Stereo Representation of Environments for Live In-Ear Monitoring Systems

Steven Kerry

MSc by Research

University of York

Physics, Engineering and Technology

January 2024

Abstract

In the area of on-stage monitoring for live sound, the use of in-ear monitoring may have the effect of isolating a listener from the acoustic environment due to the reduction of ambient sound and the representation of sources lacking influence from the surroundings. This thesis demonstrates how there may be ways of using stereo microphone techniques to assist in the representation of the surrounding and shows how, with the use of convolution, there are actually several appropriate configurations that could be applied. The intention of the study is not to suggest the use of convolution as a sole method of monitoring a performance, but is instead used to gain insight into the appropriateness of stereo configurations for monitoring with respect to one another and their ability to immerse a performer more fully into an environment. By capturing the acoustic response of a performance area with a variety of different microphone configurations and allowing a performer to monitor a convolution of the musical source signals, it has been possible to demonstrate that other near co-incidental pair setups can potentially give a comparable result to a binaural dummy head, and suggestions are made regarding a study that may add more quantitative data in the future.

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for a degree or other qualification at this University or elsewhere. All sources are acknowledged as references.

Contents

List of Figures

Chapter 1

1 Introduction

The electroacoustic monitoring of live music for performers has undergone many developments since its introduction in the 1960s [1]. The initial issues of an artist simply wanting to hear themselves on stage has altered as systems and solutions have improved, notably with the introduction of the personal in-ear monitoring system (IEM) in the late 1980s [2]. Although monitoring with loudspeakers is still common in the industry, as shown in Figure 2, IEMs are a preferred solution for many artists due to several factors, such as the ability to monitor without introducing a feedback loop, consistency of sound irrespective of the environment, and, with the introduction of wireless systems as seen in Figure 1, freedom of movement without compromising the sound being heard. However, these improvements have led musicians to experience a performance in a different way than playing while monitoring acoustic sound sources, specifically in the way that a musician is potentially isolated from their surroundings and can lead to a feeling of detachment [3].

Figure 1: An example of the type of system required to allow for a performer to monitor their performance using in-ear monitors. Pictured is the Shure PSM 1000 In-Ear Personal Monitoring System with a transmitter, receivers and headphones. From [4].

Within the live music industry, there exist various prevalent practices aimed at mitigating the problem of detachment. These practices encompass both auditory approaches, such as the placement of ambient microphones on stage (typically positioned downstage left and right, directed towards the audience), as well as physical approaches, such as the utilisation of tactile transducers [5]. Both of these practices are intended to enhance the immersive nature of the monitoring experience. This can be seen as supporting the argument that relying solely on IEMs is inadequate in meeting a performer's needs.

Figure 2: A performer using loudspeaker monitors positioned to the bottom right of the picture projecting up towards the performers listening position.

1.1 Motivation

The research conducted in this study has been shaped by the author's extensive personal experience spanning over two decades in the live sound industry, as well as prior investigations in the field of on-stage monitoring [6]. The previous publication about the future of live sound suggested that there may be a number of factors leading to deficiencies in the practice of monitoring using IEMs regarding the element of feeling detached from the performance environment. One factor that was highlighted in this previous research was that the possible feeling of detachment may be in part due to a favouring of close microphone techniques to support isolation of elements and control over the mixing stage, plus the use of direct injection (DI) of sources and an increased use of emulation of amplification. Due to the ability to recall a consistent sound that can be created in the studio and transported in a small unit as opposed to requiring sizeable amplification setups, systems such as that offered by Kemper [7] as seen in Figure 3 and Neural DSP [8] have become more popular in recent years and have become preferred over loudspeaker amplification for many artists and engineers. From a live mixing point of view, the ability to have fewer sound sources on stage means that a mix can be created without having to compete with other, often loud sounds being emitted from the stage.

Figure 3: The rack in the picture demonstrates guitarist James Bourne of the band Busted's Kemper and Spare Kemper unit (the units with the pink and green lights) in use during a soundcheck at the Nottingham Motorpoint Arena. The rack is positioned on a separate riser off stage, allowing the guitar technician to be in control out of sight of the performance area. Photo courtesy of Jack Mackrill.

An emerging trend in the entertainment industry involves the expansion into binaural listening. In live music, new technologies such as KLANG's 3D in-ear mixing technology [9] are being used in an attempt to aid the performers listening experience. The Klang system introduces a novel method for panning sound sources. Instead of the conventional one-dimensional method that only adjusts the volume of the left and right speaker levels, this system positions the sound sources as objects in a three-dimensional listening space, aligning with the natural perception of sound by listeners. Due to this, the research in this paper aims to consider some of the specifics of listening to spatial representations of audio. The use of binaural audio with head-tracking has been demonstrated as being preferable to regular stereo by studies such as that of Tomasetti and Turchet [10] and the systems have started to be seen in some live productions such as that of Ricky Martin, Pitbull and Enrique Iglesias, to name a few [11].

The motivation for this research is derived from the author's extensive professional experience working with performers on stages since 2002. After many years of collaborating with numerous artists across various settings, it has become evident that there are recurring themes related to performers' challenges in managing their monitoring setups. A common anecdotal comment regarding the use of IEMs is that they isolate a performer from their environment. Sometimes this can be in terms of a feeling of separation from the crowd in front of them, other times, this may be not feeling the energy of the backline instruments on the stage. A common method of reintroducing this sense of surroundings to the performer is to use ambient microphones. This can be a pair of condenser microphones located downstage right and downstage left, often at the height of the crowd facing away from the stage. In an effort to try and make this introduction of ambience more from the point of view of the performer, there has been an

effort to try and use the IEM itself as the method of receiving and delivering audio alongside the dry monitor feed, most notably perhaps with JH Audio's "Ambient Pro" model [12] as seen in Figure 4. The monitor engineer can receive a feed of the ambient sound directly from the performer's listening perspective in real time by placing an omnidirectional microphone capsule from DPA Microphones on the outer area of each IEM casing. This can then be fed back into the performers mix at the desired level, aiming to remove the occlusion of the IEM in the ear.

Figure 4: JH Audio's Ambient Pro IEM with omnidirectional DPA microphones mounted in the casing. From [13].

Other companies, such as ACS [14] have aimed for a similar result but in a potentially simpler arrangement by utilising filters from their line of hearing protection to add an element of ambient pass-through into the mould of the IEM as well as the sound being produced by the internal speaker drivers. This option is less adaptable than the offerings of JH Audio in terms of alteration of the ambient level due to it being based on the passive filter involved. Using a passive filter also creates other drawbacks from the point of view of the performer. Although the filter gives some sense of ambience, the rest of the elements are still localised in the head of the performer because they are being monitored over headphones.

The transmission method for the essentially binaural microphone setup, as employed with the JH Audio Ambient Pro IEMs, necessitates supplementary equipment for establishing a wireless connection and receiving the audio signal at the monitor mixing console. It is possible that a comparable condition can be replicated by positioning a simulated head in close proximity to the performer's location. Considering the dimensions of the Neumann KU100, as depicted in figure 5, it is not an inherently inconspicuous microphone for positioning in front of a performer. Consequently, alternative microphone techniques that emulate a similar concept to the intended objective of the KU100 may be more favourable if they offer comparable response.

Figure 5: A Neumann KU100 Binaural Dummy Head. From [15] .

The capture of acoustic properties of environments is an area that is often effectively utilised in the field of film and television. In order to enhance the authenticity of automated dialogue replacement (ADR) during the process of overdubbing vocals in a recording studio with a lack of natural reverberation, measurements of the acoustic response are frequently conducted on the filming set. These measurements serve the purpose of enabling the application of convolution reverb at a later stage [16]. This research will discuss the theory of creating this type of reverb in section 2.4 and demonstrate how it could be used to assess the appropriateness of representation of space using different stereo microphone techniques.

1.2 Statement of Hypothesis

To evaluate the research question, the hypothesis was:

When using in-ear monitoring, convolution of musical source signals with spatial acoustic measurements can be used to make a sound that is true to life on stage. Stereo microphone techniques can give results similar to those of a binaural head when used to record a room's acoustic response.

1.3 Objectives

This thesis aims to explore and discuss the suitability of a variety of stereo microphone configurations to measure acoustic responses that could be convolved with musical instrument signals to support a performer on stage. This refers to a convolution of the environment based on a fixed listening position with the localisation of a set number of positions within the performance area. A key area of interest in this study is regarding the microphone configuration being used to create the convolution of the space, with a focus on preference of stereo configurations and the reasons that these preferences may be attributed to. It may be the case that there are some stereo techniques using more commonly found condenser microphones that give a comparable result to that of a binaural head microphone, such as the Neumann KU100 [15], that may encourage a wider adoption of the practice.

In order to define the suitability of room impulse response (RIR) capture techniques, this study has three main objectives:

1. To create a set of measurements of a performance space from the perspective of each performer in a musical ensemble.

2. To create an authentic representation of the space by removing factors of the measuring process, such as the coloration of the frequency response of the system used to take the measurements and also the playback system (including the in-ear monitor transducers).

3. To create a monitoring environment using the captured data set that offers a fair and comparable listening experience in order to isolate the stimuli in question.

Due to the time constraints of the thesis, a further objective of the study will be to inform future research, better understand the limitations of a qualitative study, and help in understanding how a more quantitative study could be created in the future that would provide further empirical evidence. As such, this thesis should be considered a pilot study for future work.

1.4 Structure of the Report

The structure of this report can be separated into five sections.

- **Chapter 2** contains an overview of the literature relevant to the study up to this point. An overview is provided of the areas that pertain to the propagation of sound and are relevant to the principles of acoustics in a performance environment, as well as the human hearing mechanism. This overview sets the stage for further discussion on psychoacoustics and the various factors that are involved. This is followed by an investigation into the use of acoustic measurement and convolution techniques to capture and represent the sense of space and finally how this may impact performer experience.

- **Chapter 3** discusses the process of taking measurements of the performance space, going into detail on the equipment used, the variables decided upon for the testing phase of the investigation and the justification for doing so. The chapter goes into detail about the deconvolution process and the method of equalising both the test equipment and the monitoring apparatus in an attempt to remove them from having an influence on the findings of the study.

- **Chapter 4** details the methodology of the testing phase of the study, giving insight into the setup being used and the rationale behind its suitability for the study. Methods of data capture are discussed, and an initial pilot test is used to ensure the robustness of the test setup and suitability of the proposed style of experiment. The results of this are analysed to suggest where improvements can be made to the main pilot testing phase.

- **Chapter 5** analyses the results of the main testing session in the study and gives visual representations to aid the understanding and comprehension of the data.

- **Chapter 6** Evaluates the overall findings of the study, drawing conclusions in reference to the original research questions. Additionally, there are recommendations on the methodology for conducting a subsequent quantitative study and the potential implications of the study on future research.

Chapter 2

2 Literature Review

2.1 Introduction

This literature review can be broken down into four sections: Principles of acoustics in performance spaces, perception of space, convolution of space, and performer experience. The research conducted for this literature review encompasses relevant textbooks, academic journals, and applicable ISO standards.The first section contains an exploration of research and literature related to the principles of acoustics in performance spaces. This section includes studies relating to propagation of sound and reverberation. There is also a discussion of the properties of enclosed spaces relevant to this thesis. The second section contains research into the area of perception of space. This includes an insight into the human hearing mechanism and how perception of sound results in the ability to localise sources. The third section contains an exploration of studies into the convolution of space. In addition, research has been done on the area of capturing acoustic measurements and the process of converging these measurements with musical sources. The fourth and final section contains research on the experience from the point of view of the performer. This concluding section includes studies related to the factors that contribute to the performer's listening experience.

2.2 Principles of Acoustics in Performance Spaces

When discussing acoustics, it is appropriate to note the context in which the sound resides. Specifically, this refers to how sound propagates and interacts in an enclosed space as opposed to within a free field. A free field refers to when sound is able to propagate in straight lines unimpeded [17]. In this study, the acoustic reflections and reverberation will not operate in a free field, but rather will be subject to other factors that will affect how a performer hears the resultant sound.

2.2.1 Propagation of sound

In practice, a performer in a space is subject not only to the direct source sound of an instrument or amplifier but also to the later reflections of the original sound. As sound propagates towards the listener, it spreads out in three dimensions as it gets further from the radiating source. Over distance, the sound becomes weaker because of radiation over a wider surface area. As discussed by Howard and Angus [18], sound can be considered to propagate spherically, the area (A) when the radius (R) is known is given by Eq.(1).

Eq. (1)

$$A = 4\pi r^2$$

Due to the relationship between the surface area of a sphere and the geometry of its expansion, it can be stated that for every doubling in distance traveled, the surface area increases by a factor of four, an inverse square relationship. If the sound intensity at the wavefront (I) is given as a measurement of power per unit of the power at the source (W), then the relationship between the intensity of the sound and the distance traveled can be shown, as demonstrated in Eq. (2).

Eq. (2)

$$I = \frac{W}{A} = \frac{W}{4\pi r^2}$$

Another factor to be taken into consideration is that not only is the energy reduced due to the inverse square law, but also that energy in a frequency-dependent respect will be lost due to atmospheric absorption [19]. What this means in a practical sense is that as sound travels over greater distances, due to atmospheric absorption, there will be a proportionally greater reduction in higher frequency content, specifically that over 2 kHz. This is a factor that can be largely ignored in small rooms [17] [20].

2.2.2 Properties of reverb

After the listener perceives the direct source sound, there will be subsequent reflected signals that can be categorised into two distinct areas. The first category of reflections is known as early reflections. These reflections are comprised of the initial reflections that occur when sound waves encounter boundaries or surfaces within the environment. Early reflections are characterised by their delayed arrival at the listener's location, which is a result of the extended distance they have travelled. Eventually, the reflections overlap in time and continue to lose energy due to the distance travelled and being absorbed by surfaces and in the air. This is referred to as the tail of the reverb, as seen in Figure 6 [21] .

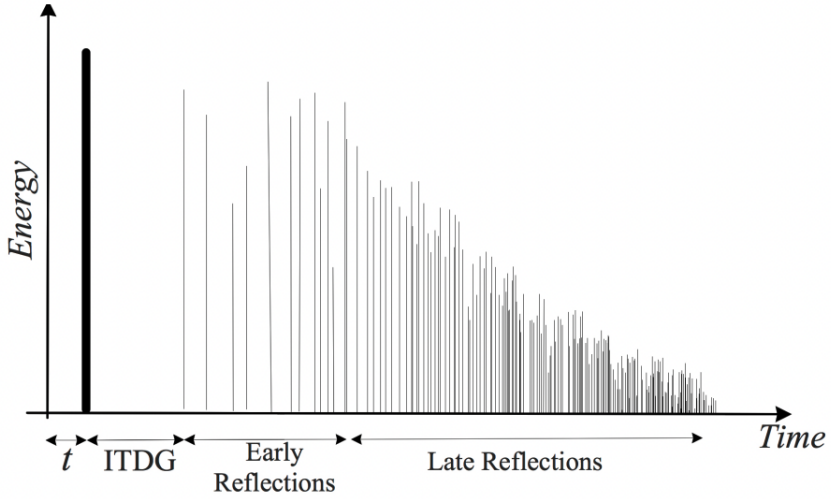


Figure 6: The initial time delay gap (ITDG) is shown in the diagram above, giving an indication of the distance to the first boundaries the sound source reflects from. Over time, the sound is shown to continue reflecting around the space, forming a period of early and late reflections. From [22] .

The time taken for the sound to reach the boundaries is dependent on the speed of sound, which in itself is dependent on environmental factors such as temperature, pressure and humidity [20]. If an estimate is given of the environmental factors of the space being used for the experiment (roughly 20 degrees Celsius and relatively dry), then a speed of between 340 m/s and 344 m/s can be assumed which gives an indication as to the amount of time that the first reflections will take to reach the listener.

If a space is sufficiently large, then the time taken to travel to and reflect back from the initial boundaries is long enough to be perceived as a separate event. Studies have demonstrated this just noticeable difference (JND) to be in the region of around 10 ms. For example, Reichardt and Schmidt [23] suggested as low as 7 ms ±0.6 ms, whereas J. Atagi, R. Weber, and V. Mellert suggested 10.6 ms [24]. These studies in particular contain some slight variables in test conditions but demonstrate a contextually similar result. A result that emerges from the JND, also known as the echo threshold, is the precedence effect, also known as the Haas effect [25] . If there are two acoustic events within this threshold, only one sound source is localised in the direction of the leading wavefront [26].

2.2.3 Properties of enclosed spaces

As discussed earlier, a free field describes a scenario where sound travels unimpeded [17]. This can be approximated to an extent, for example, in an anechoic chamber, but practically, it is not applicable to the test scenario in this thesis. In contrast to the free field, a space that has many reflective boundaries creating reflections with a similar intensity, that is, reflections with equal probability [27], would be referred to as a diffuse field [28].

2.2.4 Room modes

Room modes are created when reflections of a source interact with each other within a space where the wavelength is related to the distance between the reflective boundaries. The transitions from creating individual resonances to creating a region of diffuse behaviour are known as the "critical frequency" or "Schroeder frequency" after Manfred Schroeder, who explored the relationship between frequency and the transition from the low frequency region occupied by modes to the region of dense overlap [29]. In a very large room, such as a concert hall, this frequency may be below the lowest sound created in the space, whereas in a smaller environment, such as the suggested experiment, for example, the critical frequency will occur within the range of sounds that are being produced within the space [30]. More information on the room modes specific to the experiment can be found in Section 6.1.1.

To calculate where this frequency resides requires knowledge of the surface area and volume of the room, the speed of sound (C), and the average absorption coefficient of the space. For a quick approximation that assumes that the modal behaviour dominates after the mean free path (MFP) is equal to one and a half wavelengths rather than taking absorption coefficients into consideration [18], the equation numbered eq. (3) can be used.

Eq. (3)

$$f_critical \;=\; (\frac{3}{2})\frac{c}{MFP} \;=\; (\frac{3}{2})\frac{344ms^{-1}}{MFP}$$

In the context of this thesis, the performance space has a surface area (S) of $120.868m^2$ and a volume (V) of $80.277m^3$ which gives us the following estimation as shown in eq. (4).

Eq. (4)

$$f_critical \;=\; (\frac{3}{2})\frac{c}{MFP} \;=\; (\frac{3}{2})344ms^{-1}/\frac{4V}{S}$$

$$= 1.5 \times; \frac{344ms^{-1}}{2.657m} = 194.2Hz$$

Based on the available information, it is anticipated that in larger venue spaces, the factor of critical frequency would occur at a lower frequency that is less significant. This aspect should be duly considered when analysing the findings of this study. However, it is worth noting that the size of the space does share comparable characteristics with smaller performance spaces and stages. Where this may have an impact is if there is a position in the room that is of importance, either from a listening perspective or the perspective of a sound source, that is overly excited or, inversely, reduced in an area fundamental to the harmonic content of certain instruments.

2.2.5 Frequency content of the suggested instrumentation

The instruments suggested for this experiment comprise a drum kit, bass guitar, and electric guitars. Each of these elements encompasses a wide frequency range, but the fundamental frequency that determines pitch will vary depending on the tuning of the instruments. The drum kit itself contains multiple sound sources that have a variety of fundamental frequencies. According to a study by Hewer, the impact of a stroke on the bass drum results in a sound with an extremely low frequency. The frequency spectrum exhibits two prominent frequency peaks at 86.1 Hz and 172.3 Hz, along with a minor peak at 258.3 Hz [31]. The tom-toms examined in the study exhibited fundamental frequencies of 193.8 Hz, 172.3 Hz, and 108.1 Hz, as determined by the utilisation of two rack toms and a floor tom. The fundamental frequency of the snare drum was measured to be 194.6 Hz. It is important to acknowledge that these values are dependent upon the tuning of the drums and the tension of both the resonant and batter heads. The study conducted by Hewer identified shared characteristics among the cymbals, with a notable concentration of frequency peaks occurring within the frequency range of 0 Hz to 100 Hz for low-intensity strokes and within the frequency range of 2000 Hz to 5000 Hz for high-intensity strokes [31].

The bass guitar is a musical instrument with four strings. It is typically tuned to the notes E1, A1, D2, and G2, which correspond to frequencies of 41.2 Hz, 55 Hz, 73.4 Hz, and 98 Hz, respectively. Harmonic content for the bass guitar is created up to approximately 4 kHz [32]. In standard tuning, an electric guitar's strings resonate at fundamental frequencies: E2, A2, D3, G3, B3, and E4, which correspond to the frequencies of 82.41 Hz, 110 Hz, 146.83 Hz, 196 Hz, 246.94 Hz, and 329.63 Hz, respectively. The aforementioned frequencies provide us with an initial comprehension of the frequency spectrum generated by the guitar while playing open strings. Nevertheless, it should be noted that the absence of sounds above those frequencies does not imply that no such sounds will be generated. Electric guitars are capable of generating harmonics and overtones that can extend up to a frequency of approximately but not limited to 15 kHz. However, it is worth noting that the frequency range of a guitar generally extends from approximately 82 Hz to around 5 kHz [8].

2.3 Perception of space

The physics of sound pressure levels interacting within a space is an objective and measurable process. From the point of view of the listener, there are, however, other factors that contribute to the perception of sound in terms of its perceived tonality and location, which will be discussed in the following section.

2.3.1 The Auditory System

The human auditory system as shown in Figure 7, consists of three areas: the external ear, the middle ear and the inner ear. The external ear (sometimes referred to as the outer ear) consists of the pinna, ear canal and ear drum.

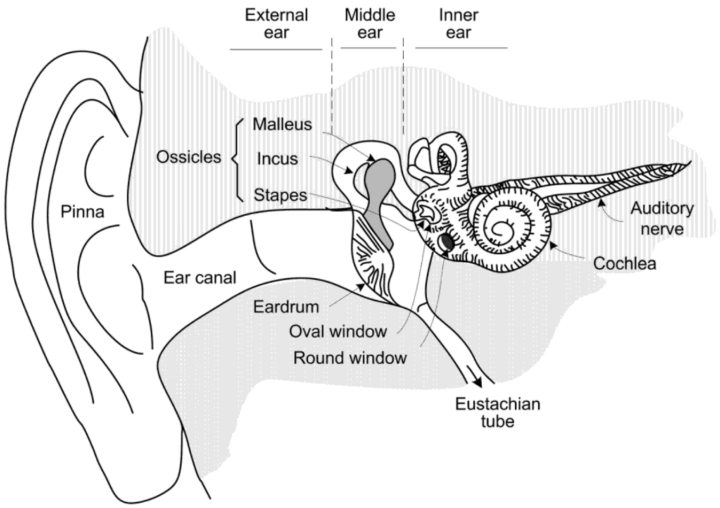


Figure 7: Cross section of the human auditory system. From [33].

The pinna, an irregular-shaped ovoid highly variable in size, is an external appendage consisting of skin covered elastic cartilage [34]. The pinna has been suggested to introduce, by means of delay paths, a transformation of the incoming signal which assists in the process of sound localisation [35]. Sound travels down the ear canal, which is generally not straight and rather S-shaped; varying in size in both its length and width, to the eardrum (tympanic membrane) [34].

Fluctuations in sound pressure level acting upon the ear drum cause it to vibrate. These vibrations are then transmitted through the middle ear via three small bones (Ossicles), the Malleus, Incus and Stapes [36]. This process acts as a mechanism for efficient energy transfer while also providing some protection from loud sounds damaging the inner ear [17].

The inner ear houses the Cochlea; a sound analysing organ whose role is to convert mechanical vibrations into nerve firings sent to the brain [18]. Once movement of the eardrum has activated the Ossicles, the motion of the Stapes causes the Oval window to move, in turn, causing vibrations through the fluid in the inner ear [17]. The Oval window is at the base of the cochlea, which is a tube coiled into a spiral of approximately 2.75 turns [18]. The area of the Cochlea responsible for carrying out a frequency analysis of incoming sound is the Basilar Membrane. Vibrations through the incompressible fluid here allow for transfer of vibrations to nerve firings by the organ of Corti, which consists of a number of hair cells that trigger nerve firings when bent. The nerves from these hair cells form a spiral bundle known as the "auditory nerve" [18].

2.3.2 Perception of volume

The sensation of hearing occurs when frequencies within a certain frequency range reach the listener above a minimum required intensity. The lowest tone that is audible as a frequency is at a frequency of approximately 16 Hz for most humans [37]. The upper end of the audible range is said to be 20,000 Hz (20 KHz), although this varies and declines with age [37]. A number of experiments have investigated the relationship between frequency and perceived

volume since the empirical formula developed by Fletcher and Munson for Bell Labs in 1933 [38] and a current demonstration of the difference in sensitivity across the frequency range can be seen in Figure 8. This will be discussed further when analysing the data in Chapters 5 and 6.

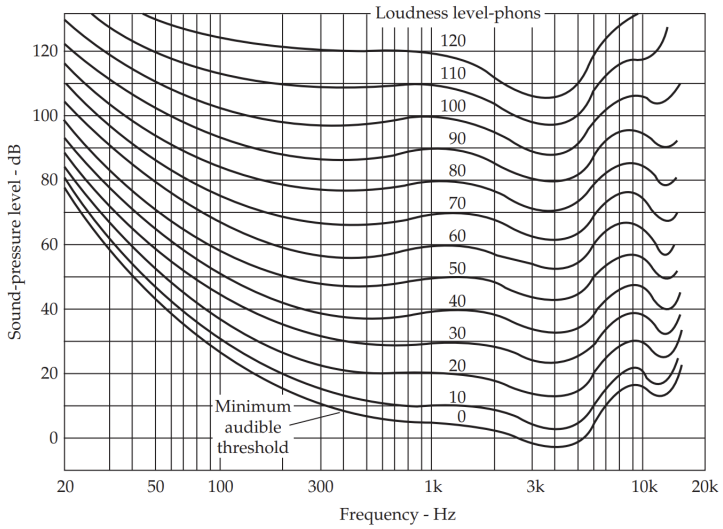


Figure 8: Equal loudness contours of the human hearing mechanism for pure tones developed by Robinson and Dadson and adopted as an international standard (ISO 226). This demonstrates that there is not equal sensitivity to all frequencies, with a reduced sensitivity, particularly at lower frequencies. From [17].

2.3.3 Localisation of Sources

The ability to create audio with spatial characteristics is dependent on the ability to localise sound sources. The term localisation refers to the ability for the listener to perceive the direction and distance of the sound source [36].

Two of the cues that are responsible for the listener's ability to localise the sound are the inter-aural time difference (ITD) and the inter-aural level difference (ILD). This refers to the comparison of the time and the intensity with which the sound source reaches the left and right ear of the listener. Further investigation into these factors reveals that the two factors are dependent on the content of the sound source, specifically in relation to frequency content. Lower frequencies with much longer wavelengths do not have as much head shadowing as higher frequencies with shorter wavelengths because of the relative length of wave forms in comparison to the size of the head. The result of this is that when localising sound sources, the ILD is most useful at higher frequencies, whereas the ITD is more useful for lower frequency content. This concept is referred to as the Duplex Theory of Sound Localisation [39].

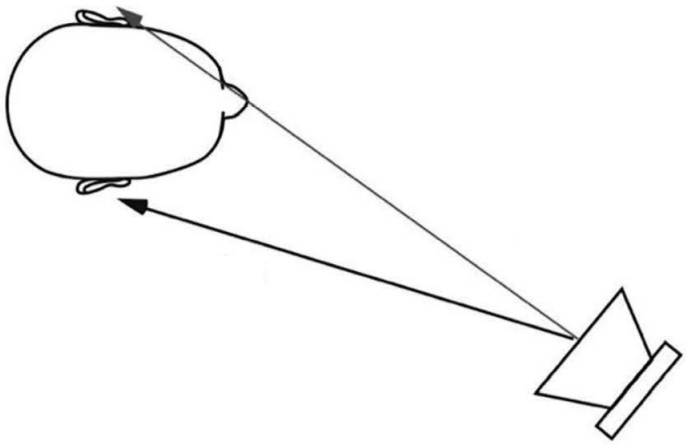


Figure 9: Sound located to one side of the listener on the horizontal plane will reach the ears at different times and intensities. Adapted from [39]

The Duplex Theory, discussed by Lord Rayleigh early in the 20th century [40] has been used as a basis for the application of spatial practices and has been improved upon due to acknowledged deficiencies in the concept. One of these deficiencies is the method used by Rayleigh, as it was dependent on single-frequency sine waves as a source [41]. Due to this methodology of testing, other studies need to be taken into account for complex wave forms to understand the relationship between the ITD and ILD in real-world scenarios.

When it comes to complex sounds, contrary to the behaviour of individual tones, lateralisation studies have demonstrated that listeners are sensitive to ITDs in high-frequency broadband sounds [42] where the auditory system is able to extract timing information from the envelope of complex higher frequency sounds.

In addition to the lateral perception of sound sources, there is also distance and elevation to consider. The inter-aural cues of time and level are not able to resolve elevation and often leave ambiguities in whether a sound is located in front of or to the rear of the listener [18]. When considering two sound sources from the perspective of a listener, assuming a spherical head with no pinnas, if the sounds are symmetrically placed in front and behind the listener, there would be little or no difference in ILD or ITD, meaning the location of where the sound source emanates from can not successfully be resolved. This area of ambiguity is referred to as the cone of confusion and is displayed in figure 10 [19][41].

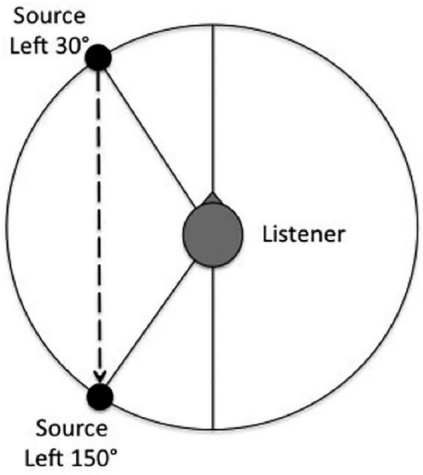


Figure 10: The image demonstrates how two sound sources, although in different locations in front and behind the listener, would display identical or near identical ILDs and ITDs. These ambiguities are in part resolved by other factors discussed in Sections 2.3.4 and 2.3.5 . From [41].

2.3.4 Additional Spatial Cues

A study by Wightman and Kistler discussed the filtering of sound at each ear individually by the pinnae, which results in so called "monarual spectral cues" [43]. Each ear individually has the ability to localise sound sources independent of the factors of ITD's and ILD's incorporated in the duplex theory. In addition to these spectral cues, small asymmetries in the head of the listener may provide spatial cues, specifically a spectral notch that moves from approximately 5KHz to 10Khz as the sound source moves vertically from 0° directly in front of the listener to 90°, over the listeners head [41]. In addition, Larsen et al. have shown that the ratio of direct to reflected sound from the sound source can also result in the creation of monaural cues [44].

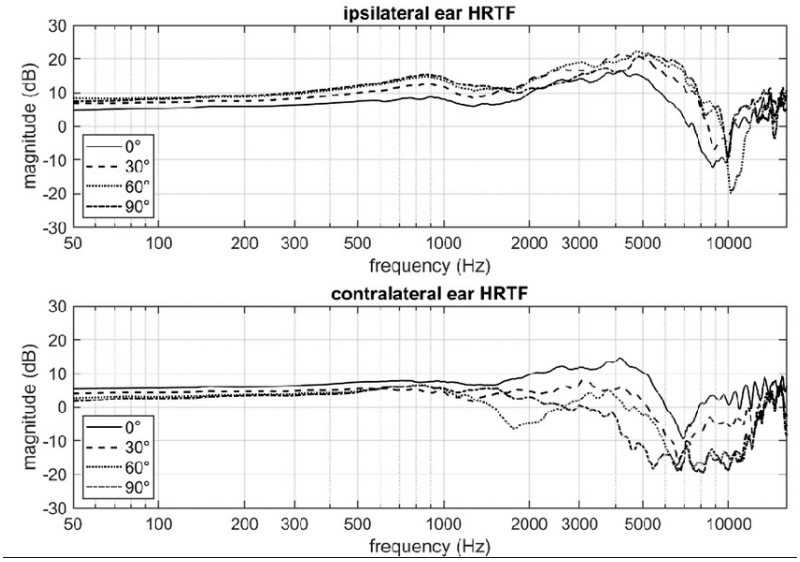


Figure 11: Head related transfer function (HRTF) measurements taken with a source 50 cm from the ear on the same side as the source (ipsilateral) and opposing side of the head to the source (contralateral), which demonstrates how the frequency response of a sound source differs with a varied reduction above 1KHz for the ear on the opposing side of the head to the sound source depending on the angle of projection of the source. From [45].

There are inherent weaknesses in any monauralisation experiment looking to isolate one ear

in the listening environment in that the other ear can not be completely removed and level can only be reduced by some form of plugging of the entrance to the ear canal. Where this represents the largest issue is in the lower frequencies, as they are the frequencies attenuated least by an ear plug and also the frequencies stated earlier to be most affected by ITDs so even at lower levels, these could be a factor.

A powerful method of resolving localisation ambiguities is to move the head to face the direction of the sound source. Moving the head alters the location of the perceived sound source relative to the location of the listener; sounds from infront of the listener change differently to sounds located behind or above. When considering the cone of confusion discussed in Section 2.3.3 and displayed in Figure 10, the movement of the head would therefore create differences in the ILD and the ITD between the two sources. This is also one explanation as to why sound heard through headphones is located internally in the head on a horizontal plane between the ears. As the head moves, the sound tracks with the head, a relationship meaning that the sound cannot be located externally [18].

2.3.5 Head Related Transfer Function

Head-related transfer functions (HRTFs) describe the acoustic impact that human anatomical structures such as the head, torso and pinnae have on a sound source reaching a listener [33]. The main components are above 200 Hz due to that being the point where linear sound field distortion because of diffraction becomes significant [46]. As frequencies become higher into the mid-frequency range, the head and torso affect the transmission of sound to the ear canal until approximately 3 kHz where the Pinna contributes to distortions [46]. These alterations in the perceived sound of a source are a factor in the listeners ability to localise sources and although generalised HRTF filters can be used to some effect, each individual will have their own personal transfer function.

2.3.6 Principles of stereo microphone configurations

In keeping with the theories relating to the relationship between the two ears of a listener, there are relatable methods of utilising pairs of microphones that make use of some of the principles, such as the ITD and ILD. When capturing audio with a pair of microphones, directional information can be encoded as inter-channel time differences (ICTDs) and/or inter-channel level differences (ICLDs) [41] which, when replayed using headphones, create some differences compared to speaker playback due to the lack of crosstalk as shown in Figure 12.

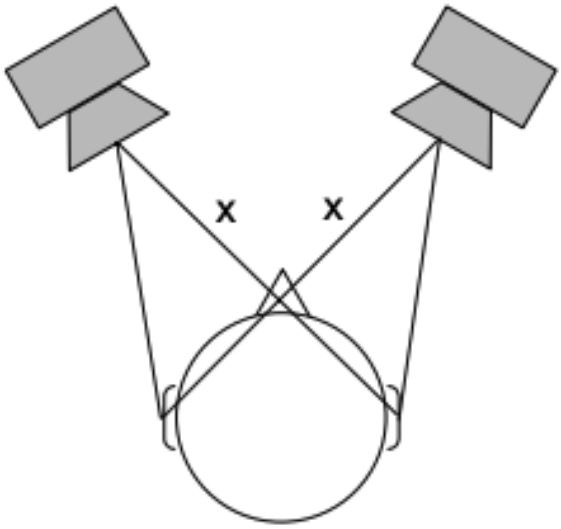


Figure 12: Speaker listening with a placement of +/- 30°, the x donating the crosstalk path to the contralateral ears. From [47] .

Using speaker playback, lower frequencies (where the ITD is the major cue) will take a longer path to the contralateral ear than the ipsilateral ear and the signal will appear delayed in time but not changed in amplitude. This is due to the wave diffracting around the head. This means that what would be heard over headphones as a level difference, when replayed over speakers, will be heard by the listener as a phase difference, resulting in a directional auditory sensation [41] [47]. An example of this is shown in Figure 13. This conversion from ICLD to ICTD cannot occur in headphone listening. How a stereo configuration of microphones captures the ICLD and ICTD in a way that resembles ILDs and ITDs will be a factor in this study.

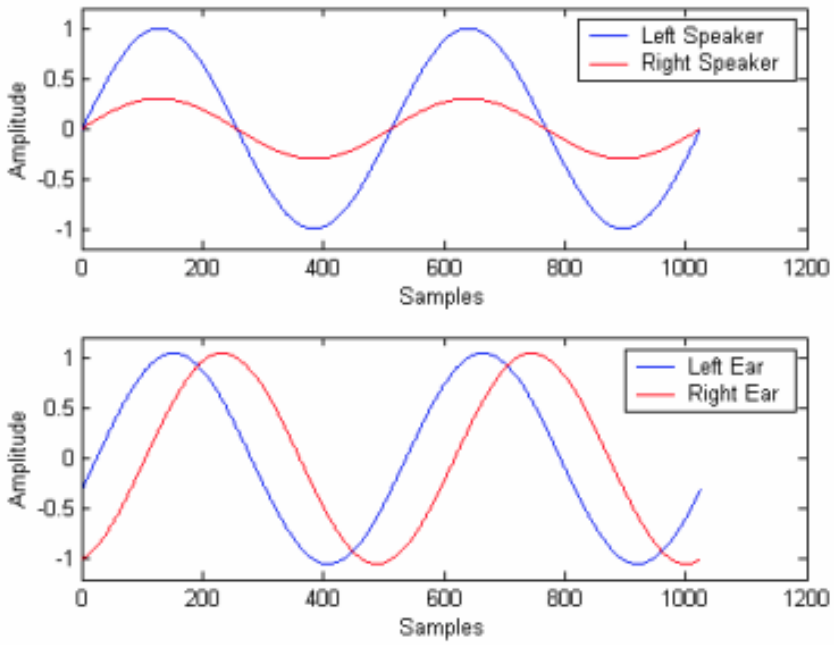


Figure 13: Representation of a source simulated to be left of centre in a pair of speakers as a level difference resulting as a phase difference to the listener. From [47] .

Stereo microphone techniques can be separated into four main categories :

1. Spaced pair
2. Co-incidental pair
3. Near co-incidental pair
4. Baffled omni/dummy head

[48]

Spaced pairs of microphones, such as the AB setup depicted in Figure 14, utilise both the ICLD and ICTD to effectively capture a stereo image.

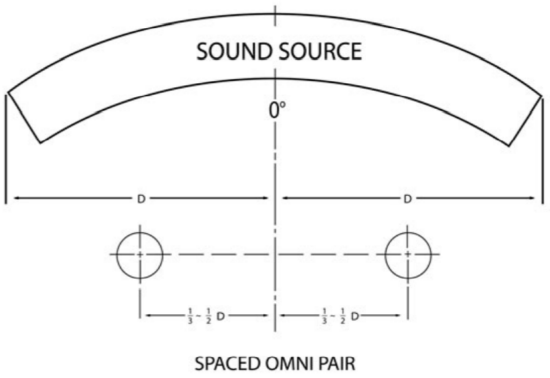


Figure 14: A Spaced omnidirectional pair of microphones placed perpendicular to the source. From [48] .

Each microphone will capture sources at varying intensities and times, depending on their relative distance from the source. This occurrence can be compared to the relationship between the ears of a listener, although there are also notable distinctions. One factor to consider is the spatial separation between the two microphones in relation to the listener's ears. The placement of two microphones further apart can create an exaggerated stereo spread, leading to notable constructive or destructive interference due to the phase relationship between the two signals in certain areas of the environment.

The concept of a co-incidental pair seen in Figure 15, is based on the utilisation of the ICLD as the sole means of representing the stereo image. The term"co-incidental" is used to describe the close proximity of the two diaphragms in physical space. If it were possible for the two diaphragms to occupy the exact same space, then the ICTD would no longer be a factor. Having them in as close proximity as the mechanics of the transducers allow minimises any impact of the ICTD as a result. Studies such as that of Pulkki [49] have demonstrated that co-incidental pairs of microphones demonstrate a consistent and stable representation of a virtual source in comparison to spaced pair techniques, which tend to leave ambiguity around the exact position of the source, albeit with a more pronounced spread of the environment [50].

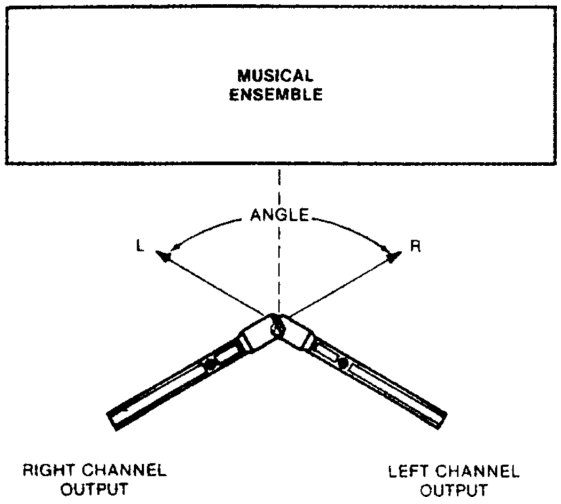


Figure 15: An XY co-incidental configuration of microphones. From [51] .

Using a cardioid polar response and aiming each of the capsules +/- 45° from the centre gives a very consistent delivery of the location of sources directly in front of the microphones. It is expected that the source will remain between the extreme left and right, therefore giving a narrower representation on the whole than that of the spaced pair discussed previously. Due to the localisation being based on a difference in level between the microphones as opposed to any time difference, mono compatibility is better with a co-incidental pair, meaning that if the left and right channels are summed to mono, there is little destructive phase cancellation of frequencies as sound arrives at the two diaphragms at close to the same time.

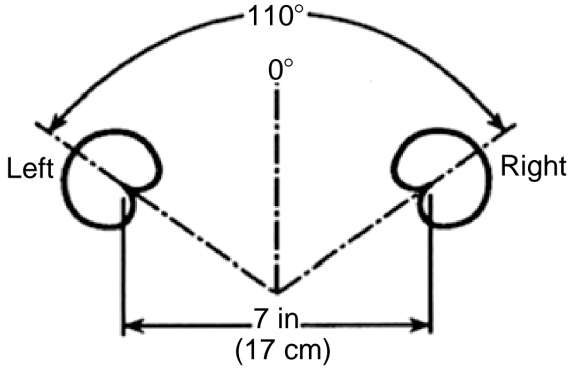


Figure 16: An ORTF near co-incidental configuration of microphones. From [52] .

A near-coincidental pair of microphones makes use of both the ICLD and the ICTD similar to the spaced pair, to capture stereo information, but due to the closer proximity of the microphones, the relationship between the two is different. One configuration that will be explored in this thesis is the ORTF configuration shown in Figure 16. ORTF stands for Office de Radiodiffusion Télévision Française; the name of the French Government Radio Station responsible for creating the technique, and consists of two cardioid microphones positioned 17 cm apart (at the diaphragm) and angled outwardly 110°, so each diaphragm is therefore 55° away from the

centre [52]. This is said to mimic the position of a listeners ears in respect to their head [53] and give a stereo image that sounds natural and from a listeners point of view.

The result of this configuration maintains a large area of the image for localisation between the left and right, similar to the XY pair but to a lesser extent, while reintroducing some of the extreme left and right width that the spaced pair demonstrates. Due to the ICTD being a factor in creating the localisation in this configuration, mono compatibility is not as coherent as the XY configuration.

Figure 17: A pair of Omnidirectional condenser microphones baffled by a Jecklin Disc. From [48] .

The baffled omni approach can be achieved in a number of ways. In perhaps its most basic form; two omnidirectional microphones are positioned approximately the same distance as a listeners ears, and then a baffle is placed between the two microphones to obstruct in a similar way that the head of the listener does [54]. The Jecklin Disc, shown in Figure 17, is one such method, where a hard disc covered in absorbent foam is placed between two omnidirectional microphones. Similar to this is the Schoeps Spherical Mic shown in Figure 18, where instead of a disc separating the two microphones, a hard sphere separates them, with the two microphones flush-mounted at either side of the surface [48].

Figure 18: A Schoeps KFM360 Spherical Mic. From [48] .

Moving on from these more basic baffled scenarios are configurations incorporating a dummy head such as the earlier mentioned Neumann KU100 and the more extensive GRAS 45BB-4 KEMAR Head and Torso, the latest model of the original KEMAR head and torso simulator first introduced in 1972, which has been used extensively in hearing aid and audiology research [55].

While these more complex configurations offer a potentially more realistic HRTF than the other simply microphone based setups, it is still not the HRTF of the listener that will eventually monitor the signal and therefore will potentially not translate directly. The Neumann KU100 that will be used in the testing for this thesis also lacks any depth to the ear canal, which will have an impact on the resultant frequency response as it will lack the usual resonance located somewhere between 2 kHz and 4 kHz [56].

Overall, it would appear that there are potential advantages and disadvantages to the differing configurations and based on the information presented in this section, it could be argued that a dummy head may be the most accurate configuration to measure the acoustic response of an environment to then translate to a listener. It is plausible, however, from what has been discussed that some of the other stereo configurations may present a similarly appropriate localisation of sources.

2.3.7 Supporting immersion with spatial audio

The term immersion in relation to audio applications has been linked with a number of different terms, such as realism, naturalness or a sense of being surrounded, which has led to an ambiguity in a specific definition [57].A definition proposed to the Audio Engineering Society by Agrawal et al in 2020 is as follows:

Immersion is a phenomenon experienced by an individual when they are in a state of deep mental involvement in which their cognitive processes (with or without sensory stimulation) cause a shift in the attentional state such that one may experience disassociation from the awareness of the physical world.[57]

The implication of this definition, within the framework of this research, is that the concept of a performer being immersed in the environment entails their detachment from the physical surroundings and their active involvement with a reimagined version of the environment, primarily through auditory means.

Spatial audio refers to a format capable of reproducing spatial cues for the listener. Many formats throughout the twentieth century have had the ability to reproduce spatial cues to some extent; however, spatial audio in this context refers to the localisation of sources in three dimensions around the listener. In a setting where an environment is implied, the aim of spatial audio could be said to be achieving a feeling of "being there" for the listener [58]. There has been a great deal of research into how the use of spatial audio can support a sense of immersion, and there are some significant factors that arise that would suggest how a performer could benefit from this in a monitoring environment.

Elias Zea [59] published a paper studying the use of binaural methods to provide real-time low-latency monitoring for performers. The reason for the use of binaural methods was due to the belief that traditional stereo monitoring now allows for the correct spatialisation of sources. An interesting quote from this paper is in reference to interaural cues, Zea states that there are four cues that allow our auditory system to locate a sound source in space: interaural cues,

pinnae differences, head movements, and sight. Up to this point, the discussion has focused on the hearing mechanism and the localisation of sources from a purely auditory point of view, but the implication that sight is a factor may have implications in this study, particularly when it comes to representing sources that can be seen and raising questions about whether a source cannot be seen.

The study by Zea is incomplete in the sense that many of the ideas discussed remain in the realm of theory due to the insufficient capability of technology at the time. There is a comment about the three main factors a dynamic tracking device needs to achieve to satisfy auralisation, "latency (29 ms), frame rate (60 Hz/FPS), and smallest spatial measurement (1 degree)," and it is these factors that create the limitations in the technology due to the amount of computational power required to satisfy these during the processing of the signal. In the present day, it is highly likely that these factors can be more feasibly attained due to advancements in technical capabilities.

A factor that is discussed in Blauert's publication Spatial Hearing, originally published in 1985 and later revised in 1997, [60] that is potentially crucial to the research being undertaken is the type of headphone being used. As demonstrated by Villchur[61], the use of circumaural headphones (that surround the pinna) creates difficulties regarding calibration due to the location of the speaker relative to the ear and the enclosed space in which it resides. A critical factor that is relevant to this study is the quantity of gain added by the pinna when using this type of headphone design, particularly at higher frequencies [62]. As already discussed in Sections 2.3.4 and 2.3.5, the information in this area is essential to sound localisation and therefore its accurate portrayal in the context of this study is of extreme importance. The use of circumaural headphones, however, although often used in experimentation, is not the relevant method of listening in this study. Instead, intra-aural headphones, inserted directly into the ear canals, are the preferred method of audio delivery for on-stage performers. In the research methodologies employed in this study, the calibration of headphones will be explored in an attempt to mitigate this issue. There is perhaps some argument to suggest only conducting experimentation using intra-aural headphones to reduce the number of artefacts from the calibration of less relevant headphone designs.

Head-tracking is a process that describes a system where the perception of the localisation of sources is altered based on head movement [63]. Head-tracking is often used to aid immersion and is common to many modern media formats and in virtual reality devices such as the Occulus Rift [64] shown in Figure 19. The implication of a listener moving their head and the source tracking correctly to correspond with that movement is fairly obvious in terms of adding to the realism of the experience but there are other implications to the application of head-tracking or more importantly for the case of this thesis, implications for the lack of head-tracking. Due to the test method of the study discussing a static listening position, head-tracking will not be used; however, this has the unfortunate effect of limiting the area inside the listener's head for where localisation occurs. As demonstrated in figure 20, without head tracking, localisation is limited to between the ears and to the rear of the head, with difficulty representing sources in front of the listener.

Figure 19: An Oculus Rift headset which incorporates head tracking for both the audio and visual element. From Engadget [65].

A 2022 study by Bauer et al discussing the use of binaural headphone monitoring to enhance musicians' immersion took the approach of both experimenting with and without head tracking [66]. A finding of that study was that the reduction in masking of elements that the binaural representation provided over a regular stereo representation allowed for elements such as the click track to be presented at a much lower level, which is beneficial both in terms of practical factors such as spill from headphones being captured in microphones and also from the aural health point of view of the performer.

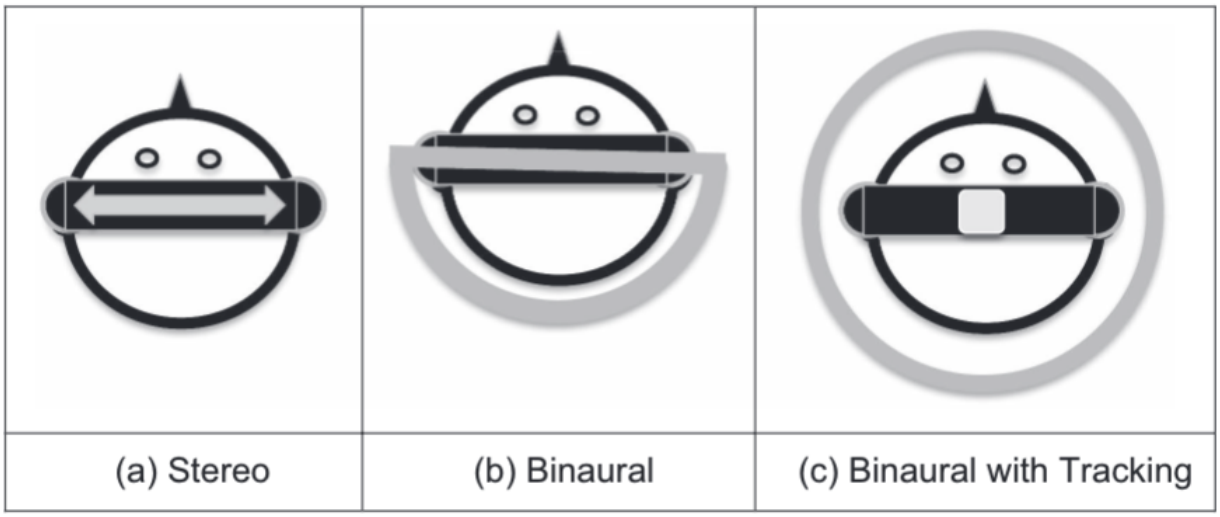


Figure 20: A visual representation of how head tracking can contribute to better representation of space in headphones and resolution of front and back ambiguity. The thick grey line shows the image locations heard by the listener. From [59]

2.4 Convolution of space

This study aims to find out the perceived benefit of introducing the element of space to in-ear monitoring, and as such, the method of creating the spatial element is to be discussed. A process that can be used to process a signal with the properties of a space is convolution. What this refers to specifically is the marriage of two signals—a dry, unprocessed signal with the measured response of something else. In the case of convolution reverb, this typically involves combining the dry, unaffected audio signal with the impulse response (created based on the acoustic response) of the desired environment.

2.4.1 Fast Fourier Transform

The Fast Fourier Transform (FFT) converts a signal into individual spectral components and, as a result, provides frequency information about an audio signal. Filtering in the frequency domain uses a point-by-point multiplication method, whereas in the time domain it uses the more complex process of convolution [41] [67] [68].

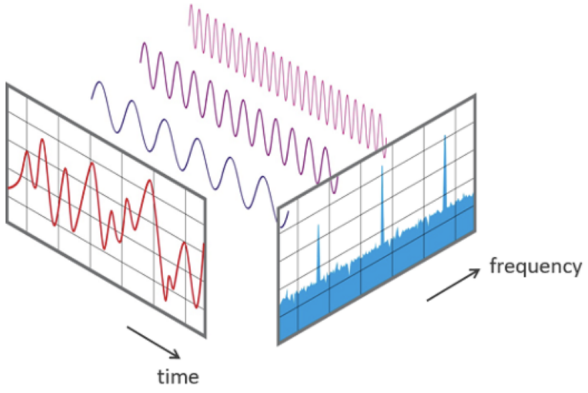


Figure 21: The diagram demonstrates how a signal is sampled at periods over time and separated into its component frequencies, each with their own amplitude and phase. The diagram shows how, over time, in this hypothetical example, there are three dominant frequencies in the signal. From [68] .

The first stage of this measurement consists of two elements: the sample rate (fs) and the selected number of samples or blocklength (BL). The BL is always an integer power to the base 2 and the range of frequencies measured; the bandwidth frequency (fn) is subject to the Nyquist Theorem [69] which states that a minimum of two samples per frequency cycle are required to represent a waveform in a digital system. This is demonstrated by the equation

Eq. (5)

$$fn = fs/2$$

In turn, the duration of measurement (D) can be shown by division of the BL by the samplerate. For example if a 96 kHz sample rate is being used, the Nyquist Theorem dictates the highest possible frequency of 48 kHz (in practice, due to filter slopes, this is generally lower than this), and with a BL of 1024 samples, the result would be

Eq. (6)

$$D = BL/fs$$

$$D = 1024/96000$$

$$D = 10.66ms$$

To demonstrate the frequency resolution (df) or the spacing between two measurement results, the following equation (eq.7) applies:

Eq. (7)

$$df = fs/BL$$

So, for the 96 kHz, 1024 sample example, this would give a frequency resolution of

Eq. (8)

$$df = 96000/1024$$

$$df = 93.75Hz$$

The result of this is that, using a sample rate such as 48 kHz or 96 kHz, by altering the size of the BL, the measurement duration and frequency resolution can be altered; smaller BL values result in faster measurement repetitions with a coarser frequency resolution than a longer BL, which will achieve slower measuring repetitions with a finer frequency resolution [68]. For the purpose of this study, the sample rate that will be used will be 48 kHz with a BL resolution of 1024.

2.4.2 Windowing

When using the FFT to measure the frequency component of a signal, the process of windowing is used to create a finite-length window with an amplitude that gradually reduces to zero at the edges. This creates a finite sequence and can improve a signal with a non-integer number of cycles where the frequency response would otherwise be smeared [70].

2.4.3 Room Impulse Responses

The standard for creating impulse response by acoustic response measurements is set out by ISO:3382-1 [71] however, there is an element of flexibility in this description and the practical application of the process is often left up to the interpretation of the user [72]. As found by Angelo Farina [73] the use of logarithmic sine sweeps is suggested as being appropriate to the measurement of acoustic spaces due to an increased signal-to-noise ratio (SNR) due to all of the energy at any one time being located at a single frequency [74] and also the lack of requirement to sync the generator of the sweep and the device capturing the signal. In addition to SNR improvements, there is also the factor that the environment is excited equally across all frequencies, therefore demonstrating a response to how each frequency would respond. This differs to more transient events that are limited in the bandwidth that is exciting the space. Due to this, for the experimentation in this study, the same logarithmic sine sweep will be used for all measurements in keeping with Farina's suggestion and to maintain consistency across the study.

Figure 22: An NTi Audio DS3 Dodecahedron Speaker is suggested as one option for omnidirectional amplification of the logarithmic sweeps. As can be seen in the picture, due to the construction and requirement for the number of speakers for this to be achieved, the size of the speaker drivers are limited, and in turn, the amplitude able to be achieved, especially at lower frequencies, is compromised. From [75] .

The ISO:3382-1 standard outlines that an omnidirectional sound source, such as that shown in Figure 22, should be used for the acoustic response measurements. As discussed by Papadakis and Stavroulakis in their review of alternative sound sources to dodecaheadron speakers [76] in a space comparable to one being used in this study, little variation in higher frequency content was found when using different variations of directional and omnidirectional sound sources, but there were some inconsistencies in the lower frequencies. Lower frequencies are often considered to be omnidirectional in nature, but it demonstrates that the reality of speaker cabinet construction does mean that there is a skewed response around the cabinet, which differs dependent on cabinet design.

The area important to this study is the method of capturing the sine sweeps to measure the acoustic response of the room and generate the impulse responses. Studies such as that conducted by Gelen have used a mono capture, and although this is not conducive to the way that we perceptually hear things, it demonstrates that a mono representation may offer some useable representation of the environment that the stereo configurations could be compared to. A study conducted by Stade, Bernschütz and Rühl [77] gives some insight into the different

recording methods that can be used that are in keeping with what has been suggested for this study. The study, similar to this one, considers the use of a dummy head, specifically in this case the Neumann KU100, and also some stereo-based microphone techniques such as a variety of spacings of an AB setup, a co-incidental pair or XY, a near-coincidental ORTF setup, and also the addition of a mid-side pair (MS). The study, in addition to others such as that by Villchur [61] highlights the requirement to compensate for the entire transducer chain, with mention of the significant impact resulting from the compensation for the headphones in particular. This will be discussed in Section 3.2 regarding how it has been addressed for this study. A method of replacing the headphones on a KU100 dummy head a total of 12 times in order to capture a stable and representative transfer function is applied. This is something that will be a factor when conducting any testing via the use of in-ear monitors and will certainly need to be addressed in the methodology of this study.

Once the sine sweeps have been used to capture the acoustic response of the environment, the process of deconvolution is necessary to create an impulse response such as the one shown in Figure 23. This deconvolution of the signal is obtained by convolution of an inverse filter, which is the time reversal of the test signal with compensation equalisation for the 6 dB per octave falloff caused by the log sweep [78] with the captured sweep signal.

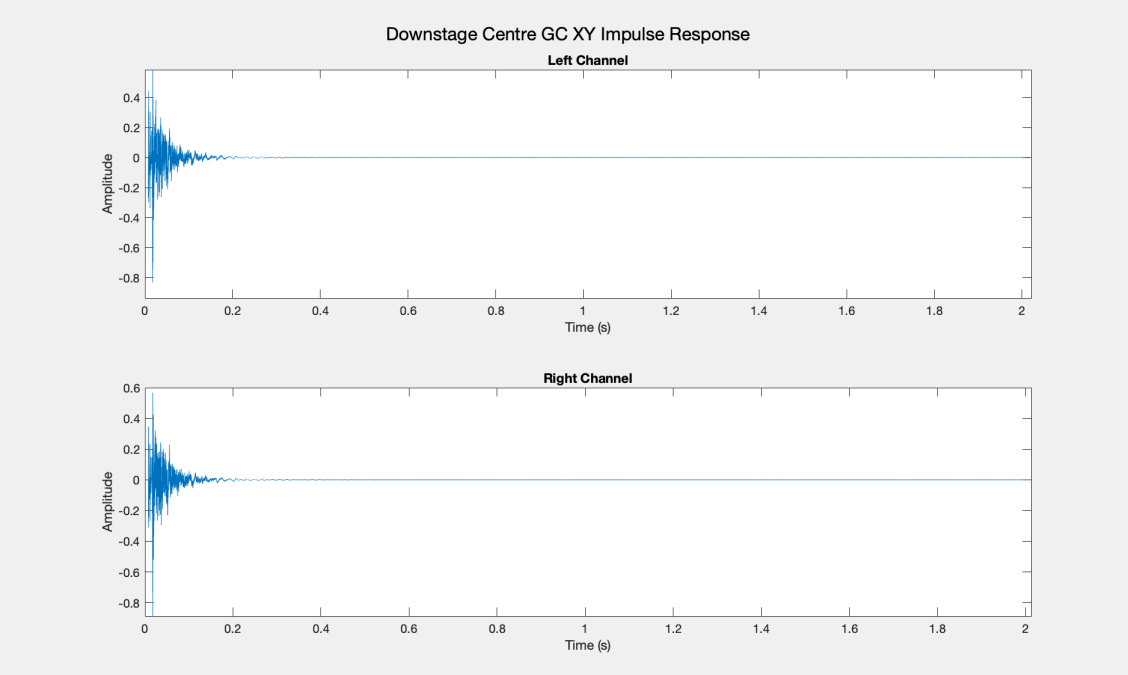


Figure 23: A plot of one of the IRs created for the purpose of this thesis. This shows the IR from the perspective of the downstage centre position of the performance space of a sound source being emitted from the location of the upstage centre guitar amplifier using an XY setup to capture the acoustic response.

The study conducted by Stade, Bernschütz and Rühl [77] was in the context of studio control rooms, and the context that this thesis wishes to address is different in the sense that it is in relation to monitoring on a stage rather than in a control room. There are many ways in which impulse responses could be used on a stage; however, at the time of writing, there is no evidence to suggest this is a practice common to any area of the industry. It is completely plausible that a touring production of a show has implemented the spatial characteristics of the surroundings in this way, but finding evidence of this has been unsuccessful beyond the use of ambient microphones, as discussed earlier in Section 1 of this thesis.

An interesting thesis by Engin Gelen [79] about how pre-auralisation can be used to help with monitoring in various settings brings up some interesting points that could be useful in this study. The thesis suggests creating impulse responses in spaces that a performer will play in and the effectiveness of creating mixes based on this convolution of space as opposed to creating dry mixes that do not factor in the environment. There are a number of limitations to this study, and therefore the conclusion is something that should be approached with caution, specifically that the impulse responses are created from a single mono capture in an individual location. What the study does demonstrate, however, is the assumption that the space that the performance takes place in has an impact on the monitoring of the performance. The idea upon reflection is novel but not practical, as if it were deemed to be important that the specific space being played in had influenced the monitor mix, then from a logistical point of view this method falls apart for a touring production as the environment being performed in would be changing regularly. Anything being performed in a single space, however, such as some sort of residency or long-term theatre production, could potentially make good use of this concept.

2.4.4 Analysis of Measurements

In Blauert's publication, Spatial Hearing, there is a useful insight into the research processes regarding the methods of experimentation. Nominal and Ordinal judgements are suggested as being the most important for auditory experiments because they are "especially well adapted to determine thresholds of perceptibility, difference thresholds, and points of perceptual equality" [60]. By using this format, the aim is to be able to further the research in a way that is both useful and contextually relevant to other research in this area.

To ensure the consistency of the analysis, it will be important that the volume is kept as static as possible . If a performer is to be subjected to different stimuli with a view to assessing their preference, then it is important that each stimuli be represented at a comparable level. If this is not the case, then a favouring of a louder mix may cloud the judgement of another factor such as tonal or spatial characteristics that may have been considered differently at a matched level due to changes in representation of frequency content as discussed in Section 2.3.2 and a general favouring of louder programme material as suggested by practitioners such as Bob Katz [80].

2.5 Performer experience

The idea of situating a performer on stage and them being immersed in the environment is a proposition with many variables. It is important to recognise what is being suggested as the "environment". In a small venue with a standard back-line (a term given to the musical ensemble of instruments on stage behind performers) of instruments consisting of drums, bass guitar and electric guitars, it is easy to suggest where the location of the sound source should be perceived to be coming from. However, this idea is not consistent with a stage that does not contain stationary traditional sound sources. In addition to this, when considering the monitoring of vocals, the position that the voice should appear to come from has other factors to consider that the other sources do not. It could be assumed that factoring in the use of floor monitoring and replicating this with binaural audio is a method that could be used, or alternatively, the idea of traditional monitoring environments may be cast aside in favour of a new representation of the audio on stage. There are potential positives to this approach, specifically around the negative impact of comb filtering created by various sources on a stage being captured with varying amounts of delay due to their relative distance from each other and their respective microphones [81]. The experience of the performer in enabling a good performance is well documented; factors such as the audio element not interfering with the

element of performance as discussed by Evans [82] demonstrate how the replication of audio is successful when the engineer and techniques employed are invisible to or go unnoticed by the performer. This assists the argument towards a realistic approach towards in-ear monitoring, suggesting that if it is possible to immerse the performer in the environment, then potentially the engineer has made this process "invisible".

A study by Merchel and Altinstoy discussed how haptic sensations interact with the human hearing mechanism, and in this study, some key findings were presented regarding the latency times of stimuli for performers [83]. This study will not be discussing haptics, but the information about latency is relevant. The use of a 0 m/s latency time between stimuli in experimentation was suggested to be flawed, and this suggestion is supported by other studies such as that by Lester and Boley [84]. The 2007 study by Lester and Boley suggests that vocalists are much more sensitive to the effect of latency in IEMs than in loud speaker wedges. As an overall general comment, the study concludes that "latency values greater than 16 ms for wedges and greater than 6.5 ms for IEM would likely produce some audible delay for some instruments." The result of this conclusion will be a significant factor when applying processing to the monitoring system being used for the purpose of this thesis and will instruct on some limitations that will need to be adhered to, specifically an ideal latency time of 6.5 ms or below.

When considering the perspective of the drummer, there are a number of factors that pose questions about methods of monitoring. A major difference when using loudspeaker amplification is that the drummer would be the only performer in this scenario hearing those sources, as seen in Figure 24, potentially from behind the cabinets or in line with them to the side, as opposed to being seated in front of the speaker cabinets as other performers are. Although it would be conceivable to create a stimulus for the drummer, taking that altered frequency response and directionality into account, it would arguably be less effective than allowing the drummer to hear the sound of the instrument from its location as if it were played towards them. This is something that may be looked at in the future; however, it could be suggested that this sort of adjustment would be reality for reality's sake, rather than any form of improvement.

Figure 24: The view from the drummers perspective during setup. As can be seen, from the perspective of the drummer in this experiment, they are behind the amplifiers, which brings in to question localisation and realism

2.5.1 Evaluating performer experience

The measurement and evaluation of audio can be categorised as either objective or subjective. Objective tests are ones based either on a defined algorithm or a measured value of a feature of the audio signal. Where testing involves participants listening to and evaluating audio stimuli, it results in subjective testing [85]. Another publication by Schöffler [86] suggests taking into account the overall listening experience when evaluating audio quality-related attributes. The aforementioned statement will be duly considered during the analysis of this study.

2.5.2 Statistical analysis

As the intention of this study is in part to inform future quantitative research into the area, it is important to recognise how the proposed qualitative data should be approached and also what would be required to create a more robust quantitative example of the research.

A method of evaluating differences in audio signals, such as the perceptual difference in audio codecs, is the use of Multiple Stimuli with Hidden Reference and Anchor (MUSHRA) tests. MUSHRA tests, although being used to compare audio, have very specific requirements to be used within the recommendation set out in ITU-R BS.1534-3 [87] and the purpose of them is suited to codecs because they compare imperfections in signals rather than what is being discussed in this thesis, where it is different representations [88]. Due to this, a test such as the MUSHRA test cannot be used for the kind of experiment suggested in this thesis in its current form; however, there may be scope for the adoption of this type of testing in future research. An example of a MUSHRA test is shown in Figure 25.

The standard set out by ITU-R BS.1534-3 states that the MUSHRA test should use the original full-bandwidth unprocessed audio as a reference signal, as well as some other mandatory hidden anchors. In the pilot tests conducted for this thesis, it was acknowledged that the unprocessed signal would exhibit significant dissimilarity from all other variations of the test stimuli, potentially compromising the evaluation of other representations of the monitoring environment due to the stark contrast. However, it is conceivable that a modification to the MUSHRA test could enable the inclusion of an alternative stimulus as the reference. Based on the fact that the spatial attributes of the stereo configurations are in question of the stereo RIR measurements, it stands to reason that the mono RIR could be used as the reference to then base the comparisons. MUSHRA tests and how they can be modified for this specific research will be discussed in greater detail in Sections 6.2.4 and 6.2.5.

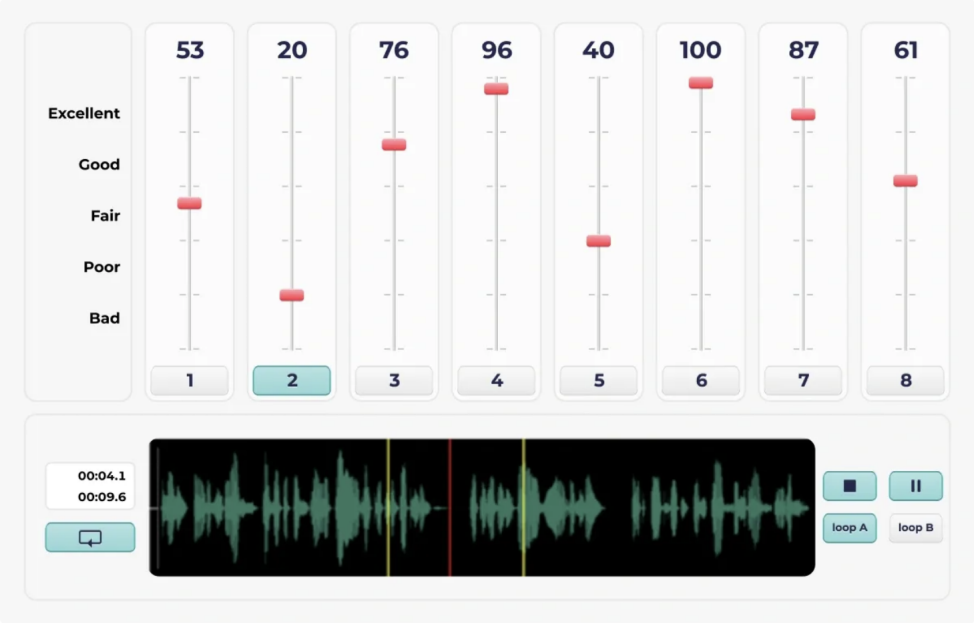


Figure 25: An example of a MUSHRA listening test. This is representative of the common use of prerecorded information to assess differences in audio signals. When each of the numbers beneath the sliders is selected, a different audio stimulus will play, and then the score is selected. For any future use in this research, the idea would need to be adapted to allow for the selection of live monitoring variables as opposed to prerecorded audio. From [89].

The t-test is another useful tool that could be used in future research to gain more quantitative data. These types of tests can be used to test whether a correlation coefficient is different from 0, whether a regression coefficient is different from 0 or if two group means differ [90]. The tests are classified as independent-means or dependent-means based on whether there are different participants assigned to each condition. If they are different, this would be independent; if the same participants took part in both conditions, it would be dependent.

Although it may be possible to carry out multiple t-tests and combine the data to be analysed, another test referred to as ANOVA (analysis of variance) [90] [91] is a test for an overall experimental effect. Because this type of test is assessing variation or deviance, it would be necessary to define a standard to which the deviation is being acknowledged. This will also be discussed further in relation to the future application of this research in Section 6.2.4.

Chapter 3

3 Acoustic Measurements of the performance space

In order to add to secondary research pertaining to suitable techniques for capturing acoustic responses and their efficacy in conveying musical signals through convolution, an experimental design was formulated with a specific emphasis on the variables associated with stereo capture methods. This next chapter details the process of preparing for the experiment and the justification for the decisions made in the process.

3.1 Physical attributes and justification of the performance area

The purpose of this experiment was to consider the monitoring variables from the perspective of performers and as such, the performers or performance positions need to be defined. The context of this experiment is for live music with a band playing as opposed to any sort of classical ensemble, and as such, it is intended that there will be a drummer and amplified sources. For a basic but fairly standard setup, there would need to be a drummer, a bass guitar player, an electric guitar player or two, and a vocalist. The intention was to use a four-piece band, which would mean having a drummer and then three downstage musicians. This would mean that, as well as addressing the drummer's perspective, the other performers would be dealt with from the centre and either side of the downstage performance area.

The chosen venue for the experimentation was a live room of moderate size that was connected to a recording studio. Equipped with a drum platform (or riser), the space in the room was comparable to a stage in a small venue. Due to the number of instances in which the space would need to be used for the preparation, capture of acoustic properties, and conduct of the experiment, it was of the utmost importance that there was a space available that would remain consistent in all of its spatial attributes.

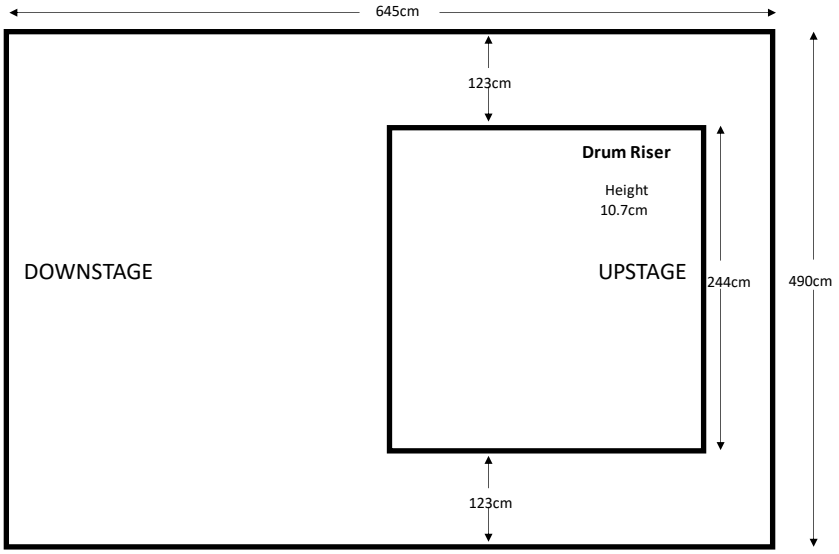


Figure 26: Performance area showing where the downstage area is located and the elevated drum riser

As can be seen in Figure 26, the width of the performance area is approximately 5 metres, with a length of just under 6.5 metres. The height of the room is 2.54 metres. The area chosen gave the required space to situate 3 performers to the left of the room as pictured above (labelled as 'DOWNSTAGE') with amplifiers situated to the side of and in front of the drum riser.

3.2 Acoustic Measurements

In order to streamline the experiment, it was imperative to take into account the specific positions of the performers and the perceived locations of the potential sources. While it is possible to broaden the range of stimuli used in this experiment, the chosen minimal setup provided enough diversity to examine the disparities in stereo methods used for convolution without excessively complicating the auditory experience for the performer. To support the performer in their mix choices and to be able to alter the balance of elements, there is a minimum number of source positions that need to be captured. The location of the drums is where there is room for interpretation. In a live sound scenario, there would generally be a significant number of microphones used on a drum kit to allow for individual control and processing of elements. What is important in this experiment is not creating a hyper-realistic response from each element of the drums but allowing the performers to hear the location of where they reside. As such, the kick drum and snare drum are to be considered key to the mix, requiring an individual element of control to raise and lower volumes, but the rest of the kit can be dealt with using a traditional overhead setup.

Using stereo overhead microphones on a drum kit in an experiment to assess the suitability of stereo configurations for representing space raises questions about which stereo configuration should be employed for this purpose. Considering what has been found in the literature review about how stereo configurations are perceived and how they are portrayed over headphones due to the ICLD and ICTD, the method of using a spaced pair was decided upon. The justification for this is the spread of elements when using this setup. What has been discussed in the literature review is that, where a spaced pair may have less phase coherency, it has an enhanced stereo spread compared to the co-incidental and near-coincidental pairs. The hope is that the use of close mic's on the kick drum and snare drum will allow for a representation of the drums that has a concise response for these key central elements with an easily identifiable width to it. It may be the case that in future tests, other stereo configurations will be considered for this role. From a practical point of view, this raises some more questions about how the acoustic response of the location would be captured for a co-incidental pair, another argument for the use of a spaced pair. With a co-incidental or near-coincidental pair, the location that the acoustic response was measured to be emitted from would be essentially the same place. It would be interesting to see how this is translated, but it is beyond the scope of this thesis.

To represent the microphone positions during the performance, seven locations were therefore decided upon as shown in Figure 27. This allowed for a minimal microphone setup that was sufficient for the purpose of the experiment. These location are:

1. Kick drum
2. Snare drum
3. Stage right drum overhead
4. Stage left drum overhead
5. Amplifier/Guitar stage right

6. Amplifier/Guitar stage centre

7. Amplifier/Guitar stage left

By using three amplifier positions, this gave the option to use any one of these for bass amplifiers or guitar amplifiers, depending on what the setup of the group of musicians happened to be. The drums are the point where minimalism has been preferred, as discussed.

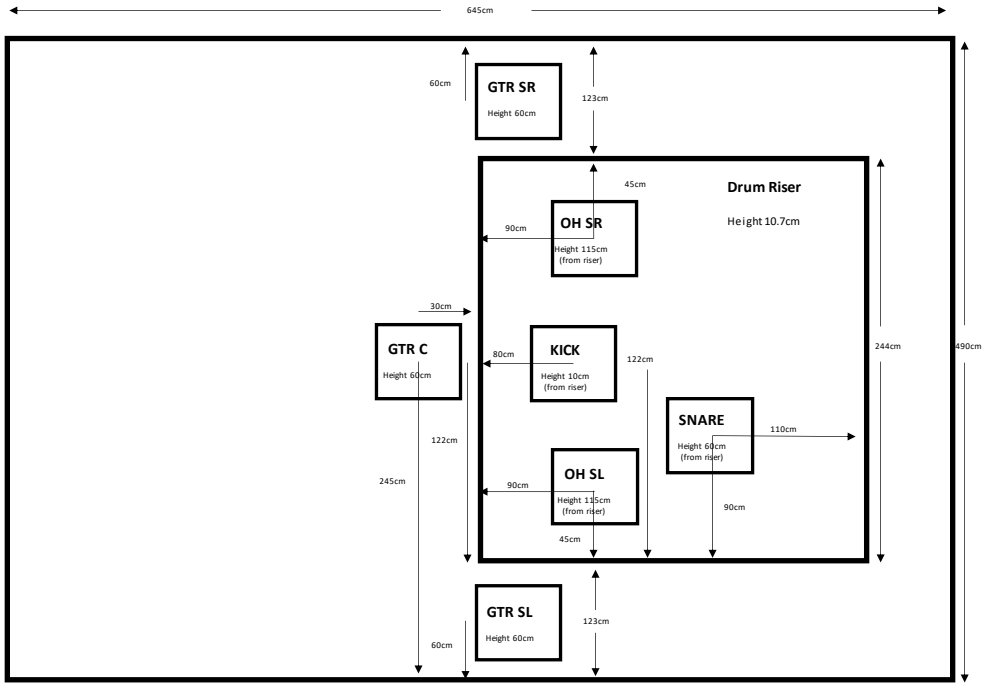


Figure 27: Diagram showing the locations of the sound sources in the experimentation

These seven positions were then considered from four different performance perspectives within the performance space as shown in Figure 28. These positions consisted of:

1. The Drummer's upstage listening position

2. Downstage right listening position

3. Downstage centre listening position

4. Downstage left listening position

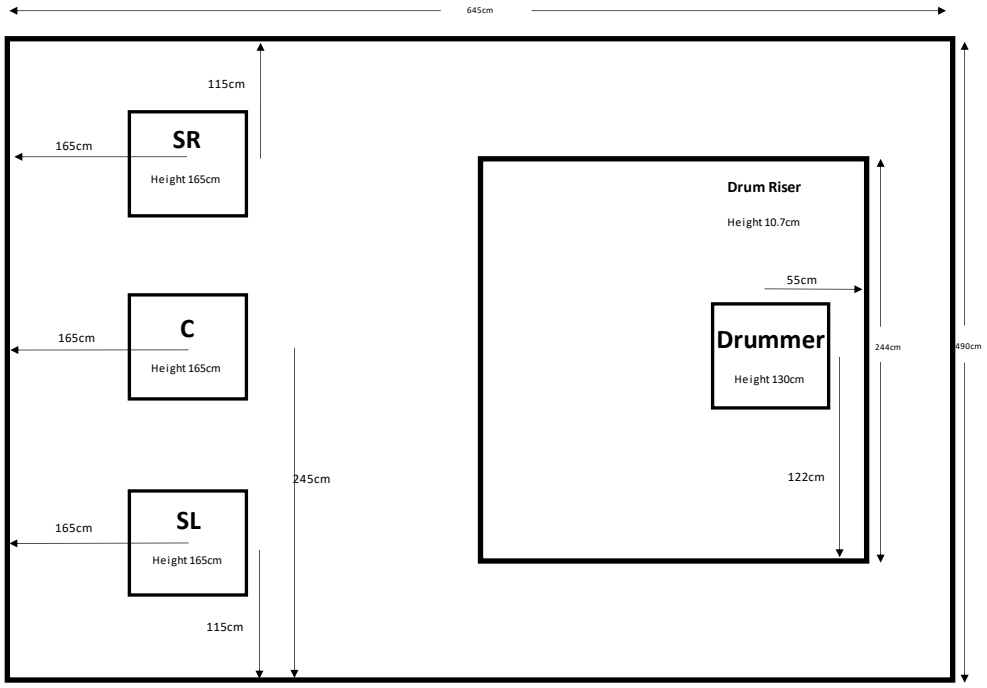


Figure 28: Diagram showing the locations of the performers perspective in the experimentation

The position of each performer was intended to be forward-facing, and in the experiment, this was adhered to. Microphone stands were used to identify the position of where the performers should stand, and performers used a static position of facing forward as they would towards a crowd to analyse the differences between the different convolutions of the space that they were listening to. To allow the performer to move around, as discussed earlier, would require head tracking, which is beyond the scope of this experiment and also not the purpose of what the study is attempting to ascertain.

3.2.1 ISO-3382-1 Impulse Response Measurements

To create the RIR's, based upon previous research from Farina [73] a logarithmic sine sweep needed to be created plus an additional inverse sweep for the deconvolution process. This was achieved using the software Matlab [92]. The sweep created had a duration of ten seconds based upon the principle of being approximately ten times the reverb time of the space as outlined by Farina, with a lower frequency of 20 Hz and an upper frequency of 20 kHz. The sample rate chosen for the sweep was 48 kHz, as discussed in Section 2.4.1, a sample rate that would be used consistently throughout the experiment. Details of the script used to create this sine sweep can be found in the appendix, Section 8.3.

The state of occupancy of the performance area is a factor within the creation of the RIR's but should have a negligible impact on the results in this situation, arguably less so than if this experiment had been conducted in an empty venue. Due to the performance space being contained to a smaller area than a venue that extends out to the viewing area, the addition of 4 musicians and instruments is potentially less intrusive than the difference between an empty venue space and the addition of a crowd. Nevertheless, this is still a factor that may need to be acknowledged in the results of the experiment. The rationale for refraining from conducting the recording of the sine sweeps in an area that houses musical instruments is based on the consideration that the presence of speaker cabinets and drums in that space could introduce additional noise in the form of resonances, which may not be accurately reproduced

when changing the equipment. Through the process of measuring an unoccupied space, the experiment can be reproduced with different levels of occupancy while maintaining the same RIR that is equally applicable to all performance configurations.

An advantage of the space that was used for the experiment was its inherent isolation. Due to the acoustic treatment of the performance area and the decoupling of the room from the structure of the building, the ambient noise was measured to be 28 dBa. With a peak level from the speaker source measured at 92.7 dBa at a distance of 2 metres, this allowed for a signal-to-noise ratio that exceeded the requirement of the ISO-3382-1 (Acoustics — Measurement of room acoustic parametres) standard [71] while being able to represent the sine sweeps at a level that did not introduce distortion in the speaker.

3.2.2 Variables of Performance Perspective

To cover the different variables discussed earlier in terms of stereo microphone techniques, the microphone configurations decided upon for the experiment were:

1. Binaural stereo pair: Neumann KU100 Dummy head microphone (Omni polar response)
2. Mono: AKG c414 (Omni polar response)
3. ORTF stereo pair: Neumann km184 matched pair (Cardioid polar response)
4. XY Co-incidental Pair: Neumann km184 matched pair (Cardioid polar response)
5. AB Spaced Pair: AKG c414 (Omni polar response)

Frequency responses of the microphones in use can be found in the appendix Section 8.2.

From a logical standpoint, it is reasonable to argue that the binaural setup, which is based on the transfer function akin to a person's head rather than just two diaphragms, would result in a more realistic and immersive configuration. However, it is important to note that this may not necessarily hold true in practical applications. A key factor in this study is the appropriateness of the microphone configuration being used. A microphone such as the Neumann KU100 used for the binaural capture is expensive and fairly inaccessible relative to more commonly found condenser microphones so this may be a factor in the concluding sections regarding its suitability.

The mono capture should offer no representation of space in the lateral and vertical senses and instead aim to deliver only the sound of the decay of the environment. This will be an important factor when assessing the result of the experiment, as identifying this lack of space will give credence to the other stereo configurations, provided they are noted as being an improvement in that respect. As has been discussed in earlier sections, however, the practice of using a mono capture of acoustic measurements has been applied with some success in the past, so it may be that it compares more effectively than is being suggested.

The stereo perspectives obtained from the ORTF pair, XY pair, and AB spaced pair will vary depending on their near-coincidental, co-incidental, and spaced configurations. These different perspectives will be analysed and discussed in the findings of this experiment. Each of these techniques will be heard over headphones in the context of the experiment so it will be interesting to see how they translate the ILD and ITD between the capsules to the ICLD and ICTD that will be conveyed to the listener over headphones.

3.2.3 Microphone positioning

The position of the performers downstage implies a standing performer, and as such, in keeping with the ISO3382-1 standard [71], a height of 165 cm was chosen and used for the microphone position. The upstage position is that of a drummer, who would be seated and therefore listening from a lower perspective than the other performers and a height of 130 cm was used based on measuring a performer in the drummers position. This is a difficult thing to consider with unknown heights of future performers but a position was used that was deemed to be approximately suitable for most performers under consideration at the time.

In regard to positioning of microphones, the XY co-incidental pair of microphones were positioned at the stated position within the space for each performer. The diaphragms were separated by an angle of 90°; each diaphragm was therefore 45° away from facing directly forward. The ORTF pair adhered to the predefined parameters of the cardioid capsules being separated by 17 cm, facing forward from the performers perspective with an angle of 110º between the capsules. Both the mono microphone and binaural head were positioned at the suggested performer position in the room. The only configuration without a predetermined location was the spaced or AB pair. To enable the spacing to be used for all three downstage positions, a position of 25 cm on either side of the central performer position was used, for a total of 50 cm between the pair of microphones. The rationale for employing omni-directional polar responses instead of cardioid responses lies in the fact that, while a cardioid response would enable a forward-facing viewpoint, positioning the microphones directly forward on an axis perpendicular to the listener's ear direction could result in a more pronounced rejection of sound from behind, which may not align with the desired perspective of the listener. An omni-directional microphone, however, becomes increasingly directional as the wavelength of the frequency gets nearer to the diameter of the diaphragm [52]. Therefore, using a large diaphragm condenser such as the AKG c414 means that the higher frequency content will increase in directionality, which will hopefully yield a more natural result.

3.2.4 Capture of sine sweeps

The sine sweeps generated within Matlab were played back in the performance space from each of the seven locations of where instrument capture microphones would be located as shown in Figure 27. The speaker chosen for the playback of the sweeps, and shown in Figure 29 was a Dynaudio BM6a.

Figure 29: A Dynaudio BM6a Lousdspeaker as used in this study. From [93].

The Dynaudio BM6a is an active two-way nearfield monitor that contains a 7-inch woofer and 1.1-inch soft dome tweeter [93]. This results in a quoted frequency response of 41Hz to 21kHz +/- 3 dB and a max SPL level of 118 dB RMS. The reason for this choice of speaker was its ability to represent the lower frequencies at a louder volume before distortion than other speakers that were trialled. The same speaker unit was used throughout the capture of sweeps to maintain consistency and to allow the specific speaker to be equalised at a later stage through the process of inverse filtering.

Pro Tools was used to import the 48 kHz wav file of the sweep that was prepared in Matlab. The frequency sweep was then repeated four times, pausing between each pass. On a detachable, rotatable stand, the playback speaker was originally oriented forward (0°) in the performance area at the intended source location. The speaker was turned to 90°, 180°, and then 270° after the first sweep. Each source speaker location was captured using a different configuration by repeating this technique, and each of these positions was subsequently captured from a different listening perspective. 560 sweep captures in total were used to map the area from the four distinct performance locations.

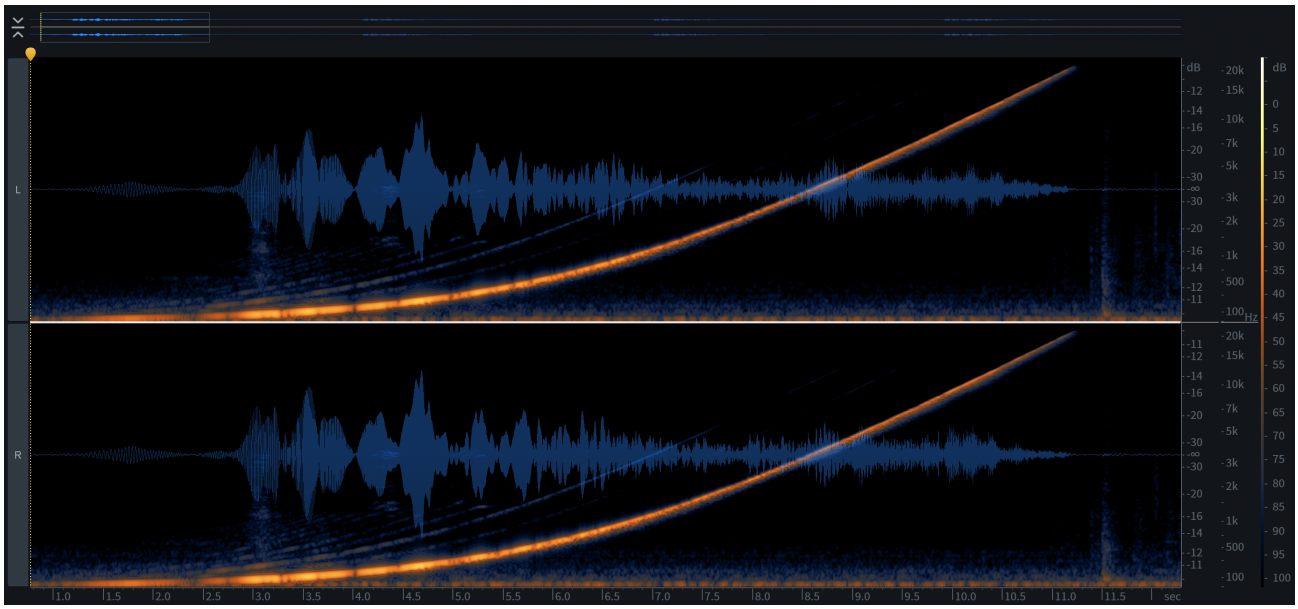


Figure 30: A visualisation using Izotope RX of the response of the kick drum location emitting forwards at 0° prior to consolidation with 90°, 180° and 270° excitations captured with a KU100 in the downstage centre performance position

3.3 Deconvolution of the performance space

After the capture of the sweeps had been completed from all listening positions with all of the different microphone configurations, the audio needed to be consolidated and exported from Pro Tools. To do this, the four directional excitation's of the space were summed within Pro Tools. As a result of this summation process, the combined signal would experience clipping if the individual files were not attenuated. Therefore, a level reduction of -6 dBFS was selected, which corresponds to half the intensity of each individual capture. This reduction allowed for the creation of a single stereo audio file representing an average while still remaining within the dynamic range of the system. This average gave a representation of how an omni-directional source sound would excite the environment as opposed to only exciting the space in one direction. This method allowed for a much greater signal-to-noise ratio because of the volume achieved at lower frequencies, which would have been an issue with a dodecahedron speaker,

as discussed in Section 2.4.3. All files were truncated to 12 seconds at the point of export to maintain the decay of the room at the end of the sweep.

This consolidation of sweep captures resulted in a set of 140 new recordings that demonstrated an omni-directional excitation of the space from 7 different locations in the room from multiple perspectives in the 4 listening positions.

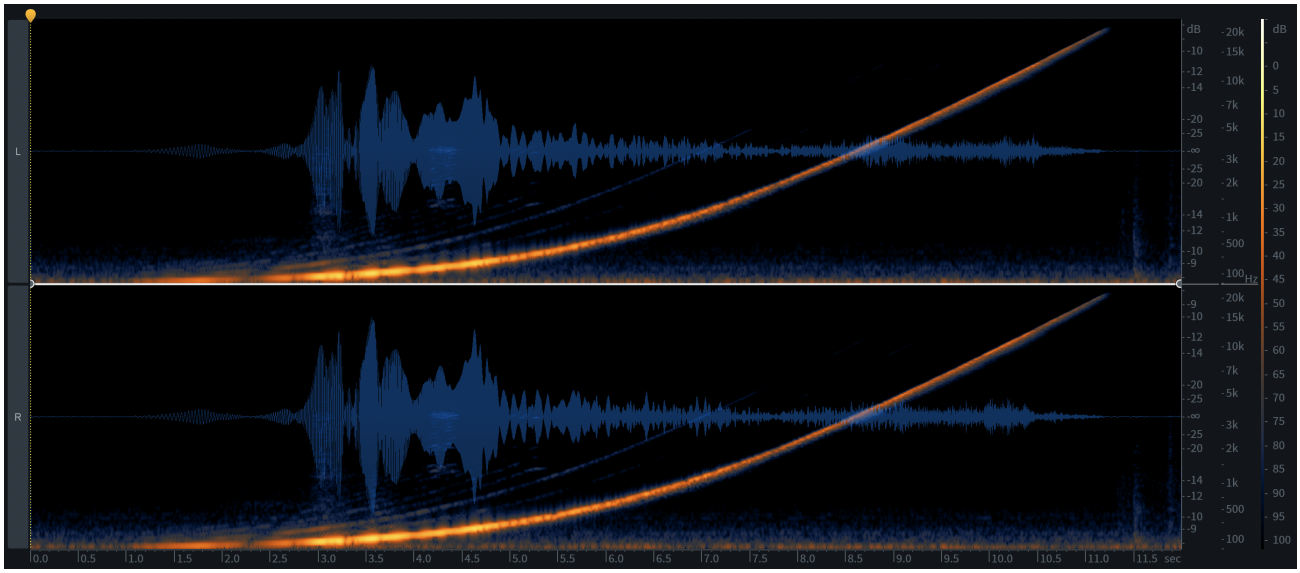


Figure 31: A Visualisation using Izotope RX of the consolidated omni-directional response of the kick drum location captured with a KU100 in the downstage centre performance position. Note the emphasis in the lower area of the frequency spectrum after the consolidation of all angles of room excitation.

This consolidated audio file was then normalised and processed using Matlab shown in Section 8.4, to create an impulse response using the process of deconvolution using the inverse of the sweep that was projected into the space. This resulted in an impulse response that could then be used in the convolution process of the experiment.

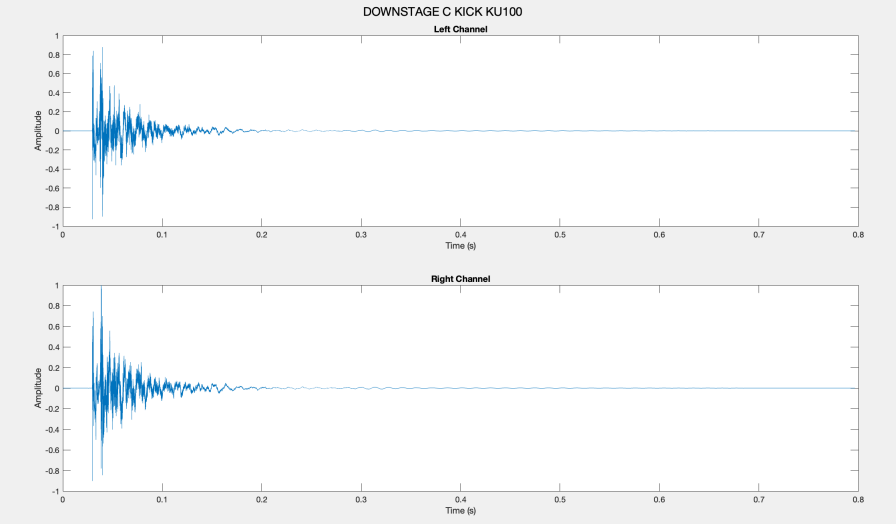


Figure 32: Impulse response in the time domain of the Kick drum location from the perspective of the downstage centre performer captured using a KU100 dummy head.

This process was repeated for all of the 7 instrument positions from the multiple perspectives of the 4 performer locations to create a data set of 140 RIRs. Each of these were checked by

running a DI'd electric guitar through a convolution reverb plug in while in the performance space to ensure that there were no obvious or noticeable errors in terms of localisation that would be detrimental to the experiment.

3.4 Measurement of test equipment

Because the sweep signal used was not equalised prior to the experiment to take into account the variation in the frequency response of the speaker being used, it was necessary to apply this filtering to the impulse response. To enable this process, the setup used to capture the RIRs in its entirety was taken and measured in an anechoic chamber.

The sine sweep created to measure the performance space was reproduced using the same signal path and captured at a distance of 1 metre using an Earthworks M30 microphone [94]. At this distance, there should be a flat response within the free field from the perspective of the microphone, with very negligible roll-off of high frequency content due to a very minimal loss of energy over that distance. The space being anechoic results in the only sound being captured by the microphone being that of those directly emitted from the source, with no introduction of later reflections. The chamber used for this as shown in Figure 33, was the isolated anechoic chamber based at the Audiolab's Genesis 6 Building at York Science park [95]. The chamber is built in a separate building to aid acoustic isolation and is a 3 m x 3 m x 3.5 m space. The combination of acoustic treatment on all surfaces, including underneath a mesh floor, enables, via use of fixed mounting points, the ability to emit and capture sound in an environment free from any reflection.

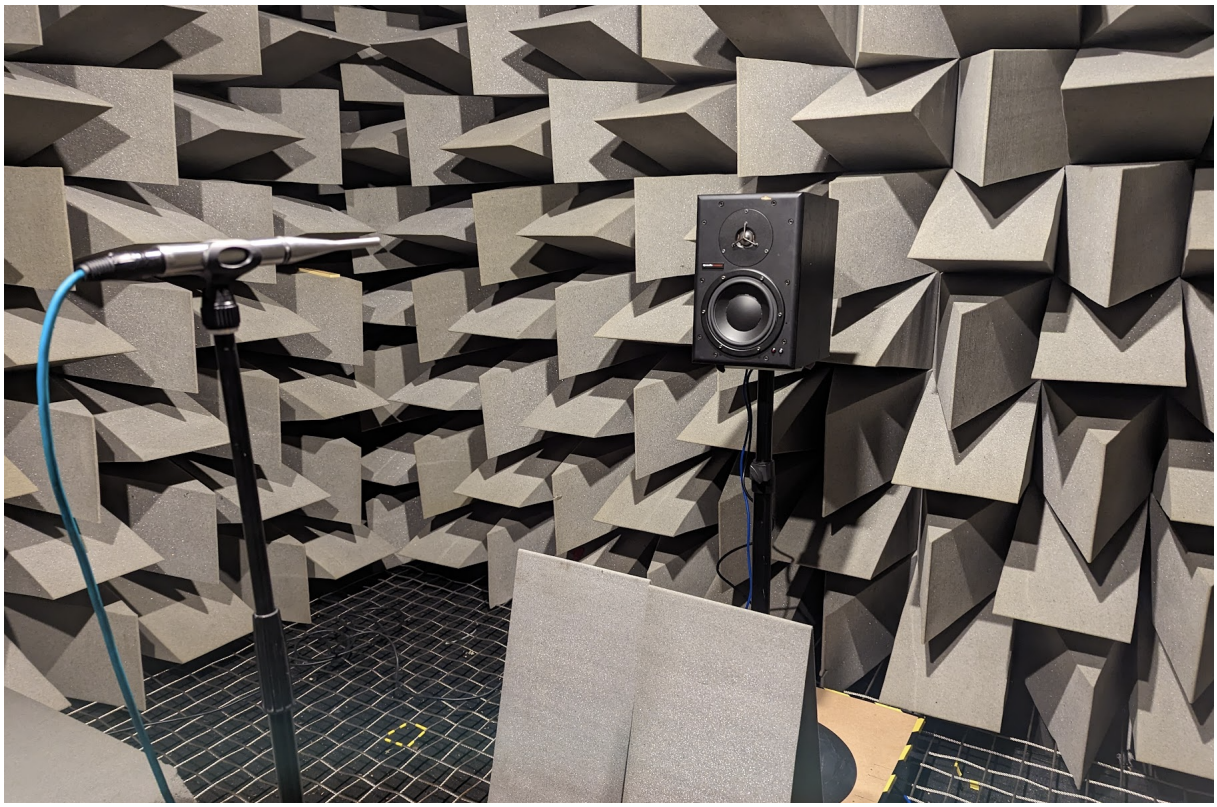


Figure 33: Measurement of the Dynaudio BM6a speaker and recording system in an anechoic chamber

After the successful capture of the sweeps, an inverse filter was generated using Matlab, the code for which can be seen in Section 8.4. The diagram presented below in Figure 34

illustrates the response of the Dynaudio BM6a loudspeaker when utilised for projecting the sine sweeps within the performance space. Additionally, it showcases the corresponding filter that is necessary to equalise the influence of both the speaker and the capture system.

The filter created was a minimum-phase filter using a non-linear fast fourier transform (NFFT) resolution of 1024. The experiment involved testing different frequency ranges for the NFFT. It was determined that the most natural-sounding and transparent response was achieved when using a bandwidth of 60 Hz to 17 KHz. Within this range, a maximum inversion of 30 dB was allowed, while outside the range, a lesser inversion of 20 dB was permitted. A Hanning window (64 point) with a length of 512 samples was applied to the IR for a natural-sounding fade in and out around the IR.

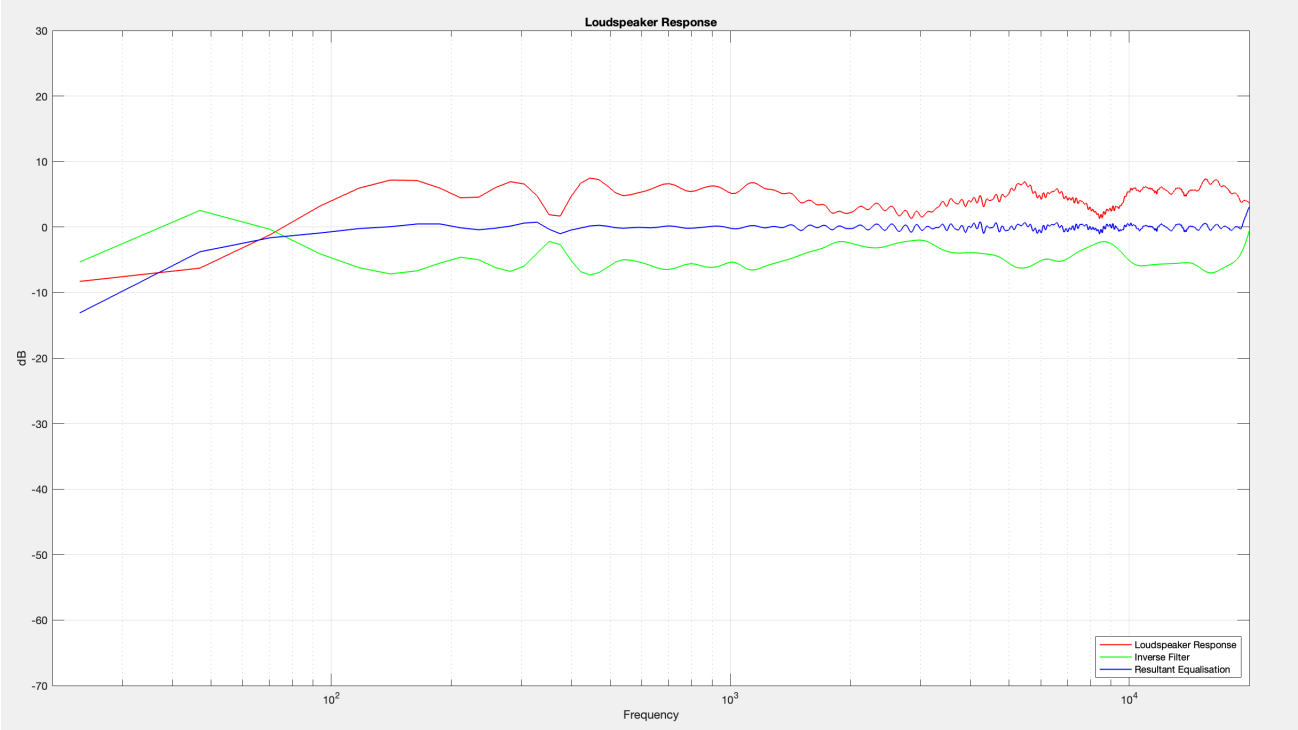


Figure 34: Response of the Dynaudio BM6a Speaker and recording system in an anechoic chamber and inverse filter created using Matlab. The red line shows the original measured response of the loudspeaker, the green line shows the inverse filter created to equalise the signal and the blue line shows the resultant equalisation which can be seen to be much flatter after the application of the inverse filter.

The frequency plot of the loudspeaker and system demonstrates a relatively consistent response above 100 Hz. However, occasional deviations of approximately 6 dB can be observed. In the context of speaker equalisation, these deviations would be noticeable and would impact the emitted sine sweep in the performance space. The speaker accentuates frequencies on both sides of approximately 350 Hz. Therefore, the use of the inverse filter is crucial in mitigating the prominence of these frequency ranges and preventing the RIR from sounding excessively inflated in the lower midrange. Due to the entire signal chain used in the earlier measurement process being part of the output signal, any impact of said components would, in theory, be removed from the impulse response.

The inverse filter that was developed for the speaker and measurement system was subsequently applied in the deconvolution process using the previously generated RIRs. This resulted

in a new set of 140 RIRs that effectively eliminated the influence of the speaker and measurement system on the room's response. One of these RIRs is demonstrated before and after the process in Figures 35 and 36, and from a frequency perspective in Figures 37 and 38.

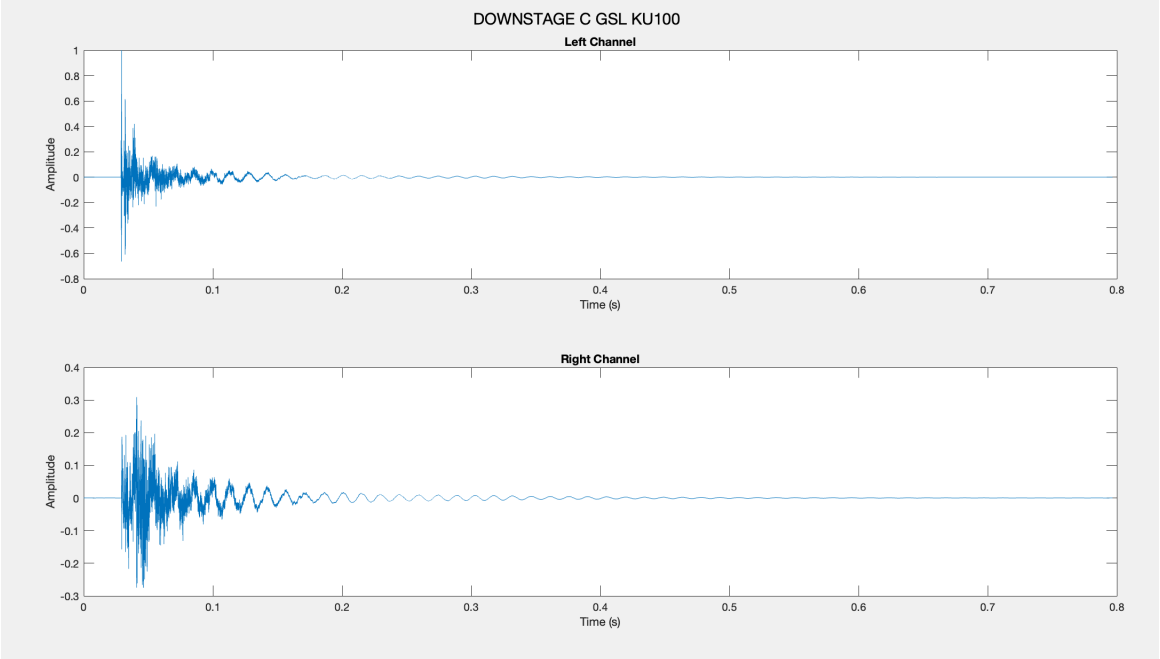


Figure 35: The impulse response of the guitar stage left position from the perspective of the KU100 in the downstage centre performance position before deconvolution of the inverse filter created in the anechoic chamber

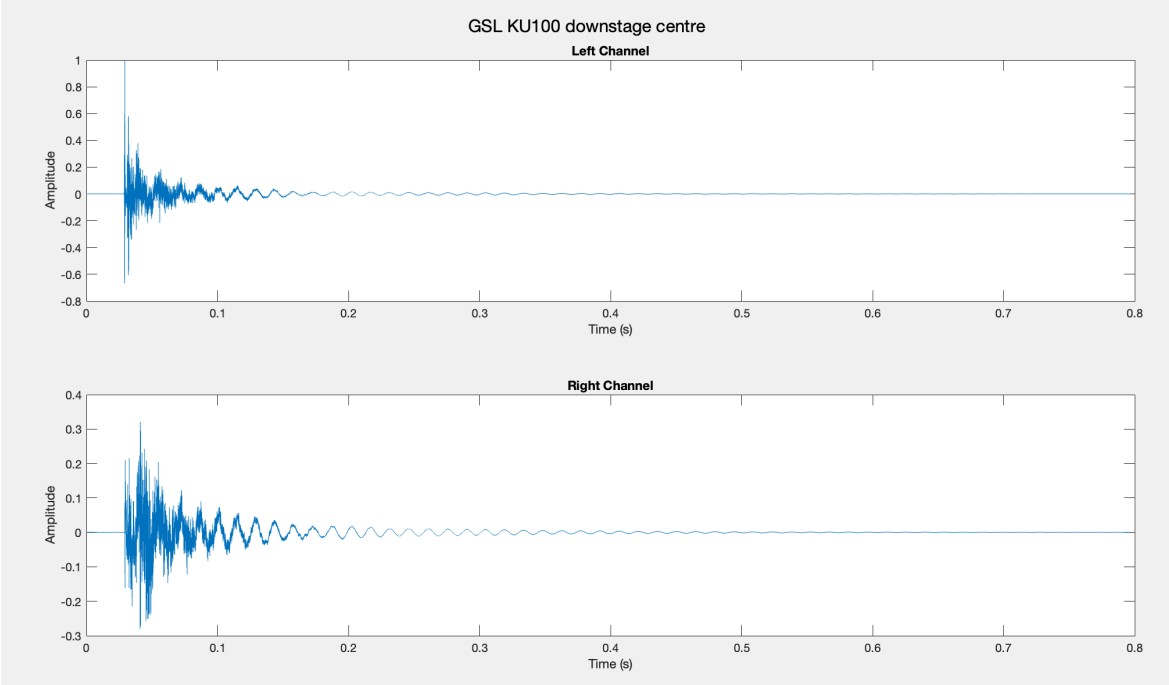


Figure 36: The impulse response of the guitar stage left position from the perspective of the KU100 in the downstage centre performance position after the deconvolution of the speaker and recording system.

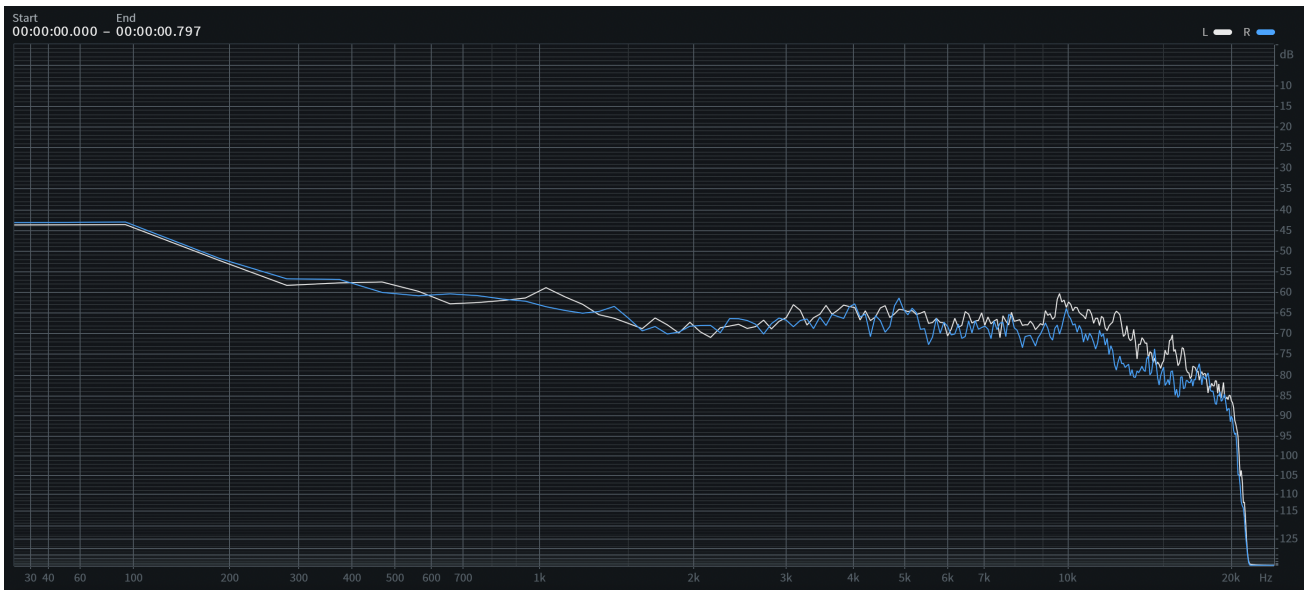


Figure 37: The frequency response extracted from the RIR of the guitar stage left position from the perspective of the KU100 in the downstage centre performance position before deconvolution of the inverse filter created in the anechoic chamber. The white curve shows the left channel, while the blue curve shows the right channel.

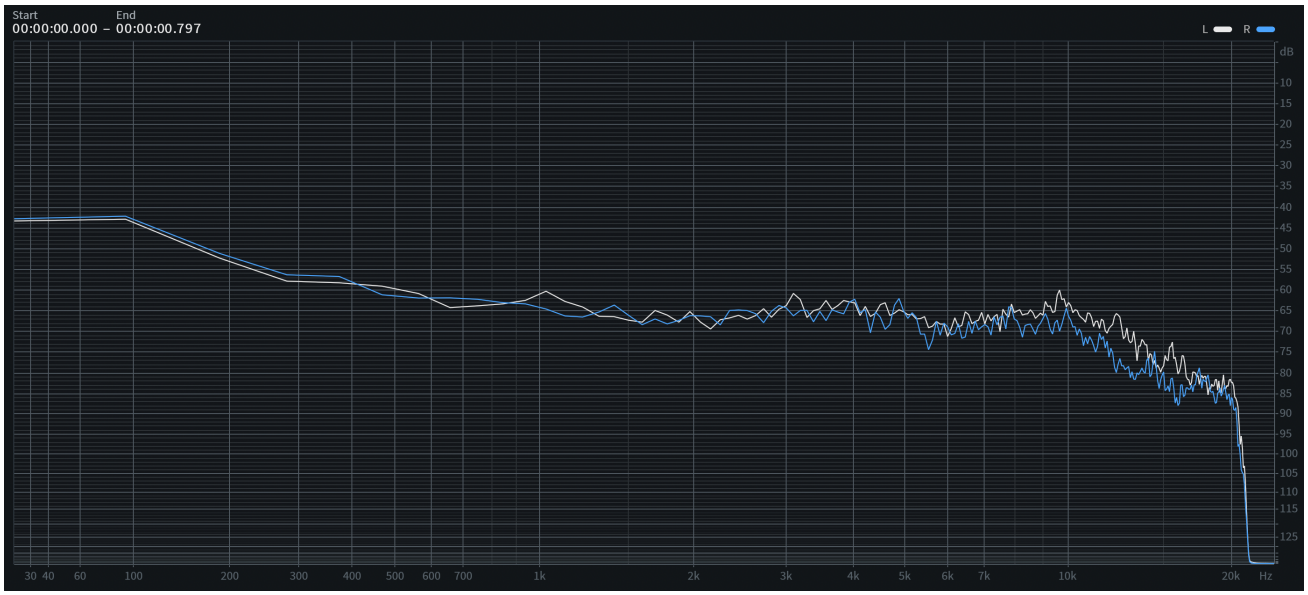


Figure 38: The frequency response extracted from the RIR of the guitar stage left position from the perspective of the KU100 in the downstage centre performance position after the deconvolution of the speaker and recording system. The white curve shows the left channel, while the blue curve shows the right channel.

The differences present in these RIRs are subtle, but what can be seen and heard is understandable when considering what the inverse filter was aiming to achieve. The example demonstrated in the plots shown is that of the guitar amplifier position situated stage left. This is from the perspective of the central downstage performer, and therefore, as the performer is forward-facing, the stimuli are located behind and to the left-hand side. This stereo information can be seen to be retained with the application of the inverse filter, and although the relationship between left and right alters equally in amplitude in the RIRs with the inverse filter applied, the distribution of frequency content is such that the intelligibility of the signal is potentially increased. There has been a reduction in amplitude in the areas between 450 Hz

and 1.5 kHz after deconvolution of the measurement speaker, which, coupled with a slight increase in amplitude in the region of 1.5 kHz to 4 kHz, has resulted in a more balanced midrange response as well as a more general balancing of frequencies overall.

3.5 Measurement of In-Ear Monitors

The measurement of the IEMs to allow for inverse filtering was a key part to the process of creating the final RIRs. It took a number of attempts to find a method of inserting the IEMs into the dummy head due to the difficulty in ensuring a seal. This is due in part to the lack of an ear canal in the KU100 itself, which gives it a limited ability to seal into the ear canal and relies much more heavily on the seal being correct around the outer part of the entrance to the ear. To ensure the IEM drivers were seated correctly; in addition to listening to the signal, Sound Performance Lab's "HawkEye" software was used to monitor the frequency response in real time [96]. When the seal is broken, there is a significant drop in lower frequency content, which is not only audibly apparent but is also able to be seen in the analyser as well as the amplitude of the captured waveform. A number of different generic inserts (sleeves) were trialled, both rubber flange style and foam style, in a variety of sizes. One of the large foam sleeves proved to be the most effective, with foam being the style of choice for the actual experiment because of its isolating properties. The Shure SE215 monitors chosen for the experiment were sourced for a number of factors, such as familiarity with the range and prior experience working with artists and musicians with these in-ear monitors in particular, as well as from a budget point of view. The price point of the SE215 allowed for the funding of four sets; one for each performer, which is very important for the consistency of the experiment. The monitors use a "single high-definition microdriver" [97] that offers up to 107 dBSPL measured at 1 kHz within a range of 22 Hz to 17.5 kHz. Shure claims that the foam sleeves can reduce external noise by up to 37 dB. There are two versions of the SE215, and it is the "pro" version selected for this experiment. There is no difference other than the cable that is supplied; the pro version has a standard headphone TRS cable (tip, ring, sleeve) made using a rugged Kevlar casing as opposed to a rubber cable with communication features terminated with a TRRS (tip, ring, ring, sleeve) connection.

The list of equipment, as shown in Figure 39, used to measure the IEMs was:

- Four pairs of Shure SE215 IEMs [97]
- RME UCX II audio interface [98]
- Neumann KU100 dummy head microphone [15]
- Macbook pro M1 Pro using Pro Tools 2022

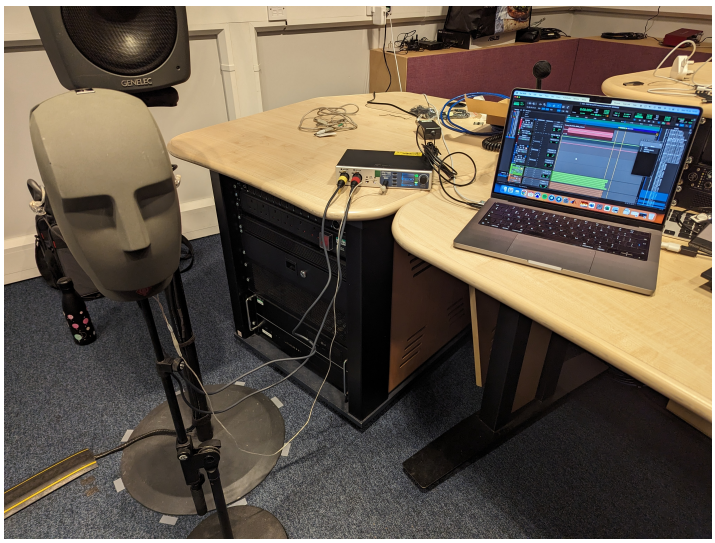


Figure 39: Measurement of the Shure SE215 IEMs using the Neumann KU100

When attempting to generate an average response, it is necessary to reposition headphones multiple times. This is because the response will vary slightly with each insertion, owing to the driver's placement in relation to the eardrum and foam sleeve. To create an average filter appropriate to all of the IEMs rather than one specific set, each pair of IEMs was measured a total of six times, and then an average of the twenty-four captures was created. This was done within the Pro Tools software as seen in Figure 40, where a sum of the sweep captures was combined with a reduction in gain to ensure the signal remained within the dynamic range limitations of the system. The four sets of headphones were all the same model and, as such, should be very similar and within the tolerances dictated by Shure, but by measuring all of them and creating an average, the aim is to achieve an inverse filter that is applicable and effective to be used with them all. The idea of creating a filter specific to each set of IEMs was considered but it was considered to be too exhaustive a process within the time constraints of the experiment for the significant number of variables that would need to be deconvolved and then managed in the sessions.

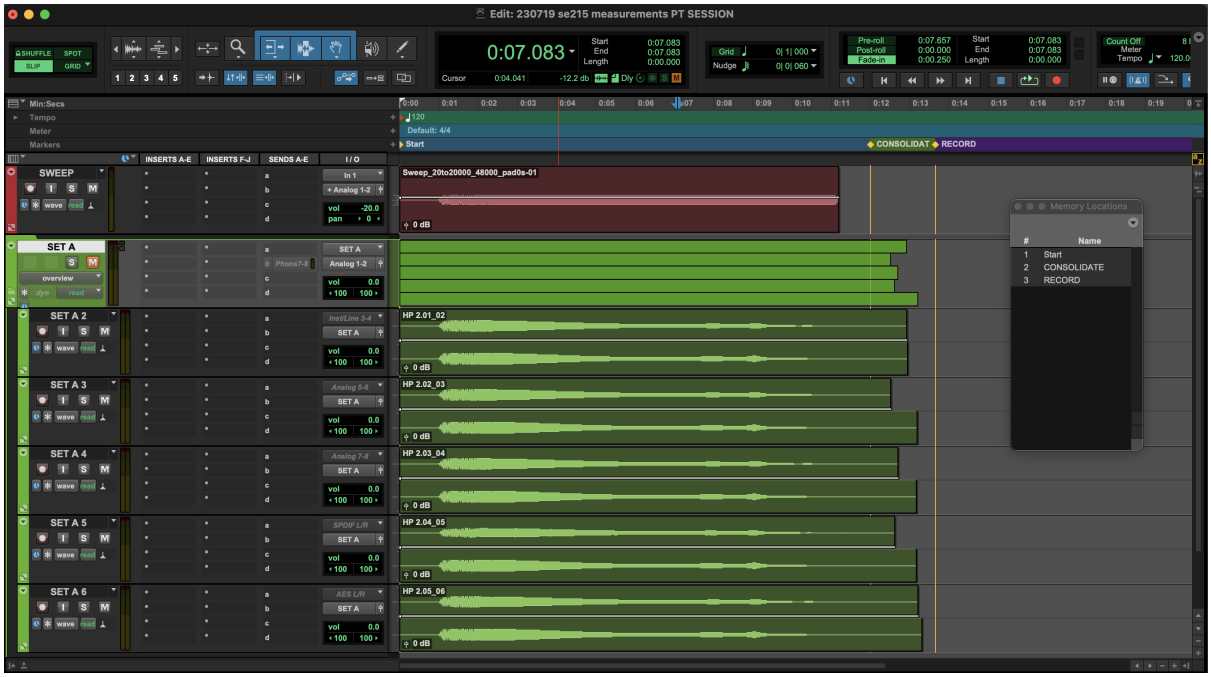


Figure 40: The Pro Tools Session shows some of the captures of the SE215 IEMs before consolidation

At this point, the average response of all four sets of IEMs was imported into Matlab and an inverse filter was created to assess the response of the IEM and to enable the impact of the IEMs to be equalised from the RIR data set that had been created.

The filter created was a minimum-phase filter using an NFFT resolution of 1024. The frequency range of the NFFT was experimented with, and a bandwidth of 60 Hz to 17 kHz, where the maximum allowance of inversion within the band was 30 dB and outside the band, a lesser amount of 20 dB, was still applicable and suitable to this scenario. A Hanning window with a 512 sample length was applied to the IR for a natural-sounding fade in and out around the IR, and the response of the IEMs and the inverse filter created were plotted. Details of the Matlab code used in this process can be found in the appendix in Section 8.4.

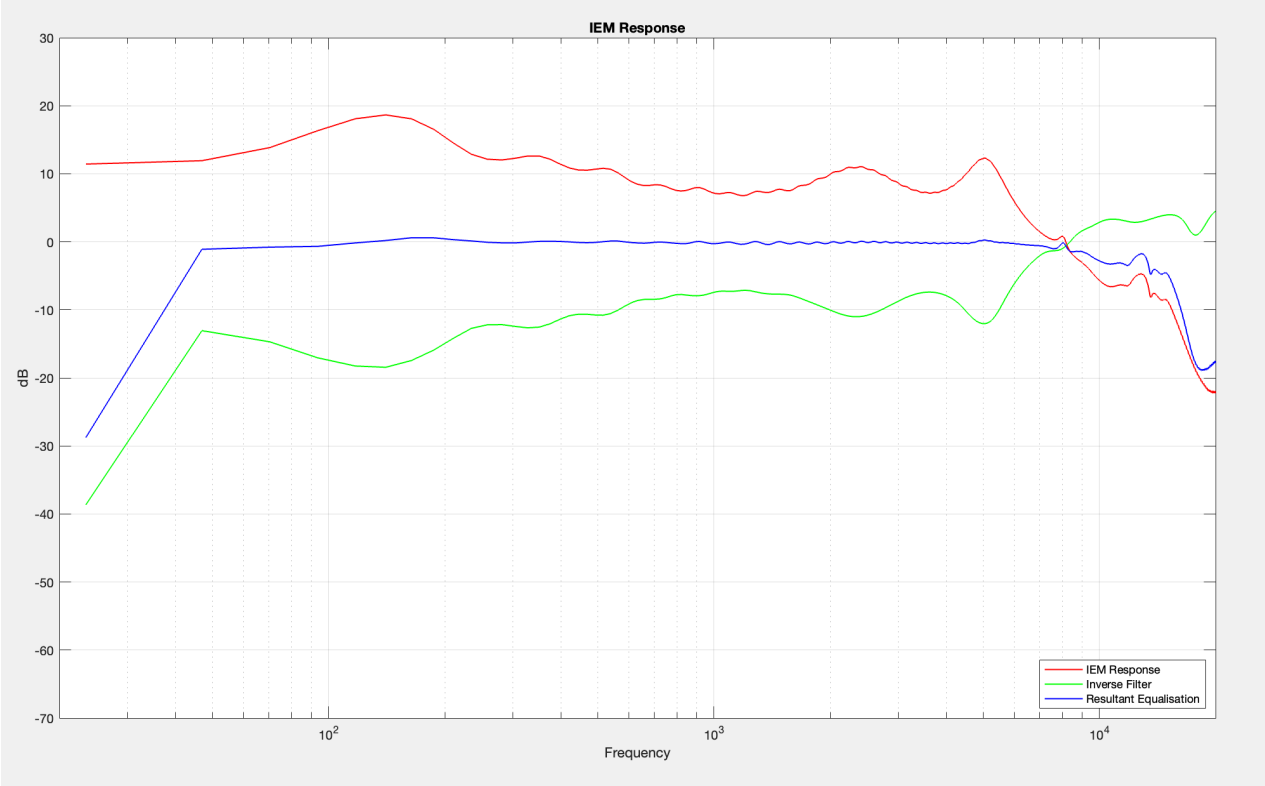


Figure 41: Response of the Shure SE215 IEM's and the inverse filter created using Matlab

As can be seen from the visualisation of the response in Figure 41, the SE215s extend quite low and consistently in terms of frequency response, but with an overemphasis in comparison to the rest of the frequency range. There is a particularly significant amount of energy centred around 140–150 Hz that could potentially cloud judgement on spatial properties by inhibiting the listener's ability to define higher frequency content. There is another significant peak centred at 5 kHz, and from the point of view of the manufacturer, it could be hypothesised that this is by design to aid intelligibility in the presence range of many sources, but for the sake of this study, a transparent response will be preferred.

From the perspective of the inverse filter, the outcome is a decrease in the lower frequencies and a limited decrease in the upper midrange, accompanied by an enhancement of the midrange overall and the very high frequencies in order to achieve a more balanced response that counteracts the impact of the IEM on the auditory perception.

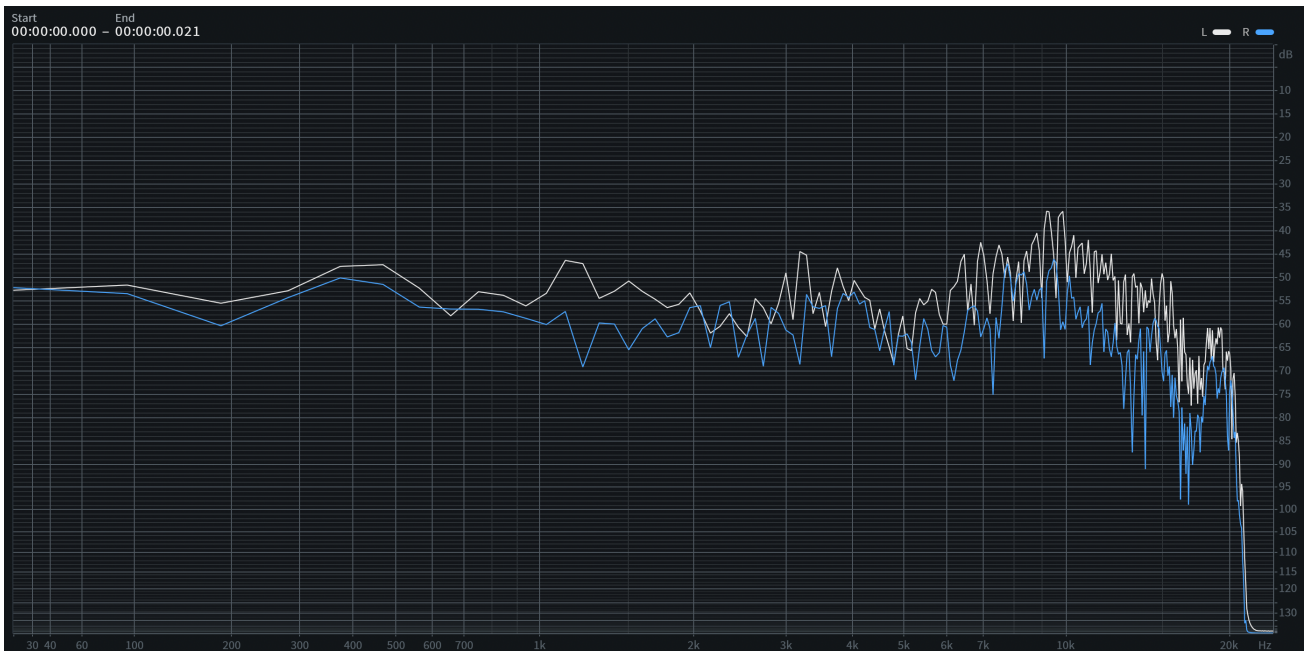


Figure 42: The frequency response extracted from the RIR of the Guitar Stage Left position from the perspective of the KU100 in the downstage centre performance position after the deconvolution of the IEMs. This is also inclusive of the previous filtering applied to remove the measurement system.

When considering the same RIR as when looking at the application of the inverse filter for the measurement system, it is evident that the inverse filtering of the IEMs has had a more significant impact on the RIR, as was suggested in the literature review of this thesis in Section 2.3.7.

The response demonstrated in Figure 42, has removed the abundance of lower and lower mid range frequencies that were unrepresentative of the space that has been captured and has allowed for a more transparent experience of hearing the original response of the room. This inverse filter was applied to the entire data set of RIRs to be used for the experiment and each one was trialled in the space to check for any obvious anomalies. The process was deemed as a success and after taking all three data sets to the performance space and listening to the response of the original capture, then with the addition of an inverse filter of the measurement system and finally with both the inverse filter of the measurement system and the inverse filter of the IEMs applied, it was apparent that each stage had offered improvement. The most noticeable difference came when applying the inverse filter created from measurement of the IEM and from a performers point of view, this is the one that initially seemed to offer the most perceived improvement of tonality and also localisation of sources.

Chapter 4

4 Methodology

In order to address the research question of this thesis, an experimental design was developed to specifically examine the variables of interest while keeping other potential variables constant. The purpose of the study is to look at plausible methods of representing space with a view to identifying the most suitable stereo configurations as opposed to an objective best, and as such, the testing needed to allow for multiple stimuli to potentially be of use. The context in which the information is relevant is live performance for a group of amplified musicians, such as a rock band for example, as opposed to a classical or acoustic setting so this was a factor implied within the creation of the testing scenario.

4.1 Latency considerations

At the point of using a system in a practical sense for real-time monitoring as opposed to just capturing audio to be played back later when capturing the acoustic response of the environment, latency becomes a consideration. As discussed earlier in this thesis in Section 2.5, a time of 6.5 ms or less to mitigate any audible delay in the IEM's is preferable. The audio interface that was used to capture the RIR's did not offer low enough latency times, even at very low buffer sizes so an alternative was sourced for the experiment. The portable interface used for the measurements was ideal due to it's ability to operate from wherever the acoustic source was based; however, this portability was no longer a benefit once the experimentation was taking place. As such, an Avid HDX system [99] was sourced to be used for the main testing as it offered much a much lower amount of latency.

4.2 Testing Setup

The system setup for the initial pilot test was as follows:

- SSL AWS900+ Console for preamplification and routing [100]
- 2 X AVID HD IO 16 Channel Interfaces [99]
- 4 X Yamaha MG12XU mixing consoles to facilitate headphone amplification in the live room [101].

4.3 Test Methods

4.3.1 Consistency of Performance Variables

There are many variables that contribute to the appropriateness of a monitor mix for a musician. Two key factors are the balance of elements in terms of relative loudness between elements and the overall loudness of the mix itself. Within this study, the factors under discussion are the spatial and sonic representation of the sources, so it is important to remove the other volume-related factors from the equation so they do not influence the results. During the first pilot test, a method of monitoring mix creation and a way to switch between them were evaluated to aid this element of the study. Initially, the method used was to create a mix using one of the variables applied over each channel, then maintain the mix balance and switch only the variable applied. However, the existing approach proved inadequate in maintaining balance due to minor level variations among the convolution methods. Consequently, it was decided to generate separate Pro Tools sessions for each variable and conduct independent mixing for each test in order to preserve the perceived relationship between the channels.

4.4 Data Collection

For the experiment, the main method of data capture was through questionnaires given to each performer to answer after every performance using different monitoring stimuli. They also had the ability to log their comments to aid in the context of judgements when it comes to dealing with the data captured. The variables were assessed in random order and repeated at irregular intervals throughout the experiment to ensure consistency of responses. The reason for this is to recognise where an artist simply prefers one variable to the previous as opposed to treating each stimuli as an individual event. The performance taking place during the experiment was also captured in a multi-track format so that any anomalies or queries could be listened to or investigated at a later date, should it be necessary.

4.5 Initial Pilot Test

An initial pilot test was created to understand the limitations and any potential issues with the methodology suggested. Through the university network, several potential candidates were spoken to about conducting the test, with an emphasis on finding musicians with an appropriate setup for the experiment in terms of instrumentation. It was also essential to source musicians that were capable of playing a chosen piece of music repeatedly without any major timing or other performance issues and as such, rather than suggesting a piece of music to be played, the musicians sourced where to play their own material that they were already comfortable with. It was vital that performance, or more specifically, the quality of performance, was not a contributing factor in the preferences of the musicians and their listening experience.

The areas of interest for the pilot test were:

1. System capabilities
2. The consistency of mixes for performers
3. The length of performance required to justify the results
4. The suggested statements for the questionnaire

Figure 43: The view from the control room during the first pilot test

The setup of the ensemble used during the first pilot test consisted of:

1. A drum kit
2. A bass guitar (through an amplifier located centrally in front of the drum kit)
3. An electric guitar stage right (through an amplifier)
4. A keyboard stage left (through an amplifier)
5. A Vocalist (downstage centre)

Due to scheduling limitations, it was necessary to conduct the first pilot test prior to acquisition of the 4 sets of SE215 IEMs. As such, only one pair of IEMs was used, and only the vocalist in the experiment was provided with an in-ear monitor mix for the first pilot test. Where this provided a limitation regarding the thorough analysis of the plausibility of the experiment was in the area of latency consideration. To remedy this, a process was created within the Pro Tools system to simulate the processing requirements of the full experiment. Within a Pro Tools system, the hardware buffer size that can be selected in the playback engine window refers to the low-latency hardware buffer domain and is applicable to live input channels [102]. By using auxiliary inputs to feed all of the sources to the four monitor mixes and the rest of the routing to simulate the full experiment, each of these elements was placed into the low-latency hardware buffer domain, and as such, the system was running at full capacity with the additional three monitor mixes being terminated once back in the analogue domain.

The amount of latency was measured to be 9 ms, which included the entire processing chain. Although 9 ms is higher than the 6.5 ms that the research in Section 2.5 of this thesis outlined, the latency was deemed to be indistinguishable by the performer. There were no issues with latency and no other methods of processing needed to be explored, and as such, that area of the first pilot study was deemed to be successful.

Another important factor to take into consideration with the first pilot test was that, due to the event being prior to the acquisition of the four sets of IEMS, the test was conducted with the set of impulse responses where the measurement system had been compensated for with an inverse filter but prior to the significant improvement of the responses when the inverse filter for the IEMs had been applied. Therefore, any data harvested in terms of the preference of the performer will not be directly used in the latter stages of this thesis.

Having the vocalist as the performer with the in-ear monitor mix ended up being very helpful in the sense of recognising the difficulty of that vocal being represented within the monitor mix itself. Throughout the planning of the test, the question had been raised about the difficulty of how to represent a vocal that is essentially an internal source in the method that is being used within this experiment. For the purpose of the initial pilot test, the method of transmission requested and preferred by the vocalist was to have his own voice fed back completely dry to his monitor mix with no spatial qualities present. If this were to be used in the main pilot test and further studies, it would raise questions about the context of other sources, and due to what is being questioned within the study, the relevance of the vocal being present within the monitor mix is not relevant to what is being asked of the performer. The method of testing is not to suggest a working process but to facilitate an understanding of the stereo configuration variables. Undoubtedly, however, this has raised a question that demands some further consideration for any exploration of the spatialisation of live monitoring systems with respect to vocal representation. This is outside of the scope of this thesis to resolve; therefore, the omission of the vocal element will be the preferred method to assess the other variables.

Having previously noted the importance of ordinal scales [60] the questionnaire was set out in the form of statements that were ranked using a Likert scale [103]. Statements that were strongly disagreed with would score 1, while statements that were strongly agreed with would score 9. The statements ranked at the end of each performance were:

1. The monitor mix immersed me in the environment
2. It was plausible that I was listening to the instruments in their positions in the room
3. The monitor mix allowed me to perform better
4. The monitor mix enabled us to play together effectively
5. The monitor mix sounded professional
6. The monitor mix sounded tonally correct

The idea behind the questioning was to try and ascertain how well the monitor mixes; using the different representations of the musical sources through convolution, could represent a realistic replication of the space. To enable the musicians to contextualise what was being asked of them, the statements were made very literal so there was little room for confusion or incorrect interpretation. There were three main areas that needed to be assessed. The localisation, the impact on the performance and finally the tone or general sound of the reproduction. Localisation was covered in the first two questions. The first question bears resemblance to the subsequent one, albeit with the aim of adopting a broader, less specific perspective on the participants' sense of immersion in the environment. Rather than using terms like localisation, the second question was simplified to be more literal, focusing on whether it was plausible that the sound source was coming from the position of the element in the room, something that could be understood just by being present in the situation. The next two questions related to the impact on the performance and whether it enhanced the performers ability to play and whether it enabled them to play together as an ensemble. Finally, the last two questions covered the area of tonality, relating what they had heard to other professional scenarios and whether the sources sounded tonally correct. In both the initial pilot test and the main pilot test, the questions were discussed with the performers prior to the experiment so there would be minimal room for misinterpretation. The questions were formulated with the intention of determining specific information and were not knowingly directly influenced by any prior research.

4.6 Analysis of Findings

There were a number of important findings from the first pilot test. Firstly, from a time requirement perspective, it was reasonably quick to conduct the experiment and change the test variables in between performances. This was of interest because it was unclear as to whether looking at five different variables would be too many and would result in fatigue and potentially inconsistent results. Although it may not seem like a vast number of variables, having numerous instances of each requires many performances of the piece of music. Having a prolonged period in between these performances would result in the duration of the test being increased significantly.

After analysing the results of the first pilot test, an adjustment was made to question three, with the statement "The monitor mix did not hinder my performance " used in place of "The monitor mix allowed me to perform better". The difficulty in quantifying improvement was attributed to the initial question's emphasis on this aspect, while acknowledging the absence of hindrance aligned more closely with the feedback provided by the artist.

The performer was insistent on playing the whole song that they had prepared for each different stimulus, and this was deemed to be unnecessary. After discussing the process afterwards, as little as a chorus would have given the performer the ability to understand how the monitoring was supporting them and would have meant that there was less chance of listening or performance fatigue. For the main pilot test, a shorter amount of music would be insisted upon.

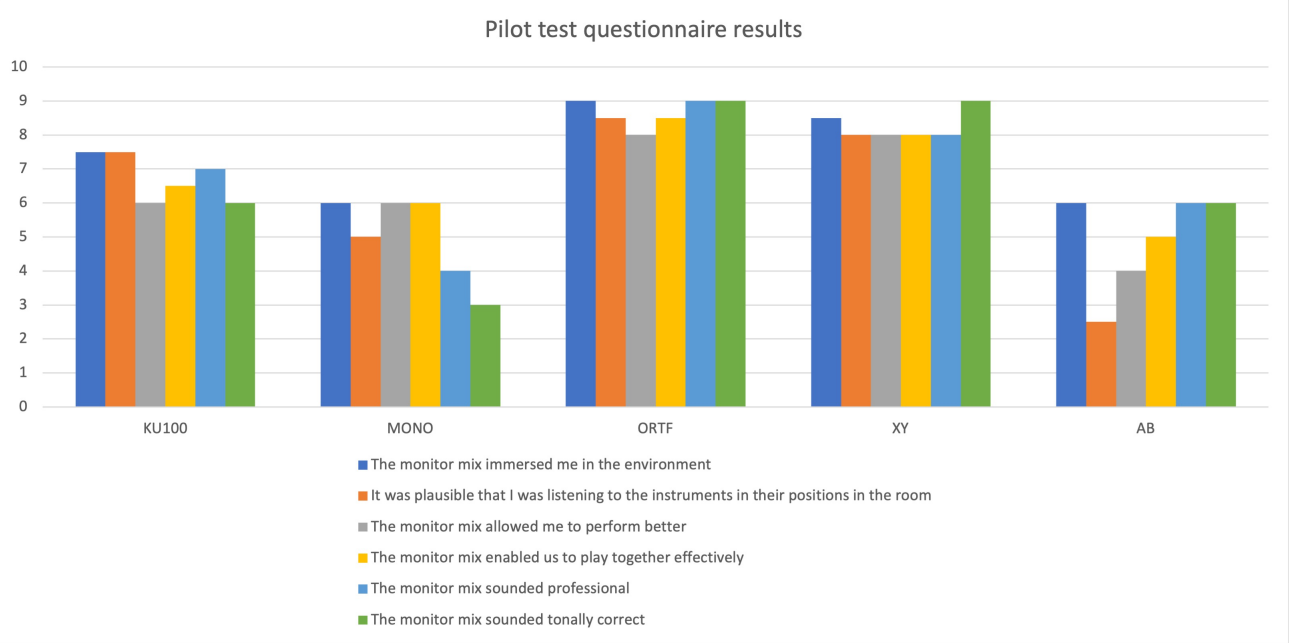


Figure 44: Results of the questionnaire given to the artist in the pilot test

4.6.1 Questionnaire Response Analysis

At this early stage in the testing process, there were some clear correlations between results that needed to be assessed to ensure the reliability and validity of further experimentation. Each of the stimuli was tested a total of 3 times in the first pilot test, and the results in the chart in Figure 44 show an average of the 3 results for each of the criteria.

As all of the criteria being remarked upon are in a positive position, for example, if the artist is being asked how much they agree something is positive rather than how much they agree on a negative attribute, then the graphs show us an overall preference by generally how high the bars are.

Something that caused concern was how poorly the AB configuration of microphones performed, specifically in comparison to the mono representation. Feedback from the artist described how the representation when using the AB capture was disorientating, leading to a requirement to reassess the RIR's to check for any errors that had been missed prior to the first pilot test. The issue was noted at the time as being specifically in regards to the stage right guitar. As the event was captured during the process, it has been possible to rerun the session at a later date and listen to the monitor mix as would have been heard at the time. In the context of the other elements monitored using the RIRs created using the AB setup, it has been found that there are no specific issues with the stage right location variable.

The collection of higher scores is also a potential issue, as scoring so highly in some areas led to little room for different ratings if there were slight preferences; however, having said that, other than the anomalies with the AB configuration RIRs, the repetition of performance variables did correlate and give a consistent result. This factor suggests it may be preferable to allow the artist time prior to the test to hear the configurations and confirm mix levels to give the performer an idea of the range of differences they will be exposed to, but obviously not give any indication in that phase as to what the differences are. This would potentially resolve the high scoring issue and also support the ability to have consistency of mixes across test configurations to ensure the thing being commented on is in fact the stereo configuration and not the balance of the mix.

4.7 Main Pilot Test Methodology

After analysis of the initial pilot test, the questionnaire was slightly altered and the statements were ranked using a Likert scale between 1 and 5 as opposed to 1 and 9. The third question about the

monitor mix allowing the performer to play better was reworded to imply that the monitoring did not take away from the performance as opposed to it being something that was giving additional support as it was difficult for performers in the pilot test to quantify improvement and feedback was more aligned with what was wrong with the various stimuli. The statements ranked at the end of each performance for the main testing were:

1. The monitor mix immersed me in the environment
2. It was plausible that I was listening to the instruments in their positions in the room
3. The monitor mix did not hinder my performance
4. The monitor mix enabled us to play together effectively
5. The monitor mix sounded professional
6. The monitor mix sounded tonally correct

4.7.1 Performance and monitoring variables

For the main test setup, there were a total of six performance variables to be used. This was due to the artist sourced for the testing using only a single electric guitar. The decision was made for the electric guitar amplifier and bass guitar amplifier to be located stage right and stage left respectively and to omit the central speaker cabinet position from the testing due to the preferred method of performance of the artist. Also from the point of view of the vocalist, due to the question of how to localise the position of a voice in the manner in which the rest of the variables were being treated, a decision to remove it from the test was made but important data was still collected from the performance position of said vocalist.

Each of the performers were using Shure SE215 in-ear monitors for their fold-back of the performance, with each performer having a dedicated mix of elements with the microphones feeding the mix all convolved individually to correspond to the perspective of the musician in their location in the room.

The specific microphones used for the experiment were as follows:

1. Kick Drum: Shure Beta 52a
2. Snare Drum: Shure SM57
3. Overhead Stage Right: AKG c451
4. Overhead Stage Left: AKG c451
5. Electric Guitar Stage Right: Shure SM57
6. Bass Guitar Stage Right: Shure SM57

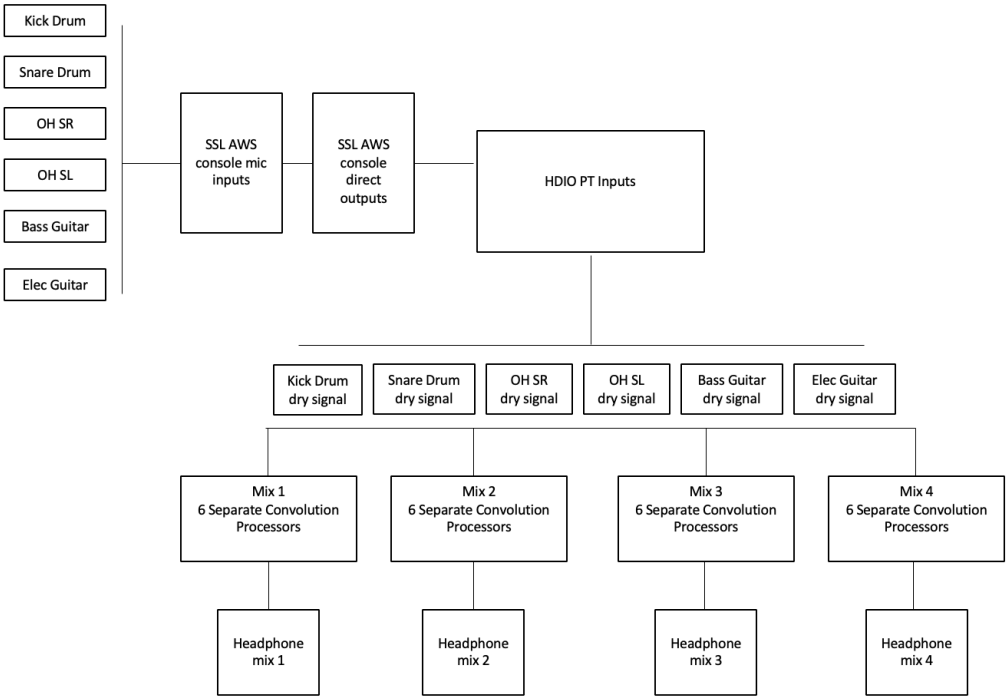


Figure 45: Signal flow diagram of the main test session

In the setup period of the experiment, the bass guitar was initially sent to the console using a DI; however, the performer requested hearing the comparison to a Shure SM57 and felt that he was getting a better response from the microphone regarding what he was expecting to hear from the bass cabinet. In terms of the experiment itself, this is arguably a preferred method of conducting the test anyway because it is taking into consideration the actual sound being emitted into the room as opposed to a dry signal of the bass guitar prior to the speaker output.

Figure 46: The artist in the performance space

The software used for the experiment was Avid's Pro Tools Ultimate version 23.3.0.89 and was used to monitor the input of each of the six source microphones via audio input channels in Pro Tools and direct, as per the signal flow diagram in Figure 45, to four internal routing folders.

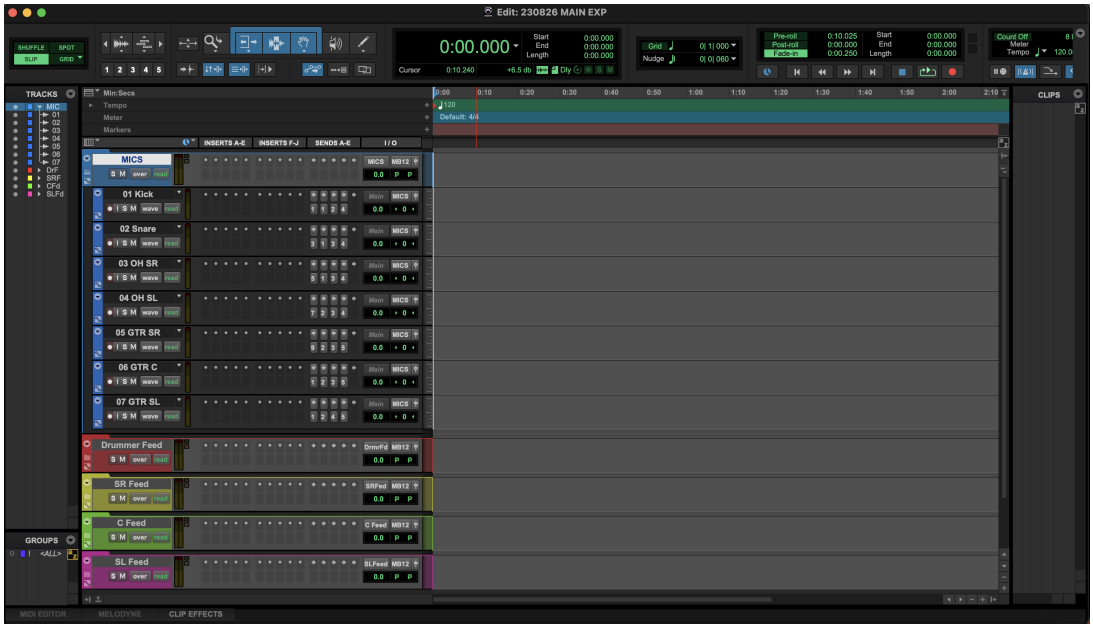


Figure 47: The Pro Tools Session showing microphone inputs

Six auxiliary inputs fed by the audio channels carrying the signals from the microphone were present in each of the four routing folders. The image Figure 48 shows seven of these auxiliary channels as the central speaker position and its affiliated information was present in case it was used in the session. This did not end up being the case and only six channels were used for the experiment.

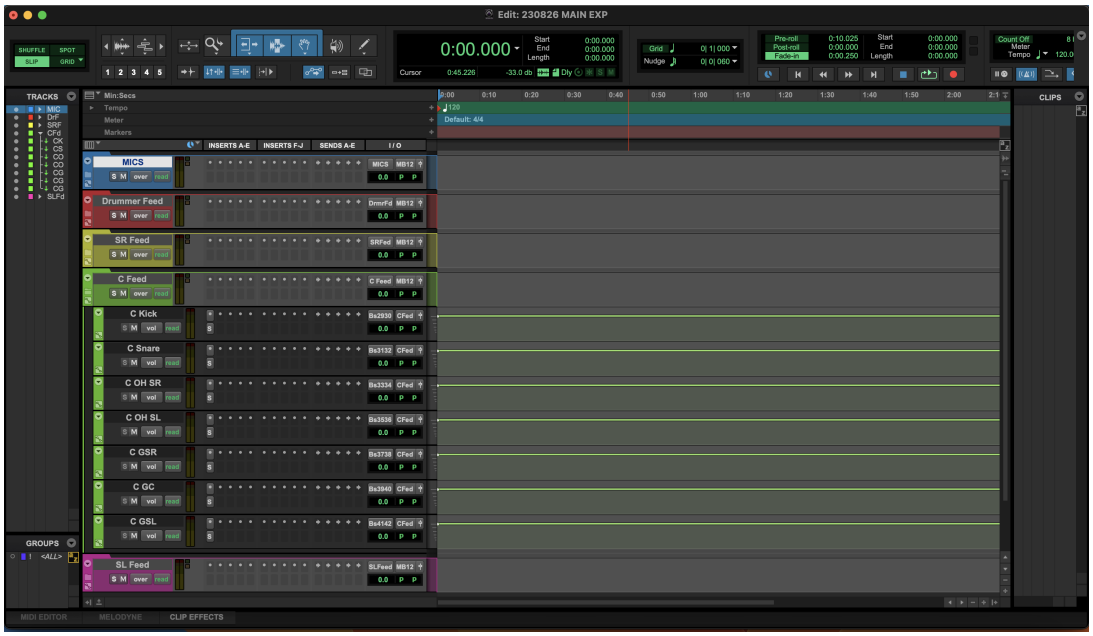


Figure 48: The Pro Tools Session showing the Routing Folders with the downstage centre expanded to show the internal auxiliary tracks

As can be seen in the Pro Tools session in Figure 48, in the first insert position on each of the channels, there is a plugin positioned over the channel. The plugin used over the live input channel

to support the convolution process was Avid's own Space plug in [104]. Each performer required an individual instance of this over each of the live microphone signals, so there were a total of twenty-eight in the session; however, due to the central speaker position not being used as previously mentioned, only twenty-four of the plugins were in use during the experiment.

Figure 49: One of the Space Plugins used in the session. This one shows the RIR used for the kick drum channel supporting the downstage centre performer with a RIR created using an ORTF microphone configuration [104].

Within the Space plugin, there is the ability to use a "wet/dry" mix of the convolution; however, the purpose of this experiment was to completely remove the dry unprocessed signal and only listen to the convolution of the sound source, so as can be seen in Figure 49, the dry signal is completely removed and the wet signal is left at 0 dB. The plugin also offers the ability to alter other parameters, such as delay times; however, the point of the experiment was to have as transparent a process as possible, so the processor was only used to convolve the signal, and no additional changes were made to any of the channels.

Chapter 5

5 Results

The main pilot study's testing provided some valuable insights from four distinct performance perspectives. These four different perspectives offer the opportunity to assess the appropriateness of the directivity of the representation of sources; however, there are other factors that may come into play, which will be discussed after demonstrating the results.

5.1 Tables and Graphics

5.1.1 Results of the qualitative assessment

The first two questions asked of participants related to the spatialisation of the sources and the resultant monitor mix. They are demonstrated as a cumulative total across all performers below to get an overall view of the effectiveness of each of the stereo configurations used. They are then individually demonstrated for each performer, so comparisons of how the performers found the stereo configurations can be assessed.

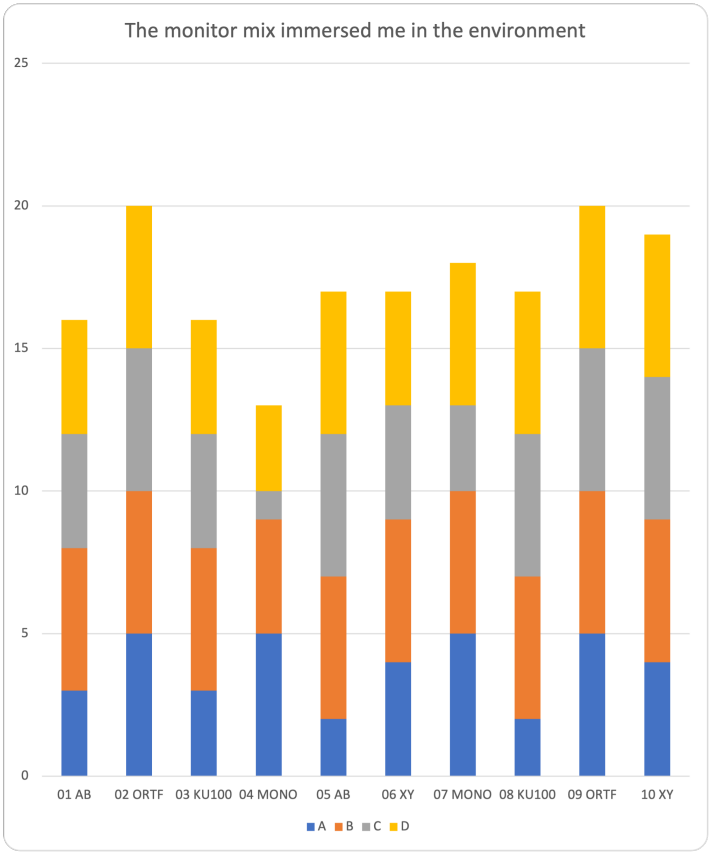


Figure 50: Combined results of all performers: A (bass guitarist), B (centre vocalist), C (electric guitarist), and D (drummer) for question 1.

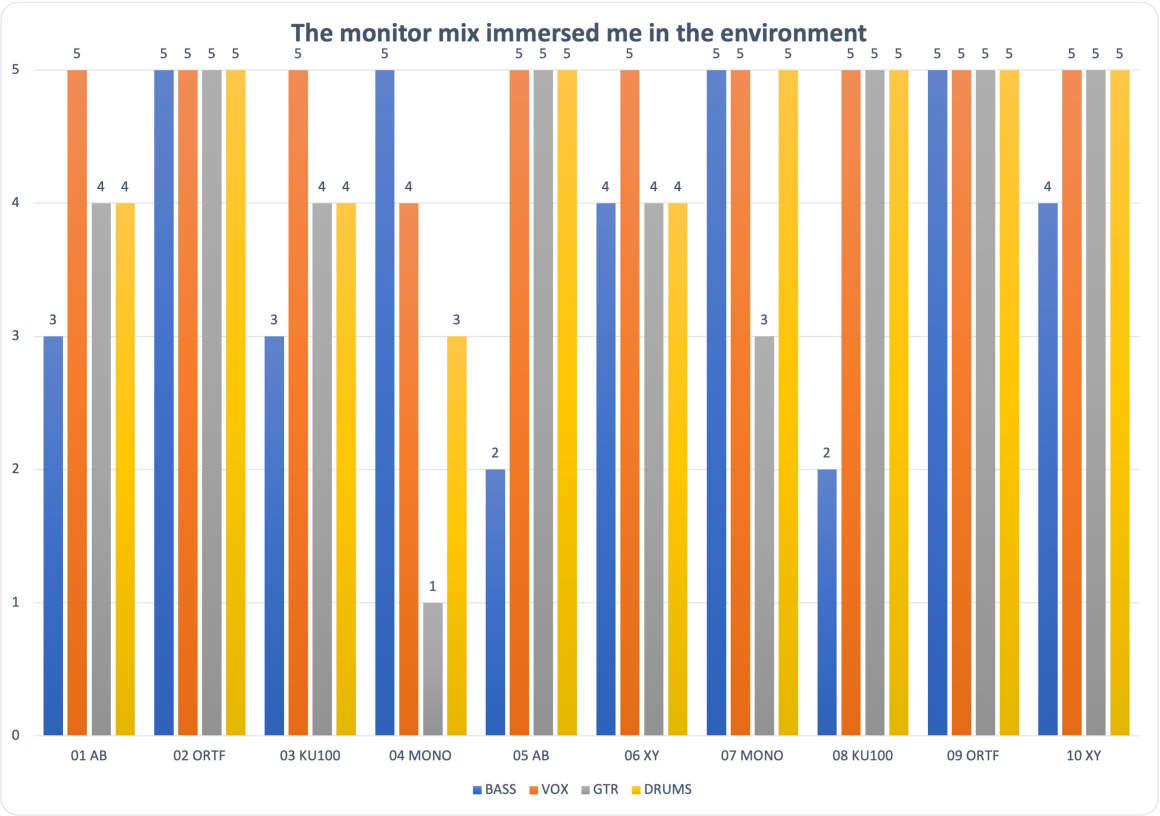


Figure 51: Results for comparison for question 1. This demonstrates clearly the overall success of the ORTF setup in particular.

Across the results of the questionnaire related to the first question about the immersive element of the mix shown in Figures 50 and 51, the standout configuration appears to be the ORTF setup, particularly as it has scored the maximum value. However, it is important to note that a large number of the results are very positive. There is also an anomaly as to why the mono signal has scored so highly, though in the context of the experiment, this can potentially be understood. Performer A, the bass guitarist, although stood downstage left, was positioned directly in front of the bass guitar cabinet. In focusing on the bass in the mix, a mono feed would potentially accurately place the performers instrument in the correct place directly behind them, or at least centrally, and therefore influence how they felt about this variable. This will be discussed in more detail when viewing the results from the perspective of Performer A.

The second question that was asked was whether it was plausible that the instrument(s) that the performer was listening to was coming from the correct location in the performance space. This is an important factor within this experiment, as it is directly tied to the aims of the thesis. The results are demonstrated in Figures 52 and 53.

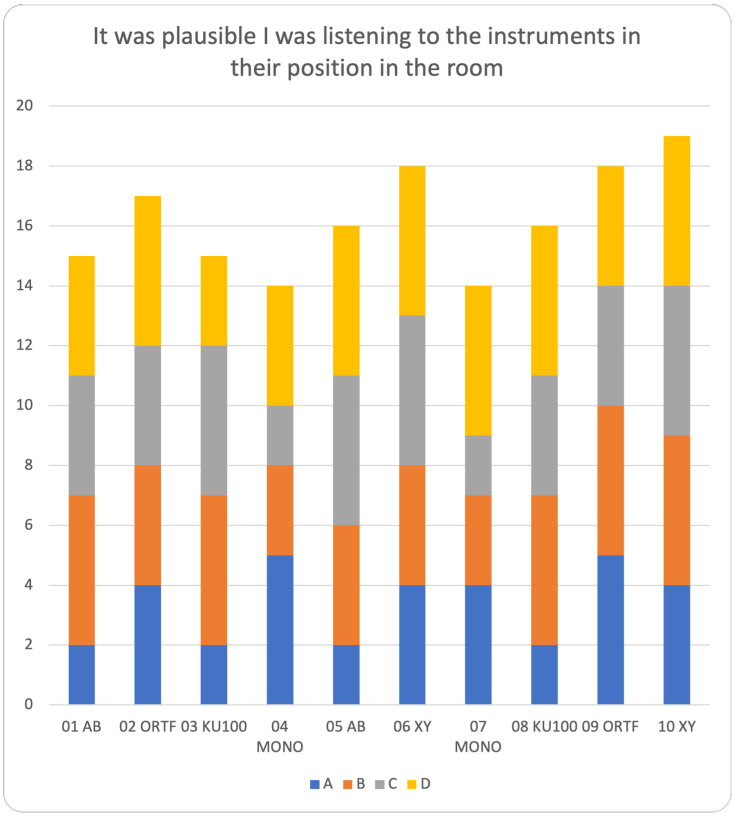


Figure 52: Combined results of all performers: A (bass guitarist), B (centre vocalist), C (electric guitarist), and D (drummer) for question 2

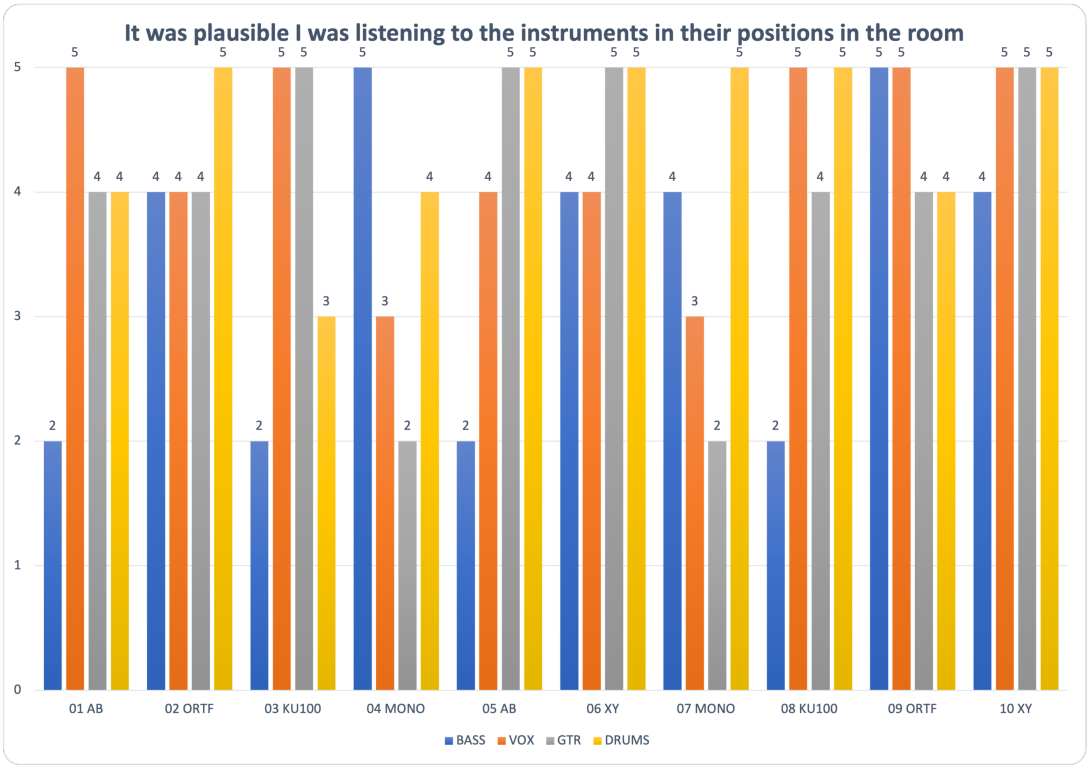


Figure 53: Results for comparison for question 2.

The result of this question on the different stimuli gave a slightly different indication of preference than the question about immersion in the environment. Overall, the XY configuration scored the

highest, followed closely by the ORTF configuration. The mono setup should have scored lower for this question, and this is the case; however, there is some support from the bass player, potentially for the same reasons as Question 1.

Question 3 in the post-performance questionnaire was related to the monitor mix not hindering the performance. The combined results for each configuration were as shown in Figure 54 and 55.

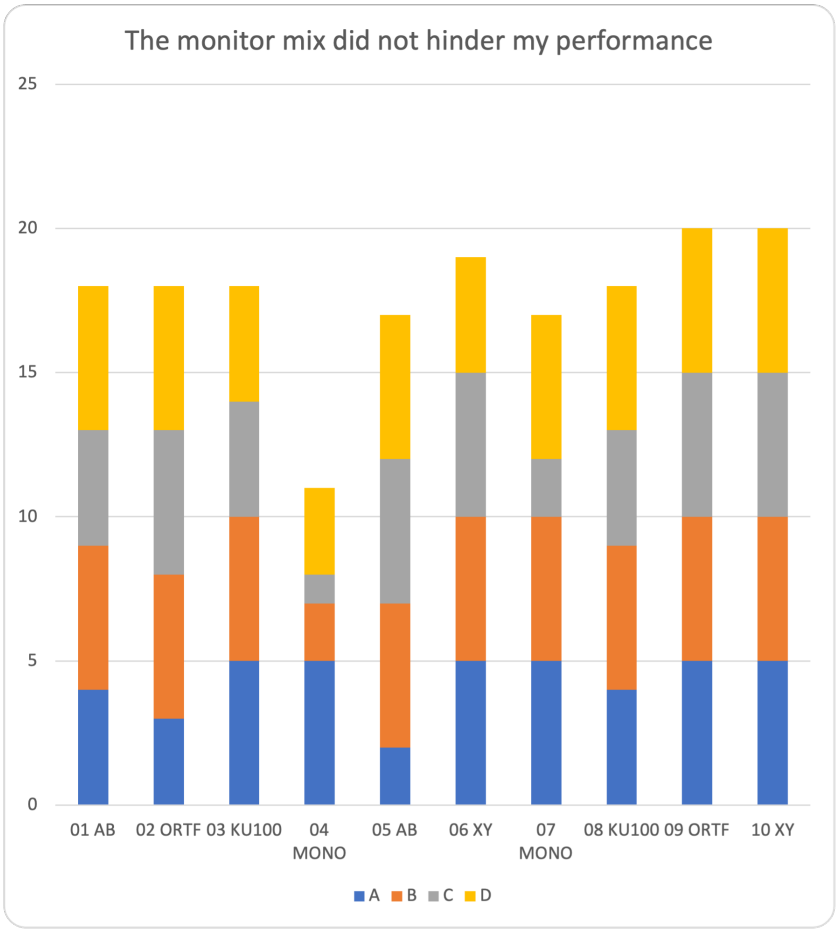


Figure 54: Combined results of all performers: A (bass guitarist), B (centre vocalist), C (electric guitarist), and D (drummer) for question 3

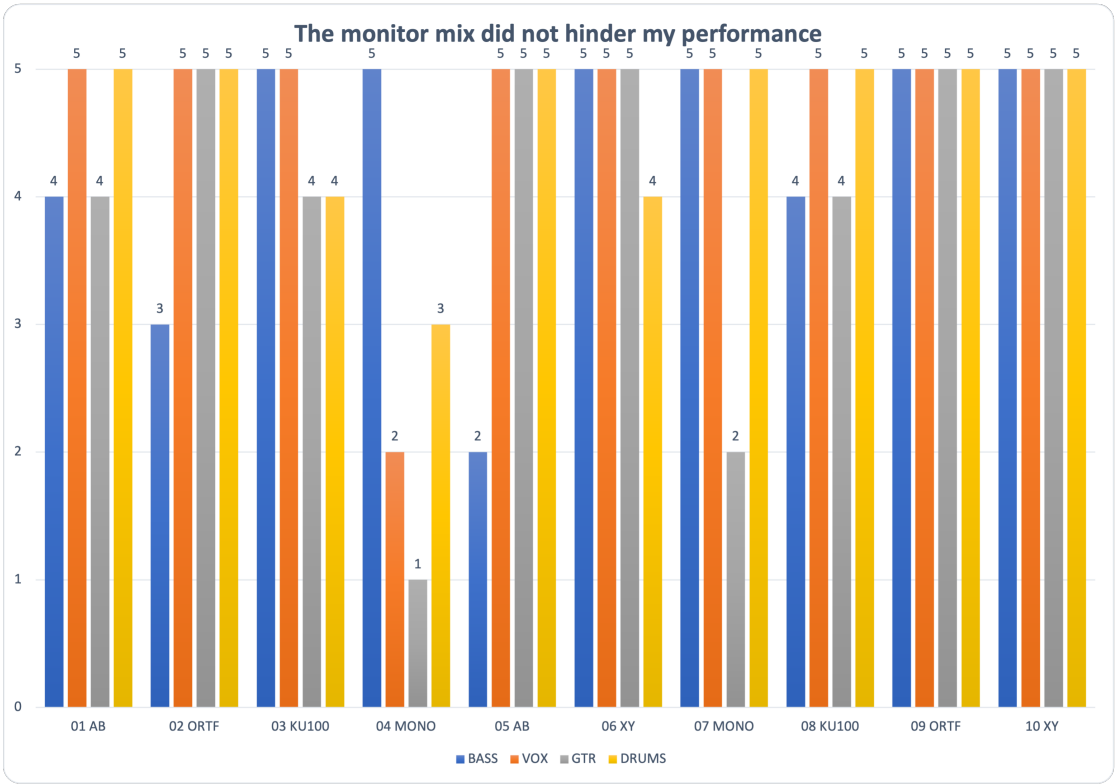


Figure 55: Results for comparison for question 3.

The results shown in relation to the question about the monitor mix hindering the performance demonstrate a reasonably wide acceptance that, irrespective of configuration, the stereo spatialisation of the mix was not something that was felt to be a negative attribute when it came to actually performing the piece. When taking the order of the stimuli into consideration, there is an anomaly with respect to the mono configuration. There appears to be a general acceptance the first time the musicians are faced with a mono convolution of the elements that this is not appropriate in comparison to the previous stereo iterations. This is true of all performers other than specifically the bass player, who then went on to find issues after playing in mono to a convolution using an AB setup. The trend of all musicians becoming more comfortable with the stimuli towards the end of the experiment suggests a strong argument toward potential listening or performance fatigue and also possibly an acclimatisation to the type of monitoring being used and the repetitive playing of the same track.

In an effort to address the idea of the isolation of performers while using IEMs, the question about the mix enabling performers to play together was asked and the results of the survey are shown in Figures 56 and 57.

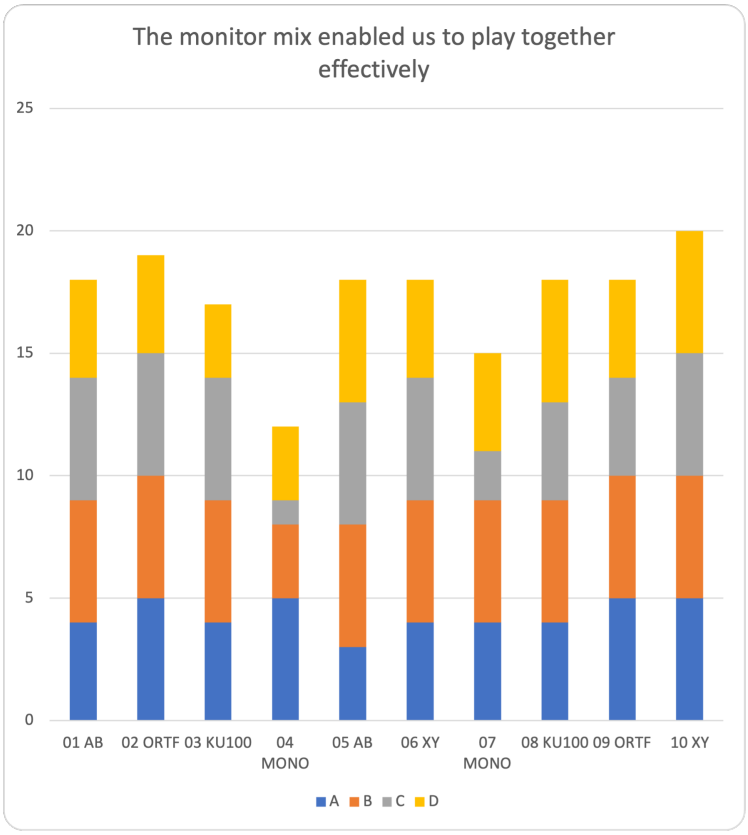


Figure 56: Combined results of all performers: A (bass guitarist), B (centre vocalist), C (electric guitarist), and D (drummer) for question 4

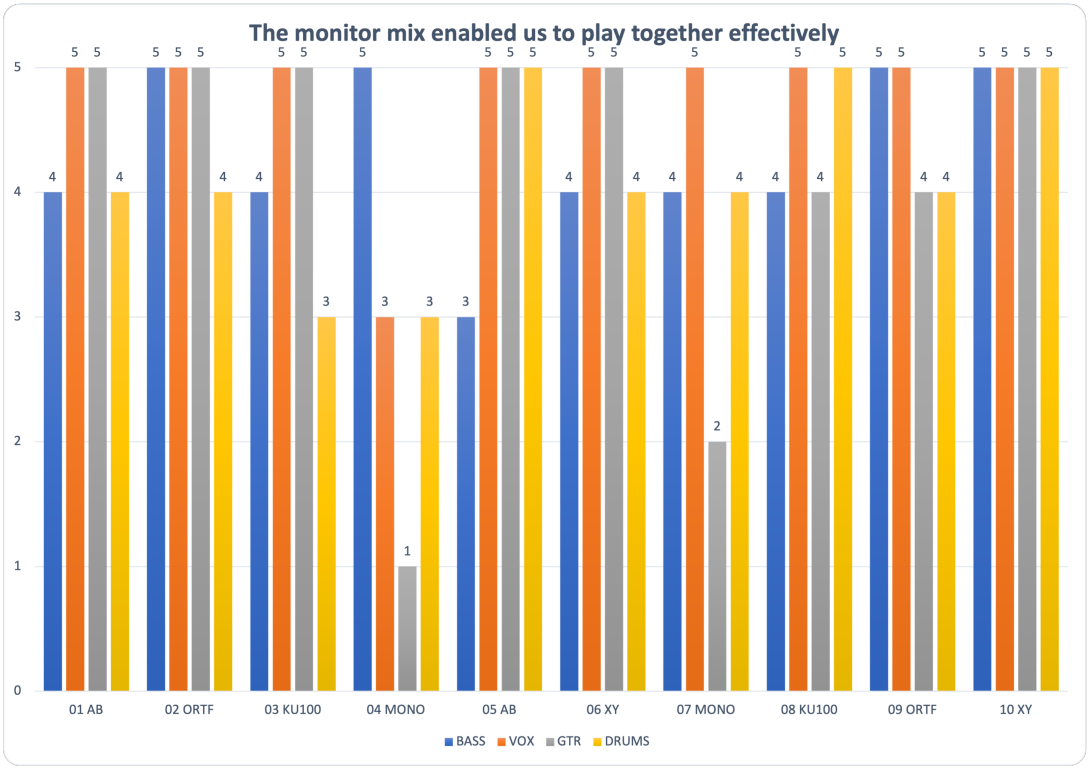


Figure 57: Results for comparison for question 4.

The strongest performers overall for this question were the XY and ORTF configurations. There

is a very similar trend with this question as there was with the previous question about the mix hindering the performance, with the exception of the mono signal being identified as the least effective but a generally strong result for all other stereo convolutions.

Question 5 related to how professional the monitor mix sounded. For an artist familiar with using in-ear monitors, this question is important because if the monitor mix demonstrates space but is unlike what the artist would generally expect and deem to be correct, then the effectiveness of the method could be brought into question. The results can be seen below in Figures 58 and 59.

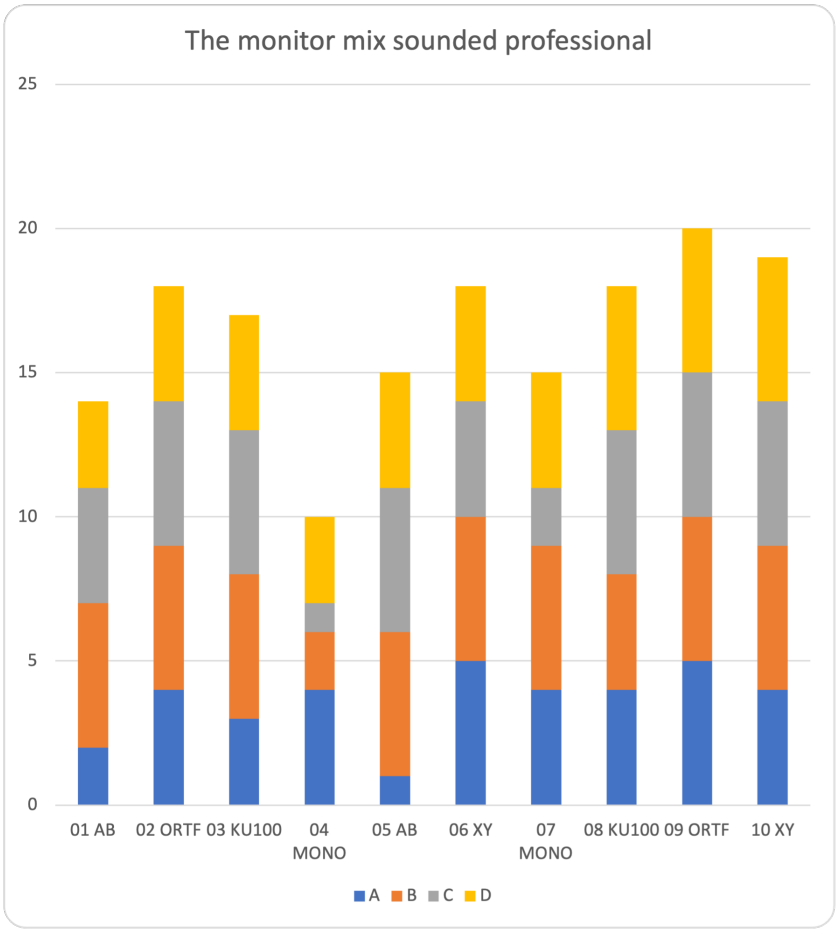


Figure 58: Combined results of all performers: A (bass guitarist), B (centre vocalist), C (electric guitarist), and D (drummer) for question 5

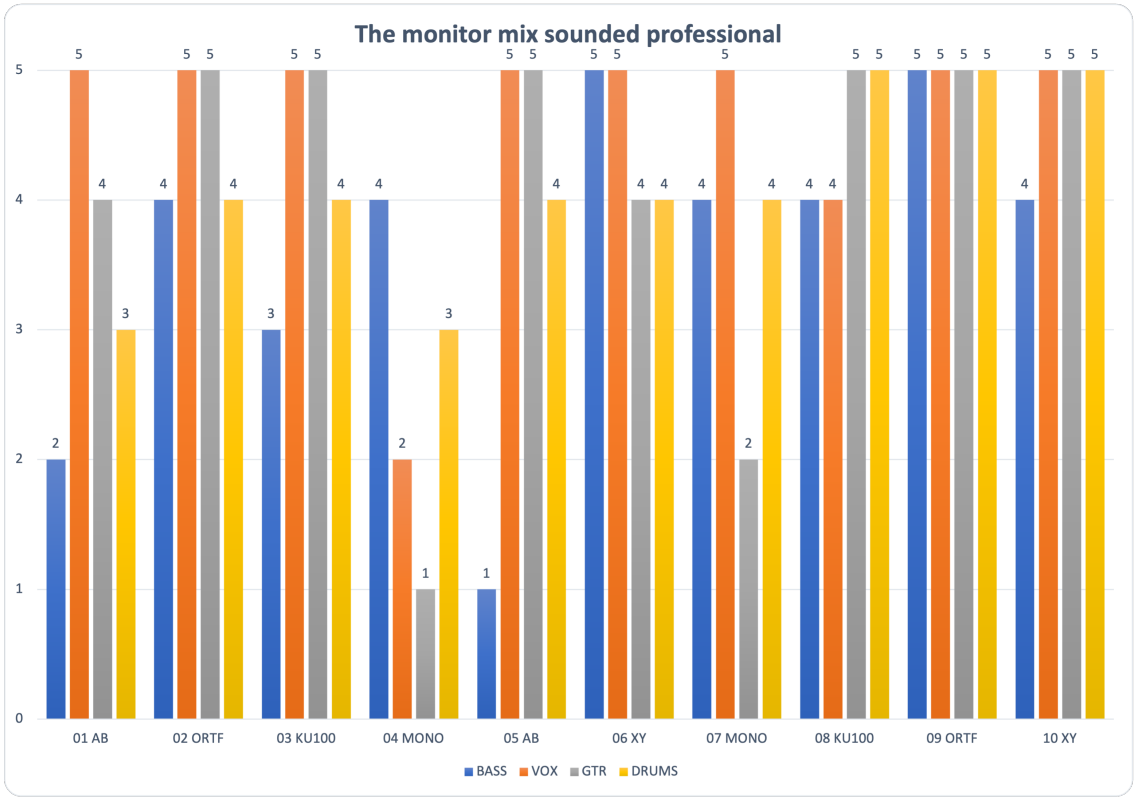


Figure 59: Results for comparison for question 5.

The lowest scoring stimuli for the question regarding how professional the mix sounded were the mono and AB convolutions. There is a general preference for the ORTF and XY convolutions, with the KU100 convolution also scoring strongly for this question. The final question, the results of which can be seen in Figure 60 and 61, was related to the tonal distribution of the monitor mix. This relates to the frequency response of the stimuli, and the hope is that they will not be lacking or over accentuated in any area.

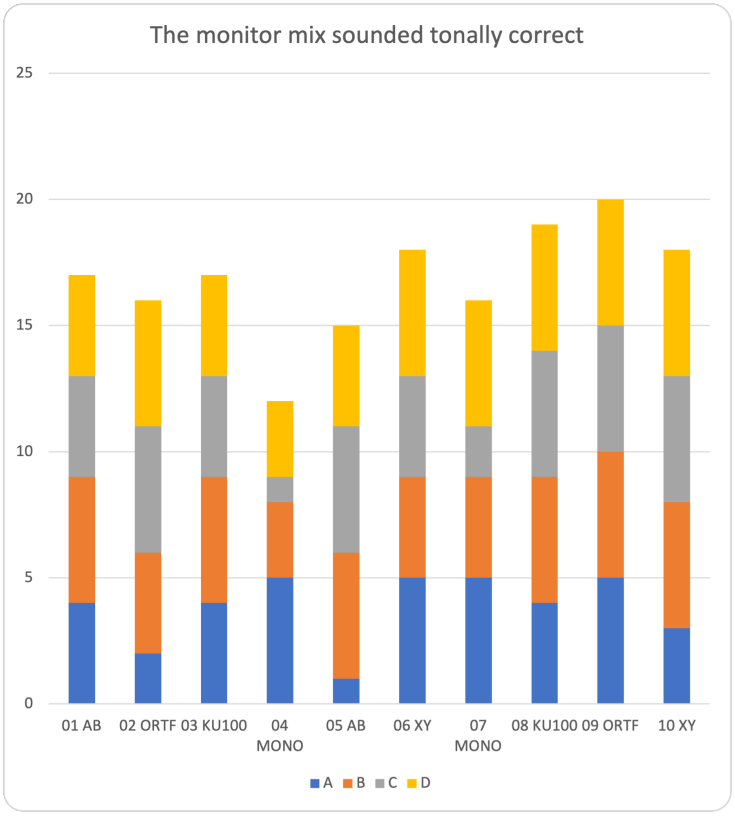


Figure 60: Combined results of all performers: A (bass guitarist), B (centre vocalist), C (electric guitarist), and D (drummer) for question 6

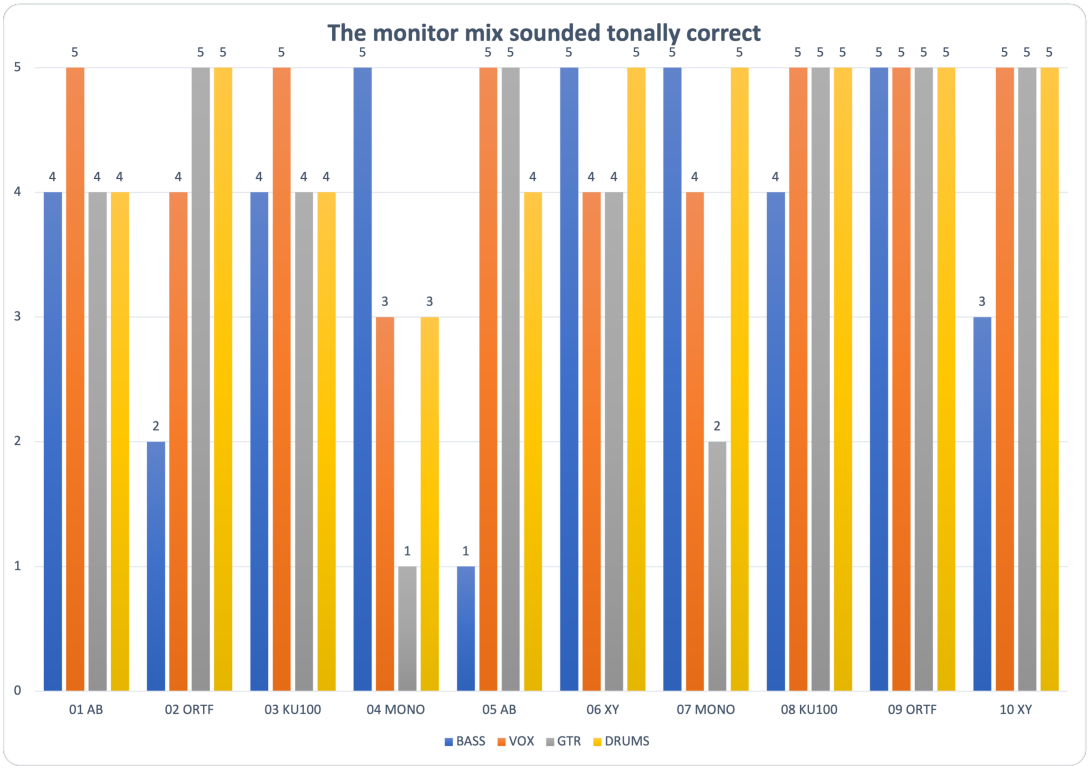


Figure 61: Results for comparison for question 6.

The configuration that scored the highest for these stimuli was the ORTF convolution, closely

followed by the KU100 and XY convolutions. As with other test questions, the stereo configurations all performed well, with the mono convolution scoring lowest for performers other than the bass player.

5.1.2 Results from individual performance perspectives: Bass Guitarist

Performer A assumed the role of the bass guitar player during the experiment and was positioned in the downstage left area of the performance space. The appendix, Section 8.1, contains the individual results for each of the questions from the performers' perspective. In this section, instead of focusing solely on the performers' responses to questions, the analysis will shift to examining the performance of the configuration in relation to the performer.

The AB setup demonstrated a reasonable level of consistency in terms of the repetition of results throughout the test, with the initial iteration yielding slightly higher scores. One possible explanation for this occurrence is that the first stimulus presented to the participants was not contextualised in relation to subsequent stimuli. As a result, questions regarding the potential hindrance to performance and the ability to effectively play together have not yet been framed in comparison to other stimuli. Upon examining the feedback regarding the AB setup, there are supplementary details pertaining to the spatial aspect of the configuration. The guitars are reported to have inadequate spatial positioning within the room, while there are also observations regarding the cymbals appearing to be located in front of the performer. As a result, the performer frequently expressed dissatisfaction with the perceived fidelity of the instruments, suggesting that they did not faithfully reproduce the spatial positioning of the instruments within the room. Consequently, an essential aspect of the monitoring objective was not achieved successfully. The performer observed that the tonality of the AB setup was characterised as 'quite woolly', a term commonly employed to describe either an excess in the lower midrange or a deficiency in the presence range of the instrument.

The ORTF setup was received positively upon initial evaluation, with performer A expressing general agreement and strong approval of the immersion and spatialisation of the stimuli. The setup was observed to experience a decrease in performance as the volume levels in the room increased. The artist specifically commented on the diminishing clarity of tonality as the volume levels rose. The performer's subjective response to various stimuli appears to have influenced the low tonality score obtained initially. Following exposure to the preceding stimuli, the ORTF setup was perceived favourably during the later stages of the test, generating strong agreement across all categories. The performer noted that it represented the highest quality of bass tone achieved throughout the test.

The perceived lack of success of the KU100 dummy head, as observed by the bass player, could potentially be attributed to suboptimal placement of elements within the room. On every occasion that the stimuli were utilised, the performer either expressed disagreement with the plausibility of instrumental placement or, at best, remained neutral toward the question. Performer A raised a general concern regarding the deterioration of tonal response in elements as the volume levels in the room increased. Contrary to expectations, the KU100 did not exhibit the same behaviour. The performer noticed that at higher volumes, the tone of the KU100 "held better", which is a favourable result for this particular configuration. Similar to the AB response that resulted in the perception of the cymbals in front of the performer, the KU100 had a similar effect on this particular performer, causing him to perceive himself as being positioned behind the drum kit rather than in front of it. Another factor that may have contributed to the challenges faced by the performer in this particular configuration is the significance of the lower frequencies in the localisation of the instrument. As stated in Section 2.3.6, the overall impact of the ITD and ILD when using the binaural head would have been minimal for specific frequencies, primarily due to the longer wavelength. Furthermore, it is possible that they may have encountered a certain degree of discrepancy in levels as a result of these factors.

The XY setup exhibited strong performance, especially during its second usage, when it was referenced and analysed in relation to previously heard stimuli. The observation was made that the sound

of this configuration was perceived as more forceful compared to the mono setup. This comment was likely intended to highlight the areas in which this particular configuration had surpassed another successful stimulus, rather than express dissatisfaction with the sound of the mono stimulus. With regard to the mono setup, it was also noted that it exhibited a comparable performance sensation. It was unanimously acknowledged that the instruments were being perceived from the appropriate position within the room. However, neither experiment yielded a definitive viewpoint on this matter, indicating that further enhancements are still necessary. This feedback is consistent with the theoretical concept of monocompatibility in the XY setup and provides evidence for the effectiveness of replicating spatial sound using these setups. The positioning of sources in the XY setup on the second listen exhibited some discrepancies. However, it was noted that everything else about the setup was considered "great".

The mono representation initially poses a challenge in terms of its effectiveness in addressing certain concerns regarding spatial accuracy. The mono configuration performed favourably on both occasions, second only to the ORTF setup. However, it is crucial to consider several factors in relation to this outcome. Firstly, the positioning of the bass player in relation to the bass amplifier is a crucial aspect to consider. The musician, despite being situated stage left, was positioned directly in front of the bass amplifier. From a localisation perspective, the individual anticipates perceiving the sound of his bass guitar as originating directly behind him, while in terms of left-right positioning, he expects it to be centrally located. One could argue that a monophonic representation would provide a succinct and focused depiction of the bass instrument, which might have been deemed more desirable. It is also worth commenting on the context of the mix when comparing the levels of elements. Regarding the bass player, it can be observed that the bass guitar held a prominent position in the overall mix, being significantly louder in proportion compared to its presence in the mixes of other performers. The result may have led to a potential obscuring of other elements and spatial cues, or perhaps even a partial disregard for them.

5.1.3 Results from individual performance perspectives: Vocalist

Performer B, located downstage and in the centre, was the vocalist that, for the purpose of the experiment, was used to listen from the perspective as opposed to singing. Something that could be considered in the future is the addition of a vocal; however, the questions regarding how to represent the vocal spatially and the conclusions drawn from the first pilot study meant that this was not explored at this time.

There is only a limited amount of information that can be gathered from the downstage centre position due to their general satisfaction with all of the spatial renderings of the monitoring environment. This is frustrating as being the central performer meant that they were in a position to potentially discern some interesting differences about the localisation of either side. That being said, it should also be considered that, in regard to the question of how appropriate it is to use the various stereo and mono techniques to demonstrate the spatial properties of the environment, there could be many appropriate methods.

When looking at how the performer felt about the AB configuration, there is a slight alteration from the first time they heard it to the second in terms of their feedback. The first AB example was also the first stimulus performed with, and it was commented on by the performer that the bass and the guitar lacked some clarity in terms of tone. By the second time hearing the AB configuration, this opinion had changed to the clarity being great, which would suggest that after hearing some of the other configurations, the performer based the clarity of sources in reference to the others. The second AB test was directly after the mono configuration, so it may be that in relation to that one specifically, the AB configuration had more clarity. Both times the AB configuration was monitored, the performer was very happy with all aspects questioned and agreed with the success in terms of immersion, tonality, and support of performance. The only area of question was that the guitar felt

quite behind the performer and not out to the side as much as it actually was in the room, which resulted in a slightly lower mark on the plausibility of the position of instruments.

The ORTF setup was very successful for the downstage centre performer. The first time of listening prompted the feedback that the balance and tonality of elements were "near perfect" and on the second time around, very close to the end of the session, it received top marks in all categories and was noted as being "one of the best overall".

The KU100 had the highest scoring of the configurations, achieving strong agreement in all categories for both instances of the configuration being used. The performer noted that it sounded remarkably similar to the ORTF stimuli (they only mentioned the test number because they were unaware of the configurations), and they also mentioned that they perceived themselves to be slightly further back in the room when listening to the monitor mix but felt fully immersed in the performance on both occasions. The fact that there is a discrepancy in the happiness of the performer does raise questions about the way that the configuration was marked, but overall, it is clear that they were comfortable with this configuration.

The XY configuration was received well but does raise questions about the consistency of these variables because of the comments made by the performer. The second instance was noted as having the most "energy in the room" something that, in speaking with this performer later, was defined as being that they felt like the in-ears were transparent or "invisible" to quote them directly. This, however, is not fully supported by the first instance of hearing the XY configuration, as at that time it was specifically noted that it was great but not as good as the test immediately before, which was the AB setup. It may be that something about the performance itself is the factor that is changing within these variables. There is the plausible explanation that upon running through the same thing ten times, there is a chance that the performance would suffer the further into the test they got due to fatigue, or conversely, that the more the performers played, the more practiced and comfortable they would become, having the opposite result of improving the performance element.

The mono representation was noted as being incorrect, especially on the first instance of hearing it. The performer noted what had been intended to be the case with the mono representation, which is very useful in supporting the aims and hypotheses of the thesis. There was a note of the balance being OK, but everything was somehow "wrong" and making the situation off-putting. The only thing that changed on the second occasion was that the ability to play together and the mix not hindering the performance were not noted as being detrimental. This is not particularly helpful in supporting any dismissal of the appropriateness of the mono capture. The second instance of the mono configuration did lead to the remark that the guitar felt like it was in the performers head, which supports that they were hearing a lack of space between the left and right and, as such, supports their lack of agreement on the instruments sounding like they were coming from their location in the room.

5.1.4 Results from individual performance perspectives: Electric Guitarist

Performer C was the last of the three downstage performers located on stage right, playing electric guitar through an amplifier also located on stage right. The feedback provided by this performer was highly beneficial as it encompassed written feedback for each test, occasionally incorporating relevant and specific information applicable to other scenarios.

For performer C, the AB configuration was very well received. Their approval of the setup improved from the first iteration to the second but only from the respect of agreeing more strongly. On the first test, it was noted that everything felt that it was panned correctly and the guitar sounded directly behind. The positioning of instruments in the room was only agreed upon at this point, which may be down to either not wanting to answer too strongly before hearing other versions or to finer details such as the guitar not being quite directly behind in reality. By the second testing of this setup, the opinion had been strengthened to strongly agree to all questions and a specific note about how it was

better than the second test, which was the ORTF setup, with a note that clarity of elements was comparable between the two but the panning with this AB setup was "better".

The ORTF setup was also very well received by this performer. The notes on the first instance of test mention that the mix was fantastic and allowed the performer to focus, presumably on playing along with the other performers as well as his own performance. The bass was noted as being possibly too quiet; this, however, is arguably a mix correction rather than something to do with the convolution. There is also the note that the "drums had punch!" which is interesting and supports some of the results about how the ORTF setup differs in terms of representation of the lower midrange, which will be addressed in the discussion sections later in this thesis. Upon listening again, the ORTF continued to perform well across all the questioned areas.

The KU100 configuration gave some interesting insight into the performers thoughts on the tonality of the representation and the immersion that it offered. The areas questioned were agreed with, but in the feedback, the immersion could be questioned as they noted that although it felt good, it felt more like they were playing along to a good mix rather than playing in the room. They also commented on how the drums now lacked the punch that was present when listening to stimulus number 2 (the ORTF setup). The second time the KU100 configuration came around, again it received a general agreement to the questioned areas but was noted as sounding very full range but to the extent that it was "almost punishingly". This may simply be a level issue with the playback of the scenario, and by reducing the level a little overall, they would have been more happy, as they did go on to say that it sounded great; they just couldn't imagine having it like that for a whole set.

It could be suggested that the XY setup was perhaps the most successful for this performer, or at least on a par with the most successful. At the end of the final test, it was noted as sounding superb, with the panning of elements being "uncanny". The performer also had a strong agreement to all statements for the second testing of this setup, a slight improvement of their first listen where they felt the bass was getting a little lost in the midrange frequencies. This may suggest that the reason for the improvement was something about the performance of the bass player, potentially them playing a little harder, which would give a more percussive and midrange response from the strings of the instrument.

The mono configuration was not approved of by this performer, who had strong disagreements across nearly all statements on the first time of listening. They noted that the bass and the attack of the drums were in the way of everything else in the mix, in addition to disagreeing with the plausibility of space and immersion in the environment. On the second time of listening, they also correctly identified that the bass sounded like it was coming from directly behind them, something also noted by the bass player but in this case, not supported by actually being stood in front of the bass amplifier.

5.1.5 Results from individual performance perspectives: Drummer

The last performer, referred to as performer D, assumed the role of drummer within the band. This particular arrangement provides an opportunity to examine the application of convolution to the stereo components from a unique standpoint. In this case, the drummer is positioned behind the amplifiers, and it is the front of the amplifiers that serves as the source of their most comprehensive frequency output. The feedback provided exhibited narrative-like qualities, offering insights into the participants' perceptions of the experiment's progress and their comparative experiences with the various stimuli.

The AB configuration was noted as allowing the drum kit in front of the drummer to sound very natural spatially; however, they remarked that the kick drum sounded like a kick drum through a mic in front of the kick drum, as opposed to it sounded like they were just sitting behind the kit. They noted that maybe this was down to the mic choice, and what could be suggested from that is that the tonal response of the kick drum sounded emphasised in some way, or perhaps it sounded restricted in terms of frequency content and therefore unnatural. The second time that this setup was performed,

the participant was much happier with it as a monitoring setup, which could be down to having the context of some of the other variables or perhaps just due to comfort in the playing. There was a fairly extensive warm-up period to mitigate that factor, so it is perhaps more likely that context was the reason.

The ORTF configuration seemed to work well from the drummers perspective, with them being in agreement across all statements. When monitoring using the ORTF setup directly after the AB setup, it was noted that the ORTF version made the drums in front of the performer sound more natural and that the positioning of the instruments was much clearer.

The KU100 gave the joint best monitoring experience, with the drummer stating at the point of the second test that this was the preferred method so far. On the first instance of hearing this setup, the bass was noted as being a little unclear as to where it was positioned, but this was not mentioned the second time around.

Additionally, there was a strong consensus in favour of the XY setup, in addition to the drummer's preference for the KU100. The performer expressed a sense of being immersed in their drumming experience and praised the exceptional balance achieved. Whether this XY version superseded the earlier KU100 test is unclear, but in the context of the experiment, it is not particularly important. Across both experiments, it is worth noting that the XY setup scored more highly and received praise both times.

The mono configuration scored lower than other setups and was noted as having a "muddy" sound, masking certain elements. The drummer did still feel that elements seemed to be coming from the right locations, which, when considering topics discussed in the literature review, may be because of the visual stimuli that wouldn't have been present for other performers being located in front of the instruments. There are issues present when looking at the later experiment using the mono capture, as it scored higher than expected, and this is something that will be discussed in the next chapter.

As evident from the findings, when considering the viewpoint of each performer, there is no distinct distinction among any specific setup. The identified trends will be further examined in the subsequent section of this thesis.

5.1.6 Summary of individual performers responses

Overall, the discussion of each individual performer reveals certain trends in their preferences, as well as some notable differences. The bass player exhibited a favourable reaction towards the ORTF, XY, and Mono convolutions, while expressing a negative sentiment towards the KU100 representation. The vocalist expressed overall satisfaction with the stereo convolutions, showing a slight preference for the KU100 while acknowledging the strong performance of the ORTF and XY configurations. The individuals expressed some dissatisfaction with the Mono convolution of space. The guitarist exhibited the most robust reaction to the XY configuration, while the ORTF and AB setups also received high scores. The KU100 elicited enquiries from the performer, and the Mono configuration was not well received. The most robust response, as perceived by the drummer, was observed in the XY and KU100 configurations, while the ORTF configuration also yielded an acceptable response. Both the AB configuration and mono configuration have raised questions regarding their suitability in supporting this performer.

Chapter 6

6 Discussion

6.1 Discussion

After conducting this test, it is evident that the proposed method is not suitable for implementation in real-time monitoring systems. The process of mapping out this small space, using only 7 locations and 4 view points, necessitated a significant investment of time. This involved several hours of recording in an environment that was carefully isolated from external noise. The process of capturing sweeps in isolation, utilising a more extensive configuration and potentially from multiple perspectives, could require several days prior to any subsequent processing. Furthermore, the outcomes derived from this endeavour would solely be applicable to the particular location under consideration. Considering the aforementioned, the primary discovery derived from the investigation and experimentation pertains to the comparative analysis of stereo configurations in relation to their capacity to capture and portray spatial attributes when observed through IEMs and the potential implications this may have on subsequent research endeavours.

The experiments conducted for this study not only yielded information regarding stereo microphone arrangements but also offered valuable insights into the monitoring requirements and preferences of performers. By examining the results from both a holistic and individual perspective, it has been possible to identify certain factors that influenced the preference for certain setups over others. Additionally, cases where there may have been deficiencies in the preferred setups have been analysed.

In hindsight, assembling the monitor mixes for every performer during the primary pilot test setup was a potentially flawed aspect of the study. The mixes were only altered based on level; no other processing was used. However, requiring the performer to hear every variable prior to being questioned on them could have potentially impacted perception. The repetition of the procedure, as well as the anonymity and randomisation of the test stimuli throughout both the setup phase and the test, helped to mitigate this. It would not have benefited the performance to be exposed to a dry monitor mix with no convolution. It was omitted since it would have stood in sharp contrast to the experiment's monitoring technique. As a result, the performers heard just signals processed with a convolution reverb for the whole time they were in the testing area.

Following the initial setup, practice, and warm-up, each performer received their own set of in-ear monitors. A rough mix or balance was applied to each setup and the stimulus sent was altered periodically to ensure that, as each performer became comfortable, they did not stay on any one stimulus for an extended amount of time, allowing for habituation. No matter which version of the convolution was sent, it appeared that the performers were all astonished by how realistic the material was, which was encouraging and suggested that the test would be successful. The drawback of this was that it would make it more difficult to distinguish the subtleties of the various setups than if there were some more pronounced variations.

There were advantages, even though it was acknowledged that it was not ideal to give the performers access to all of the variables before the testing phase. In addition to the warm-up period during which they had to play with each other in the room, the performers had to go through all of the mixes to verify the correct mix balance. This gave them more opportunity to become accustomed to using the monitoring, which should result in a more consistent response of judgement throughout the testing process and prevent the earlier tests from being compromised due to unfamiliarity with the monitoring setup and scenario. In future experiments to collect more quantitative data, this is one area that will need to be addressed to ensure it does not impact on the results of the study and will be discussed in section 6.3.

6.1.1 Comparison of Stereo Configurations

Each of the stereo setups will be covered in this section, with an emphasis on any potential benefits and drawbacks. Frequency charts for the same perspective and source are included with each stereo capture to provide some context. This is specifically of the centre position downstage, where the source is emanating from the guitar amp's central point. The tone discrepancies in the stereo capture and representation in the RIR can be seen by examining these charts.

The primary goal of employing a mono capture of the room response was to make sure that a lack of space was recognised; however, the manner in which it was used did not produce that response in such a noticeable way. Although the mono capture of the sweeps and creation of mono RIRs contained equal representations in both the left and right channels and therefore no left-to-right difference to satisfy the ITD and ILD's ability to perceive direction, it did contain all of the time-related response of the environment, creating what ended up being a somewhat plausible reverberation that could be related to the space that the performer was in. The monitored signal should have been audible in isolation according to the IEMs' claimed 37dB of external sound suppression [97], but it is possible that this was not the case. It may be the case that some of the ambient sound is still influencing the listener, or perhaps it is just that the visual cues and possibly the tactile cues were giving the listener some directional influence from the monitoring setup that is in reality not really present. The preference for mono convolution may be attributed to the factors discussed in Sections 2.2.4 and 2.2.5, which pertain to the presence of room modes and the frequency characteristics of the instruments employed. It is possible that the bass guitarist's preference for these fundamental frequencies is influenced by a bias towards this configuration due to an accentuation of frequency information as can be seen in Figure 62, that is relevant to the music being played within the range of their own instrument.

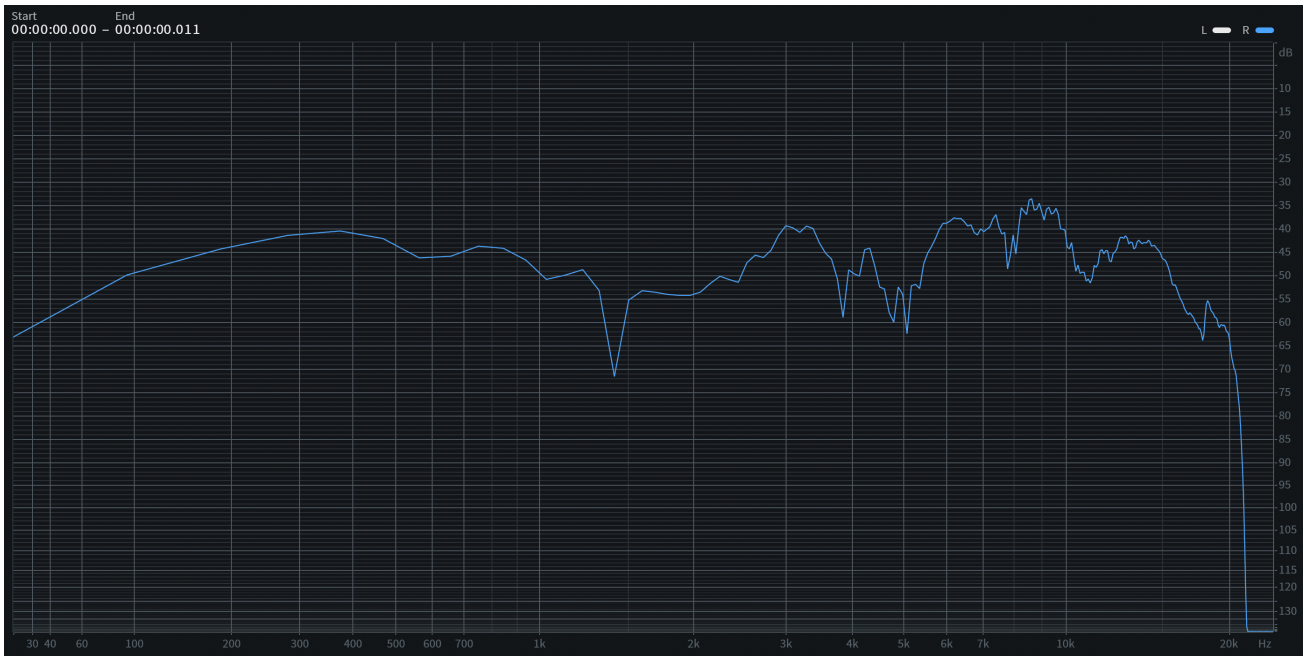


Figure 62: RIR from the Downstage Centre perspective of the central guitar source as captured by the mono configuration.

One primary finding of the pilot study was that the KU100 did not exhibit a distinct advantage in relation to all performance measures. Regarding the KU100-based RIRs and their overall reception, there was no unanimous response observed among all performers. A noteworthy distinction is that the two artists positioned in the centre of the room appeared to have a more favourable encounter with it compared to the performers positioned towards the downstage left and right. This may be due to the close proximity of the boundaries, which causes the perspective to be distorted when using the KU100

to record the room's acoustic response, en example of which can be seen in Figure 64. This distortion did not effect other stereo microphone techniques in the same manner. Regarding the literature review, it was observed that the Schroeder frequency of the room was relatively high, measuring around 194.2 Hz. This implies that some modes within the spectrum of fundamental frequencies may have been exerting influence on those specific sections of space. The occurrence of prominent modal resonances at frequencies around 105 Hz, 130 Hz, and 140 Hz, as illustrated in figure 63, could have had a substantial impact on the issue at hand. Similar to the mono configuration, these resonances may have influenced the musicians' performance by affecting how the frequency characteristics of their respective instruments responded to these factors.

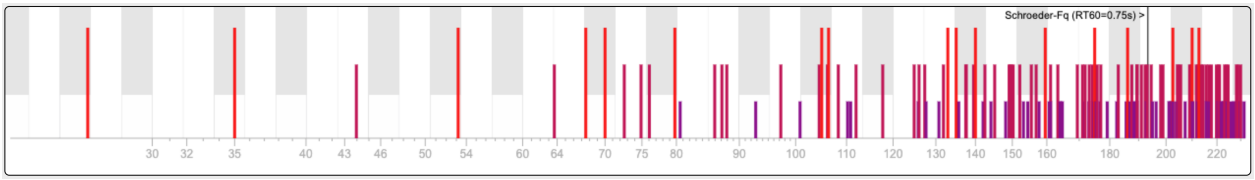


Figure 63: The chart demonstrates the modal response of the room used for the pilot study created using amcoustics room mode calculator [105]. The resonances shown here may have contributed to the listeners perception of frequencies affecting the overall balance and tonality of the monitoring mix.

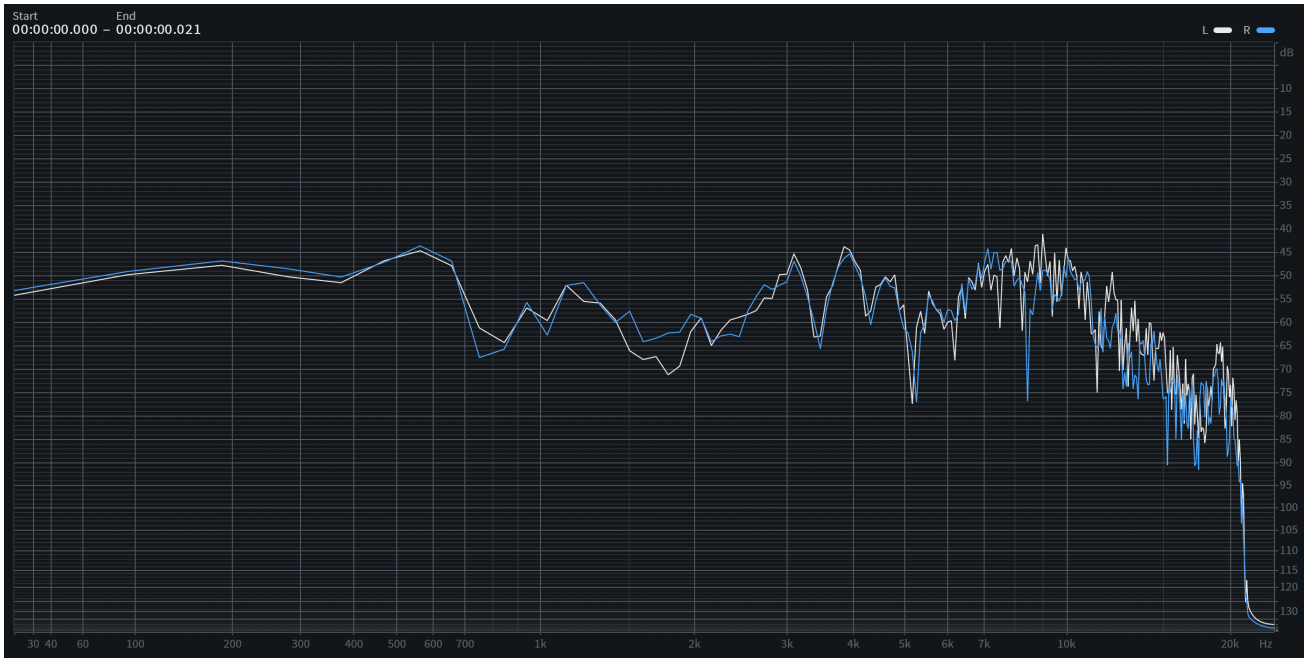


Figure 64: RIR from the Downstage Centre perspective of the central guitar source as captured by the KU100 configuration.

The bass guitarist experienced challenges in localising sources, potentially due to disparities between the performer's ear and head response and the dummy head utilised. These difficulties may be attributed to issues with spatial information in lower frequency content or the equalisation of room impulse responses. Despite the aforementioned issues, there were some highly positive responses to the KU100-based RIRs. Something that was consistent across all performers interpretations of the KU100 was the tonal response; however, this demonstrated how the subjective element of preference can colour the response of the questionnaire in the sense that there is personal preference regarding the tonality of sources. The KU100 based RIRs provide a precise and realistic representation of the lower

mid-frequency response of a space. However, when these RIRs are convolved with musical sources, although technically accurate, they may not always be the preferred choice. Particularly when it comes to percussive elements, it is common practice to reduce a substantial portion of the lower mid range in order to create room for the lower and upper frequencies to enhance the"Punch" effect. This term or attribute is used to describe the amplification of the transient element of sources within a mix [106].

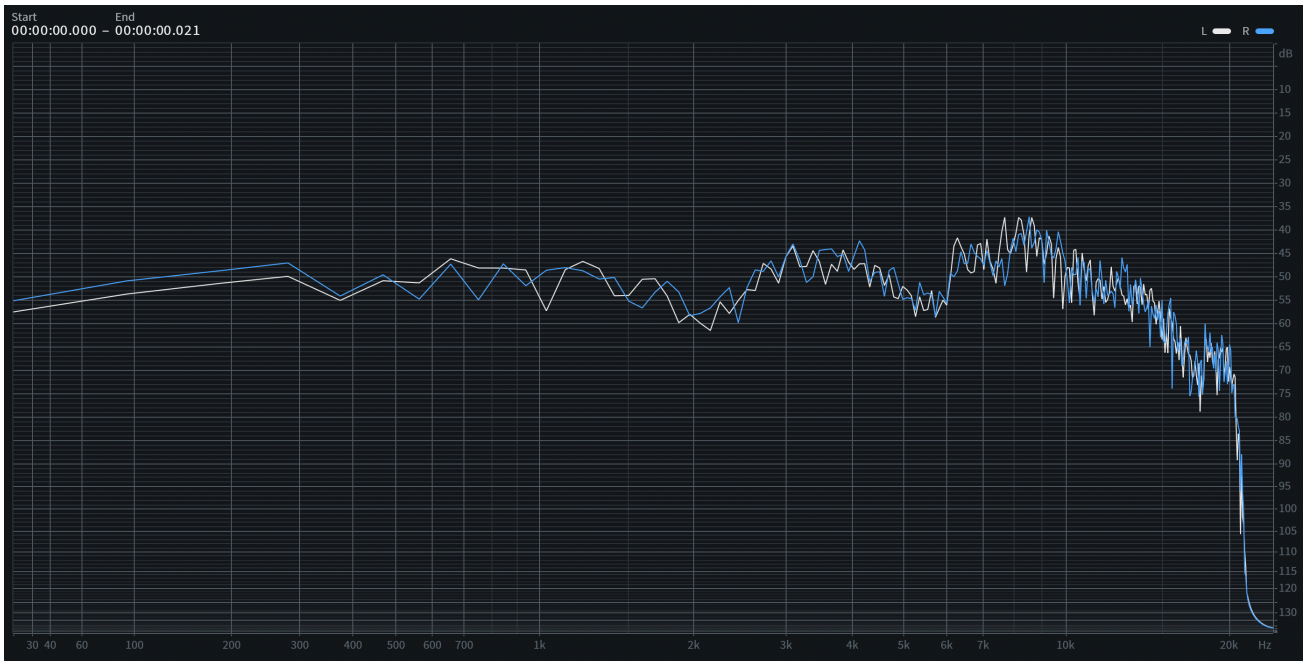


Figure 65: RIR from the Downstage Centre perspective of the central guitar source as captured by the ORTF configuration.

Based on the feedback obtained from this test, it seems that the ORTF and XY pairs are considered the preferred methods for stereo capture of acoustic environments. The offered localisation was superior overall compared to the other three methods of capturing acoustic spatial measurements, resulting in a greater sense of immersion in the environment on average. Neither configuration appeared to impede performance at any stage, and they exhibited satisfactory tonal qualities, particularly the ORTF setup in that regard. Upon examining the frequency response of the RIR depicted in Figures 65 and 66, it is evident that both the XY and ORTF microphone techniques exhibit a commendable ability to produce a consistent and uniform response in the lower frequency range and lower mid range. The validity of this statement is particularly evident in the case of the ORTF setup, as the frequency plot provides substantial evidence to corroborate the feedback provided.

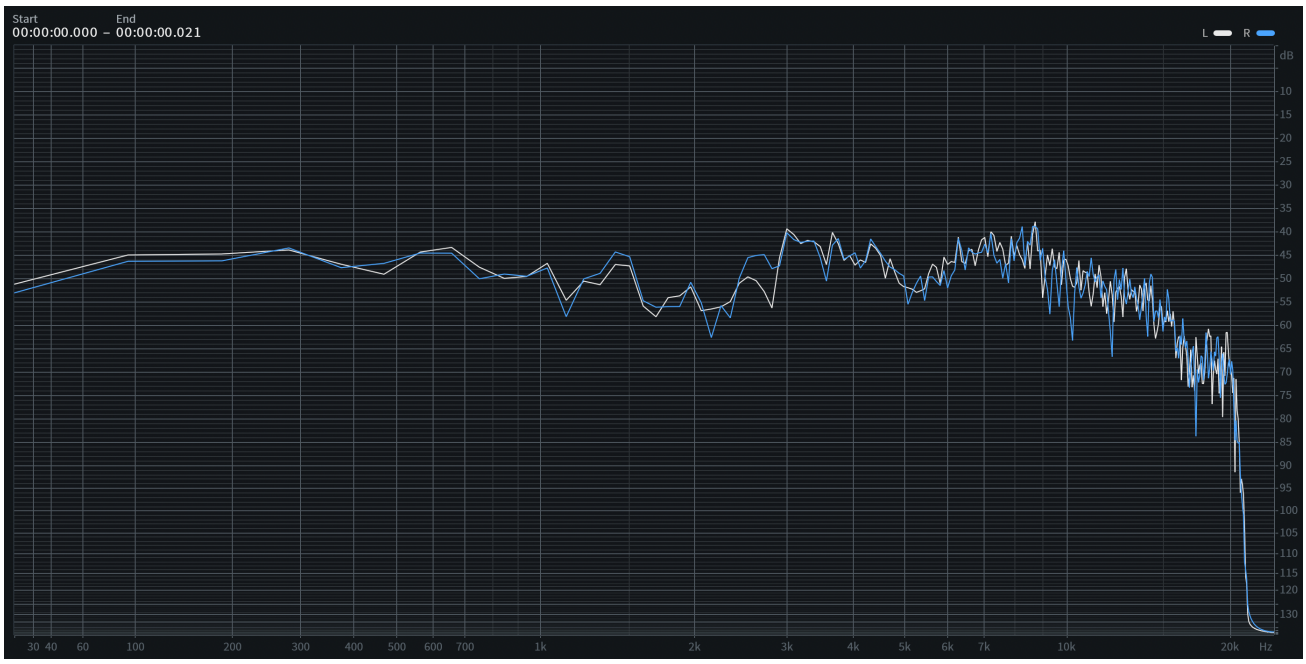


Figure 66: RIR from the Downstage Centre perspective of the central guitar source as captured by the XY configuration.

The AB method employed for capturing the acoustic environment exhibited some degree of inconsistency, yet it did manage to effectively represent the spatial characteristics to a certain extent. It is evident that the AB version depicted in Figure 67, exhibits a significant variation in its frequency response when compared to the other RIRs. The overall precision of localisation was found to be less accurate compared to the other configurations, with the exception of the mono capture of the acoustic environment. The tonal variation observed across the tests and the superior performance of the ORTF and XY configurations suggest that the AB pair method may not be as suitable.

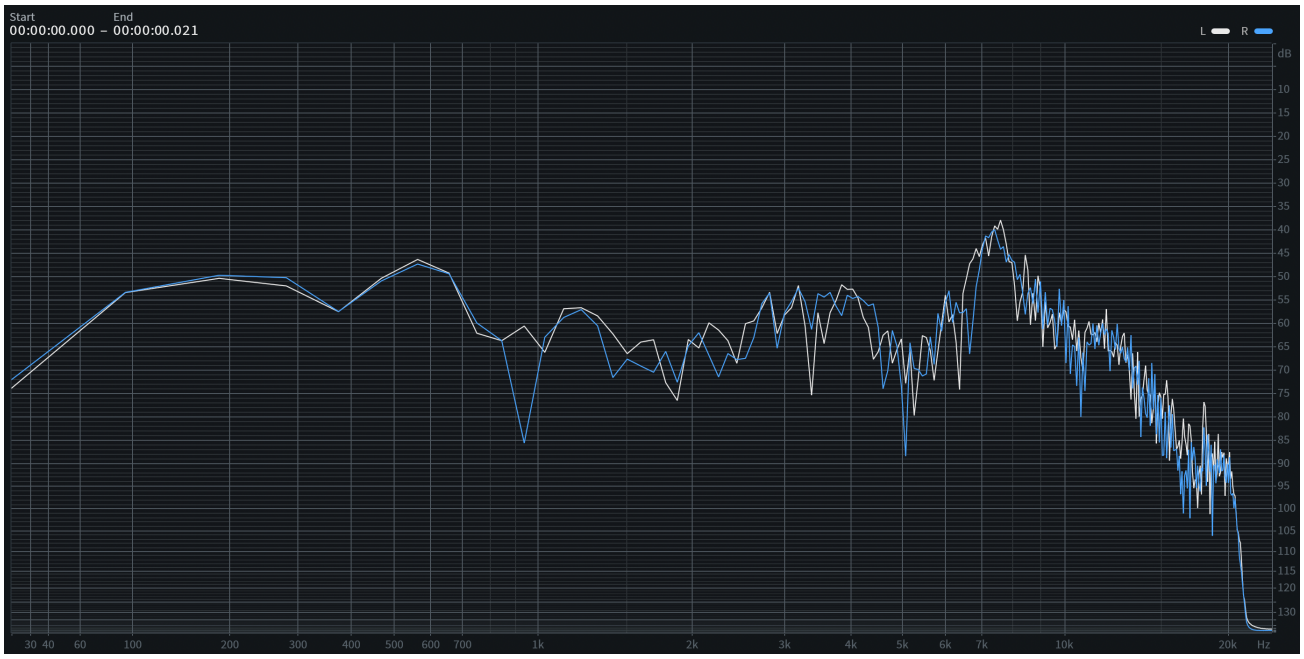


Figure 67: RIR from the Downstage Centre perspective of the central guitar source as captured by the AB configuration.

6.2 Conclusions and suggestions for further work

The present study has effectively developed a collection of RIRs that encompass a designated acoustic environment for a quartet of musicians to perform in, considering various stereo and mono perspectives. The objective of this study was to reduce any potential alteration in the captured and played back audio resulting from the recording and listening devices. The study has presented evidence to support the initial hypothesis through the use of subjective, qualitative evaluation, which has prompted the need for a more comprehensive objective, quantitative evaluation of the criteria in the future.

6.2.1 Evaluation of the stated hypothesis

When embarking on the study, the original hypothesis stated:

When using in-ear monitoring, convolution of musical source signals with spatial acoustic measurements can be used to make a sound that is true to life on stage. Stereo microphone techniques can give results similar to those of a binaural head when used to record a room's acoustic response.

Based on the existing literature and empirical investigations in this field, it can be contended that the convolution of space has the potential to generate a lifelike portrayal of sound on stage. Nevertheless, the evaluation of whether utilising spatial convolution is more beneficial than directly monitoring sources cannot be supported solely based on the findings of this study. This aspect warrants further examination in future testing of the concept. Regarding the appropriate stereo techniques, based on the findings of the thesis's pilot testing phase, it is reasonable to infer that the use of stereo techniques is not restricted solely to a binaural head. In fact, it is possible that an alternative approach may be more advantageous in this particular scenario.

The preliminary findings from the pilot test phase suggest that employing an ORTF pair of microphones or a co-incidental XY configuration may be the most suitable approach for capturing the acoustic characteristics of a room, which can then be convolved with musical signals. Furthermore, the technique's preference for use is not only attributed to its ability to yield favourable results based on the configuration's presentation of the ICLD and ICTD to the listener but also to the accessibility of the necessary equipment.

From the point of view of the artist, it has been suggested so far that the use of convolution gives a pleasing response, but this does not take into account how it would compare to other spatial representations of the sources or, indeed, a more basic stereo representation of the monitor mix. This is not necessarily a shortcoming of the pilot study, as the variable in question was in relation to the comparison of stereo capture techniques of convolution and was not intended to suggest a new method of monitoring for musicians.

6.2.2 Limitations

One limitation that emerged during the pilot tests was the establishment of performance levels for the experiment. Given the time limitations of this qualitative testing phase, it was considered essential to permit the performer to audibly perceive the stimuli before the evaluation in order to facilitate the adjustment of mix levels. Despite efforts to optimise effectiveness, reliance on quantitative data is not feasible when employing this approach. In order to ensure that this particular aspect of the process does not affect the performer, it is necessary to set aside more time for future quantitative testing. In practical terms, there are several limitations associated with replicating monitoring environments. The replication process heavily depends on numerous variables that are unlikely to be precisely replicated. However, if replication were to be pursued, it would be crucial to ensure accurate measurements of equipment placement and microphone positioning. This has the potential to allow for the adjustment of monitoring levels several weeks before the main testing, thereby mitigating this issue.

A potential enhancement to the testing procedure could involve the use of a different microphone specifically designed for capturing the sound of the bass guitar. As previously stated in Section 4.7, the

choice to use a microphone instead of employing direct injection for the bass guitar was proposed and determined to be more favourable based on both theoretical and practical considerations. In order to enhance the recording of the bass guitar for subsequent convolution in future experiments, it may be more suitable to utilise a large diaphragm omni-directional condenser microphone, such as the AKG c414. Its more linear frequency response and the fact that the proximity effect has no impact on its polar response support this choice. Consequently, it enables a closer and more isolated capture of the speaker cabinet. A comparable argument could also be developed regarding the acquisition of electric guitars; however, additional investigation into these variables would be necessary. Currently, it is not deemed feasible to implement this method in its present state due to the laborious process of mapping the performance space. However, it is important to consider that microphones employed in live sound are typically not separated for monitoring purposes and transmission to the primary speaker system. The front-of-house engineer may well be the one who chooses the microphones that will capture the sound of the stage's components, giving consideration to the microphone's tonal qualities rather than its ability to faithfully reproduce the original sound.

There is also the factor of considering the drummer's perspective, specifically in terms of the directional response of the sources and the creation of an RIR based on the listener's position behind the source. There are several potential methods to accomplish this objective, such as directing the room's excitement towards the source instead of employing an omnidirectional approach and applying an equalisation filter to the RIR. This process has the potential to introduce a sense of authenticity. However, it is important to note that this aspect does not align with the primary objective of the experiment conducted in this research. One could argue that the resulting reality may be deemed undesirable, thus negating any notion of improvement.

With respect to the various configurations utilised and deemed appropriate, an untested aspect pertained to LCR recording, which involves the incorporation of a third microphone to capture an accurate mono or centre image instead of relying on a stereo technique to replicate a phantom centre. The resulting effect presents a distinct and clear central image that may be more desirable for representation. However, the limitation of the experiment lies in the number of variables, which raises concerns about the validity of the results due to testing fatigue.

6.2.3 Further Research

As mentioned in section 1.3, this thesis serves as a preliminary investigation, suggesting the need for a more comprehensive study to determine a definitive answer from a quantitative perspective. However, conducting such a study is beyond the scope of this thesis. One potential avenue for enhancing the existing experimentation is to carry out additional iterations of this experiment, which would enable the aggregation of responses across a larger and more diverse sample of performers and music genres. In order to draw any definitive conclusions from the research, it would be necessary for this event to occur. While the suitability of the ORTF and XY microphone techniques has been proposed, further experimentation is necessary to validate their effectiveness.

6.2.4 Proposed Future Testing

This section of the thesis will outline the precise methodology that should be employed for future testing in order to obtain a greater amount of quantitative data. It is recommended that future research adopt the strategy of identifying a control stimulus, which represents a departure from the current approach employed in this preliminary phase of investigation. To quantitatively assess changes in attributes, it is necessary to establish a baseline level. According to the existing research, there is evidence to support the use of the mono representation of the environment as a control.

In order to eliminate a substantial number of variables, it is recommended that future testing involve only one musician as the test subject and that the type of musician remain consistent throughout the testing process. Based on the preliminary pilot tests conducted, it is recommended that the

quantitative tests be approached from the perspective of the guitarist. This would enable a direct comparison to be conducted with respect to the stereo configurations being utilised, while excluding any other variables that could influence the data. The rationale for selecting the guitarist to be positioned downstage right is based on the findings of the pilot study, which indicated that this particular instrument and placement effectively detected variations in stimuli to a greater degree. Additionally, this positioning was found to be less susceptible to anomalies compared to the off-centre bass guitar player and the drummer, particularly in relation to low-frequency positioning and the alignment of the performers with the direction of amplifier propagation.

A proposed variation on the MUSHRA test is suggested as the preferred style for future study. The conventional approach described in Section 2.5.2 involves utilising preexisting material to enable the listener to switch between audio sources. However, given the live nature of this scenario, a modified approach will be necessary. Specifically, participants will be obligated to receive various audio stimuli in real-time, which can be alternated between rather than switching between prerecorded materials.

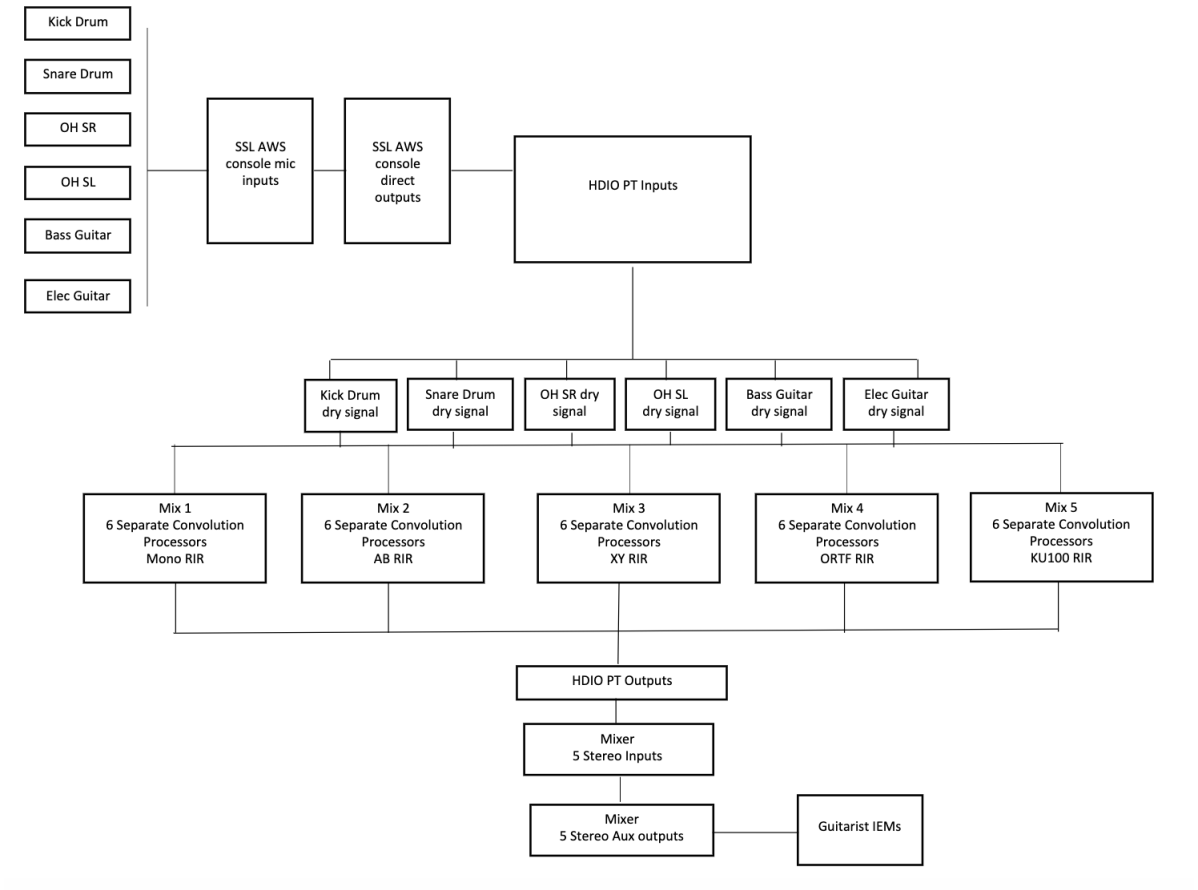


Figure 68: The adapted signal flow diagram demonstrates the methodology of the quantitative testing. Five different convolution reverb processors will each process a live microphone input before sending the results as five distinct stereo headphone sends to a mixer in the live room for the performer to choose from.

As depicted in Figure 68, rather than receiving a single monitor feed specifically designated in the control room, the performer will be provided with five monitor feeds in the performance area. In order to accomplish this, a setup that closely resembles the pilot tests is necessary, although there are some subtle distinctions that should be acknowledged.

Since there is only one performer receiving the convolved monitor mix, the processing required is only marginally increased. The primary pilot test necessitated the utilisation of six instances of convolution reverb, which were distributed among four performers. Consequently, a total of 24

processors were engaged, resulting in an acceptable latency of 9 milliseconds for the performers. In order to conduct the proposed testing, it would be necessary to have six instances of the convolution reverb running concurrently for each of the five different perspectives. Therefore, the total number of active processors required would amount to thirty. It is necessary to measure the latency generated by this in order to ascertain that it does not have a detrimental effect on the monitor feed.

Each of the monitor mixes from the 5 different perspectives will be sent through to the performer into a console, enabling the performer to select the desired stimulus to be heard. The proposal is to utilise a letter or number system to designate the monitor mixes, implying that these 5 feeds can serve as an alternative to the standard procedure of employing prerecorded material in a MUSHRA test.

An aspect that requires careful consideration and additional pilot testing is the technique for transitioning between stimuli, given that it is intended for the guitarist to perform. The ITU-R recommendation [87] asserts that it is advisable to allow for a duration of 10 to 12 seconds for the signal to be audible before transitioning to a different stimulus. It is possible that this requirement can be fulfilled by assigning an individual to control the feed from the performance area, or alternatively, the entire system can be streamlined and made more efficient by implementing automation within the playback system. If the level of organisation were to adhere to the described structure, which includes mandatory alternation between the mixes, the utilisation of automation could result in a substantial reduction in processing requirements, with only 6 convolution reverb units operating simultaneously.

This also raises the question of when the performer would reach a point where they could assign a rating to each mix on a scale of 1 to 100, as this would require them to interrupt their performance. If a simple visual display, such as an iPad, were at the disposal of the performer, they could potentially assess each event efficiently during the intervals between short performances. Alternatively, it is possible that another individual could be present alongside the performer to input the information onto a visual display, allowing the performer to continue playing without interrupting the performance. By utilising the same framework as demonstrated in the example presented in Section 2.5.2, individuals would be able to visually perceive the relative rankings of the previous combinations, thus enhancing their ability to make comparisons and distinctions. An advantage of running all mixes simultaneously, as opposed to automating them, is that the performer would have the ability to choose a mix from the performance area based on an arbitrary label and compare them in any desired order. This would allow them to ensure that they have been able to appropriately compare each stimulus.

The conventional MUSHRA test necessitates that the evaluator assess the audio in terms of its quality. The concept of quality will need to be modified for this particular test method, as it is not the subject of inquiry. Given the provided information, it is suggested that the evaluation be conducted using criteria such as *realism*, *authenticity* or potentially *spatial representation*. As such, the question that the MUSHRA test response will be asking in the future study will be: Which monitor mix do you prefer for performing live with a band?

6.2.5 Analysis of MUSHRA test data

To attain a desirable level of confidence, it is necessary to consider the findings of previous studies, such as that conducted by Bauer et al., which involved 13 performers [66]. There have been other studies discussing latency in performance, such as that of Bartlette et al., who used as few as 5 groups of performers [107] and Hupke et al., who used 5 performers in total [108]. According to this analysis, the proposed study would necessitate a total of 10 to 20 performers. Furthermore, given the performance-based nature of the experiment, it would be required to enlist 10 to 20 full ensembles of musicians, thus potentially influencing the outcome. The rationale behind the maximum limit of 20 performers is derived from the recommendation provided by the ITU-R, which asserts that, based on empirical evidence, it is generally unnecessary to solicit input from more than 20 assessors in order to reach a reliable and trustworthy conclusion [87].

By employing a statistical technique such as ANOVA, the preferences of the performer can be examined in a quantitative manner. This would improve the precision of hypothesis testing compared to the current pilot tests. The mono RIR would serve as the reference for all other variables, enabling the measurement of deviation from the monaural source. This deviation can then be analysed using a statistical model.

6.2.6 Expectations of Future Testing

Based on the observed patterns from the qualitative pilot testing conducted thus far, it is anticipated that the ORTF and XY configurations will maintain a high level of performance compared to alternative methods of capturing RIRs. What remains uncertain based on the current testing is the extent to which the monitor mixes produced using various methods of capturing RIRs would differ from a dry monitor mix or one generated using a binaural representation of the acoustic environment, such as the system provided by KLANG as discussed in section 1.1. This statement departs from the primary inquiry of this thesis, which focuses on the comparison of stereo microphone techniques employed for the capture of RIRs. However, it presents a potential avenue for future research, exploring the perception of performers in relation to this particular method of capturing acoustic space.

6.2.7 Other Applications and Areas of Research

The study has shed light on another potential area for future investigation, which pertains to live audio but with a focus on remote live audio mixing rather than the performer. The concept of remotely operating a front-of-house console is not widely adopted in the industry and may never become a standard practice. However, there are certain scenarios in which a live engineer may desire the capability to remotely control the system. By employing stereo configurations and convolution, it may be possible to capture the mixing position. Throughout the final stages of composing this thesis, an extensive UK arena tour was undertaken, encompassing a total of eighteen different arenas, including the renowned O2 Arena in London. The data for this study was collected from the perspective of the front-of-house engineer, situated at the listening position. The stereo configuration employed in this study was an ORTF setup, which was selected based on the research findings and its capacity to accurately capture the surrounding environment. This will facilitate the exploration of potential avenues for future research in this field, aiming to identify the shortcomings and limitations of current practices. The primary concern that can be anticipated at present is the challenge of precisely capturing the variations in acoustic characteristics within the venues and achieving a balanced response from the speaker system. However, it is worth noting that this thesis has revealed an encouraging finding: the most significant factor influencing the frequency response of the final RIR is the characteristics of the listening device, rather than the response of the speaker employed to emit the acoustic sweeps.

Another method in which this has been considered to be used would be to convolve the sound of the arena using the IR created from the information captured with the ORTF setup with the main output of the console that is sent to the speaker system. This may enable the front-of-house engineer to monitor the mix of IEMs at a lower and safer level. If the IEM being used was also equalised in the IR, there is the potential that this could be an effective way to have a useful headphone alternative when mixing a live performance.

In revisiting the motivation section of this thesis, one topic that was addressed pertains to the operational practices of numerous large-scale productions, which involve scenarios where the presence of amplified sources on stage is rendered unfeasible. It would be intriguing to employ the methodologies utilised in this thesis in scenarios involving DI'd sources where there is no primary source position to be determined. It is possible that this approach serves as a means of incorporating a sense of realness into a context where it was previously absent.

There is a potential application of the findings that the author will explore in subsequent professional work. The research has demonstrated how the spatial characteristics of the space can best be

represented from a stereo configuration point of view, and it will be interesting to see how this can be incorporated into a live monitoring setup. Much in the same way as was discussed earlier with the use of the JH Audio "Ambient Pro" IEMs [12], a discrete pair of microphones could be positioned at the location of the performer to be fed back to the IEM mix. From what has been found so far, this could be most appropriately done with an ORTF pair or an XY pair, positioning the microphones around the area of the performer's mic stand. Some experimentation with how this can contribute to the sense of space captured in the monitor feed could result in a simple, workable solution for artists not happy with the more common practice of using a spaced pair of microphones downstage left and right that do not directly relate to their monitoring position. This would be more effective for performers that perform from a reasonably static position rather than performers that move around a lot with radio mics, for example, an ensemble where each performer plays an instrument and uses a vocal mic where a pair of microphones could be mounted.

Finally, the author intends to investigate potential applications of the findings in future professional endeavors. The study has provided evidence on the optimal representation of spatial characteristics in a stereo configuration. It will be intriguing to observe the potential integration of these findings into a real-time monitoring system. Similarly to the previous discussion on the use of the JH Audio"Ambient Pro" in-ear monitors [12], it is possible to place a discreet set of microphones at the performer's location to provide audio feedback to the IEM mix. Based on the available evidence, the most suitable approach for this task would involve using either an ORTF pair or an XY pair of microphones, strategically positioned in the vicinity of the performer's microphone stand. Investigating the potential influence of this approach on the perception of spatiality conveyed in the monitor feed may provide a direct and efficient alternative for artists who are unsatisfied with the current practice of using a spaced pair of microphones placed downstage left and right, which do not align directly with their monitoring position. This approach would likely yield greater efficacy for performers who maintain a relatively stationary position during their performances, as opposed to performers who engage in significant movement while utilising radio microphones. For instance, in the case of an ensemble where each performer plays a musical instrument and employs a vocal microphone, it would be feasible to mount a pair of microphones at the position of the microphone stand.

7 Bibliography

[1] B. Pell. *Hear At Last: A History Of Stage Monitoring*. URL: https://www.prosoundweb.com/hear-at-last-a-history-of-stage-monitoring/. (accessed: 22.06.2023).

[2] J. Burton. *An Introduction To In-ear Monitoring The Sound In Your Head*. URL: https://www.soundonsound.com/techniques/introduction-ear-monitoring. (accessed: 22.06.2023).

[3] J. Burton. *An Introduction To In-ear Monitoring The Sound In Your Head*. Feb. 2013. URL: https://www.soundonsound.com/techniques/introduction-ear-monitoring.

[4] Shure. *Shure Prodcut Guide. PSM1000*. URL: https://www.shure.com/en-GB/products/in-ear-monitoring/psm1000. (accessed: 18.01.2024).

[5] Porter Davies. *Porter Davies. World leaders in tactile monitoring*. URL: https://www.porteranddavies.co.uk/.

[6] S. Kerry. "The Silent Stage: The Future of On Stage Sound Systems. The Future of Live Music". In: Bloomsbury Academic, 2021. Chap. 3.

[7] Kemper Amps. *Kemper*. URL: https://www.kemper-amps.com/. (accessed: 18.01.2024).

[8] Neural DSP. *Neural DSP*. URL: https://neuraldsp.com/. (accessed: 18.01.2024).

[9] KLANG. *KLANG: Immersive In Ear Monitoring*. URL: https://www.klang.com/en/home.html. (accessed: 21.10.2022).

[10] Matteo Tomasetti and Luca Turchet. "Playing With Others Using Headphones: Musicians Prefer Binaural Audio With Head Tracking Over Stereo". In: *IEEE Transactions on Human-Machine Systems* 53.3 (2023), pp. 501–511. DOI: 10.1109/THMS.2023.3270703.

[11] Klang. *Testimonials - What You Think Of 3D In-Ear Monitoring - KLANG:technologies — klang.com*. https://www.klang.com/testimonials/. [Accessed 21-05-2024]. 2022.

[12] Jerry Harvey Audio LLC. *JHA-CIEM-AMBPRO*. URL: https://jhaudio.com/iem/JHA-CIEM-AMBPRO. (accessed: 03.12.2023).

[13] JH Audio. *JH Audio. Products. Ambient Pro*. URL: https://jhaudio.com/all-item-groups/products/custom-fit/ambient-pro-9tv2n. (accessed: 18.01.2024).

[14] ACS Custom. *Ambient Series*. URL: https://www.acscustom.com/uk/products/in-ear-monitors/ambient-series. (accessed: 03.12.2023).

[15] Neumann. *KU100 Dummy Head*. URL: https://www.neumann.com/en-en/products/microphones/ku-100/. (accessed: 20.10.2022).

[16] Sound On Sound. *Convolution Processing with Impulse Responses*. URL: https://www.soundonsound.com/techniques/convolution-processing-impulse-responses. (accessed: 03.12.2023).

[17] F. Alton Everest K.C. Pohlmann. *Master Handbook of Acoustics, Fifth Edition*. McGraw Hill, 2009.

[18] D. M. Howard J.A.S. Angus. *Acoustics and Psychoacoustics, Fourth Edition*. Focal Press, 2009.

[19] G. Ballou. *Handbook for Sound Engineers, Fourth Edition*. Focal Press, 2008.

[20] D. A. Bohn. *Environmental Effects on the Speed of Sound**. Presented at the 83rd Convention of the Audio Engineering Society, New York, 1987 October 16-19, 1987.

[21] J. Traer J. H. McDermott. *Statistics of natural reverberation enable perceptual separation of sound and space*. Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, 2016.

[22] N. Kaplanis S. Bech S. H. Jensen T. V¿ Waterschoot. *Perception of Reverberation in Small Rooms: A Literature Study*. AES 55TH INTERNATIONAL CONFERENCE, Helsinki, Finland, 2014 August 27–29, 2014.

[23] W. Reichardt W. Schmidt. *Die Wahmehmbarkeit der Verand-erung von Schallfeldparametem bei der Darbietung von Musik*. Acustica18, 274–282, 1967.

[24] J. Atagi R. Weber V. Mellert. *Effect of modulated delay time of reflection on autocorrelation function and perception of echo*. Journal of sound and vibration 258.3, 2002.

[25] H. Haas. *The influence of a single echo in the audibility of speech*. Building Research Station (Great Britain) Library Communication 363, 1949.

[26] Andrew D Brown, G Christopher Stecker, and Daniel J Tollin. "The precedence effect in sound localization". In: *Journal of the Association for Research in Otolaryngology* 16 (2015), pp. 1–28.

[27] J. A. S. Angus et al. *The Effect of Acoustic Diffusers on Room Mode Decay*. In: Audio Engineering Society 99th Convention, 6-9 October 1995, New York, USA. , pp. 1-13., 1995.

[28] B. Truax. *Handbook for Acoustic Ecology. Second Edition*. World Soundscape Project, Simon Fraser University, and ARC Publications, 1978.

[29] M. Skålevik. *Schroeder Frequency Revisited*. In: FORUM ACUSTICUM 2011, 27. June - 1. July, Aalborg, 2011.

[30] P. J. Shalkouhi S. M. Khezri. *The Schroeder Frequency of Furnished and Unfurnished Spaces*. RJAV vol IX issue 2. Department of Environmental Engineering, Graduate School of the Environment, Energy, Science, and Research Branch, Islamic Azad University, Tehran, Iran, 2012.

[31] Katrin Hewer. "Drum Sound Analysis". In: (2015).

[32] Michael Zevin. "Resonance and Harmonic Analysis of Double Bass and Bass Guitar". In: (2012).

[33] B. Xie. *Head-Related Transfer Function and Virtual Auditory Display: Second Edition*. J Ross Publishing, 2013.

[34] S. A. Gelfand. *Hearing An Introduction to Psychological and Physiological Acoustics*. Taylor and Francis, 2018.

[35] D. W. Batteau. *The Role of the Pinna in Human Localization. https://doi.org/10.1098/rspb.1967.00* Royal Society, 1967.

[36] B. C. J. Moore. *An Introduction to the Psychology of Hearing*. Emerald, 2013.

[37] J. Meyer. *Acoustics and the Performance of Music. Manual for Acousticians, Audio Engineers, Musicians, Architects and Musical Instrument Makers. Fifth Edition*. Springer, 2009.

[38] H. Fletcher W. A. Munson. *Loudness, its definition, measurement and calculation. The Bell System Technical Journal (Volume: 12, Issue: 4, October 1933)*. Nokia Bell Labs, 1933.

[39] W. A. Yost A. N. Popper R. R. Fay. *Human Psychophysics*. Springer, 1993.

[40] L. Rayleigh J. W. Strut. *XII. On our perception of sound direction , Philosophical Magazine Series 6, 13:74, 214-232, DOI: 10.1080/14786440709463595*. 1907.

[41] A. Roginska P. Geluso. *Immersive Sound: The Art and Science of Binaural and Multichannel Audio*. Routledge, 2018.

[42] J. C. Middlebrooks E. A. Macpherson. *Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited*. In: Kresge Hearing Research Institute, University of Michigan, 1301 East Ann Street, Ann Arbor, Michigan 48109-0506, 2002.

[43] D. J. Kistler F. L. Wightman. *Monaural Sound Localization Revisited*. In J. Acoust. Soc. Am. 101 (2),Acoustical Society of America, 1997.

[44] E. Larsen N. Iyer C. R. Lansing A. S. Feng. *On the minimum audible difference in direct-to-reverberant energy ratio*. In J Acoust Soc Am. 124 (1),Acoustical Society of America, 2008.

[45] Yuqing Li, Stephan Preihs, and Jürgen Peissig. "Development and Validation of a Sound Source for Near-field HRTF Measurements". In: Aug. 2021.

[46] M. Vorländer. *Auralization. Fundamentals of Acoustics, Modelling, simulation, Algorithms and Acoustic Virtual Reality. Second Edition.* Springer, 2020.

[47] B. Wiggins. *An investigation into the real-time manipulation and control of three-dimensional sound fields*. University of Derby, 2004.

[48] B. Owsinski. *The Recording Engineers Handbook. Second Edition.* Pro Audio Press, Artist Pro Publishing, 2005.

[49] V. Pulkki. *Microphone techniques and directional quality of sound reproduction.* Audio Engineering Society Convention Paper 5500 Presented at the 112th Convention 2002 May 10–13 Munich, Germany, 2002.

[50] B. Bartlett. *Stereo microphone techniques. Boston, MA*. Focal Press, 1991.

[51] J. Bartlett B. Bartlett. *Practical Recording Techniques: A Step-by-Step Approach to Professional Audio Recording. Fifth Edition.* Focal Press, 2009.

[52] G. Ballou. *Electroacoustic Devices: Microphones and Loudspeakers*. Focal Press, 2009.

[53] DPA Microphones. *DPA Mic Dictionary. What is ORTF?* URL: `https://www.dpamicrophones.com/mic-dictionary/ortf`. (accessed: 03.12.2023).

[54] R. Toulson. *Drum Sound and Drum Tuning: Bridging Science and Creativity*. CRC Press, 2021.

[55] GRAS Acoustics. *KEMAR for sound quality recording*. URL: `https://www.grasacoustics.com/products/head-torso-simulators-kemar/kemar-for-sound-quality-recording-2-ch/product/502-45bb-4`. (accessed: 03.01.2024).

[56] Ah-Hyun Choi et al. "Resonance changes in the external auditory canal associated with the ear canal volume". In: *Phonetics and Speech Sciences* 1.3 (2009), pp. 151–154.

[57] Sarvesh Agrawal et al. "Defining immersion: Literature review and implications for research on immersive audiovisual experiences". In: *Journal of Audio Engineering Society* 68.6 (2019), pp. 404–417.

[58] F. Rumsey. *Spatial Audio*. Taylor Francis, 2012.

[59] E. Zea. *Binaural Monitoring for Live Music Performances*. Stockholm: Royal Institute of Technology School of Computer Science and Communication, 2012.

[60] J. Blauert. *Spatial Hearing: the psychophysics of human sound localization*. MIT Press, 1997.

[61] E. Villchur. *Free-Field Calibration of Earphones, The Journal of the Acoustical Society of America 46, 1527-1534 https://doi.org/10.1121/1.1911897*. 1969.

[62] E.A.G Shaw. *The ExternalExternal Ear. In: Keidel W.D., Neff W.D. (eds) Auditory System. Handbook of Sensory Physiology, vol 5 / 1*. Springer, 1974.

[63] Johahn Leung et al. "Head Tracking of Auditory, Visual, and Audio-Visual Targets". In: *Frontiers in Neuroscience* 9 (2016). ISSN: 1662-453X. DOI: 10.3389/fnins.2015.00493. URL: https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2015.00493.

[64] Steven M. LaValle et al. "Head tracking for the Oculus Rift". In: *2014 IEEE International Conference on Robotics and Automation (ICRA)*. 2014, pp. 187–194. DOI: 10.1109/ICRA.2014.6906608.

[65] Devindra Hardawar. *Oculus rift review: High-end VR is here – if you can pay*. Mar. 2016. URL: https://www.engadget.com/2016-03-28-oculus-rift-review.html?guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAAGz-n0GCyEwE0bC7I_cPmsaSg12XP9-kCzbW9QYKDvGdVhwHxKlB1GCRlyVkzN73C7vA1g_ECrCJN2Iaa8J0pDMkZFmXX7htG72O1CqDkTvnHtuKlo5ePC15amY1on1mx2rU9b8gjD3LA85s4Bw2FhTAcNa_VkK7&guccounter=2.

[66] Valentin Bauer et al. "Binaural headphone monitoring to enhance musicians' immersion in performance". In: *Advances in Fundamental and Applied Research on Spatial Audio* (2022).

[67] C. Roads. *Musical Signal Processing. Sound transformation by convolution*. Routledge, 1997.

[68] NTi Audio. *Fast Fourier Transformation FFT - Basics*. URL: https://www.nti-audio.com/en/support/know-how/fast-fourier-transform-fft#:~:text=The%20%22Fast%20Fourier%20Transform%22%20(,frequency%20information%20about%20the%20signal.. (accessed: 03.01.2024).

[69] K.C. Pohlmann. *Principles of Digital Audio, Sixth Edition*. McGraw Hill, 2011.

[70] National Instruments. *Understanding FFTs and Windowing*. URL: https://download.ni.com/evaluation/pxi/Understanding%20FFTs%20and%20Windowing.pdf. (accessed: 03.01.2024).

[71] ISO 3382-1. *Acoustics—Measurement of Room Acoustic Parameters—Part I: Performance Spaces*. International Organization for Standardization, 2009.

[72] Wenmaekers Luxemburg Hak. *Measuring room impulse responses : impact of the decay range on derived room acoustic parameters*. Acta Acustica united with Acustica, 2012.

[73] A. Farina. *Simultaneous measurement of impulse response and distortion with a swept-sine technique Audio Engineering Society Convention 108*. Audio Engineering Society, 2000.

[74] I.H. Chan. *Swept Sine Chirps for Measuring Impulse Response*. Stanford Research Systems, Inc, 2010.

[75] NTi Audio. *DS3 Dodecahedron Speaker*. URL: https://www.nti-audio.com/en/products/noise-sources/ds3-dodecahedron-speaker. (accessed: 18.01.2024).

[76] N. Papadakis G. Stavroulakis. *Review of Acoustic Sources Alternatives to aDodecahedron Speaker*. MDPI Applied Sciences Journal, 2019.

[77] P. Stade B. Bernschütz M. Rühl. *A Spatial Audio Impulse Response Compilation Captured at the WDR Broadcast Studios*. Cologne University of Applied Sciences, Germany - Institute of Communication Systems, Technical University of Berlin, Germany - Audio Communication Group, 2012.

[78] R. Farina A. Ayalon. *Recording Concert Hall Acoustics For Posterity*. AES 24th International Conference on Mulitchannel Audio, 2003.

[79] Engin Gurur Gelen. "Convolution an approach for pre-auralization of a performance space". 2019.

[80] B. Katz. *Mastering Audio. The art and the science*. Focal Press, 2016.

[81] A. Clifford J. Reiss. *Calculating time delays of multiple active sources in live sound. Convention Paper 8157*. Audio Engineering Society, 2010.

[82] B. Evans. *Live Sound Fundamentals*. Course Technology, 2011.

[83] M. Merchel S. Ercan Altinsoy. *Musical Haptics. Audio-Tactile Experience of Music*. Springer Open, 2018.

[84] M. Lester J. Boley. *The Effects of Latency on Live Sound Monitoring. Convention Paper Presented at the 123rd Convention 2007 October 5–8*. Audio Engineering Society, 2007.

[85] Michael Schoeffler et al. "Towards the next generation of web-based experiments: A case study assessing basic audio quality following the ITU-R recommendation BS. 1534 (MUSHRA)". In: *1st Web Audio Conference*. 2015, pp. 1–6.

[86] M. Schöffler. *Overall Listening Experience - a new Approach to Subjective Evaluation of Audio*. International Audio Laboratories Erlangen, 2017.

[87] B. Series. "Recommendation ITU-R BS.1534-3 Method for the subjective assessment of intermediate quality level of audio systems". In: *International Telecommunication Union Radiocommunication Assembly* (2015).

[88] Slawomir Zielinski et al. "Potential biases in MUSHRA listening tests". In: *Audio Engineering Society Convention 123*. Audio Engineering Society. 2007.

[89] TestDevLab. *How We Conduct Listening Tests Based on MUSHRA Methodology — testdevlab.com*. `https://www.testdevlab.com/blog/how-we-conduct-listening-tests-mushra-methodology-from-itu-r-recommendations`. [Accessed 16-07-2024]. 2023.

[90] A. Field. *Discovering statistics using IBM SPSS statistics*. Sage publications limited, 2024.

[91] Lars Sthle and Svante Wold. "Analysis of variance (ANOVA)". In: *Chemometrics and Intelligent Laboratory Systems* 6.4 (1989), pp. 259–272. ISSN: 0169-7439. DOI: `https://doi.org/10.1016/0169-7439(89)80095-4`. URL: `https://www.sciencedirect.com/science/article/pii/0169743989800954`.

[92] The MathWorks Inc. *MATLAB version: 9.8.0 (R2020a)*. Natick, Massachusetts, United States, 2020. URL: `https://www.mathworks.com`.

[93] Thomann. *Dynaudio BM6A Classic*. URL: `https://www.thomann.de/gb/dynaudio_bm6a_monitor.htm`. (accessed: 28.01.2024).

[94] Earthworks Audio. *M30 — earthworksaudio.com*. `https://earthworksaudio.com/measurement-microphones/m30/`. [Accessed 28-06-2024]. 2020.

[95] Audiolab York University. *Isolated anechoic chamber*. URL: `https://audiolab.york.ac.uk/facilities-anechoic-chamber/`. (accessed: 18.01.2024).

[96] SPL electronics GmbH. *HawkEye. Precision Audio Analysis Software*. URL: https://spl.audio/en/spl-produkt/hawkeye-plugin/. (accessed: 03.01.2024).

[97] Shure. *Shure SE215 earphone user guide*. URL: https://pubs.shure.com/guide/SE215M/en-US?_gl=1*17dlrpv*_ga*MTkxODk3OTc2Ny4xNjgxNDY5MDky*_ga_DB3CR9SF0C*MTcwNTY1NTk5Ny4yLjAuMTcwNTY1NTk5OS42MC4wLjA.*_gcl_au*MTk2NzA0ODEwNi4xNzA1NjU1OTk5&_ga=2.151717113.196315262.1705655999-1918979767.1681469092. (accessed: 18.01.2024).

[98] RME. *Fireface UCX II - RME Audio Interfaces, Format Converters, Preamps, Network Audio MADI Solutions*. 2024. URL: https://rme-audio.de/fireface-ucx-ii.html.

[99] Avid. *Pro Tools HD IO. Premium, Customizable Pro Tools HD Series INterface*. [Accessed 21-05-2024]. 2014. URL: https://cdn-www.avid.com/-/media/avid/files/resources-pdf/hdiodsa44.pdf.

[100] Solid State Logic. *AWS 900 + SE Owners Manual*. [Accessed 21-05-2024]. 2009. URL: https://www.solidstatelogic.com/assets/uploads/downloads/aws/AWS900-SE-Owners-Manual.pdf.

[101] Yamaha. *Yamaha Professional Audio. MG Series*. URL: https://europe.yamaha.com/en/products/proaudio/mixers/mg_series_xu_model/index.html. (accessed: 12.06.2024).

[102] A. Hagerman. *Pro Tools Fundamentals II. PT 110*. Avid Official Curriculum. Avid Learning Services, 2022.

[103] Rensis Likert. "A technique for the measurement of attitudes." In: *Archives of psychology* (1932).

[104] Avid. *Space - Reverb Plugin - Avid*. URL: https://www.avid.com/plugins/space/. (accessed: 20.10.2022).

[105] Amcoustics. *Amroc. The Room Mode Calculator*. URL: https://amcoustics.com/tools/amroc?l=645&w=490&h=254&r60=0.75. (accessed: 20.06.2024).

[106] Steven Michael Fenton. "Audio Dynamics: Towards a Perceptual Model of'punch'." PhD thesis. University of Huddersfield, 2017.

[107] Christopher Bartlette et al. "Effect of network latency on interactive musical performance". In: *Music Perception* 24.1 (2006), pp. 49–62.

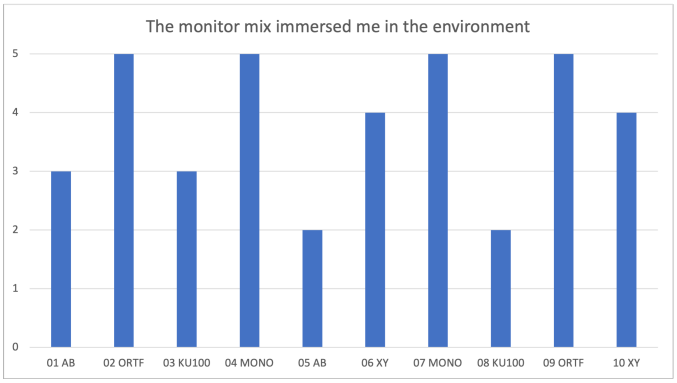
[108] Robert Hupke et al. "Latency and quality-of-experience analysis of a networked music performance framework for realistic interaction". In: *Audio Engineering Society Convention 152*. Audio Engineering Society. 2022.

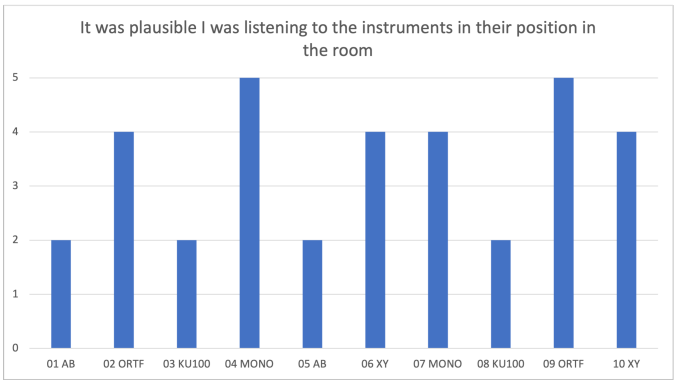
[109] AKG. *AKG c414 XLS Reference Multipattern Condenser Microphones*. https://uk.akg.com/on/demandware.static/-/Sites-masterCatalog_Harman/default/dw0d863e9c/pdfs/AKG_C414XLS_C414XLII_Manual.pdf. [Accessed 05-07-2024].

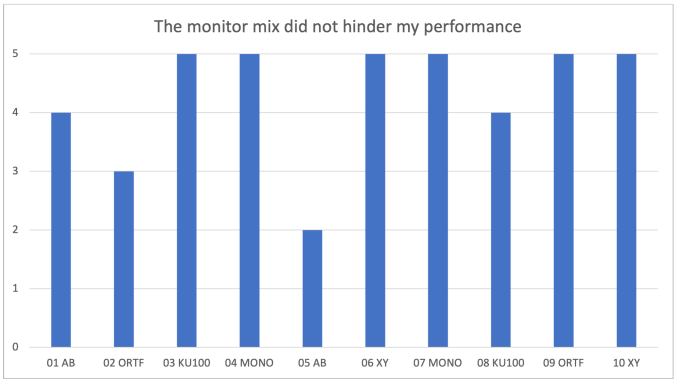
[110] Neumann. *KM184 (Series 180)*. URL: https://www.neumann.com/en-en/products/microphones/km-184-series-180/. (accessed: 20.10.2022).

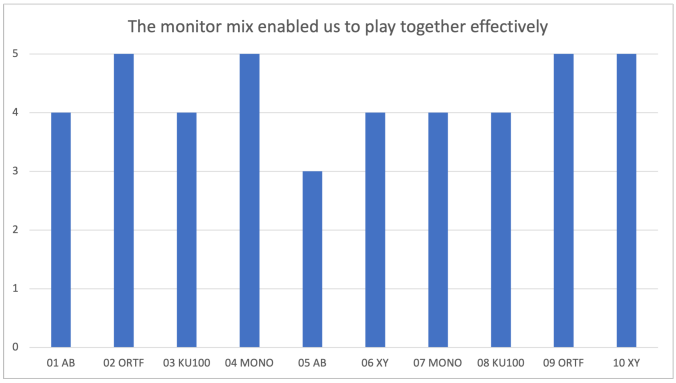
8 Appendix

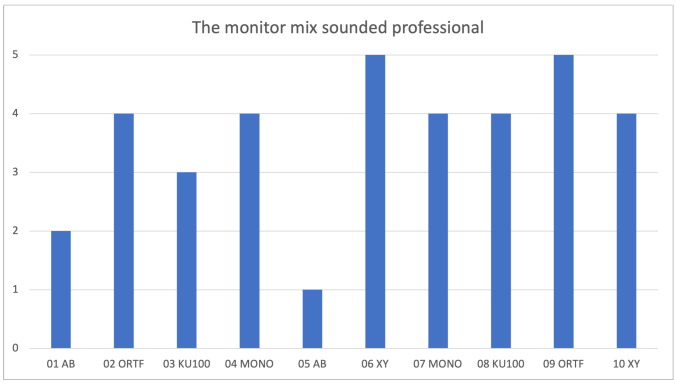
8.1 Testing Results

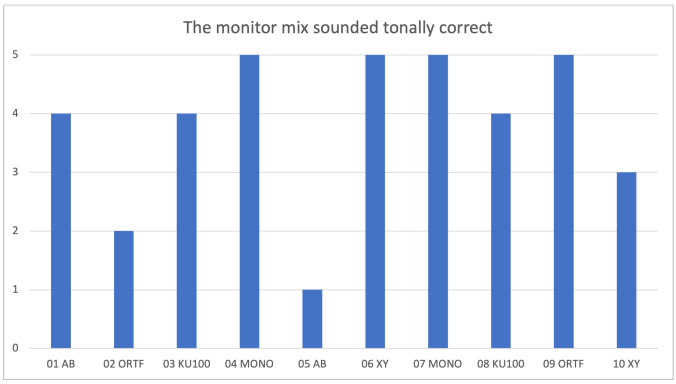


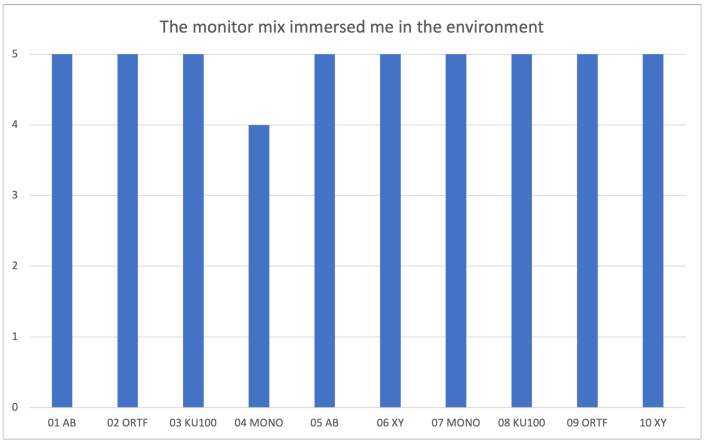


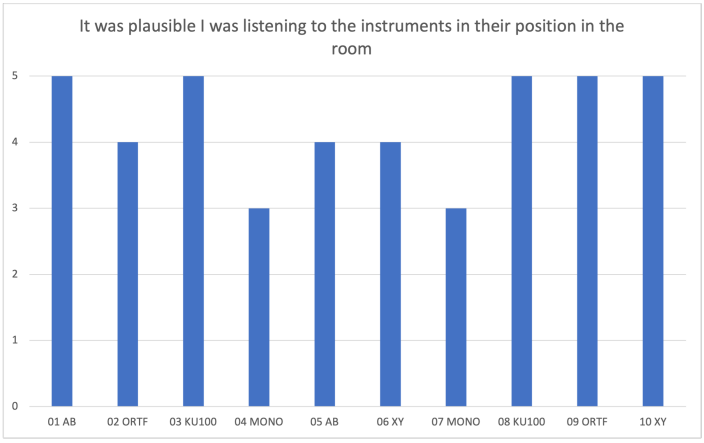


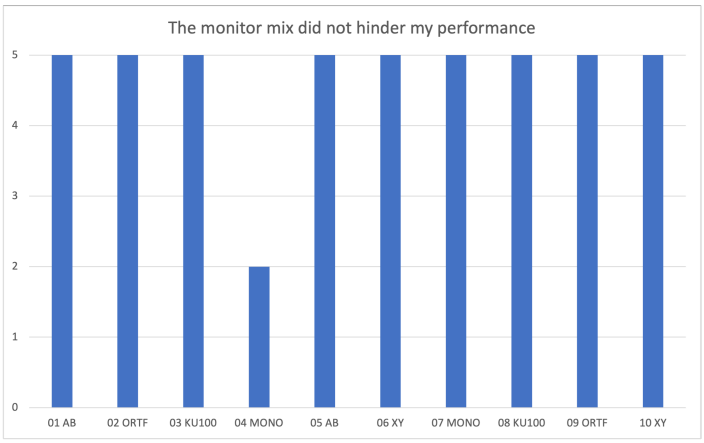

The monitor mix enabled us to play together effectively
5
4
3
2
1
0
01 AB
02 ORTF
03 KU100
04 MONO
05 AB
06 XY
07 MONO
08 KU100
09 ORTF
10 XY

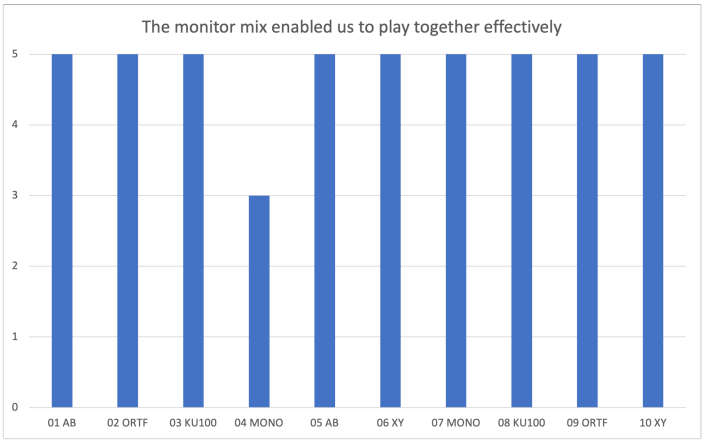

The monitor mix sounded professional
5
4
3
2
1
0
01 AB
02 ORTF
03 KU100
04 MONO
05 AB
06 XY
07 MONO
08 KU100
09 ORTF
10 XY

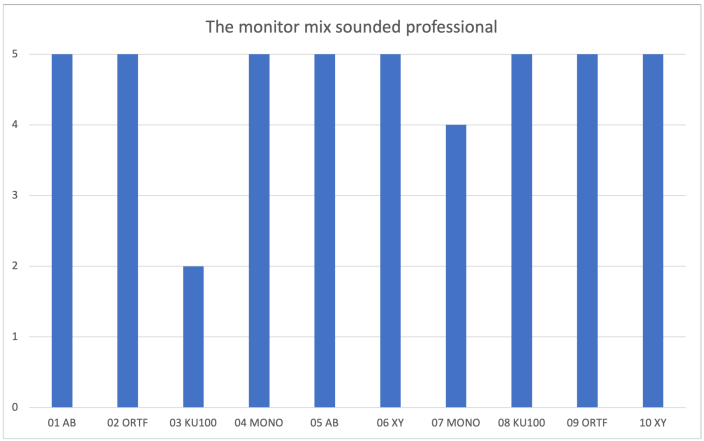

The monitor mix sounded tonally correct
5
4
3
2
1
0
01 AB
02 ORTF
03 KU100
04 MONO
05 AB
06 XY
07 MONO
08 KU100
09 ORTF
10 XY

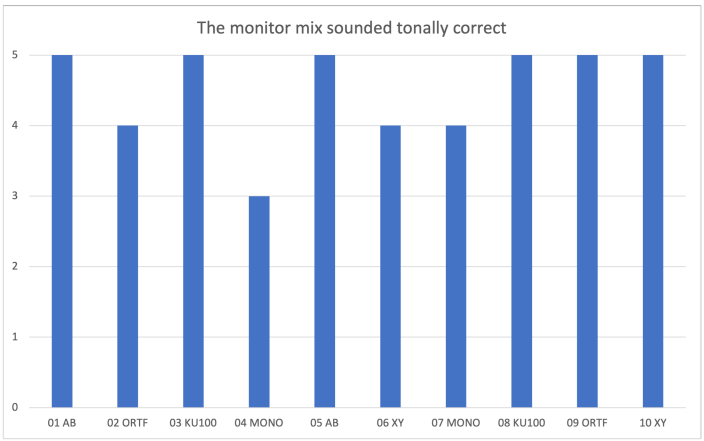

The monitor mix immersed me in the environment
5
4
3
2
1
0
01 AB
02 ORTF
03 KU100
04 MONO
05 AB
06 XY
07 MONO
08 KU100
09 ORTF
10 XY

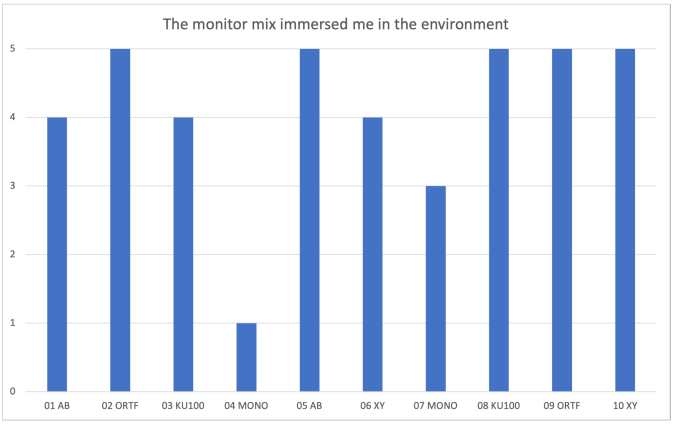

It was plausible I was listening to the instruments in their position in the room
5
4
3
2
1
0
01 AB
02 ORTF
03 KU100
04 MONO
05 AB
06 XY
07 MONO
08 KU100
09 ORTF
10 XY

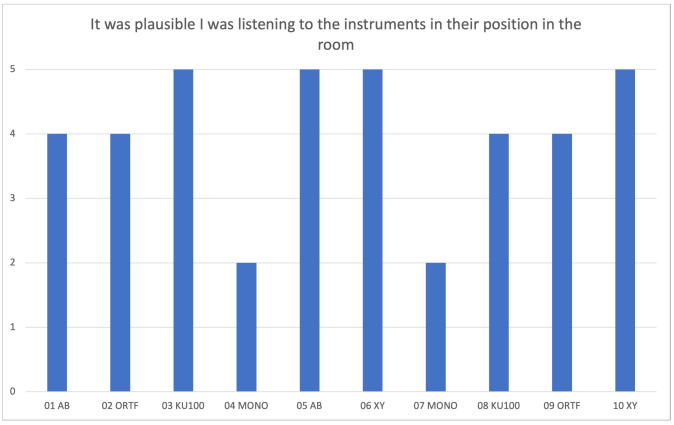

The monitor mix did not hinder my performance
5
4
3
2
1
0
01 AB
02 ORTF
03 KU100
04 MONO
05 AB
06 XY
07 MONO
08 KU100
09 ORTF
10 XY

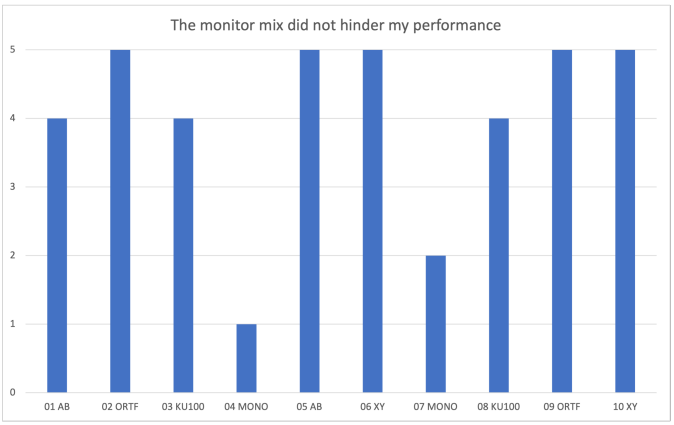

The monitor mix enabled us to play together effectively
5
4
3
2
1
0
01 AB
02 ORTF
03 KU100
04 MONO
05 AB
06 XY
07 MONO
08 KU100
09 ORTF
10 XY

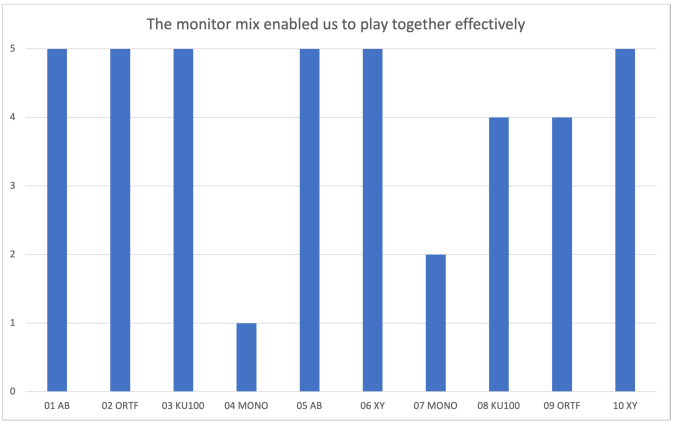

The monitor mix sounded professional
5
4
3
2
1
0
01 AB
02 ORTF
03 KU100
04 MONO
05 AB
06 XY
07 MONO
08 KU100
09 ORTF
10 XY

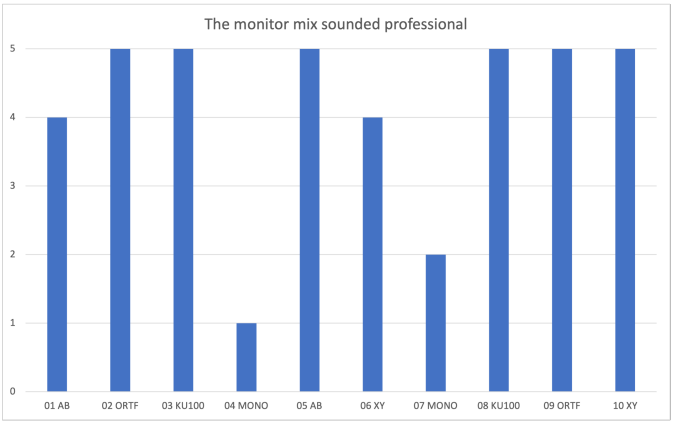

The monitor mix sounded tonally correct
5
4
3
2
1
0
01 AB
02 ORTF
03 KU100
04 MONO
05 AB
06 XY
07 MONO
08 KU100
09 ORTF
10 XY

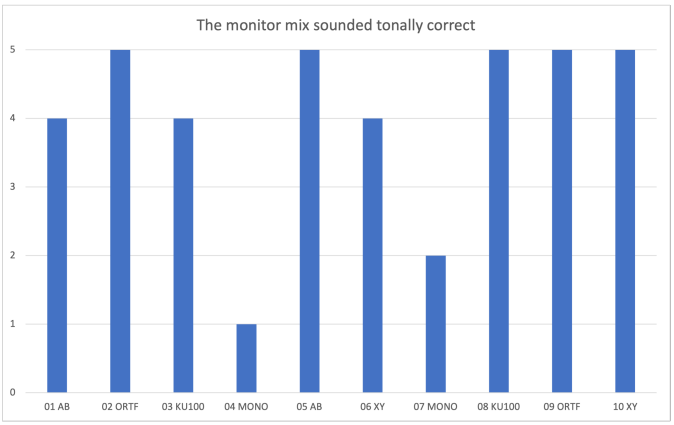

The monitor mix immersed me in the environment
5
4
3
2
1
0
01 AB
02 ORTF
03 KU100
04 MONO
05 AB
06 XY
07 MONO
08 KU100
09 ORTF
10 XY

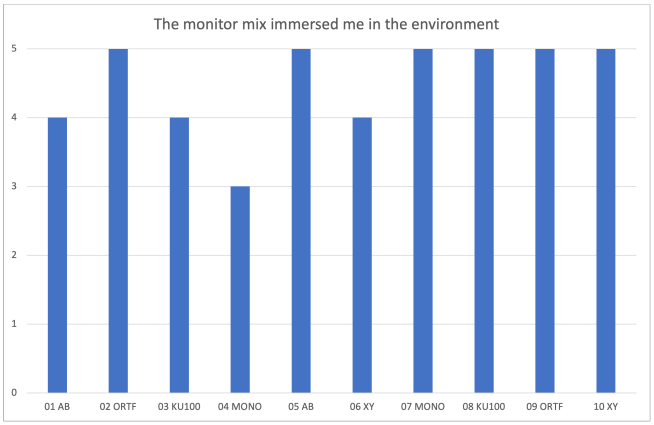

It was plausible I was listening to the instruments in their position in the room
5
4
3
2
1
0
01 AB
02 ORTF
03 KU100
04 MONO
05 AB
06 XY
07 MONO
08 KU100
09 ORTF
10 XY

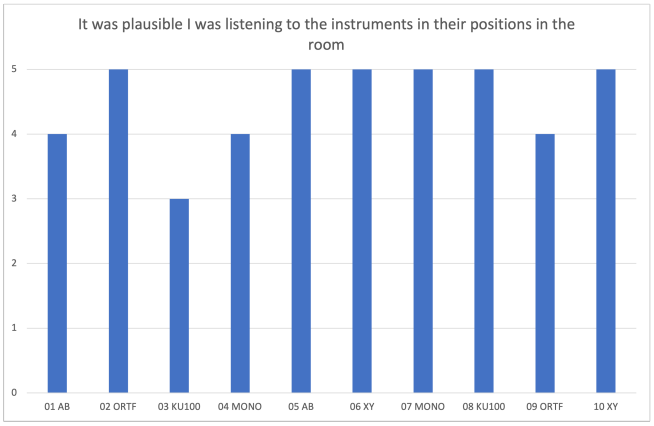

The monitor mix did not hinder my performance
5
4
3
2
1
0
01 AB
02 ORTF
03 KU100
04 MONO
05 AB
06 XY
07 MONO
08 KU100
09 ORTF
10 XY

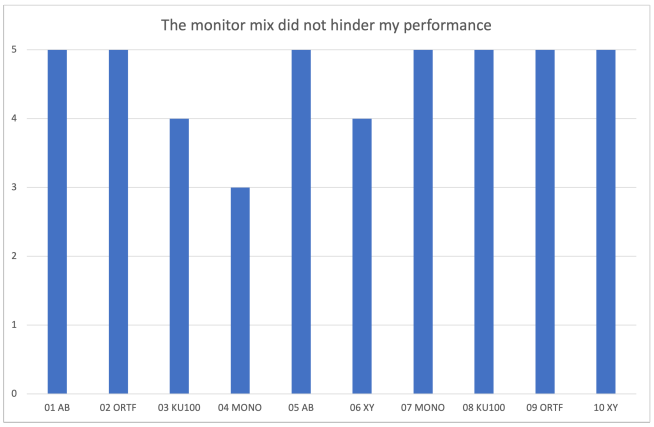

The monitor mix enabled us to play together effectively
5
4
3
2
1
0
01 AB
02 ORTF
03 KU100
04 MONO
05 AB
06 XY
07 MONO
08 KU100
09 ORTF
10 XY

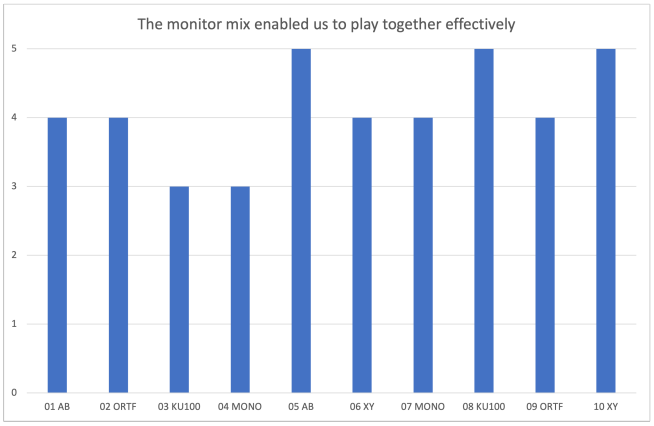

The monitor mix sounded professional
5
4
3
2
1
0
01 AB
02 ORTF
03 KU100
04 MONO
05 AB
06 XY
07 MONO
08 KU100
09 ORTF
10 XY

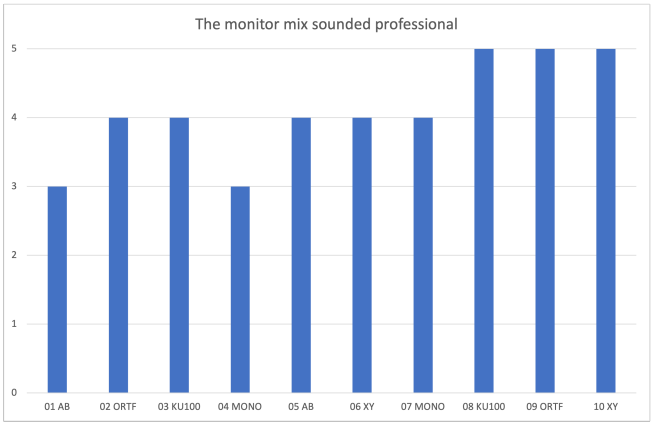

The monitor mix sounded tonally correct
5
4
3
2
1
0
01 AB
02 ORTF
03 KU100
04 MONO
05 AB
06 XY
07 MONO
08 KU100
09 ORTF
10 XY

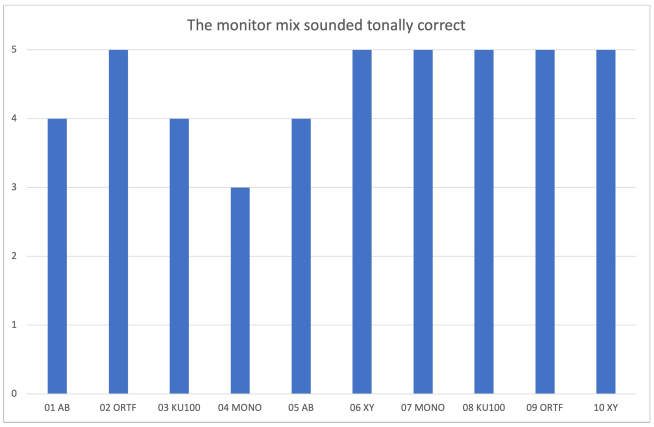

The monitor mix immersed me in the environment
5
4
3
2
1
0
01 AB
02 ORTF
03 KU100
04 MONO
05 AB
06 XY
07 MONO
08 KU100
09 ORTF
10 XY


It was plausible I was listening to the instruments in their positions in the room
5
4
3
2
1
0
01 AB
02 ORTF
03 KU100
04 MONO
05 AB
06 XY
07 MONO
08 KU100
09 ORTF
10 XY


The monitor mix did not hinder my performance
5
4
3
2
1
0
01 AB
02 ORTF
03 KU100
04 MONO
05 AB
06 XY
07 MONO
08 KU100
09 ORTF
10 XY


The monitor mix enabled us to play together effectively
5
4
3
2
1
0
01 AB
02 ORTF
03 KU100
04 MONO
05 AB
06 XY
07 MONO
08 KU100
09 ORTF
10 XY


The monitor mix sounded professional
5
4
3
2
1
0
01 AB
02 ORTF
03 KU100
04 MONO
05 AB
06 XY
07 MONO
08 KU100
09 ORTF
10 XY


The monitor mix sounded tonally correct
5
4
3
2
1
0
01 AB
02 ORTF
03 KU100
04 MONO
05 AB
06 XY
07 MONO
08 KU100
09 ORTF
10 XY

8.2 Microphone Frequency Response Charts

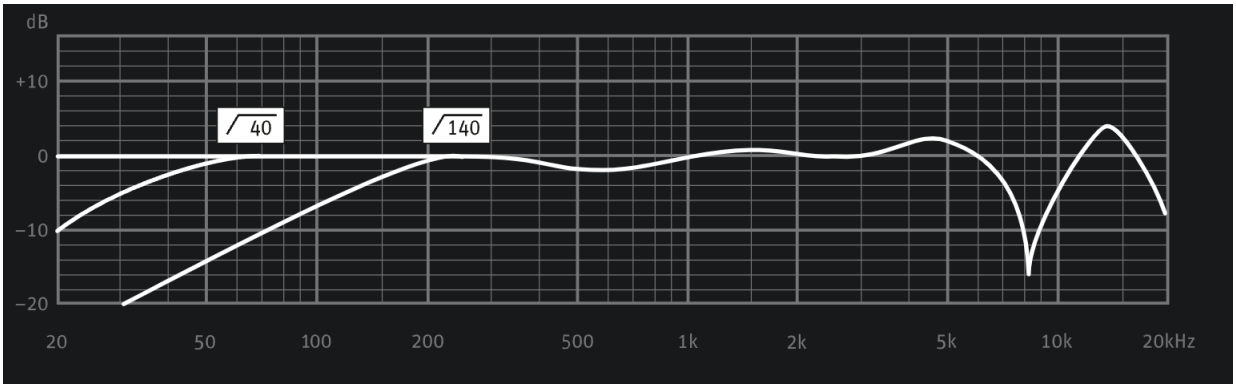


Figure 69: The frequency response of the Neumann KU100 from [15]

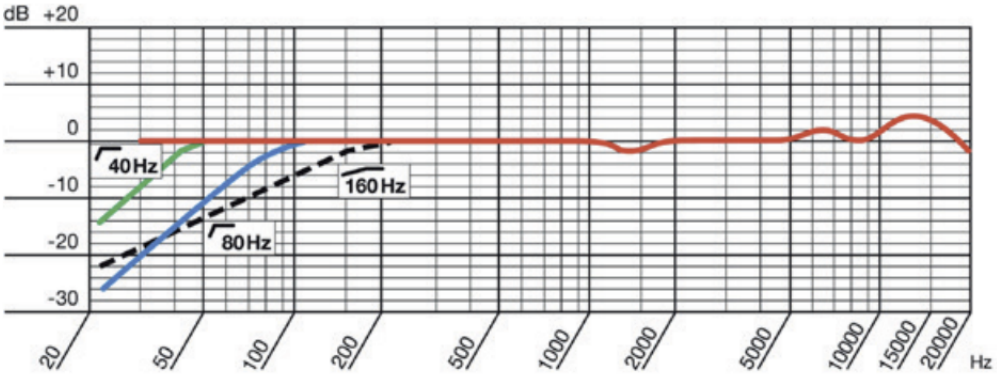


Figure 70: The frequency response of the AKG c414 with the polar response set to omni-directional from [109]

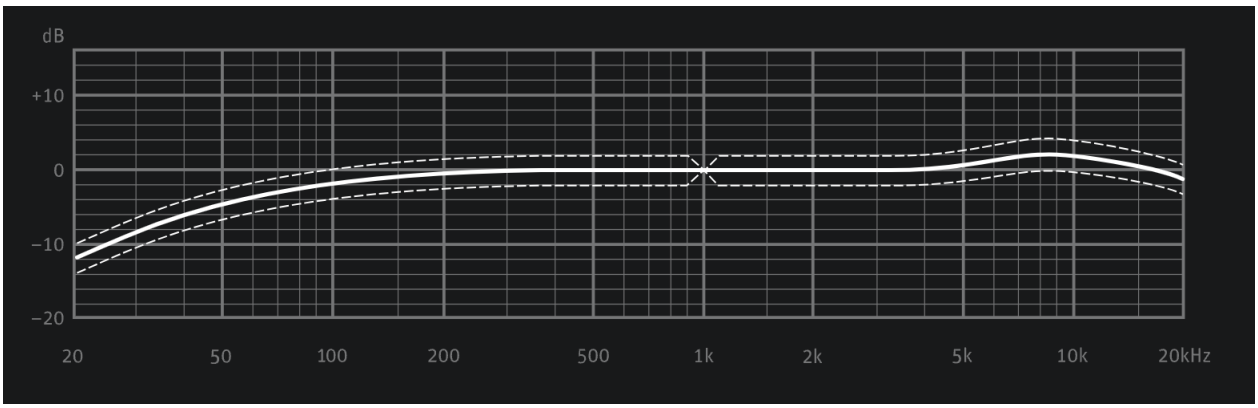


Figure 71: The frequency response of the Neumann km184 from [110]

8.3 Ethics Approval Application

THE UNIVERSITY of York

Application Form for Physical Sciences Ethics Committee Approval

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

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

elec-ethics@york.ac.uk

Title of project: **Stereo Convolution of Space for Live Performance Monitoring**

SECTION 1 DETAILS OF APPLICANTS

Details of principal investigator (name, appointment and qualifications)

Steven Kerry – Lecturer of Audio Engineering and Production at Futureworks Manchester. MSc by research student at University of York.

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

7th May 2015

Location(s) of project

SSL Studio
Futureworks University Education, Mamchester. M3 5FS

SECTION 2 FUNDERS

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

There is no external funding for the project however the MSc is being funded by Futureworks (Employer).

Please answer the following:

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

YES		NO	x

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

YES		NO	x

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

YES		NO	x

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

YES		NO	x

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

7th May 2015

SECTION 3 DETAILS OF PROJECT OR OTHER ACTIVITY

Aims (100 words max)

The aim of the project is to evaluate the use of convolution of space to immerse a listener in a performance environment in real time. The variables in question relate to the use of different microphone configurations to create the RIRs (room impulse responses).

Background (250 words max)

In a live sound environment, performers will often use in ear monitors to hear themselves and other instrument sources on stage. Although there are many benefits to this practice due to the isolation it achieves and inherent reduction of noise for the listener, this also results in an unrealistic separation from the performance area.

The ability to capture the acoustic attributes of a space is not a new practice however it is not currently something commonly used within the area of live sound. The aim of this project is to situate the performer back into the performance area by processing the in ear monitor feed with the acoustic attributes of the space.

The method of capturing the acoustic attributes is the area in question. The use of different microphone techniques to capture the RIRs (Room Impulse Responses) gives a different result due to the space being captured from a different 'perspective'. The use of a binaural dummy head would in theory give the most relatable response due to it being modelled on a human head however there may be other appropriate methods of capturing the RIRs with more accessible stereo microphone techniques.

Brief outline of project/activity (250 words max)

The participants will first read through an information sheet and sign a consent form. These forms are included in this application.

Participants will be set up in the performance area and then will be asked to perform a pre-arranged piece of music. While playing the performers will either be wearing IEMs, or if for any reason they are not currently being fed a monitor feed via IEMs, earplugs will be supplied for them to wear to reduce noise levels.

The performers will repeat the performance with different test conditions used on each subsequent performance. At the end of each performance a survey will be taken in the form of a Likert scale questionnaire.

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

The method of study will be a randomised control trial whereby the performers will have different stumuli sent to them at random, so they are unaware of the test conditions for each performance. A mono test stimuli will be used as an anchor.

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

7th May 2015

Initially a group consisting of four performers will be recruited for a pilot test followed by four different performers for the actual study. Depending on the results of the experiment and time available a repetition of the experiment may be conducted with a further four performers. So in total between eight and twelve people will be recruited.

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

The study requires at least four performers because the aim is to analyse the usefulness of the technique from the different perspectives of performers on a stage. Although there may be situations in practice where there are more or less performers on a stage than this, having four allows the consideration of performers centrally both in upstage and downstage positions and then also to the right and left of the downstage area.

SECTION 4 RECRUITMENT OF PARTICIPANTS

How will the participants be recruited?

Participants will be recruited by advertising within the Futureworks University campus.

What are the inclusion/exclusion criteria?

Participants will be over the age of 18.

Will participants be paid reimbursement of expenses?	YES [ ]	NO [x]
Will participants be paid?	YES [ ]	NO [x]
If yes, please obtain signed agreement		
Will any of the participants be students?	YES [x]	NO [ ]

7th May 2015

SECTION 5 DATA STORAGE AND TRANSMISSION

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

Manual files	Consent forms/Questionnaires
University computers	
Home or other personal computers	
Laptop computers, tablets	All digital files
Website	

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

Data will be anonymised as there is no requirement to store any personal information. A unique user ID will be created for each participant so that if the data needs to be deleted or is requested at any point, it can be located effectively. Digital information will be stored on secure encrypted devices which are password protected.

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

Steven Kerry

Anonymised date will be accessible to MSc Supervisor Dr Gavin Kearney and TAP advisory panel member Dr Helena Daffern

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

Steven Kerry

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

All data will be analysed and dealt with by Steven Kerry

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

Data will be anonymised and put into password protected Zip files and stored in a private online storage area for the duration of the study.

SECTION 6 CONSENT

Is written consent to be obtained?

YES	x		NO	

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

If no, please justify

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

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

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

SECTION 7 DETAILS OF INTERVENTIONS

Indicate whether the study involves procedures which:

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

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

List any potential hazards:

The main potential hazard present within the study is the noise levels in the performance space. The use of IEMs will give a reduction of ambient noise of approximately 37dbSPL. For any

performer not using IEMs while performing, ear plugs will be provided offering a suggested 39dBSPL of reduction. This brings all exposure levels to under the government recommendations for exposure levels in a work environment of 85dBSPL.

List any discomfort or distress:

What steps will be taken to safeguard

(i) the confidentiality of information

Participant data will remain confidential, participant information will be anonymised with the use of individual identification codes.

(ii) the specimens themselves?

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

Nothing specific applies in this study.

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

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

Is there a secrecy clause to the research? YES [ ] NO [x]
If yes, please give details below

7th May 2015

SECTION 8 SIGNATURES

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

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

Signature of Investigator(s): 

……

Date: ……30th March 2023……

Responsibilities of the Principal Researcher following approval

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

8.4 Matlab Code

```
%% Sweep Parameters
freq_lower = 20         % Lower frequency of sweep in Hz.
freq_upper = 20000;     % Upper frequency of sweep
duration = 10;          % Duration of sweep. Usually at least 10 times the (approx) reverb time of the room
fs = 48000;             % The sampling frequency in Hz
padstart = 0;           % Place some silence at start if desired
padend = 0;             % Place some silence at end of sweep if desired
nowav = 0;              % 0 for wavfiles output, 1 for wav files.
nometa = 1;             % 0 for no metadata, 1 for metadata.
ReverbSpaceTime = 1;    % Rough estimate of reverb time of the room


%% Now run the generatesweep function. This will create a sweep and its inverse filter.
[ sweep, inv_filter ] = generatesweep( freq_lower, freq_upper, duration, fs, padstart, padend, nowav, nometa, ReverbSpaceTime);
```

Figure 72: The Matlab code used in the sweep generation process including the creation of the inverse filter of the sweep. Originally created by Simon Shelley, 2012

```
%% Load in the files

rec_filename = 'example.wav'; % Name of recorded sweep

inv_filename = 'Inverse_Sweep_20to20000_48000_pad0s.wav'; % Name of inverse sweep

out_filename = 'example_IR.wav'; % Name of output IR file

[rec] = audioread(rec_filename); % Load recording and note its sample rate

[inv] = audioread(inv_filename); % Load inverse and note its sample rate

%% Perform deconvolution

ir_L = fftfilt(inv, rec(:,1));     % Deconvolve left channel
ir_R = fftfilt(inv, rec(:,2));     % Deconvolve right channel

ir_out = [ir_L ir_R]; % Make a stereo IR from both L and R IRs

%ir_out  = ir_out(Fs*(Sweeplen + 1):end,:); % Remove some silence from start of IR (Sweep length + 1 sec)

ir_out  = ir_out./max(abs(ir_out(:)));  % Normalise the output

%% Write the output to a wav file

audiowrite(out_filename, ir_out, Fs, 'BitsPerSample', 24); % Write the output to a wav
```

Figure 73: The Matlab code used in the deconvolution process. Originally created by Simon Shelley, 2012

```
%% Windowing function for quick fade in and out around IRs

myhan = hanning(64); % Creates a hanning window
mywin = [myhan(1:length(myhan)/2); ones(512-length(myhan), 1); ...
    myhan(length(myhan)/2+1:end) ]; % Fade in, signal part, fade out

figure(1)
plot(mywin)
xlabel('Samples');              % x-axis label
ylabel('Nominal amplitude');    % y-axis label
xlim([0 512]);                  % x-axis limits
ylim([0 1]);                    % y-axis limits
```

```
%% Load the deconvolved impulse responses

[ir_1, Fs] = audioread('SE215 Sweep 230706.wav');  % Load IR, sample rate
```

```
%% Tidy up IRs

[m, I] = max(abs(ir_1)); % Find point of max value i.e. peak of IR

ir_1 = ir_1(I-100:I+411,:).*mywin; % Apply 32 sample fade in and fade out.
```

Figure 74: The Matlab code used in the inverse filtering process. Originally created by Simon Shelley, 2012

```
%% Compute inverse filter

% Filter parameters
type = 'minphase';          % Minimum Phase
Nfft = 4096;                % Frequency resolution
Noct = 2;                   % Setting for 1/2 octave smoothing
L = Nfft;
range = [60 17000];         % In-band frequency range
reg=[10 20];                % Amound of out-band and in-band frequency
                            % inversion permitted in dB
window = 1;                 % Apply windowing

[ih]=invFIR(type,ir_1(:,1),Nfft,Noct,L,range,reg,window, Fs);
                            % Compute Inverse Filter

resp = conv(ih(:,1), ir_1(:,1));     % Resultant frequency response after EQ
                            % (for plotting only)

audiowrite('DELETETHISFILE.wav', ih./max(abs(ih))*0.99, Fs);
                            % Write out the (normalised) inverse filter to
                            % a file.
```

```
%% Plot IRs in the time domain

figure()
subplot(2,1,1)
hold on
t = 1/Fs:1/Fs:length(ir_1)/Fs; % Time index for plotting
plot(t, ir_1(:, 1));        % Plot ir against time
xlabel('Time (s)');
ylabel('Nominal amplitude');
title('Left channel response');
```

Figure 75: The Matlab code used in the inverse filtering process. Originally created by Simon Shelley, 2012

```
subplot(2,1,2)
hold on
plot(t, ir_1(:, 2));        % Plot ir against time
xlabel('Time (s)');
ylabel('Nominal amplitude');
title('Right channel response');
```

```
%% Plot IRs in the Frequency domain

figure()
% Left
%subplot(2,1,1)
GK_Freqplot(ir_1(:,1),Fs, 2048, 'k:', 1, 8, ...
    'IEM Measurement', 'Frequency', 'dB',  [-70 30], ...
    [20 20000],  0, 0);

% % Right
% subplot(2,1,2)
% GK_Freqplot(ir_1(:,2),Fs, 2048, 'k:', 1, 8, ...
%     'IR Measurements - Right', 'Frequency', 'dB',  [-70 30],  ...
%     [20 20000],  0, 0);

% Plot average response and inverse
figure()
GK_Freqplot(ir_1,Fs, 2048, 'r', 1, 10, 'IEM Response Equalisation',  ...
    'Frequency', 'dB',  [-70 30], [20 20000],  0, 0);
hold on
GK_Freqplot(ih, Fs, 2048, 'g', 1, 10, 'IEM Response',  ...
    'Frequency', 'dB',  [-70 30], [20 20000],  0, 0);
GK_Freqplot(resp, Fs, 2048, 'b', 1, 10, 'IEM Response',  ...
    'Frequency', 'dB',  [-70 30], [20 20000],  0, 0);
legend('IEM Response', 'Inverse Filter', 'Resultant Equalisation',  ...
    'location', 'southeast');
```

Figure 76: The Matlab code used in the inverse filtering process. Originally created by Simon Shelley, 2012