterns and recordings of musical instruments in auralizations,” in *Proceedings of the MOSART workshop on Current Research Directions in Computer Music*. Citeseer, 2001, pp. 230–232.

[16] M. C. Vigeant, L. M. Wang, and J. H. Rindel, “Investigations of multi-channel auralization technique for solo instruments and orchestra,” 2007.

[17] F. Otondo and J. H. Rindel, “A new method for the radiation representation of musical instruments in auralizations,” *Acta Acustica united with Acustica*, vol. 91, no. 5, pp. 902–906, 2005.

[18] M. C. Vigeant, L. M. Wang, and J. Holger Rindel, “Investigations of orchestra auralizations using the multi-channel multi-source auralization technique,” *Acta Acustica united with Acustica*, vol. 94, no. 6, pp. 866–882, 2008.

[19] J. Brereton, "Singing in space (s): singing performance in real and virtual acoustic environments: singers' evaluation, performance analysis and listeners' perception," Ph.D. dissertation, University of York, 2014.

[20] T. Lokki and J. Pätynen, "Applying anechoic recordings in auralization," in *The EAA symposium on auralization, Espoo, Finland*, 2009.

[21] M. C. Vigeant, L. M. Wang, and J. H. Rindel, "Room acoustics computer modeling: study of the effect of source directivity on auralizations," pp. 1–9, 2006.

[22] ——, "Objective and subjective evaluations of the multi-channel auralization technique as applied to solo instruments," *Applied Acoustics*, vol. 72, no. 6, pp. 311–323, 2011.

[23] J. H. Rindel, F. Otondo, and C. L. Christensen, "Sound source representation for auralization," in *International Symposium on Room Acoustics: Design and Science. Hyogo, Japan.* Citeseer, 2004.

[24] I. Iaird, D. Murphy, P. Chapman, and S. Jouan, "Development of a virtual performance studio with application of virtual acoustic recording methods," in *Audio Engineering Society Convention 130.* Audio Engineering Society, 2011.

[25] S. V. Amengual, D. Eddy, M. Kob, and T. Lokki, "Real-time auralization of room acoustics for the study of live music performance," *Fortschritte der Akustik–DAGA*, vol. 42, 2016.

[26] M. Gorzel, G. Kearney, F. Boland, and H. Rice, "Virtual acoustic recording: An interactive approach," in *Proc. of the 13th internation conference on Digital Audio Effects (DAFX). Graz, Austria*, 2010.

[27] M. Vorländer, *Auralization: fundamentals of acoustics, modelling, simulation, algorithms and acoustic virtual reality.* Springer Nature, 2020.

[28] M. A. Gerzon, "Practical periphony: The reproduction of full-sphere sound," in *Audio Engineering Society Convention 65.* Audio Engineering Society, 1980.

[29] B. Wiggins, "Has ambisonics come of age," *Proceedings of the Institute of Acoustics*, vol. 30, no. Pt 6, 2008.

[30] D. Barry and G. Kearney, "Localization quality assessment in source separation-based upmixing algorithms," in *Audio Engineering Society Conference: 35th International Conference: Audio for Games*. Audio Engineering Society, 2009.

[31] H. Wierstorf, A. Raake, and S. Spors, "Binaural assessment of multichannel reproduction," *The technology of binaural listening*, pp. 255–278, 2013.

[32] A. Lindau, H.-J. Maempel, and S. Weinzierl, "Minimum brir grid resolution for dynamic binaural synthesis," *Journal of the Acoustical Society of America*, vol. 123, no. 5, p. 3498, 2008.

[33] T. Musil, A. Sontacchi, M. Noisternig, and R. Höldrich, "Binaural-ambisonic 4. ordnung 3d-raumsimulationsmodell mit ortsvarianten quellen und hörerin bzw. hörer fürr pd," tech. rep., IEM Report 38/07, Tech. Rep., 2007.

[34] M. Noisternig, T. Musil, A. Sontacchi, and R. Höldrich, "A 3d real time rendering engine for binaural sound reproduction," in *Proceedings of the 2003 International Conference on Auditory Display, Boston, MA, USA, 6-9 July 2003*. Georgia Institute of Technology, 2003.

[35] D. Sun, "Generation and perception of three-dimensional sound fields using higher order ambisonics," 2013.

[36] Z. Ben-Hur, F. Brinkmann, J. Sheaffer, S. Weinzierl, and B. Rafaely, "Spectral equalization in binaural signals represented by order-truncated spherical harmonics," *The Journal of the Acoustical Society of America*, vol. 141, no. 6, pp. 4087–4096, 2017.

[37] F. Brinkmann and S. Weinzierl, "Comparison of head-related transfer functions pre-processing techniques for spherical harmonics decomposition," in *Audio Engineering Society Conference: 2018 AES International Conference on Audio for Virtual and Augmented Reality*. Audio Engineering Society, 2018.

[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.

[39] C. Schörkhuber, M. Zaunschirm, and R. Höldrich, "Binaural rendering of ambisonic signals via magnitude least squares," in *Proceedings of the DAGA*, vol. 44, 2018, pp. 339–342.

[40] J. Vilkamo, T. Bäckström, and A. Kuntz, "Optimized covariance domain framework for time–frequency processing of spatial audio," *Journal of the Audio Engineering Society*, vol. 61, no. 6, pp. 403–411, 2013.

[41] "Sadie ii database." [Online]. Available: https://www.york.ac.uk/sadie-project/database.html

[42] T. Lokki, R. Kajastila, and T. Takala, "Virtual acoustic spaces with multiple reverberation enhancement systems," in *The AES 30th International Conference, Saariselkä, 15.3.-17.3. 2007*, 2007.

[43] R. Kajastila, "Interactive multi-channel auralization with camera-based tracking," *Proc. 19th ICA, 2007*, 2007.

[44] J. Buchholz and S. E. Favrot, "Loudspeaker-based room auralization in auditory perception research," in *International Workshop on the principles and applications of spatial hearing.* IWPASH, 2010.

[45] S. E. Favrot, *A loudspeaker-based room auralization system for auditory research.* Technical University of Denmark, 2010.

[46] K. Ueno and H. Tachibana, "A consideration on acoustic properties on concert-hall stages," *Building Acoustics*, vol. 18, no. 3-4, pp. 221–235, 2011.

[47] ——, "Experimental study on the evaluation of stage acoustics by musicians using a 6-channel sound simulation system," *Acoustical science and technology*, vol. 24, no. 3, pp. 130–138, 2003.

[48] Z. S. Kalkandjiev and S. Weinzierl, "Room acoustics viewed from the stage: Solo performers' adjustments to the acoustical environment," in *International Symposium on Room Acoustics*, vol. 6, 2013.

[49] W. Woszczyk, D. Ko, and B. Leonard, "Virtual stage acoustics: a flexible tool for providing useful sounds for musicians," in *Proceedings of the International Symposium on Room Acoustics*, 2010, pp. 1–8.

[50] W. Woszczyk and W. L. Martens, "Evaluation of virtual acoustic stage support for musical performance," *Journal of the Acoustical Society of America*, vol. 123, no. 5, p. 3089, 2008.

[51] W. Martens and W. Woszczyk, "Virtual acoustic reproduction of historical spaces for interactive music performance and recording," *The Journal of the Acoustical Society of America*, vol. 116, no. 4, pp. 2484–2485, 2004.

[52] Y. H. Kim, J. Y. Jeon, and D. Cabrera, "Evaluation of stage support for musicians' performance in a concert hall," in *20th International Congress on Acoustics*, 2010.

[53] K. Ueno and H. Tachibana, "Cognitive modeling of musician's perception in concert halls," *Acoustical science and technology*, vol. 26, no. 2, pp. 156–161, 2005.

[54] W. Chiang, S.-t. Chen, and C.-t. Huang, "Subjective assessment of stage acoustics for solo and chamber music performances," *Acta Acustica united with Acustica*, vol. 89, no. 5, pp. 848–856, 2003.

[55] E. Battenberg and R. Avizienis, "Implementing real-time partitioned convolution algorithms on conventional operating systems," in *Proceedings of the 14th International Conference on Digital Audio Effects. Paris, France*, 2011, pp. 248–235.

[56] J. D. Miller, M. R. Anderson, E. M. Wenzel, and B. U. McClain, "Latency measurement of a real-time virtual acoustic environment rendering system," in *Proceedings of the 2003 International Conference on Auditory Display, Boston, MA, 6-9 July 2003.* Georgia Institute of Technology, 2003.

[57] E. M. Wenzel, "The impact of system latency on dynamic performance in virtual acoustic environments," *Target*, vol. 135, no. 180, 1998.

[58] C. Chafe, M. Gurevich, G. Leslie, and S. Tyan, "Effect of time delay on ensemble accuracy," in *Proceedings of the international symposium on musical acoustics*, vol. 31. ISMA Nara, 2004, p. 46.

[59] S. Yokoyama, K. Ueno, S. Sakamoto, and H. Tachibana, "6-channel recording/reproduction system for 3-dimensional auralization of sound fields," *Acoustical science and technology*, vol. 23, no. 2, pp. 97–103, 2002.

[60] A. Libeaux, T. Lentz, D. Houben, and M. Kob, "Voice assessment in choir singers using a virtual choir environment," in *19th International Congress on Acoustics*, 2007.

[61] N. Eley, S. Mullins, P. Stitt, and B. F. Katz, "Virtual notre-dame: Preliminary results of real-time auralization with choir members," *2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA)*, pp. 1–6, 2021.

[62] H. Møller, "Fundamentals of binaural technology," *Applied acoustics*, vol. 36, no. 3-4, pp. 171–218, 1992.

[63] T. Hiekkanen, A. Mäkivirta, and M. Karjalainen, "Virtualized listening tests for loudspeakers," *Journal of the Audio Engineering Society*, vol. 57, no. 4, pp. 237–251, 2009.

[64] B.-I. Dalenbäck, M. Kleiner, and P. Svensson, "Audibility of changes in geometric shape, source directivity, and absorptive treatment-experiments in auralization," *Journal of the Audio Engineering Society*, vol. 41, no. 11, pp. 905–913, 1993.

[65] F. Otondo, J. H. Rindel, R. Causse, N. Misdariis, and P. de La Cuadra, "Directivity of musical instruments in a real performance situation," in *Proceedings of the International Symposium on Musical Acoustics, Mexico city, Mexico*, 2002, pp. 230–232.

[66] J. Meyer, *Acoustics and the performance of music: Manual for acousticians, audio engineers, musicians, architects and musical instrument makers.* Springer Science & Business Media, 2009.

[67] Angelo Farina, "X-MCFX-CONVOLVER," http://www.angelofarina.it/X-MCFX.htm, 2021, Accessed: 25 July 2023.

[68] Matthias Kronlachner, "ambiX v0.3.0 – Ambisonic plug-in suite," https://www.matthiaskronlachner.com/?p=2015, 2020, Accessed: 15 July 2023.

[69] "ReaPlugs VST FX Suite," https://www.reaper.fm/reaplugs/, 2023, Accessed: 25 July 2023.

[70] trsonic, "nvsonic-head-tracker," https://github.com/trsonic/nvsonic-head-tracker, 2021, Accessed: 02 July 2023.

[71] L. Beranek, "Acoustics and musical qualities," *The Journal of the Acoustical Society of America*, vol. 99, no. 5, pp. 2647–2652, 1996.