The phenomenal rise of periphonic record production: a practice-based, musicological investigation of periphonic recomposition and mix aesthetics for binaural reproduction

The Phenomenal Rise of Periphonic Record Production

A practice-based, musicological investigation of periphonic recomposition and mix aesthetics for binaural reproduction.

Jocelyne Lord

A thesis submitted in partial fulfilment of the requirements of The University of West London for the degree of Doctor of Philosophy

February 2022

This Ph.D thesis is dedicated

to

Akis Kollaros

(1980 – 2015)

A true friend, a knowledgeable mentor and an exceptional human being. I am forever grateful for the positive influence, for all he taught me and for the joy he brought into my life.

Acknowledgments

Firstly, I'd like to give a huge thank you to Andy Caldecott – my best friend, bandmate and musical collaborator of almost two decades. Without you so many things would not be possible. Thank you not only for the music but for your friendship, support, patience and resilience.

Another huge thank you to all of those who also contributed their music to this project; Joey Clarkson, Beautiful Thing, Jimmy Logic and Jerome Thomas. A thank you also extends to those whose music did not make it into the final research project – thank you to every one of you for offering your artistic endeavours, they were a valuable contribution to helping further my research.

Thank you to Professor Simon Zagorski-Thomas for his musicological influence, breadth of knowledge and guidance.

Thank you to Dr Andrew Bourbon for fostering my interest in binaural audio.

An important shout out to Denise Humphreys for opening the academic gate to me and for being the most brilliant and one of very few female inspirations in a male dominated field. I have so much respect for you – my music production and business sister.

A shout out to Romeo aka 'Pops' for keeping it real and for his absolute persistence in checking in on me and keeping me in check. My sound system Dad, big love.

A huge thank you to Mike Exarchos aka 'Stereo Mike' for his unwavering support and encouragement, the lengthy late night Ph.D. conversations and for his consistent love, patience and understanding.

A thank you extends to all of my colleagues, friends, family and loved ones who have supported and encouraged me over the years. A special thank you to those who bore the brunt of my stress, the incessant neglect and hermitism – I love you all dearly.

Thank you to Ryohki for being a good girl, for not deleting my thesis when you sat on the keyboard and for only scratching and biting me when the lack of attention really became too much.

Thank you to Professor Justin Paterson and Dr Hyunkook Lee for the publication opportunity.

Finally, a big thank you to my examiners Dr Sara McGuinness and Dr Paul Thompson for their absolutely invaluable feedback.

Contents

1 – Abstract

'The Phenomenal Rise of Periphonic Record Production' is a practice-based, musicological research project investigating the musicality of a non-front orientated approach to spatial music sound staging, posing the question 'How can non-front orientated sound stages for music be approached and structured?'

The thesis argues that with integration of periphony (height and surround) there will be a requisite change in the way we actively listen to recorded music, facilitating new approaches to sound staging and record production. Further, in taking an ecological, embodied approach to production, a periphonic sound stage provides more creative agency than that offered through stereophonic or surround sound productions, and that without a visual informing the auditory perception the additional sonic dimensions may be enhanced beyond what current approaches to production can afford.

The topics of study are explored through creative research practice and applied development of contemporary music production technique, drawing upon phenomenological method, and adopting practice as research and critical theory as research paradigms. The study constructs, collates and assesses spatial sound staging and production approaches for binaural encoded 3D audio arrangements and provides a framework for conceptualising and interpreting musical structure and lyrical narrative to spatial sonic schema using a non-front orientated approach to production.

The techniques constructed within the scope of this project address key issues pertaining to periphonic sound staging and production, offering

solution through a non-traditional, unique and democratic approach to spatial music production and creative research practice.

The study collates primary research though practice and corroborates this data through focus group sessions that explore the perceived efficacy of the staging constructs and a non-front orientated approach to production.

The work herein has been circulated through oral presentation at a variety of conferences, seminars and workshops over the last 6 years. Most recently, elements of Chapter 6 have been published and can be found in Chapter 13 of 'Perspectives on Music Production – 3D Audio' (Lord, 2021). The research presented in this thesis has also received citation in undergraduate, post-graduate and PhD level studies pertaining to spatial music production.

2 – Introduction

Multi-channel surround sound has existed in varying formats for three quarters of a century and for most of this time with undulating salience and favorability. Through all of its developments, surround sound has managed to survive this tumultuous vacillation of popularity versus purpose and today has a continued success, well-established in theatre, cinema, live performance DVD recordings and gaming audio, amongst other sound-to-picture applications. However, despite the achievements of researchers and the success garnered across the audio-visual industries, two-channel stereo continues to be the medium of choice for the delivery of recorded music (Torrick, 1998) (Farina et. al., 2001) (Wuttke, 2005) (Holman, 2008). This is due to the versatility of the format and a century of applied research and technological developments informing the scope of approaches to creative practice.

Stereo remains largely successful, as over time we have defined ways to utilise the relationship presented between the two channels to construct sonic projections of recorded ‘sound stages’. In a contemporary music production context, sound staging can be defined as the organisation of sound-sources within the boundaries outlining a schematised perceived performance environment (PPE) (Moylan, 1992) (Moylan, 2012) (Lacasse, 2005) (Dockwray & Moore, 2010) (Moylan, 2020). Dockwray and Moore define the boundaries of the perceived stereo performance environment as the ‘sound box’; “a four-dimensional virtual space within which sounds can be located” (see figure 1). This study from Dockwray and Moore defined sound source positioning across lateral placement within the stereo field; foreground and background placement due to volume and distortion; height according to sound vibration frequency; and time. They ultimately presented a taxonomy of common mix approaches; clustered, triangular, diagonal and dynamic (see figures 1-4). These approaches all centre on how the two-channels of stereophonic audio may be linked to create a schematised virtual performance environment, presented through a cohesive front-projected phantom image. However, when it comes to spatial audio, specifically multi-channel speaker-based surround sound, there was and still is no definitive approach to surround sound music staging - with most surround sound staging practice adopting a primarily front-orientated, stereophonic approach that often resulted in the surround channels being utilised for ambience, special FX and decorrelated sound source placement. This presents an approach far more suited to audio accompanying picture (allowing for the mapping of diegetic and non-diegetic audio between the front-rear channels) - an approach which is, by nature, much less viable for music record production where there is no secondary perception reinforcing or even influencing the surround placement.

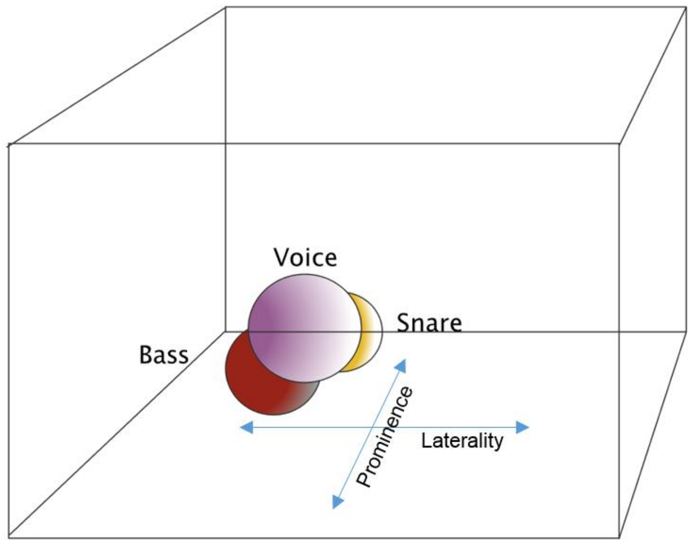


Figure 1: A schematic representation of Dockwray and Moore's (2010) 'Sound box' showing a clustered sound source placement within the perceived performance environment relative to the two dimensions; width (laterality) and foreground-background placement (prominence).

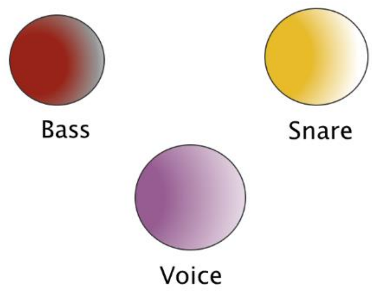


Figure 2: A schematic example of Dockwray and Moore's (2010) 'Triangular' mix taxonomy with a front-centred voice, as viewed from above (bird's eye).

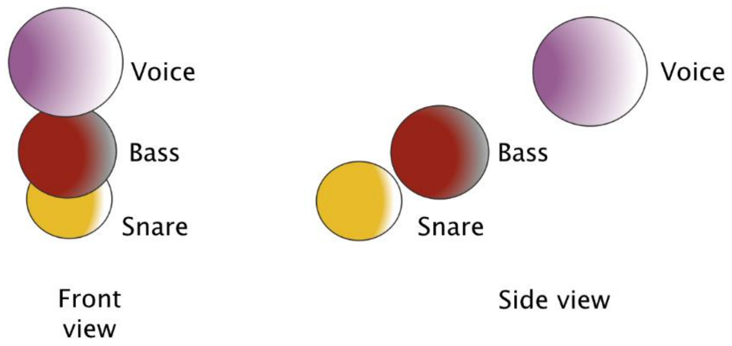


Figure 3: A schematic example of the 'diagonal' mix taxonomy, defined by Dockwray and Moore (2010).

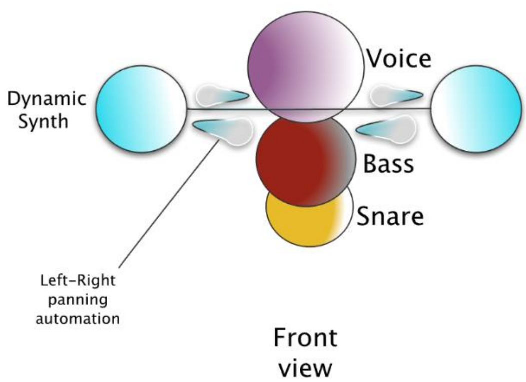


Figure 4: A schematic example of Dockwray and Moore's (2010) 'dynamic' mix taxonomy showing left-to-right-to-left panning movements framing a 'diagonal' mix.

The lack of spatial audio assimilation within record production is due to perceptually and technologically limited creative practice, format wars and the difficulties associated with multi-channel consumption (Guttenberg, 2009). It could be said that even the most creative recording producers were never fully able to exploit the surround medium, locking themselves into a rigid four corner reproduction and a front-reliant sound stage. Due to the cost and complexity of traditional multi-speaker systems, issues surrounding consumer delivery have long impacted the development and integration of surround sound record production practice. One of the most definitive contemporary delivery challenges has been the change in consumer listening behaviour moving away from speaker-based presentations in favour of headphone-based consumption; a more personal and mobile format (Torrick, 1998, pg.31) (Guttenberg, 2014; Holman, 2008, pg. 11; Marsden, 2015). This shift to predominantly headphone-based consumption is something that speaker-based surround sound had difficulty keeping up with, until most recently.

The last few years have seen a considerable boom in the development and popularity of Virtual Reality, Augmented Reality and immersive interactive experiences that has informed a resurgence in the use of binaural audio, particularly as a spatial audio delivery format. Binaural is a term used to define the phenomena of listening with two ears and it features heavily throughout this study. In an audio production context it refers to a legacy recording approach designed to capture-reproduce recorded audio with the localisation cues and spatial content as if perceived through two ears separated by a head. This is traditionally done in the recording stage using a dummy head / ears microphone, in-ear binaural microphones or a Jecklin disc. Binaural recording has long been known for inconsistency in perception across listeners and incompatibility with speaker-based reproduction, both of which have negatively affected its use and popularity over several decades. The incompatibility of binaural recording with speaker-based reproduction is due to the inherent differences presented between speaker-based and headphone-based modes of listening, and the way in which binaural filtering works to generate a spatial perception. When we consider these differing characteristics of listening medium and how binaural recordings are generated we can better understand why binaural audio has traditionally been overlooked by the commercial music market.

In speaker-based listening sound propagates from a distance and there is cross-talk between channels which allows the direct sound of all speakers and the acoustic reflections to be heard by both ears (see figure 5). Although much of this sound enters the auditory canal directly, some also reflects off of the ridges and curves of each pinna (the fleshy outer part of each ear) creating micro-delays before entering the canal (see figure 6) (Doctorlib.info, 2019). Each of our two pinna interact differently with the mix of incoming sound waves which generate spectral, phase, time and level differences between our two ears. These differences create interference patterns known as ‘Head-Related Transfer Function’ (HRTF) which provide our ears with the spectral information characteristic of the elevation of the sound source, and the level and timing information necessary in determining a sound source’s position relative to our bodies.

Speaker-based listening allows for our natural HRTFs to affirm that the sound source is propagating from the speakers within the room, and specifically which speaker is generating which sound. This presents the perception of an externalised sound stage with size and inter-linked laterality relative to the speaker placement.

However, in headphone-based listening the above modal characteristics do not apply; there is no cross-talk or acoustic interference and we cannot use our pinnae as the left and right speaker outputs are isolated and fixed to each respective ear (see figure 7). This creates an inability to localise source positions using natural inter-aural localisation cues, which in contrast to speaker-based listening results in a reproduction presenting as more internalised. Stereophonic headphone reproduction may also present an exaggerated laterality and less phantom image interlinking than when reproduced over speakers.

In the case of binaural audio, artificial or recorded HRTFs are used to synthesise the inter-aural spectral, phase, time and level differences pertaining to each pinna. This spatial information is imprinted upon the respective left-right audio streams, comprising a mix of direct and ambient sound capture perceived by two-ears. Binaural audio works best over headphones as the specific inter-aural differences can be delivered directly to each auditory canal without acoustic interference. This provides phantom inter-aural cues which result in an externalised perception of the reproduced sound sources and a perceived surround spatiality. When applied to speaker-based listening, the recorded binaural image distorts and the spatial effect is lost because of cross-talk, the acoustic interference of the listening environment and our own natural HRTFs of which we are using to listen. The perceived surround spatiality and the perception of elevation is impossible when the source audio is localised as emanating from the speakers. Although an exaggerated laterality may still be perceived and the image may be adapted and stabilised with cross-talk cancellation processing[1], the aforementioned factors prevent the artificial binaural phenomena from being properly experienced when reproduced on speakers.

Traditional speaker-based listening and production culture has consequently limited the use of binaural audio as a format for commercial production. However, the shift toward headphone-based listening – further supported by the resurgence in VR - affords an opportunity to explore the application of binaural audio in wider commercial contexts.

[1] Cross-talk cancellation is a process which applies filtering to the binaural signal before it reaches the speakers to better deliver each left-right signal to the respective ears of the listener. Phase and delay mechanisms are deployed to cancel out and reduce the cross-spill information from each channel signal. However, this may lead to a compromised and degraded reproduction with a loss in audio quality and dynamic range (Hamdan and Fazi, 2021).

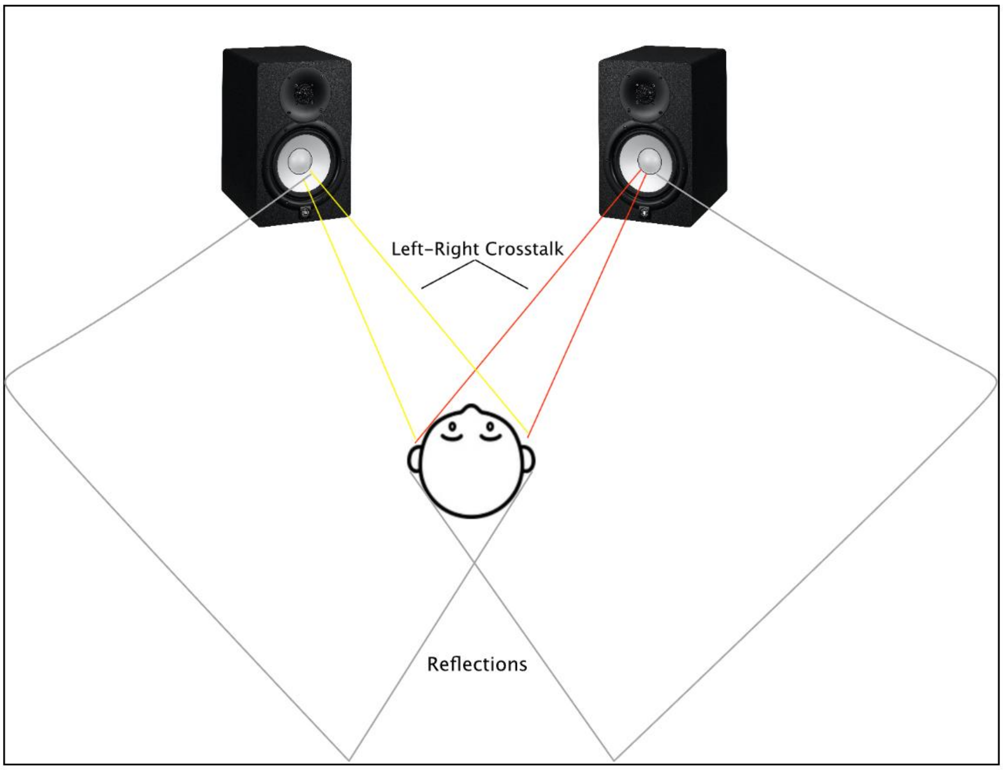


Figure 5: A schematic representation of cross-talk and acoustic interference associated with speaker-based listening.

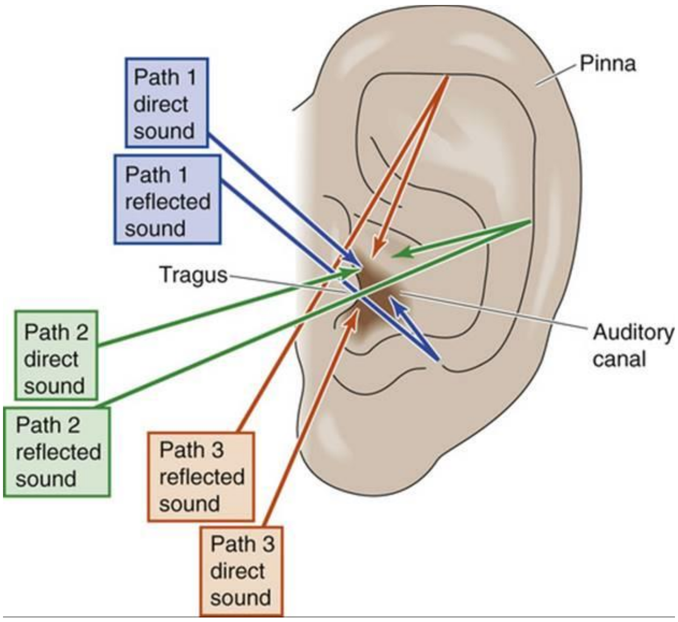


Figure 6: A representation of the pinna function in the detection of sound on the vertical plane using direct and reflected pathways (Doctorlib.info, 2019).

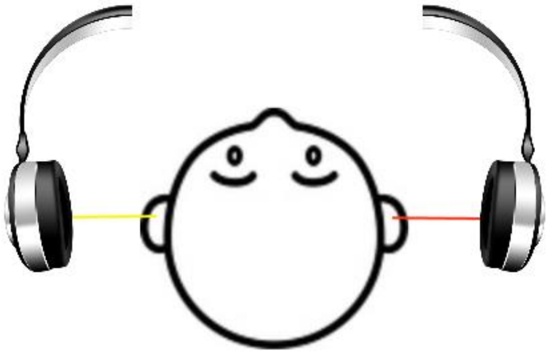


Figure 7: A schematic representation of the direct and isolated delivery of left and right channels associated with headphone-based listening.

In the context of this project, binaural refers to the headphone-based delivery format of the associated spatial audio. This use of binaural decoding for delivery applies these same localisation cues and spatial information to the sonic objects that comprise a large-scale surround sound stage in order for it to be reproduced over headphones whilst retaining its spatial characteristics. Importantly, it is through a headphone-based approach that this study circumnavigates a democratic resolution to the common problems associated with spatial music staging in an attempt to address the limited creativity in production approach that is deeply rooted in spatial audio's multi-channel surround sound history.

This resurgence in binaural audio has helped to bring immersive audio formats back into the public interest and has further necessitated the development of a number of pioneering tools for VR spatial audio production, such as the object-based binaural processing tool used within this project

(Dear VR Pro) and the ground-breaking standard known as MPEG-HD (3D audio) ISO/IEC 23008-3 - a codec for non-physical digital delivery of 3D audio in audio channel, audio object and higher order Ambisonics (HOA) based formats, also supporting binaural rendering. Ultimately, these developments and others, have made the production and delivery of spatial music much more democratic and accessible. However, though the technology and delivery methods for spatial audio have evolved, the limited creative approach to spatial music production practice remains much the same today: in its infancy.

Historically there is myriad literature exploring the creative ‘art’ of stereophonic mixing and record production as articulated and referenced through the works of Moylan, Lacasse, Gibson, Dockwray & Moore and Zagorski-Thomas, to name but a few (Moylan, 1992; Lacasse, 2000; Lacasse, 2005; Gibson, 2008; Dockwray & Moore, 2010; Moylan, 2012; Zagorski-Thomas, 2014, Zagorski-Thomas, 2018, Moylan, 2020; Lord, 2021). Yet in contrast to this abundance, the majority of literature pertaining to surround and binaural mix technique focus on constructing ‘sonic realism’ for sound-to-picture applications, with notable but sparing exploration of creative mix technique for audio-only music in surround or binaural mediums being confined to a short list of experimental albums spanning the later 20th and early 21st centuries.

To reiterate in summary those points made earlier, the entrenchment of surround sound within audio-visual application is due to past technology and delivery methods limiting the creative agency and specifying a particular mode of employment in securing audience reach. Firstly, the size and cost of these systems meant that until the low cost commercialisation of surround sound systems in the late 90’s there were few other places one could listen to multi-channel surround - aside from large-scale auditoriums such as cinemas, theatres and arenas and often these spaces fundamentally existed to provide an audio-visual experience, be it on-screen or on-stage performance. Secondly, the experiential limitations imposed through the visual performance itself and by the surround technology (a single-tiered, horizontal reproduction with limited output channel matrixing) meant that physically there was less accommodation for creativity, and subsequently creative research, in surround sound mix practice – panning and movement had to be horizontal and the audio had to reflect the associated visual to be perceptually convincing - if one could see the sound on screen it would have to be localised in the front.

However, with the integration of periphony (surround sound with captured or artificial height) there is a requisite change in the way we actively listen to reproduced and recorded audio and this change subsequently facilitates new approaches to sound staging and record production. The term periphony was first introduced by Gerzon in the 1970s to describe the phenomena of both vertical and horizontal sound reproduction around a listener, and though this study does not work specifically with loudspeaker configurations or native Ambisonic technology, the term periphony is used in much the same manner (Gerzon, 1973). The etymology of periphony comes from the Greek words ‘Peri’ - meaning ‘around or about’ and ‘Phone’ - meaning voice or sound. Therefore, periphony can be defined as literally meaning ‘sounds from around' and is used in this study to refer to surround or binaural audio with height.

The aim of this research project is to explore mix techniques that relate to spatial audio without a visual and that translate for binaural reproduction. Ultimately the research goal is to create the same type of non-realistic yet evocative staging techniques for binaural-decoded 3D audio that exist in popular music for stereo and in doing so contribute and expand the limited research in this field. In seeking to answer the primary research question: 'How can non-front orientated sound stages for music be approached and structured?' this study redefines our approach to spatial staging and proposes, through creative research and practical exploration, a presentation of key considerations offering a suggested approach to non-front orientated periphonic staging for binaural record production.

Why binaural record production?

A headphone-based periphonic sound stage may provide more creative agency than that offered through a stereophonic or single-tier surround sound production system, whilst fundamentally providing a democratic spatial music delivery method well suited to current consumer listening behaviour. This study explores how those additional dimensions afforded through a periphonic sound stage can be further exploited using binaural phenomena, proxemic theory (Hall, 1966), sonic embodiment and metaphor (Zagorski-Thomas, 2014, p.80) as a theoretical framework to enhance the representation and expression of the musical concepts presented through the staging.

The techniques discussed within this thesis mostly focus on exploiting binaural perceptual phenomena as an aesthetic enhancer for spatial music production, although it can be said that not all of the techniques defined herein are strictly reliant on a specific binaural synthesis tool (such as Dear VR Pro, used within this project) as a means of generating this perception. In many instances the same or a closely similar experience may also be garnered through our natural binaural perception when listening upon a domed, tiered multi-speaker system (such as Auro-3D 13.1). Thus, these techniques are considered *transferable* - defined in this instance as being system agnostic and therefore not reliant upon or governed by any specific tool or playback system. There are, however, some specific techniques that are designed to employ proxemic reaction through binaural sensation (Hall, 1966) (using socio-spatial association to influence Human emotion), and so are implicitly reliant on the binaural phenomena as presented over headphones in order to convey a more embodied experience, especially where the staging explores intimacy and in/externalisation.

Why a non-front orientated approach?

A cultural shift toward headphone-based listening has altered the way that mix engineers and listeners think about stereo, with a more internalised presentation becoming the norm. Engineers have been slower to respond to these cultural changes because the standard in professional practice for mixing requires a speaker-based approach that uses headphones only for referencing. With a much more democratised industry and with more people mixing from home, on headphones, it seems as though there is a cultural shift to headphones in both the consumption and also the construction of music records. This effects mix philosophy[2] and presents an opportunity for a psychoacoustic shift towards a more headphone-based approach to production.

Due to the additional dimensions afforded through a headphone-based spatial sound field it is possible to approach staging in this environment in a different manner than that which we approach stereo. A binaural sound field encompasses the listener, placing them inside the phantom-centre. The entire production is built upon a perceptual illusion, with the listener central to the illusion. This automatically affords an increased sense of depth, height and width reinforced through a shift in phantom-image projection. With stereo, as aforementioned, the illusory phantom image is projected in front of the listener rather than around them and so depth, height and width

[2] This is because people are now more used to listening with the internalisation presented through headphone listening.

enhancements need to be considered relative to the limitations presented through a 2D horizontal-vertical projection.

The shift in phantom-image projection and the additional perceived dimensions binaural-decoded 3D audio affords presents important questions concerning production approach; how does this new surround sound with height presentation affect mix practice? Does it correlate to or deviate from the sound box framework proposed by Dockwray and Moore? How could we use this newly available height within the context of musical sound staging? And why continue to structure front-respecting spatial sound stages when we have a much expanded and surrounding virtual performance area now available to us?

To approach 3D sound staging as we approach a stereo production would defy and deny the purpose of the format – to provide periphony, and in doing so, further stifle creative practice and application. The past several decades have seen surround sound production often approached with fundamentally stereo technique, and it only truly works for sound-to-picture reinforcement. If there is no front-orientated image or pre-requisite for representation of a live performance stage, then sound source placement remains undetermined and should be free to flourish in the spatial environment. To make optimal use of the periphonic sound field and present a cohesive use of the binaural phantom imaging, non-front orientated approaches to staging are a consideration and a prime candidate for research investigation.

Within the context of the primary staging question a number of subsequent questions arise, including but not limited to; ‘How could non-front orientated vocal staging be approached?’- this is a particular area of interest addressed within this project and a problem surround sound struggled to resolve (commonly reverting to a stereo approach using centre-mono placement of the lead vocal with a stereo or surround placement for the chorus/backing voices). Underpinning these and other lines of inquiry in praxis are questions pertaining to our musical, sonic and ecological spatial perceptions, how they may be reinterpreted through the musicological study of spatial record production, and explored and applied in praxis through the affordances of current spatial production technology; conceptualisation and delivery methods.

Following this introduction, the thesis is broken down into the following chapters;

The Literature Review (Chapter 3) presents critical discourse on key theoretical concepts underpinning the research as practice framework. One primary theory discussed is that of *proxemics*, as mentioned above. Proxemics is considered a ‘non-verbal communication’ system – much like music, and is a branch of knowledge that studies the human use of space, how it influences human emotion and how we arrange objects and beings in relation to space. This could be viewed as an integral theory in the study of spatial music production, especially a study such as this that examines the possibilities in arranging sonic objects within space as a means of communicating a concept (message, thought, feeling, environment etc.) and presenting cultural information through a musical arrangement. The literature review also critically highlights *memetic theory*, which ties into this project in much the same way – acting as a means for presenting cultural information through *evolutionary* models of communicative transfer. *Memetic theory* allows us to consider certain concepts as resonating within a particular culture and that, like language, it allows these concepts (or percepts in this case) to suggest shared forms of interpretation. This promotes the notion that staging concepts can develop as a characteristic of thought, informed through the cultural signifiers suggested by the sonic context and sonic content of the audio sources. Though the memetic staging ideas may propagate successfully, this does not necessarily imply any particular percept is factual but promotes the notion of memes existing as an advantageous means of resolving a particular staging query or promoting a particular cultural or musical concept because of their usefulness in suggesting metaphorical structures. This applies to the practical framework in the sense that both the sonic context and content of any given musical production will be the initial driving factor behind the conceptualisation of the musical narrative (as later chapters shall reveal), which may then be reconceptualised, developed and applied in practice through a series of schematised, metaphorical musical representations of any given concept's tangible form – establishing a shared ground for interpretation through the representation of a cultural truth. Zagorski-Thomas refers to these representative musical 'meme' structures as *sonic cartoons*, which along with sonic embodiment and musical cognition are critically discussed within this chapter as being integral in the conceptualisation of non-front orientated spatial music stages. Holistically, this chapter provides an ecological understanding of sonic perception relative to spatial phenomena that can be used in conjunction with the aforementioned communicative theories and classic stereophonic sound staging studies as a means for defining a framework governing praxis.

Chapter 4: Methodology – Following the Literature Review, this chapter outlines in detail the methodologies used in conducting the following body of practice-based research and explores why both qualitative and quantitative elements are used in the collection and triangulation of phenomenological and experiential research data. Creative practice-as-research is discussed as the primary methodology driving the investigation and how it offers an action-based approach that works in tandem with phenomenology in providing a means for exploring the theoretical framework through applied practice and the examination of the subjective human experience. The chapter outlines how using this mixed method approach allows for practice-based data to be collated, examined and presented through case study, and for individual subjective experience to be collated through focus group and interpreted, quantified and triangulated through thematic and statistical analysis. Further to this, the chapter details how this combination of strategies provides a means to assess the commonalities across group experience and evaluate the intention of the practice against the result of the practice in a more objective manner. Holistically, this chapter presents critical discourse as to why these methodological systems have been selected and how they were specifically deployed using a pragmatic approach that is necessary to the thorough, objective investigation of an interdisciplinary musicological project exploring creative audio production technology and the human experience.

Chapter 5: Preliminary research as practice – In the early stages of this project there were few known tools available that allowed for periphonic spatialisation within a DAW, and even fewer that accommodated a headphone-based binaural reproduction for delivery. This impacted the commencement of the primary investigation as a suitable toolkit for a headphone-based spatial reproduction had to first be sought to ensure an appropriate approach to practice could be designed. This chapter problematizes the historical approaches to spatial music making and explores the opportunities presented through a more democratic approach, further justifying this as a key element not only of this research but of any consumable spatial music production project. This was an important preliminary investigation that would dictate the progression of the entire research framework to follow and subsequently the development of practice, herein. This section critically examines the affordances and limitations of various spatial audio conventions; traditional & contemporary speaker-based multichannel surround sound formats; the historic and contemporary use of binaural audio; Ambisonic object-based approaches; and head-tracking systems, critically assessing their suitability in application to this project through a combination of historical literature analysis, tacit knowledge, observation and practical experimentation. This section outlines and critically assesses the result of the preliminary practical experimentation and evaluates the affordances and limitations presented through speaker-based and headphone-based approaches to practice available at the time, and how these may affect workflow, operability, creative agency and delivery. The chapter ultimately concludes by defining the toolkit required in undertaking the investigation discussed in the body of work to follow and presents a comprehensive and empirical reasoning through case study analysis as to why the Dear VR Pro VST was selected as the most appropriate candidate relative to the project prerequisites.

Chapter 6: Redefining the Spatial Stage – This chapter is considered the fundamental body of work and applies the theoretical framework as constructed within the literature review parallel to the defined methodological approach and the knowledge garnered through the preliminary research as practice to explore the primary and secondary research questions in practice through the medium of binaural encoding. This body of work revisits the primary line of inquiry in addressing the question ‘How can non-front orientated sound stages for music be approached and structured?’ The investigation is compartmentalised on a track-by-track basis and is presented through a series of case study pieces that focus on pertinent areas of interest and the historical challenges these areas present for spatial music making, such as; staging the voice, spatial instrumental arrangement and the use of *sonic cartoons* and metaphor within the spatial domain. Each case study presents the research as practice data as an explorative and evaluative discourse of the production approach. The efficacy of these techniques is assessed and triangulated against a small set of participant data which is discussed in Chapter 7. The data is attained through a series of focus group sessions designed to examine listeners’ perception and their experience of the non-front orientated spatial productions against equivalent fronted stereophonic versions. This triangulation of first and third person data is important in exploring the efficacy of the memetic systems applied within the staging production as a means of communicating the intended musical concepts, providing an insight as to how the binaural productions are perceived across a varying set of users. This is an important consideration when using a spatial production system that relies on generalised HRTF algorithms to encode the binaural audio and as such this triangulation of first and third person experience also acts as a quality control and a means to help further justify and support the case studies data.

The results of the research project are presented within each of the chapter investigations and are reiterated in Chapter 8, which concludes the thesis and reiterates all of the research findings.

2.1 – Related Publications of Work

Lord, J. (2021) Chapter 13 - Redefining the spatial stage: non-front orientated approaches to periphonic sound staging for binaural reproduction. In: Lee, H. and Paterson, J., (eds.) *Perspectives on Music Production - 3D Audio*, London, England, Routledge, pgs. 256-273.

2.2 – Related Presentations of Work

1st - 3rd December 2017 – *Art of Record Production – Mono, Stereo, Multi.* Kunligamusikhogskolan. Stockholm, Sweden.

8th December 2017 – *CPS2: 3D Audio Masterclass* – University of West London, England, UK.

26th – 27th June 2018 – *Sounds in Space* – University of Derby, England, UK.

3rd – 5th September 2018 – *Art of Record Production – Crosstown Traffic: Popular Music Theory and Practice*. Huddersfield University, England, UK.

23rd October 2019 – *3D Audio: Periphonic-Binaural Masterclass* – University of West London, England, UK.

29th March 2019 – *CPS2: 3D Audio Masterclass* – University of West London, England, UK.

17th – 19th May 2019 – *Art of Record Production – In C; Creation, Connectivity, Collaboration and Controllers.* Berklee College of Music, Boston, Massachusetts, USA.

16th – 19th October 2019 – 147th *Audio Engineering Society Convention – ISO2 Panel: Music Production in Immersive Formats: Alternate Perspectives.* Jacob K. Javits Convention Centre, New York City, New York, USA.

5th-7th December 2019 – Innovation in Music Conference – University of West London, England, UK.

21st July 2020 – IASPM ‘London Calling’ International Virtual Conference – Online Video Presentation.

3 – Literature Review

"…A fully fledged musicology, whether it is examining record production or any other musical practice, needs to interrogate how the value systems that underpin any aesthetics are constructed." (Zagorski-Thomas, 2014, pg. 64).

As stated by Zagorski-Thomas in the opening quote, it is fundamental that the research examines and interrogates the core systems that underpin the key concepts of study involved within any musicological project, and as such, this chapter presents a critical review of pertinent interdisciplinary literature that is necessary in establishing a theoretical framework for governing the research practice. This chapter acts as an overture to the associated topics and builds an understanding of the musicological and theoretical backbone of the project necessary in the thorough problematisation, contextualisation and critical exploration of the investigative subject matter.

Firstly, part 1 of this chapter examines the key concepts underpinning traditional front-orientated stereophonic approaches to sound staging through the critical appraisal of classic studies from musicologists such as Dockwray and Moore, Moylan, Gibson, Lacasse and Zagorski-Thomas. These studies are then later recontextualised for spatial audio through critical evaluation of the phenomena and physiological function of binaural listening and spatial audio technologies. Reflecting upon these stereo sound staging studies within the context of spatial record production provides a rationale for this research project and justifies the continued exploration and the further development of staging approaches in practice for the specific study of non-front orientated approaches to sound staging for spatial music production.

Following the contextualisation of the research problem, part two of the literature review interrogates the values underpinning the phenomena of human perception through the critical exploration of binaural listening, ecological perception and sonic embodiment theories. This understanding of the human perception system coupled with a critical evaluation of meme theory and proxemics, as discussed in part three, inform a holistic understanding of how human beings perceive, interact with and internalise reality through their [sonic] associations with the world around them. This combination of interdisciplinary theories presents a socio-ecological approach to record production that is born from the phenomena of the human experience, thus providing a robust theoretical basis for guiding the investigative practice to follow. This socio-ecological framework informs practice through the applied development and construction of representational aesthetics and culturally significant schema as a means of creatively constructing non-front orientated spatial sound stages. Before we embark on this journey into creative research practice we must first fully define the underpinning research problem and contextualise its importance within the scope of this research project. This is addressed in the proceeding section of this chapter and through exploration of the following questions; what, if anything, is wrong with a fronted stereophonic approach? And consequentially, why explore constructed non-front orientated spatial sound stages?

3.1 - Part 1 - Traditional Approaches to Record Production Sound Staging

We can open discourse on these questions by firstly reflecting on the 1960s transition from mono to stereo recording and reproduction, where technological and ideological parallels had presented a similar question in regards to the development of innovative practices. Why make the transition from monophonic to stereophonic production practice?

The answer to this question, in short, is that Alan Blumlein created the stereo format to better suit the physiology of listening. Stereo works particularly well as the two channels align closer to our two-eared way of listening than mono does. The relationship offered through a two-channel stereo speaker system affords a more ‘realistic’ presentation of an audio performance capture than a single channel mono speaker, effectively adding an additional dimension to the presentation. This affords us the opportunity to arrange sound sources between the two speaker channels as a means to create a two-dimensional, schematic representation of a musical performance and it is because of these affordances that it has also seen over 75 years of continued practice, research and development. Monophonic audio by comparison is singularly dimensional, only affording foreground-background placement as achieved through volume adjustment, spectral manipulation and spatial processing, such as reverb and delay.

In their publication ‘Configuring the Soundbox’ (2010) Dockwray and Moore refer to a multi-dimensional stereo sound staging framework based around the concept of the ‘sound box’. They define the ‘soundbox’ as a four-dimensional virtual space, or perceived performance environment (PPE), within which sounds can be placed and located. These four dimensions are as follows; “lateral placement within the stereo field; foreground and background placement due to volume and distortion; height according to sound vibration frequency; and time” (Dockwray and Moore, 2010). This multi-dimensional space exists purely on the record and provides a virtual performance to be perceived in the mind of the listener. These dimensions define the stereophonic image and are typically imparted through judicious use of basic music production processes. A matrix of some of these functions relative to the dimension they impart has been structured below for visual reference.

Lateral placement *– affecting perceived width*	***Foreground and background placement*** *- affecting perceived depth*	***Height*** *- affecting perceived verticality*	***Time*** *- affecting perceived spatiality, width and depth.*
L-C-R Panning	Volume	Spectral content of source	Time / Sample Delay
Stereophonic source recording techniques	Distortion	Equalisation adjustment	HAAS Effect (Volume/Time Delay & Panning)
M/S and Width Processing	Equalisation adjustment		Spatial Effects: True and artificial Reverberation and Echo Delay

Table 1 – The four dimensions of stereo imaging as defined by Dockwray and Moore (2010), and basic processes of affect.

Dockwray and Moore consider these four dimensions as important factors in defining a coherent and well schematised stereophonic sound stage, as without these dimensions of placement and localisation there would be little spatial separation between sources and a lack of localisation information within the mix - ultimately resulting in the production lacking the musicality and resultant illusory performance space comprising a quality stereo record. Most significantly, these dimensions construct and further enhance the relationship between the two stereophonic channels and as a result create the necessary multi-dimensions of space and time within the front-projected performance we perceive when listening to stereophonic music from within the listening position (LP) of the stereo triangle (see figure 8).

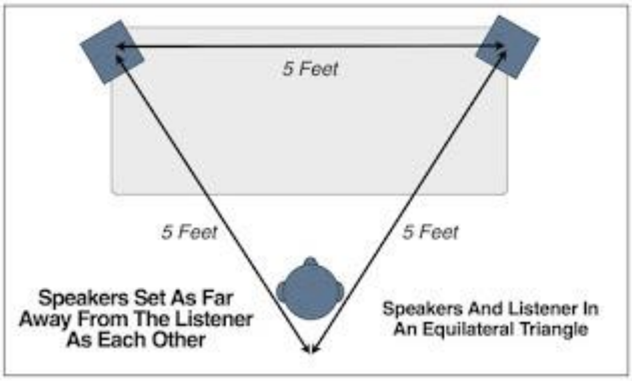


Figure 8: Equilateral triangle of a stereo speaker set up (Oswinski, 2011).

It is important to note that all of these aspects are illusory with the exception of width, which is not only determined through the lateral sound source placement afforded through stereophonic recording techniques, pan-pot manipulation and width enhancement but is fundamentally generated relative to the physical placement relationship between the two speakers of any given stereo reproduction system. Moylan (2009) states that the dimension of width is defined through the furthest left and furthest right sound (lateral localisation) and that depth is defined similarly through the most distant and closest sounds – thus creating the *two* dimensions of stereophonic sound source placement. This contradicts the four dimensions of sound source placement as outlined above by Dockwray and Moore. When referring to height in both two-channel and surround sound productions Moylan defines that sound sources are not placed in actual unique elevations, as such elevations cannot be physically reproduced by speakers placed on the median plane. Though Moylan does acknowledge that there is some conceptualisation of elevation relative to pitch/frequency level, he attributes it to limited perceptual phenomena generated through the unique head-related transfer functions[4] of any particular listener and states it as being a conceptualisation of vertical placement of pitch (representing register) and not an element of actual spatial localisation. In 'The Art of Mixing' (2018), Gibson presents a similar conceptualisation to that of Dockwray and Moore, aligning pitch/frequency to elevation. However, Moylan (2012) reiterates that these are concepts more aligned with 'pitch density' and 'timbral balance'

[4] Head-related transfer function is defined in the next part of the literature review covering binaural listening and perceptual phenomena, for now we can think of it simply as the way in which the head and ears impact our means of aural perception.

than they are with spatial locations and sound source relationships. As such, in his 2012 journal article ‘Considering Space in Recorded Music’, Moylan opts to not include elevation as an element of spatial dimension in his stereophonic and surround sound music recording explorations.

When applied to spatial audio record productions, one could argue that Dockwray and Moore’s ‘sound box’ framework and the four dimensions of placement and localisation remain unchanged in principle – all elements are relevant to the multi-dimensional schematisation of a spatial perceived performance environment (PPE). However, it should be noted that on a spatial multi-speaker system with a physical overhead reproduction, elevation is no longer a conceptual illusion achieved only through higher spectral content, as elevation, like width, may now also be physically generated due to the relative vertical and overhead placement of speakers. Often the scope of sound source placement is limited by the measured physicalities of the speaker array configuration. This is especially true of channel-based 3D audio systems, such as Auro 3D, for example. The technical and characteristic differences between channel and object-based audio systems are outlined and further discussed in more detail in the ‘Preliminary Research as Practice’ chapter further on in this thesis.

In the cases of all channel-based audio, object width and height placement are determined relative to the speaker configuration, and therefore, affordances for source positioning are secondarily dependent on the phantom imagining relationships presented between these speaker positions, where typical panning law applies. Panning law refers to the matrixing of a given signal’s amplitude between two or more channels in order to create the illusion of source position. For example, in two-channel stereo a centrally placed source would equate a 50:50 split of signal amplitude between the left and right speakers; at 25L:75R the source would appear to the right; 75L:25R and it would appear to the left. These changes in amplitude define the lateral sound source positon within the stereo image. Historically, the phantom imaging across traditional surround sound systems has been front-oriented due to the stereo configuration of the front 2 or 3 channels (Left, *Centre*, Right)[5]. These present a stable frontal phantom image that can easily be mixed down for both stereo and mono systems. The supplementary surrounding speakers are often too decorrelated due to their wide, offset rear positioning for any meaningful phantom image generation outside of this frontal stereophonic triangle. These technological limitations, coupled with the cost and size of the array configurations have made it quite difficult to creatively stage and reproduce surround sound music with commercial efficacy without defaulting to a fronted stereophonic or 'horseshoe' approach with supplementary and often erroneous rear source placements that are inevitably folded to the front when mixed down to stereo for mass consumption.

In its more contemporary 3D audio format, surround sound can now physically reproduce the dimension of height through additional layers of overhead speakers. According to spatial audio researchers such as Dr Hyunkook Lee these additional layers afford the presentation of vertical phantom imaging, whereby the signals can now be matrixed both horizontally

[5] The three channel LCR configuration was originally a cinema format to allow for central placement of dialogue.

and vertically between speakers across tiers providing more coherent surround coverage and subsequently better reinforcing the spatial surround reproduction (2017, pgs. 733 - 738). The addition of these reproductive layers suggests mixes could be presented in more of a ‘domed’ or ‘box’ shape around the listener, rather than the aforementioned ‘horseshoe’ that is typically associated with surround mixing.

These developments recall and further reinforce the underpinning questions this thesis presents regarding future mix practice; how does this new surround sound with height presentation affect mix practice? Does it correlate to or deviate from the sound box framework proposed by Dockwray and Moore? Why continue to structure front-respecting spatial sound stages when we have a much expanded and surrounding virtual performance area now available to us? And how could we use this newly available height within the context of musical sound staging?

In order to understand how future mix practice may be affected, we must first define the current paradigm of stereophonic mixing practices. In doing so we turn again to the same 2010 publication where Dockwray and Moore go on to define a set of mixing taxonomies pertaining to stereophonic staging practice; clustered, triangular, diagonal and dynamic. These have been tabled below, alongside Dockwray and Moore’s respective identifications.

Clustered	A narrow stereo image created by the central placement of the key sound sources within the sound-box.
Triangular	A triangular configuration between the vocals, snare and bass.
Diagonal	The normative layout providing the paradigm for subsequent record production. This mix identifies the vocals, bass and snare as being on a slight diagonal line in a linear configuration (relative to the vertical axis), with other instruments placed to either side.
Dynamic	Describes tracks where there is some level of movement within the sound-box; lateral and linear.

Table 2 – A tabled reference to Dockwray and Moore's (2010) Taxonomy of Mixes

In this instance we shall focus more on the latter two mixing approaches; diagonal and dynamic, as these together present the most common approach adopted in contemporary mix practice and may provide a basis for further development through spatial recontextualisation.

Dockwray and Moore (2010) describe the diagonal mix as "the normative layout that provides the paradigm for subsequent record production". Through their survey of recordings from the late 1960's and 1970's they identify the shape based on the vocals, bass and snare being positioned on a diagonal line in a linear configuration heading downward and backward, or where the vocals and bass are arranged perpendicular to one another, with other instruments placed to either side of these configurations. This practice emerged through the late 1960's and 1970's and was driven by the separation afforded by multitrack recording and the use of continuous rotary pan pots.[6] This 'accepted wisdom' of the normative stereo mix approach can be seen in application across contemporary industry practice and can be found outlined in a number of practitioner journals, such as Sound on Sound (White, 2000), as well as noted within other academic studies; "typically the bass, snare, kick drum, and vocals go to centre; keyboards and guitars can be panned left and right" (Bartlett and Bartlett 2002, p. 289).

The dynamic stereo mix defines a kinetic mix approach and can be created through the use of pan pot devices where a sound source moves laterally; it can correspondingly be created through movement in depth, where a sound becomes softer or changes its reverberation level or treatment. Sometimes sound sources are repositioned in separate takes (chops) and are then panned on separate tracks or composited in the final mix (Dockwray and Moore, 2010).

Could these mixing taxonomies be expanded upon or redesigned within the context of 3D audio as a means of constructing non-front orientated sound stages? Interestingly, in the context of this project, the dynamic mix approach seems the most adaptable and transferrable at this point, offering the possibilities of fully surrounding kinetic sound stages and presenting

[6] As opposed to the 'all or nothing' triangular approach which was mostly confined to the recordings of the 1960s when panning involved the use of LCR switches.

opportunities for lateral, vertical, linear and diagonal movements through a non-front orientated spatial phantom image. The idea of the diagonal mix could apply in the context of vertical or horizontal diagonal placements as a means of utilising the increased PPE, however whether this will apply to the vocal – bass – snare configuration as outlined by Dockwray and Moore (2010) is yet to be seen and is something this study addresses later through primary research practice. For now, the mixing taxonomies and four dimensions of placement and localisation defined within the ‘Sound box’ framework present an adaptable starting point for exploration in practice.

3.2 - Part 2 - Lend me your ears: An overview of binaural phenomena and the embodied human perception

In part one I outlined the four dimensions of sound source placement as defined by Dockwray and Moore and discussed how these may apply to mono, stereo and surround sound staging practices, further noting which elements are illusory and which may be affected by the physicalities of speaker placements. In the context of headphone-based spatial audio we determined that all dimensions must be illusory as there is no reproductive height layer as with contemporary 3D audio sound systems and so, as a result of the binauralisation process and medium of delivery both lateral and vertical placement are therefore not necessarily directly dependent on the position of each headphone speaker. This section aims to outline and evaluate the reasoning for this phenomena by examining the physiology of binaural listening and the principles that underpin the function of binaural audio technologies. The latter part of this section delves deeper into the systems that underpin our human perception and begins to introduce demonstrable elements of sonic embodiment before part three explores its importance and the role embodiment theory plays in taking an ecological approach to spatial music production.

The way that we perceive sound is imperative to understanding how to convincingly reproduce recorded audio. Understanding the physiology and function of the human auditory system is fundamental in the development of most audio technologies but particularly spatial audio technologies, such as binaural emulation, which is designed to synthesise the phenomena of a human's binaural perception; principally tricking the ear and brain into believing a virtualised aural presentation as being (almost) 'real'. A typical example of this virtual binaurality would be QSound Labs' audio barbershop experience, which can be accessed on YouTube or via David Webb's website (LovelyVirus, n.d) (Webb, 2008). This video presents a great introduction to what binaural audio is capable of presenting in terms of experiential audio illusion but though we feel it working, the video does not explain how it works to generate this sense of aural perception and spatiality, nor does it offer an insight into how this same binaural perceptual experience could be applied to music –which is much less relatable in terms of tacit and embodied world-experience (more detail on embodiment is presented later). A haircutting experience is a relatively consistent and commonplace human experience; most of us can relate to sitting in a salon chair at one time or another but music has many facets to its presentation and our consumable experience of it. Therefore, this part of the literature review is about critically defining how binaural listening works (or doesn't in some cases and why?) and how we embody our perceived experiences in real-world associations. In unpicking this understanding we will further aid in building upon the theoretical framework for research practice by generating an approach to applying binaural presentation within a musical context.

This sub-section not only details the physiology and function of the binaural auditory system and related phenomena but also critically evaluates pre-existing stereo production technique in the context of binaural listening as a means to assess whether stereophonic techniques could be adapted to serve a purpose within this project's practical research framework. As far as current research shows, at the time of writing there were no pre-defined 3D audio staging production techniques specifically relative to this project's trajectory. Therefore, much of the research herein will involve adapting stereo paradigms of technique, or at least the philosophy of it, for development in application to the periphonic-binaural model that this research project explores. Given the expanded sound field afforded through contemporary spatial audio technologies, the current paradigm of stereo staging and production techniques outlined previously may not directly apply as is. Therefore, rather than build on a forced application in attempt to make-fit pre-existing technique borrowed from another format - as has been the case with surround sound for so many years - this project seeks to invent an innovative method based off past-format understanding and philosophical exploration of the 'problems' presented in approaching a new mode of production that pre-existing approaches can functionally not accommodate to resolve.

As outlined in detail later in this thesis, exploiting binaural phenomena can be used to an advantage. However, there is a process of negotiation, as in some instances the binaural effect may not always work as intended, though once discrepancies can be defined these areas of issue can then be avoided or reconciled.

What is binaural hearing and how does it work?

The term 'binaural' refers to how we as human beings naturally perceive sound with two ears separated by the head. We naturally hear sounds from all around us but rely on a number of different auditory clues, involving inter-aural level and time differences and shifts in frequency content and phase, in order to deduce the direction from which a sound source is arriving (Howard & Angus, 2006, pg. 97). This binaural phenomena experienced through the physiological construction of our auditory system can be reproduced using audio technologies that use signal processing to synthesise the effects of this physiology and the inter-aural cues it generates, and over time these technologies have become better at doing such.
It is these measurable differences and their interactions that are utilised and employed alongside other things like spectral changes, reflections and dynamics for a more convincing binaural sound field placement (Howard & Angus, 2006, pg. 98-100). These are the fundamental principles of binaural human hearing and though each phenomenon individually has its own term they are collectively known as Head Related Transfer Functions (HRTF). Physics-governed laws, such as the law of the first wavefront, refraction,

reflection and those regarding RT60 are invaluable in creating a realistic sonic environment, as spatial hearing relies upon reflection and ambience, not only direct sound, to determine a sound source's location and the intrinsic character of the acoustic environment. As Zagorski-Thomas states "…timbre and ambience are inextricably entwined. We never hear sound without ambience…" (2014, pg.65).

The pinna (the outer ear flap) itself plays an important role in localising sound but it is that we have two ears located either side of and separated by the head, that has the most acoustic effect on how we localise sound binaurally. As mentioned briefly above, this is because there are two effects (due to the separation of our ears) on the sound wave: Interaural Time Difference and Interaural Intensity Difference (ITD and IID, respectively) (Howard & Angus, 2006, pg. 97). It was John Strutt, better known as Lord Rayleigh, a pioneer in spatial hearing research, who developed Duplex Theory and discovered the azimuth sound cues (University of California, 2011).

This is not to say, however, that the role of the pinnae is less important or of no consideration. The fleshy outer part of the ear itself acts similarly to the dish of a satellite/radar, collecting and directing the sound into the ear canal but it is the convolutions and ridges featured on this outer part of the ear that further impacts the spatial information of a sound. These ridges and convolutions are unique to an individual and aid to imprint spatial information through minute temporal changes in frequency arrival (Cheng & Wakefield, [n.d], pg. 6; Howard & Angus, 2006, pg. 103). These miniscule spectral delays create consciously unperceivable phasing that, when correlated with the unreflected direct sound at the ear canal entrance, aids the brain in deciphering the encoded spatial information of a sound source more accurately than the more conspicuous interaural differences could afford alone (Everest & Pohlman, 2009, pg.56). Through this filtering the pinnae helps the brain to interpret the localisation of sounds relative to all 3 azimuths (linear (depth), lateral (width) and vertical (elevation)). Whereas without the pinnae we could possibly still distinguish a general hard-left and hard-right placement on the lateral horizontal plane due to the separation of the ears by the head but we would find it impossible to distinguish vertical and front-back placements, both on and off-axis and oblique (Everest & Pohlman, 2009, pgs.41, 56-59). This does reinforce the notion that a good quality set of artificial HRTF's employed within the spatial production and binaural decoding processes would be beneficial in ensuring a less distorted height and front-back perception. Where possible the listeners own HRTF's would be the most appropriate option for binaural playback, offering the most accurate localisation. Though a custom set of HRTFs would no doubt aid the listeners own personal experience and localisation accuracy, in the real world this is just not feasible without the means to measure and input such data into the binaural playback system. Therefore, at the time of writing, customised HRTFs are ultimately less commercially implementable without the binaural music being presented through an app or another 'smart' and interactive medium; something that has the ability to measure and input a digital print of the pinnae and generate the user's set of HRTFs in real-time during playback. This process then becomes even more complex, and perhaps even impossible, when you have multiple users listening from the same audio output stream, i.e. splitting a headphone signal. Therefore, it could be concluded that though in theory customised HRTFs may be useful, in practice they are currently beyond the scope of this project.

Having defined the concept of HRTF and determined the importance of them and the outer ear in effecting binaural listening, we can now move on to unpicking the two interaural cues mentioned before; Interaural Intensity Difference and Interaural Time Difference. Interaural Intensity Difference is established through differing levels of sound intensity at each ear due to the scattering and shadowing effect of the head between them.

In the 1907 paper titled ‘On our perception of sound direction’, Lord Rayleigh states “the discrimination between right and left is usually supposed to be explicable by the greater intensity of sensation experienced by the ear which lies nearer to the sound… also heard very badly with the averted ear” and that when the pitch is high this explanation is correct, without doubt, and that high frequency localisation is thus easily explained using ‘Intensity Theory’ (L. Rayleigh, 1907, pg.297-298). The levels at each ear are equal when the sound source is on the ‘Z’ plane (directly in front, above, behind or below the listener), however if the sound source moves away from the Z plane, to the left or to the right (‘Y’ plane), then the sound level progressively reduces in the ear that is furthest away from the sound source and increases in the one ear closest (Howard & Angus, 2006, pg. 101; Everest & Pohlman, 2009, pgs.56-59). Rayleigh examples this initially through his 1876 studies using pure tones and further through detailing a number of experiments whereby more complex harmonic ‘real-world’ sounds are used. The sound of a hiss, whistle or running water situated either left or right of the listener is played whilst the ear closest to the source is closed; the comparatively feeble sound can then be heard in the averted ear (Rayleigh, 1907, pgs. 297-298). This physiological function allows us to infer stereo panning law and reflects how we can achieve the perception of lateral placement of sound sources within both the stereo and binaural domains using stereo recording techniques and pan-pot devices, as previously discussed in part one of this literature review. These studies also suggest that intensity differences are frequency dependent – this is discussed in more detail in the paragraphs to follow.

In the same study Rayleigh further notes that moving the head was not necessary in order to deduce the left-right localisation of pure tones but that with pure tones front-back localisation differentiation could not be made and that where a sound is positioned obliquely (at height) further confusions are presented. That being said, what we also deduce is that front-back confusions could be resolved with head movements (Rayleigh, 1907, pgs. 297). This portion of Rayleigh's studies presents both a case for and also against the use of head-tracking in this project, as to allow for head-movements could act to aid any localisation ambiguities presented within the binauralisation. However, this project will not be working with pure tones but with complex harmonic musical content and proposes that our pre-conditioned stereophonic way of listening to music will have to and can adapt to a binauralised spatial format. Furthermore musical immersion and enhanced experience do not necessitate realistic localisation accuracy when the audio is without a visual informing the perception. In this particular project head-tracking is seen as unnecessary and possibly a negative impact to data collection, as inclusion of head-tracking capabilities could distract the listener from the music and the production aesthetic in question. This is turn may possibly present too many variables and inconsistencies across each listener's experience and therefore impact the reliability of their experiential data. Therefore, to summarise, on the one-hand head-tracking would help with increasing localisation accuracy but on the other hand it may act to distract from or change the narrative of the content itself with the listeners being more attentive to the head-tracked medium than to the musical content it conveys. To circumnavigate this issue I theorise that if the head cannot move, then the sound stage should. This movement of sonic objects around the listener, rather than the listener's head moving in relation to the sonic objects, may serve to reinforce the cue changes required for a more distinct localisation. This notion of imparting movement within the mix to help with sound source localisation further reinforces the idea of using Dockwray and Moore's (2010) dynamic mixing taxonomy (as discussed in the previous section) as an adaptable means of initially bridging practice between stereo and periphonic-binaural domains.

Focussing back in on Lord Rayleigh's research, he also shows in his experiments that if a hand is held up a little distance from the averted ear then the feeble sound volume is increased. This is due to the scattered sonic content reflecting from the hand into the listener's ear, the effect remains conspicuous even when the arm is extended to full length and Rayleigh states that a larger reflector would be more effective (L. Rayleigh, 1907, pg.298). However, an object cannot shade or scatter sound significantly enough until it is at least two thirds the size of a wavelength; the scattering will begin an octave below the tonic frequency and this shows that there will be a minimum frequency below which the effect of IID is less useful for localisation. Rayleigh (1907, pg.298) states that there is nothing surprising about the observation that low frequency sounds can be heard nearly as well in the furthest ear as in the closest, as when a wavelength amounts several feet (larger than the circumference of the head) it is not to be expected that such a sound, emanating from a distance, could be limited to one side of the head alone. This is due to the wavelength being too large to refract from the head and so instead it bends around the obstacle with ease (see figure 9).

Studies show us that the approximate minimum frequency for IID to be effective is around 637Hz. This frequency is located within the middle range, therefore showing that interaural intensity difference is a cue for direction at high-mid and high frequencies, as opposed to Interaural Time Difference which works best as a directional cue at low frequencies. The cross-over of the two cues starts at around 700Hz and would continue up until the 4th harmonic of this frequency: 2.8 kHz. It is the range between these **two** frequencies that proves more difficult for humans to resolve direction (Howard & Angus, 2006, pg. 102-103).

This difficult area falls into the middle section of the human vocal range, however as the presence lies around 3-5kHz and that we as human beings are greatly attuned to the timbral characteristics of the human voice, it would present as a good candidate for periphonic placement experimentation within this project, with a female voice possibly proving to be more effective at height placement due to an increased amount of higher harmonic content usually being present in a female voice. This also counter-suggests that low frequency sounds, such as sub, kick drum and bass, may not respond particularly well in height placement unless there is enough higher harmonic content to inform the IID cue.

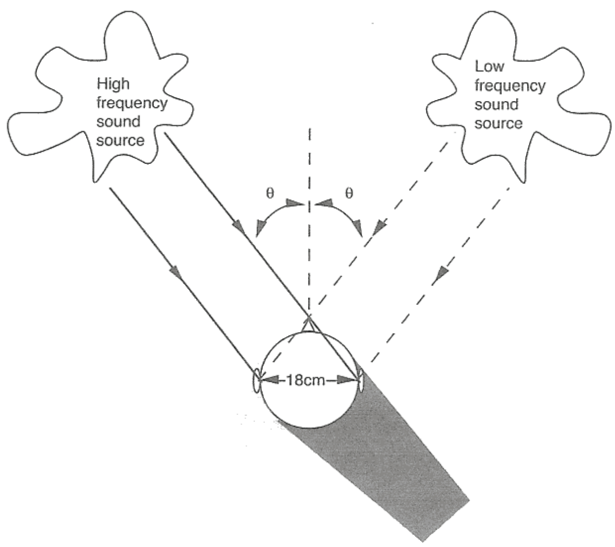


Figure 9: The effect of the head on Interaural Intensity Difference (Fielding, [n.d]).

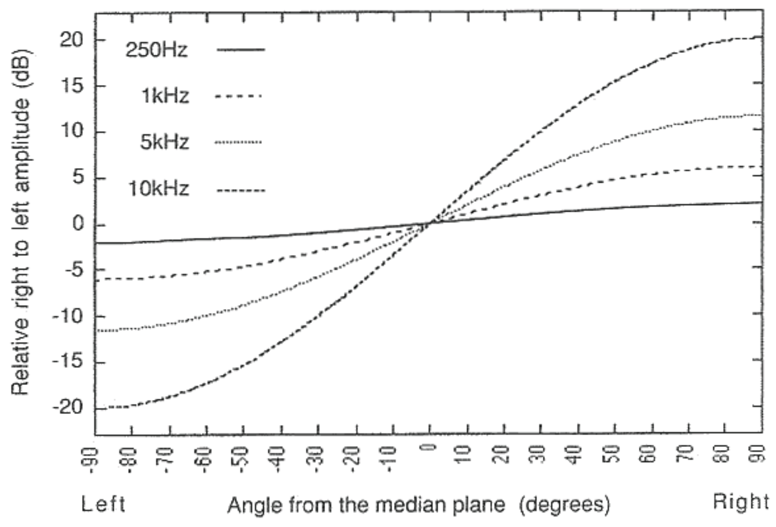


Figure 10: The Interaural Intensity Difference as a function of angle and frequency (data from Gulick, 1971) (Fielding, [n.d]).

The second interaural cue, Interaural Time Difference, occurs because the ears are separated by an approximate head diameter of 18cm, this spacing causes a time of arrival difference between the two ears if the sound source is to the right or left of the median plane (Rayleigh, 1907, pgs.299-302; Howard & Angus, 2006, pg. 97). However, if the sound source is located on any part of the median plane (directly in front, above, behind or below) the sound will arrive simultaneously at both ears. The time difference between the two ears depends on the difference in lengths that the two sounds travel (Howard & Angus, 2006, pg. 97).

Studies have shown it is possible to calculate the maximum interaural time difference taking into account that the sound must defract around the head, adding an additional delay (Howard & Angus, 2006, pg. 98-100). The maximum ITD to occur at a 90° angle (π / 2 radians) is: ITDmax = 344 ms^{-1}

The delay is minimal but the variation between this and zero helps the brain to determine the direction of sound at low frequencies.

There is no difference in delay between front and rear positions of the same angle and this means we must use a different method to distinguish between front and rear sounds. There is also a frequency limit to the way in which the sound direction can be resolved by the ear in this way. This is because the ear uses the phase shift in a wave caused by the ITD to resolve the direction (Rayleigh, 1907, pg.299-302).

When the phase shift is more than π radians (180°) there will be an unresolvable confusion in regard to the direction because there are two possible angles that could cause such a phase shift, one to the left and one to the right (Ibid.). This helps to explain why the historical issues relating to a synthesised binaural perception often occur pertaining to front-back and elevated sources. These positions present natural difficulties in localisation due to this phase shift and are often compounded further due to poor generic HRTF modelling and the inability to further resolve the anomalies with head-movements.

The previously detailed cues only explain the resolution of direction on the azimuth plane and do not cover how we resolve front, back and elevated ambiguities; there are two ways in which the human auditory system can perform these tasks; using pinnae refraction and head movement **(**Howard & Angus, 2006, pg. 103).

Spectral cues corresponding to elevation are thought to be related to the pinna; the aforementioned convoluted outer flap that is generally referred to as ‘an ear’. The first cue uses the effect of our ears on the sounds we are receiving to resolve angle and direction. The ridges upon the pinna are responsible for reflecting the received sounds into the ear canal, these reflections suffer a very small but significant delay that cause comb filter interference effects upon the sounds being received by the ear (Cheng & Wakefield, [n.d], pg. 6; Howard & Angus, 2006, pg. 103; Everest & Pohlman, 2009, pg. 41)**.**

The amount of delay experienced by the received sound wave is a function relative to the direction of arrival, in all 3 dimensions, allowing these cues to help resolve direction when the main directional hearing mechanism does not. As the delays are very small, these effects occur at frequencies typically higher than 5 kHz (Howard & Angus, 2006, pg. 103). Cheng and Wakefield state that “Consequently, frequencies near 6-8 kHz are thought to be important for elevation decoding, since these frequencies have wavelengths which are similar to characteristic lengths of the pinna, and therefore interact strongly with the pinna. There are noticeable patterns in HRTF data near these frequencies which have been shown psychophysically to be correlated with the perception of elevation” (Cheng & Wakefield, [n.d], pg. 6). This research again reinforces the theory that, due to the frequency range and higher harmonic content, the human voice could respond particularly well in a periphonic-binaural placement, especially elevated.

The resultant effect of the pinna is unique to the individual, as we all have differently shaped ears and learn these cues as we grow and develop. It is interesting to understand that we can actually become confused when radical changes to our head or ears occur; cutting very long hair short for example. We also find that if we hear sound recorded through other people’s ears, as in the case of binaural microphones and generic decoders, this may prove to impart a different ability to localise the sound, as the interference patterns are not the same as those we are accustomed to for our own ears (Howard & Angus, 2006, pg. 103). This is an important consideration when defining an appropriate method of binaural decoding for the project. A decoder using a poorly averaged generic set of HRTF’s may present as less than suitable in terms of consistent cross-user perception of the production aesthetic, however understanding the frequency content of a particular sound and using that to negotiate best placement in terms of binaural response may help to resolve this issue if it presents itself.

As, unlike some animals, we cannot move our ears, the most powerful way for human beings to resolve directional ambiguities is for us to move our heads. When we hear a sound that we wish to focus on, or resolve the direction of, we turn our head toward the sound source, in order to place it in front of us, where all of the delays and intensities are the same (Howard & Angus, 2006, pg. 103-104).

The act of moving the head changes the direction of the sound arrival and it is this change of direction that depends upon the location of the sound source relative to our own position. For example, a sound from the rear will move in a different direction from a sound above or to the side of the listener. This movement cue is one of the reasons that we perceive the sound from headphones as being 'inside the head', because the sound is statically linked to and tracks our head movements, whilst also presenting no acoustic reflection and so gives us no cue to assume the sound is coming from anywhere except inside the head (Howard & Angus, 2006, pg. 104). This suggests a kinetic sound stage could be a vital part of a periphonic-binaural production, not only providing the movement that would aid a more distinct localisation but also in conjunction with a good set of HRTF's in the decoding process this sense of 'internalisation' may be lessened, with sounds appearing as if they are coming from beyond the headphones.

ITD vs. IID Trading and The Precedence (Haas) Effect:

As both intensity and delay cues are used for the perception of sound source direction, we might expect the mechanisms of such cues to overlap and create confusion in the way in which they are interpreted by the brain, this is in fact the case and this effect is known as interaural time difference versus interaural intensity difference trading (Howard & Angus, 2006, pg. 104)**.** Interestingly, an ITD can be compensated for by an appropriate intensity difference. If the delayed sound is more than 12dB higher in amplitude than the first arrival, then the perception of sound direction will appear to come from the delay source. After 30-35ms of delay, the auditory fusion breaks down and the delayed source signal will be perceived as an echo, leaving the listener able to differentiate between the delayed and un-delayed sound sources (Ibid.). This phenomenon is called the law of the first wavefront, the precedence effect' or 'Haas effect'**,** named after Helmut Haas (Mc Carthy Iatse, 2009; Everest and Pohlman, 2009, pg. 60-62). The Haas effect is that which causes the ear to attend to the direction of the sound that arrives first, ignoring reflections provided that they arrive within 30ms of the first perceived sound. The Haas effect is often used in stereophonic record production to create a pseudo-stereophonic effect from mono source content and to widen existing stereo content. This same implementation of Haas effect for pseudo-stereo technique has been used in a couple of the project's productions, the Haas effect has also been implemented as a way to construct echo effect and placement through delay within the productions.

3.2.1 - Principles of Binaural Audio Technology

Spatial Emulation and Periphonic Binauralisation

Though binaural recording has fascinated researchers for many years, it has received little serious attention until recent times, partly because it has generally been difficult to make it work accurately for a wide range of listeners over such a range of different headphones. This is due to the variation in perception presented through an individual's fine-tuned HRTF response and the variation of specification underpinning the function of differing headphone types. It also has often shown not to reproduce well over loudspeakers without an amount of signal processing (including purposeful phasing through comb-filtering) to retain a wider stereo image, this in itself for commercial purposes can present inconvenience, as to ensure a level of quality and consistency two versions of the same recording would usually have to be made (one for headphones and one for speakers), in order to avoid inconsistent reproduction through up-down processing a mix. However, with recent DSP developments it is possible to construct hi-fidelity binaural reproductions using Ambisonic decoding systems, which allow for one mix to be upmixed or downmixed for a number of different formats including multichannel, surround, stereo, binaural and mono. This versatility in output format and the ability to decode a spatial mix to headphones affords an increased chance of success for the use of these formats in the commercial music market.

Traditional binaural approaches to 3D spatial sound representation are based upon the idea that most accurate reproduction of 'naturalistic' spatial listening cues will be attained if the ears of the listener can be provided with the same signals and cues that they would have experienced in the source environment during ‘real-life’ listening (Rumsey, 2001, pg.66-67). This can be attempted in the field through binaural recording techniques or virtually through digital emulation using 3D / spatial audio software, such as an Ambisonic-based processor or Head Related Transfer Function (HRTF) emulation system.

Research shows that recent technological developments have made the signal processing for synthesised binaural signals more commercially available, at reasonable cost (Rumsey, 2001, pg.66-67). There are also various binaural microphone sets commercially available, spanning a budget of £70-£6000; Roland CM10-ES and Neumann KU-100, respectively for binaural and in recent years there have been a number of consumer priced quad-coincident Ambisonic microphones released to market, some as affordable as £299, in the case of the Zoom H3-VR.

Head-Related Transfer Function: Measuring, Emulation and Interpolation

As previously discussed, Head Related Transfer Function or HRTF is defined as the far-field frequency response of an individual’s left or right ear, measured from a particular point in the free-field to a particular point within the ear canal (Cheng & Wakefield, [n.d], pg. 2-3). HRTF’s are measured from humans for both the left and right ears at a fixed radius from the listeners head. They are measured at several different azimuths and elevations, measured in radians or degrees and including both ITD and IID (Ibid.).

A common technique used to empirically measure HRTF's is to insert probe-tube microphones into the ear canals of a subject and to then play a known spectrum stimulus (click, pseudo-random binary sequence or Golay code[7]) through a loudspeaker situated at a specified azimuth θ, elevation Φ and known distance from the subjects head (Ibid. pg. 3). Selections of the measured transfer functions due to the measurement apparatus (microphone and speaker transfer functions), along with portions of the measured transfer functions which are the same for all locations, are called the Common Transfer Function (CTF) and are removed from the raw measurements, resulting in directional transfer function (DTF) at azimuth θ and elevation Φ. It is the DTF quantity that contains spectral cues responsible for spatial hearing and is commonly referred to as HRTF (Ibid.).

Raw HRTF's are regularly used to synthesise spatialised sound over headphones, i.e. reproducing surround sound playback over stereo headphones. We can presume that the right HRTF, the left HRTF and the ITD for a specific location can characterise the acoustic filtering of a sound originating from that location, and so assuming the auditory system relates these quantities to a specific spatial location, HRTF's and ITD's can be used to filter a monaural sound into a binaural one; sounding as if it originates from that location (Cheng & Wakefield, [n.d], pg. 4) (see figure 11). There are already many applications that involve real-time synthesis of spatial audio where the sound sources move location over time, using high speed DSP hardware to implement the convolutions with delay lines implementing the

[7] A Golay code is typically a binary impulse artefact used for error-correcting digital data. However, these can be used for measuring the impulse response of the ear canal or for generating binaural impulse responses to measure HRTF data.

time delay. In order to synthesise moving sounds, HRTFs and ITDs are updated in time to correspond with new locations.

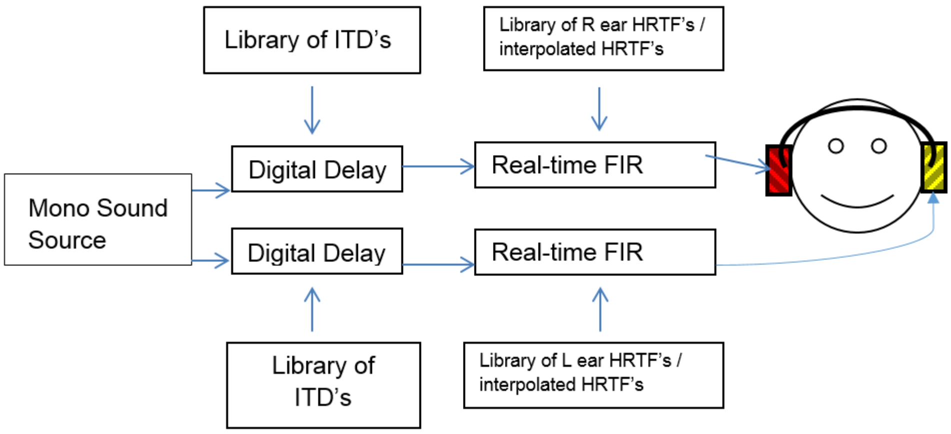


Figure 11: Block diagram of a simple HRTF-based spatial sound synthesis system (Adapted from Cheng & Wakefield, [n.d], pg. 16).

Although in theory using HRTF's to synthesise spatial audio is simple, there are still a few problems that occur regarding practical implementation. Simple HRTF-based algorithms like the one depicted in Figure 10 do not always produce sounds with the intended spatialisation effect. Subjects often report those sounds spatialised on or near the median plane sound as though they are 'inside' the head rather than outside and those synthesised to be in front, actually sound as if they come from behind the listener or vice versa (these are known as front-back confusions).

Synthesis of sounds with non-zero elevations are also quite difficult and, furthermore, since every individual has a unique set of HRTF's, a subject listening to sounds generated by “average or generalised” HRTF’s may not perceive the sound to be in the intended location. A statement made by Cheng & Wakefield ([n.d], pg. 9) is of another problem in the practicality of HRTF’s being that dedicated real-time DSP hardware is often necessary to implement the simplest of spatial algorithms. This means that high quality synthesis of virtual audio is often not possible on low-cost generic computers (Ibid.). This said, technology has evolved dramatically since the publication of that paper and though the argument of perceptual confusions remains valid the statement that binaural processing tools are costly is not. There are a number of software plugins available, to suit a wide range of budgets, the only issue is that the result between them varies dramatically, regardless of the price. One is looking at as little as £75 for a basic HRTF-based 8 channel surround to binaural encoder, such as Flux IRcam Hear V3. Then there are more complex spatial panner/binaural decoders such as Dearreality Dear VR Pro at £300 or as much as £1000 for a comprehensive suite of Ambisonics-based upmixing and decoding tools, such as Blue Ripple Audio O3A. The latter two have been used in this project and the result from each side of the spectrum has been surprising and although mostly satisfactory, they were not without some minor, inevitable anomalies. In particular the Blue Ripple tools produced some of these type of anomaly but this was more a side-effect of the less than desirable method of usage and the inability in that method for real-time binaural monitoring of the production stage.

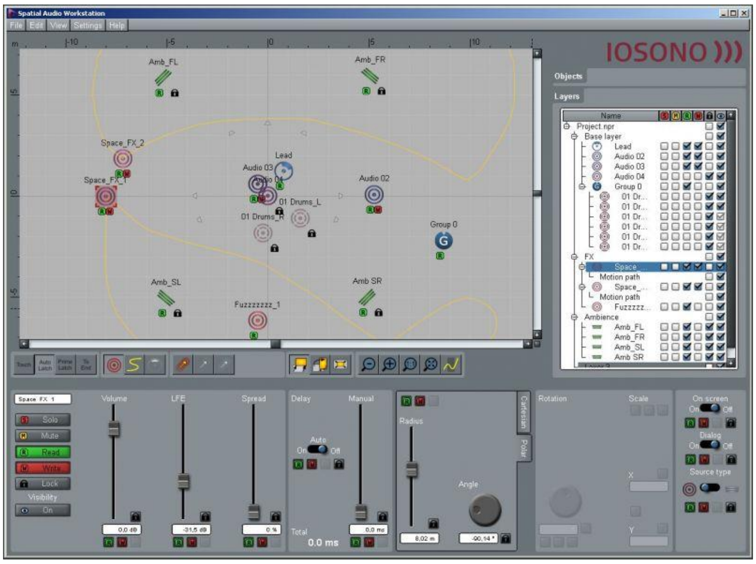


Figure 12: A spatial audio workstation, IOSONO plug in for Nuendo 5 (IOSONO, [n.d]).

Auralisation and Virtual Auditory Display

“Auralization is the technique of creating audible sound files from numerical (simulated, measured or synthesised) data… In the end, the goal is to achieve an auralization in real time, a dynamic interaction with the user, and the user’s immersion and presence in the virtual scene” – Michael Vorländer (2008, pgs.3-4)

Auralisation must cover all of the relative cognitive aspects of any specific perception; i.e. the distance and size of the sound source, direction of the event, reflection and the environment or room (size, shape and sonic characteristics), movement of the sound source and the listener’s movements. Due to this multidimensional perception and in order to allow simulation and accurate replication of psycho-acoustically relevant features, an authentic model of the sound and vibration field is required, representing the correct sound propagation constant (governed by inverse-square law), RT60 and directivity (Vorländer, 2008, pg.3).

“The full characterisation and interpretation of a sound can be achieved only when hearing and other senses are involved directly, therefore auralization offers an important extension to acoustic analysis and synthesis, prediction and rating… it represents an important component of multi-modal sensation and corresponding psychological effects” (Vorländer, 2008, pg.103).

Auralisation tools allow for the construction of hyper-real periphonic sonic spaces, whilst affording the ability to also decode this output into a binaural reproduction through applying HRTF based decoding.

The flow diagram (see figure 13) details the principle of auralisation; sound generation, transmission, processing and reproduction, and in the case of generation and transmission, the diagram details how these two models couple. This coupling is due to the acoustic response of the environment and the synthesised displacement generated relative to sound propagation and reflection within that particular audioscape. An accurate acoustic response adds psychological effect to the audio, heightening the reality and increasing immersion.

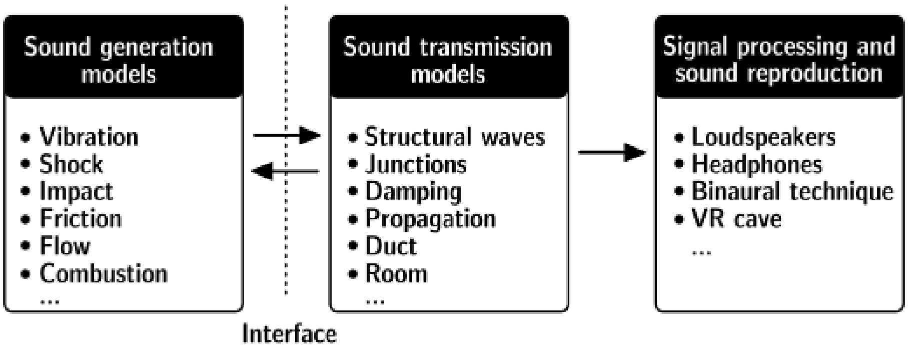


Figure 13: Principles of auralisation (Vorländer, 2008, pg.104)

Research has shown that there are many problems in achieving accurate reconstruction of HRTF's because, although they share some common features, they are unique to the individual based upon a combination of ear-shape, hair length, head mass etc. This makes it difficult to generalise HRTFs (Rumsey, 2001, pg.66-67).

As previously discussed, head movements themselves are a crucial aid in resolving locational ambiguities and are difficult to incorporate within reproduction systems. Visual cues are often missing from binaural experiences and these cues generally have quite a strong effect on the depth and realism of perception. Models of headphones differ in EQ and mounting and have been known to cause distortions in perceived HRTFs. These distortions are caused by phase and frequency response errors in the signal chain and can affect the subtle cues necessary for accurate reproduction, resulting in a less than accurate spatial interpretation whereby some sounds may appear to come from 'inside the head' and others are mis-judged (front-back ambiguity). Head-tracking may help to address this issue by allowing for head movements as a means to aid localisation. However, head-tracking goes against the grain of this project, which aims to produce spatial mixes for current playback systems, and head-tracking is not possible with a normal set of headphones.

What we can understand from this research is that using contemporary DSP technology we can synthesis the cues and spatial acoustics required to emulate a binaurally perceived sonic environment without fixing the binaural space, as with traditional ambient binaural recording methods or stereo to binaural mix-downs.

Sound sources with a strong high-mid content could work particularly well in a periphonic-binaural sound field, as the act of high frequency refraction off the head and pinnae increases the perceived directionality of sound sources, especially those between 3-8k; the range spanning the presence of the human voice and the pinnae spatial sensitivity band. This research leads us also to understand that sounds with a strong lower frequency content would be less inclined to work well in a periphonic-binaural placement and may possibly increase presentation of localisation anomalies if placed in height, the rear or the non-frontal medium plane.

“Movement can alter sensory inputs and so result in different perceptions… changes in output are merely a means to changes in input, on which perception depends directly” (Briscoe, 2015). This statement is fundamental in exampling the natural function in human sound localisation. Head movement is identified as an important factor in aiding localisation of ambiguous sound sources and so the project theorises that incorporating head-tracking as a localisation aid could help overcome some of the possible perceptual anomalies associated with binaural listening. If upon hearing the sonic information we do not fully understand it or perceptual problems persist, we use our head to help locate the sound and define further information. Research has taken this concept and immersive-headphone technology has already integrated motion-mapping technology to simulate head-turns and the relative audio response for use in VR & gaming. It won't be too long until we can sit at our laptop in a virtual reality concert online, turn our head to the virtual string section and perceive the strings to be more present (Smith, 2014). However, acknowledging the technological advances and the practicality of application, one can hypothesise that if head-tracking is not an option, and for many established industry producers and music consumers it is not currently an option, nor is it a necessary requirement to an audio-only experience whereby there is no visual cue to trigger movement informing and reinforcing the virtual auditory display, then the head cannot move and if the head cannot move then the movement could then instead be created within the musical context. A kinetic sound stage may possibly help to address any perceptual ambiguities presented through the generic HRTF's used in the binaural decoding, aiding the listener in an improved localisation of front-back and vertical sound sources, ultimately reinforcing the believability of perception and interpretation of the periphonic space, thus defining movement as a key concept when conceptualising a periphonic-binaural sound stage.

3.3 - Part 3 - Two sides to the same penny (drop): Proxemics, Spatial Memetics and the Spectromorphologic Analysis of Recorded Music.

In part two I examined the physiology of binaural listening and the principles that underpin the function of binaural audio technologies to provide reasoning and context for the conclusions ascertained in part one; that lateral and vertical placement perception must be dependent on the binauralisation process and headphone-based mode of listening, and not reliant upon the physicalities of speaker placement as seen with stereo and multichannel speaker reproductions. What was concluded is that the efficacy of a reproduced binaural perception relies upon the quality of Head Related Transfer Function synthesis applied in the capture or decoding processes of recorded audio, and that object-based DSP approaches afford reflexive binaural emulation. It was concluded that in order to generate the best perception, synthesised HRTFs need to have a close correlation with the unique HRTFs of the listener; implementing user-specific HRTFs is not currently possible in a wider commercial context but high-quality averaged HRTFs formulated from a large pool of listener data can help to overcome this limitation[8]. It was also concluded that elevation perception is heavily reliant on how well the applied HRTFs seemingly replicate pinnae interference patterns; those which generate the micro-delays in higher frequency content. As such, elevated sound sources should contain higher frequency content to help facilitate this. Laterality is dependent on the interaural time and level differences generated through either the physical

[8] The more HRTF measurement data that can be used in generating an average, the more chance there is of lessening perceptual anomalies across listeners when synthesising a binaural perception.

separation between capsules in the binaural recording process or inferred time and level differences resulting from the binaural decoding process. However, all of the above are reliant upon the headphone-based mode of listening in order to generate the binaural perceptual phenomena. Part two also introduced demonstrable elements of embodied perception; outlining how Human beings perceptually interact with their listening environment through head movements. We explored their importance in aiding localisation and the role head-tracking may play in spatial music production; concluding that a dynamic sound stage may help to lessen localisation ambiguities in music productions without head-tracking integration.

Part two established an understanding of the principles under pinning the physiologic and psychoacoustic functions of our perceptual system. In this section I build on this through exploring how we make tangible meaning from the information gathered by these systems. Part three delves deeper into the concept of sonic embodiment and the role embodiment theory plays in taking an ecological approach to spatial music production. Further, part three explores how we interpret, interact with and respond to the world around us. I evaluate the cultural communication theories; proxemics and memetics, how these inform our understanding and experience of real-world phenomena, and how these theories may be applied in constructing metaphorical spatial music schema. Firstly, however, it is imperative to outline the concept of spectromorphology and how it can be used in analysing the construct of sounds identified within a listening experience. This is fundamental in providing the analytical approach required to de-construct the sonic characteristics of any given real-world phenomena, so that subsequently we have the understanding required to effectively re-construct them within a music production context.

The Spectromorphologic Analysis of Recorded Music

Spectromorphology is defined by Denis Smalley (1986, pg. 61-63) as a tool for describing and analysing a listening experience. The two parts of the term refer to the interactions of the *spectro* – referring to the sound spectra (the elemental sonic characteristic) and its morphology – the ways in which they are shaped and change through time. The spectro cannot exist without the morphology and vice versa; a sonified something has to be shaped, and sound shape must have sonic content (Smalley, 1997, pg.107). A spectromorphological approach therefore refers to the descriptive analysis of interactions pertaining to how a particular sound or sonic shape changes over time and sets out spectral and morphological models and processes providing a framework for understanding the structural and behavioural relationships that exist in the temporal flux of the music (Smalley, 1997, pg.107). Though Smalley developed the notion to help him think about his compositional method, spectromorphology is not a compositional theory but it can be used to influence compositional practice once the concepts and semantics are consciously known to the composer. The project utilised spectromorphology as an analytical and descriptive framework and, to some extent, also as a compositional influence in some of the later pieces.

Proxemics

The sociologist Edward Hall introduced the discipline of proxemics in his 1966 book titled “The Hidden Dimension,” in which he defined proxemics as being “the interrelated observations and theories of man’s use of space as a specialised elaboration of culture.” Hall is responsible for the notion of ‘personal space’, or the invisible force field most Americans ensconce themselves in while moving through public places; a breach of implied boundaries (per Hall, the human ego extends about a foot and a half outside the body) is neither welcome nor tolerated. (Petrusich, 2016). This understanding of proxemics influences the philosophy behind the depth and intimacy production experiments in this work, implementing influence of proxemic theory through sonification of auralised space and the suggestion of breached boundaries, heightening the emotive experience and creating perceptual contrast and movement comparative to the mid-field sonic placement. This understanding of proxemics could work particularly well for vocal production work, as the listener is more likely to be reactionary to a human voice encroaching upon their personal sonisphere (the personal intimate listening space around them equating to Hall’s defined physical intimate space). In ‘The Hidden Dimension’ (1966) Hall develops his theory of Proxemics and argues that the human perception of space, though derived from the sensory apparatus that we all share, is moulded and shaped by our own cultural experiences. Hall also argues that differing cultural frameworks for defining and organizing space are internalised by people at a subconscious level and that these differences can lead to failures of communication and understanding at a cross-cultural level. Where in one culture a stranger entering this foot and a half of personal space surrounding the body may be completely unacceptable, even perceived as being shocking, intimidating or threatening, in another culture it may be wholly acceptable and a normal demonstration of intimacy and closeness. Some cultures are more tactile with gestures, whereas some cultures totally refrain from or limit bodily contact unless you have the spousal or familial relationship to warrant such. To expand on this understanding further we can look toward embodiment as a means to explain and understand these failures in cross-cultural communication in the musical realm. To bring this into a musical application we can think of some of the vocal work theorised throughout this project; the gang-style vocal production within 'Monomorphic'[9] addresses omnipresent localisation through a schematised choral structure. The metaphorical suggestion of a choir presented in the production staging works particularly well with the vocal style, phrasing and genre of music. However, if the same technique were to be applied to drill music, for example, the suggestive vocal output could possibly be considered in interpretation to be much more aggressive, darker and more intimidating given the vocal style, vocal phrasing, the street-gang associations with that genre of music and the intimidating larger-scale surrounding placement comparative to the smaller, central point of the listener. Therefore this notion of cross-cultural communicative error can be explained and negotiated using an understanding of embodiment theory and

[9] An electronic music production piece from 'Hidden Behind Static' which is used within the practice as research investigations of this thesis.

critically seems to be an important correlation to the function of proxemic theory's cultural communication systems. The interlinking of embodiment and proxemic theories suggest that if the listener had no past association with the origins of drill music or the space or environmental situations implied through the surrounding drill-vocal placement then the resultant communicative affect could possibly be less intimidating or aggressive in perception than interpreted by those that have an embodied real-world experience of the intimidating interpretation suggested through the schema. However, given the distinct spectromorphology of the hypothesised drill vocal production; the aggressive delivery, the lyrical content and the morphological structure of the intimate spatial placement may homogenously act to reinforce the intimidation and aggressive experiential association of the genre regardless of the listener's knowledge of the musical style.

To provide an example of this idea of cross-cultural miscommunication, we can look to the meme in Figure 14 for a visual reference that is easy to understand. Here it depicts the caption 'I hope your day is…' and an image of a can of cola with the word 'legend' printed on the can and a container of milk. This is not an interpretation reliant on embodied association at all but one that is reliant on a good grasp of the English language. The correct interpretation of the image in order to complete the caption would be 'legend-dairy', a play on words for 'legendary'. However, cultural miscommunication in this case can occur if the viewer does not speak English, or if they have a limited English vocabulary or are unfamiliar with such word play. An alternative interpretation would be 'coke milk' or 'cola milk', though these are valid interpretations of the image content, it does not work with the caption and so may leave the viewer confused and the meme non-functional. The same can be applied to the interpretation of the spectromorphology of the auralised spatial scenes; a lack of embodied experience of the suggested environment may cause cross-cultural miscommunication, as could an alternative interpretation of the musical context, or a pre-conditioned means of fronted-listening driving an interpretation pertaining to that which we associate with stereophony.

Figure 14: 'Legendairy' meme (rodkimblestuntman, 2014)

Figure 15: An eye or a kitchen sink? (oodlepic.com, [n.d])

Embodiment vs Memetics

Embodiment theory recognises that meaning and semantics cannot be captured by abstract, logical systems, but are dependent on an agent's environmental experiences. The representation or expression of something; quality, idea, feeling, in a tangible or *audible* form is based on the assumption that said thoughts, feelings, and behaviours are grounded in bodily interaction and past experience with the environment. This theory is well supported in cognitive science literature and is of particular importance in our understanding of music perception and cognition. "Many authors have discussed the notion of unnatural or impossible auditory scenes in recorded music, but how we interpret them stems from the way they suggest 'real' phenomena rather than the way in which they are 'unreal'" (Zagorski-Thomas, 2014, pg.80).

Embodiment theory can be used to underpin a number of the staging concepts and periphonic decision making within this project, such as the synthesis and implementation of *sonic cartoons*. In his chapter '*The Spectromorphology Of Recorded Popular Music: the shaping of sonic cartoons through record production'* within *The Relentless Pursuit Of Tone*, Simon Zagorski-Thomas (2018) defines *sonic cartoons* as "… sounds which suggest interpretation through their metaphorical relationship with our embodied experience …the notion that we understand music schematically… through association with sounds that do have identifiable causes".

To provide an example of this perceptual trickery and embodied association in a visual manner that is easier to understand, we can refer to another meme as reference. Figure 15 at first glance looks to have all the invariant properties pertaining to the necessary schema of an eye; a dark centred pupil, an iris and a moistened 'white'. However, upon reading the caption it becomes clear that in reality it is actually an image of soapy water draining through the plug hole of a kitchen sink, thus altering the way in which the viewer now perceives the image, allowing for more than one interpretation; a realistic one and a believable one. Given the cross-cultural association with both objects represented within this image it is less likely to present a cultural miscommunication in either interpretation thereof, as most people could relate to an image of an eye or a sink, schematic or otherwise, as all human beings have eyes and most people would be familiar with the topography of a sink, regardless of cultural background. We can think of the schematic sonified representations of auralised scenes in recorded audio in much the same way.

Importantly, the project explores and utilises an understanding of proxemics through an applied ecological approach to perception coupled with sonic embodiment and memetic evolution as a means for expanding creative agency in record production. These theoretical systems act as vehicles for conveying sonic space, depth and intimacy through surrealist scenarios using metaphor and *sonic cartoon* in the construction of new and unique approaches to spatial music staging and production technique.

Memetic Theory refers to the theory that knowledge and ideas can evolve more or less independently of their human-agent substrates. Though humans provide the vehicle for this cultural evolution, memetic theory states that ideas can be developed without human comprehension or deliberate interference. This applies in this particular case, as although I am the vehicle for the initial creative conceptualisation of the production techniques and staging ideas, the continued development cross-application depends greatly on the concept of the musical piece, the approach to production and the sonic content available. Therefore, all aiding to memetically evolve the initial technique (1st Meme) into something more befitting the next musical application (2nd meme, 3rd meme and so on).

Jonnie Hughes presents a great demonstration for ‘meme theory’ in his 2012 Independent article ‘Meme Theory: Do we come up with ideas or do they, in fact, control us?’ Hughes puts forward the question of ‘Where did the idea for the cowboy hat come from?' as an example. He states that the textbook answer would be that John Stetson synthesised the idea and invented the Stetson in the 1860s having joined the gold rush in Colorado and subsequently realising that the floppy felt or racoon skin cowboy hats the cowboys wore were not particularly suitable for the harsher, wetter weather or could endure the dusty, rainy plains. So, Stetson stitched together fur pelts and created a wide-brimmed, tall-crowned hat that he dubbed 'The Boss of the Plains'. However, an alternative answer would be that actually the Stetson hat did not look anything like the Stetson we know today. The crown was a uniform dome without indentation, the brims were flat with no curled edges and though incredibly popular, the Stetson hat or 'Boss' was also fairly expensive so the cowboys could generally only afford one. This meant that the hats would get worn quickly, the brims rolling inward and the crown getting punched downwards, before inevitably and eventually falling apart or requiring re-stitch. What we understand to be a Stetson hat these days is far from what it was back then or initially constructed by John Stetson, with the modern variant being a product of climate, a function of purpose and the purchasing selections of a hundred thousand cowboys (Hughes, 2012). This is how the memetic lens applies to the invention of the cowboy hat. Nobody invented the Stetson we know today, it evolved over time. The idea existed independently of John Stetson and of the cowboys, who were merely the host vehicles to the evolution and developmental implementation of the Stetson we know today. This is analogous to the way in which staging techniques and production practices were developed and shaped by technologies. This can be seen in Dockwray and Moore's (2010) survey of recordings from the 1960's and 1970's, where it is demonstrated that the aesthetics of what constituted a 'good mix' emerged through the practice of many engineers over time and were dependent on the technology of the era.

In 'The Selfish Gene', first published in 1976, it was Richard Dawkins who builds upon the principle theory of evolution and natural selection, proposing the concept of a "meme" as an element of cultural transmission, which Dawkins states is analogous to genetic transmission (Dawkins, 2016). Just like the 'survival machine' hosts that carry them, ideas are subject to evolutionary principle with humans being little more than host vehicles to this continued development and application of said ideas. Dawkins continues to put this into context of language evolution with the following statement "Geoffrey Chaucer could not hold a conversation with a modern Englishman, even though they are linked to each other by an unbroken chain of some twenty generations of Englishmen, each of whom could speak to his immediate neighbours in the chain as a son speaks to his father" Language appears to 'evolve' through non-genetic means, and at a much faster rate than genetic evolution (Dawkins, 2016).

Interestingly, though somewhat paradoxical alongside *Embodiment Theory*, spatial music memetics can too be employed in defining the project's staging work, whereby a number of environments and staging constructs are created independently of embodiment, and do not associate any sort of bodily interaction, realism or past experience in their construction but evolve dependent only on the requirements of the production context and the affordances of the technology used to implement them. The output association would be due to the listener's interpretation of the memetic suggestion and their own embodied associations pertaining to the contextualisation thereof.

This notion of socio-ideological evolution being independent of human beings is suggestively reinforced in the writings of Marshall McLuhan and Quentin Fiore in their 1967 classic ‘The Medium is the Massage’ whereby, decades ahead of their time, they discuss technology as a means to reshaping society and sensory perception, each of which have great impact on cultural evolution and subsequent synthesis of ideas. They state that the medium, or process, of our time – electric technology – is shaping our everyday lives and social independence through forcing us to re-evaluate and reconsider our thoughts and actions (McLuhan and Fiore, 1967).

> “Everything is changing and they’re changing dramatically… societies have always been shaped more by the nature of media by which men communicate” (McLuhan and Fiore, 1967, pg.8).

Acknowledging this, I presuppose that the technology assumed within the project will present as the biggest driving force behind the evolution of ideas and the development of creative staging practice, with the initial philosophical and theoretical concepts serving as a starting point for a continued memetic development.

In a Brain Pickings’ book review of David Byrne’s ‘How Music Works’, Popova (2011) states that “Among the book’s most fascinating insights is a counterintuitive model for how creativity works, from a chapter titled “Creation in Reverse” — a kind of reformulation of McLuhan’s famous aphorism “the medium is the message” into a somewhat less pedantic but no less purposeful "the medium shapes the message"". The same can be said in regard to the creative output required when structuring a periphonic sound stage. The medium shapes the message; i.e. the spatial sound field and the production tool's affordances are a creative technological extension of the artist/spatial producer, shaping the message through creative accommodation, with the intended delivery format continuing to aid the shape, and fundamentally, the interpretation of said message. However, it is the context of the message content that fundamentally and intrinsically determines the level of creative output in constructing the message.

Byrne (2012a) states that "The accepted narrative suggests that a classical composer gets a strange look in his or her eye and begins furiously scribbling a fully realised composition that couldn't exist in any other form. Or that the rock-and-roll singer is driven by desires and demons, and out bursts this amazing, perfectly shaped song that had to be three minutes and twelve seconds — nothing more, nothing less. This is the romantic notion of how creative work comes to be, but I think the path of creation is almost 180º from this model. I believe that we unconsciously and instinctively make work to fit pre-existing formats." "In a sense, we work backward, either consciously or unconsciously, creating work that fits the venue available to us… In a sense, the space, the platform, and the software "makes" the art, the music, or whatever. After something succeeds, more venues of a similar size and shape are built to accommodate more production of the same" (Byrne, 2012b). Perhaps unknowingly here Byrne is suggestive of memetic theory as a driving vehicle of creativity, alongside the output being a function of the limitation and affordance of the technology or format. This is very much reflected through surround sound's history and aids to reinforce the notion that with spatial DSP advances comes a new wave of creative opportunity for exploitation, and with the modern style of music consumption being headphone-based we are in a better position than ever before to successfully utilise and integrate spatial production within music recordings.

The Spatial Sound Institute state their research on 'Spatial Memetics' to be "a methodology that aims to develop trans-sensory language sets for use in technology-driven immersive experiences. These languages are formed by a series of sound objects with definable, recognisable syntax and dimensional and behavioural characteristics that can integrate across multiple sensory modalities." (SSI, [n.d]). They propose that spatial sound, more so than visual or any other sensory media, can offer new directions in new linguistic exploration, which may possibly return us to much older modes of communication, such as seen in oral traditions. Stating that "Written language forms are a relatively recent invention, with much communication in ancient history conveyed orally, and therefore sonically and aurally." (SSI, [n.d]) This sonic, aural oral tradition meant that consequently place and environment were fundamental to the character, nuance and cultural delivery of communication (SSI, [n.d]). This is supported by much of Edward T. Hall's theory of proxemics (as previously detailed) and both theories translate to synthesise the philosophies behind the staging approaches within this project's practical constructs. Proxemics acts more as a means to impart certain feelings, concepts and cultural messages through the depth experiments and spatial memetics as a means for creating spectromorphological structures to convey certain shapes, cultural symbolism, thoughts, feelings, and ecological concepts that can be (re)developed and (re)applied across pieces.

It should be acknowledged that ‘memes’ - be they presented in pictorial, animated (GIF) or musical form - do equate to language and when applied to text-based conversation or, indeed in this instance, music, they offer an alternate and more complex creative and cultural expression, through tangible implied metaphor, than typical words or musical stages would generally afford. The concept and application of musical memes, just as with pictorial or animated memes, allows for the same memetic structure to evolve through reapplication across the differing scenarios (in this case productions) without necessarily implying/recalling exactly the same metaphorical concept or interpretation.

Some of the sound stages, such as the vocal work discussed in detail in a later chapter in this thesis, are often surreal scenarios that are, at the very least, impractical, or indeed, improbable and in some instances even lean toward being ecologically impossible. Most of the defined techniques, especially those pertaining to the vocal staging, act as musical memes that are re-applied and continually developed and carried across the project’s practical works- sonic context and creative concept dependent. For example; the vocal work in Hidden Behind Static’s ‘Penny Drops’ begat what the project coins as the ‘omni-monophonic’ vocal tree in ‘Monomorphic’, which then begat the creation of the ‘polyperiphonic’ vocal placement within ‘Far From Here’. Further to this, in Jerome Thomas’ ‘Late Nights’ and Beautiful Thing’s ‘Waiting’ I then use an adapted amalgam of both ‘omnimonophonic’ and ‘polyperiphonic’ styles, and continue to produce periphonic-binaural spatial vocal work using adapted versions of these primary vocal staging techniques and cultural metaphors (memes).

3.4 - Part 4 – Musical Perception and Cognition

In part three I outlined the concept of spectromorphologic analysis and how it can be employed to reverse-engineer sonics in a music composition and production context. Further, part three explored how Human beings interpret, interact with and respond to the world around us. I introduced embodiment theory and acknowledged how our bodies are grounded in our past experiences and that these associations influence how we form meaning from our future experiences, including and not limited to our experience of music. I explored proxemic theory and outlined that the human perception of space is moulded and shaped by our own cultural experiences, and that differing cultural frameworks can lead to failures of communication and understanding at a cross-cultural level. This is pertinent to spatial music production, especially where the construction of metaphorical or surrealist sonic representations of certain cultural signifiers are concerned (such as depth and intimacy). I also discussed the independent evolution of ideas and drew upon memetic theory to help outline this. We ascertained that although humans may provide the vehicle for the evolution, said ideas can develop without human comprehension or deliberate interference, depending on the circumstances surrounding the inception; the evolution of the Stetson hat was presented as analogous to this notion. This can also be observed through the way in which the vocal staging ideas of this project evolve; there is a single ‘seed’ idea born from the requirement to resolve the vocal staging research problem, which grows in a variety of ways depending on the contextual circumstances (the audio content) producing different evolutionary results or ‘memes’. It can be concluded that how said ‘memes’ are interpreted will too be reliant on the underlying cultural factors relating to our embodied experiences, and that this may also result in the ‘memetic’ evolution of the interpretation of the idea, or lead to communicative failures of understanding similar to those associated with proxemics. I also pointed to two visual memes to helps reinforce this understanding of cross-cultural communication and interpretation. Together these three theories provide an interconnected knowledge of how we experience and make meaning from real-world phenomena. This understanding can then be applied in a more abstract manner to create metaphorical representations of phenomena using spatial sonic schema.

This final part of the literature review further deepens our knowledge of perception and investigates the musical part of the four-pronged conceptual approach presented throughout this literature review; Part one evaluates the four-dimensions pertaining to the virtual sound staging of real-world performance; Part two provides an understanding of how we audibly perceive real-world phenomena in three-dimensions; Part three defines how we procure meaning from our experiences of and interactions with real-world phenomena, and this final section will provide the musical understanding required to complete the framework. In the following section I examine literature revolving around pertinent areas of musical cognition and employ spectromorphologic analysis to explore how the mechanisms of musical perception may be exploited to benefit the construction and interpretation of musical schema.

Musical Perception and Cognition

Sensory analysis contains processes that help us in perceiving a unified perception, causing the various attributes of sound to not be experienced individually of one another. Though the pitch and temporal aspects of music are processed in different parts of the brain, when we listen to music we do not generate separate experiences of them or experience them independently of one another. The dimensions of musical perception merge into one integrated and unified experience and to properly understand this requires examination across multiple levels of musical structure and neurological processing spanning; perception of melody and rhythm, emotional and embodied responses, skilled performance and many and all other creative activities pertaining to music (Thompson, 2015, pg.71).

Thompson (2015, pg.82) states that “like language, music can be segmented into a series of phrases, each phrase is a digestible group of musical events…a number of cues influence our understanding of the significance of events in music.” These cues include intensity, metric strength and tonal stability and they aid to differentiate musical events as being either significant or less significant. Just as this idea of segmentation can be applied to the cognition and composition of music, the spatial stages in this project could be broken down into phrases (or mini-stages) that sum to construct the large-scale stage. Critically, using this idea of significant and less significant elements in music cognition may aid in navigating the construct of the spatial stage through defining significant and less significant staging and production focusses.

Upon listening to music produced with the Auro 3D system, Thornton states “what’s really interesting are the percussive elements - that is where you really hear the early reflections and the height is based off all that production… with the snare drum right at the end, that’s when I really became aware of the additional height.” (Thornton, 2014, 1:24). This statement, perhaps unknowingly, supports and divulges the notion that high frequency content reproduces in height periphony particularly well. This is further supported through Dr Hyunkook Lee’s research into pitch-height effect and vertical stereophony. Though Dr Lee’s experiments were based on pure tones, his research finds that high frequency content (typically of around 6-8kHz) seems to give a better vertical stereophony and some particular mid-high frequency bands gain elevation independent of speaker height. Certain frequencies also seem to retain a more stable spatial image when placed in specific relation to the human body; a relation in which we naturally perceive those frequencies to be more present (Lee, 2015, 7:24 - 31:11). This can also be justified through our understanding of stereophonic record production techniques, such as whereby EQ and level can be used to impart height into a mix, with temporal effects imparting movement and depth.
The effective response of the percussive elements in height periphony can be evaluated through an understanding of music cognition and how we embody rhythmic structure. Research states that rhythmic perception is

strongly embodied (Leman, 2008, pgs.43-49), this can be experienced in the way the tempo of the music we listen to can affect our speed of walking or exercising, the way we draw dance moves from musical rhythms and through the physical movements and expressions that produce musical sounds. This explains air drumming and air guitar actions as an experiential musical response from those who do not physically know how to play these instruments – they present an embodied understanding and association of the relative actions that constitute the emanation of that particular sound (Martens and Benadon, 2017, pg. 115). This is due to neural activation occurring even when only listening to musical rhythms. Henry and Grahn propose that this automatic activation of motor regions in the brain and the connectivity between auditory and motor regions during rhythmic listening underlies the connection between rhythm and movement, suggesting that movement can in turn reinforce the perception of rhythm (Henry and Grahn, 2017, pgs. 63, 65-69, 71).

Rhythm is created through regular sonic repetition and the perception of rhythmic groupings is influenced by IOI or inter-onset interval, the time between the onset of one tone to the next (Klein and Posner, 2019, 04:00; Thompson, 2015, pg.96). Listeners perceive temporal organization most easily when IOIs fall between a particular range. When they are less than 100 milliseconds (ms) rhythmic sequences are generally perceived as being a continuous event and whereby they exceed 1.5 - 2 seconds they perceive them to be separate events (Thompson, 2015, pg.96; Martens and Benadon, 2017, pg.119). For temporal structures within this range listeners tend to perceive rhythmic patterns with a duration of up to five seconds, which correlates to the approximate limit of auditory sensory memory (Thompson, 2015, pg. 96). This suggests a means to form new rhythmic groupings through IOI and implies that providing the usually clustered percussive elements (such as a typical drum kit) retained their usual rhythmic schema their coherence should not be affected by the periphonic panning or the deconstruction of their usually fronted group-staged cluster.

The pitch height effect as previously mentioned could contribute to this effectiveness of percussion at height, if the percussion contains enough responsive high frequency content to enhance the perception of height placement.

Melodic contour refers to the upward and downward spectromorphology of a melodic structure. Thompson (2015, pg.77) states that listeners' mental representations of melodies contains contour information and that it is retained for a longer period of time than other structural elements, like interval size or absolute pitch. This suggests that the spectromorphology (shape over time), or the movement of the melody is more memorable than some of the musical elements that form the sonification of the melody itself. This was an interesting consideration in the experimental construction of the spatial stages within this project, as research suggests the embodiment of melodic contour and the rhythmic elements could possibly help to create or enhance the movement of sound sources within periphonic placements.

3.5 - Summary and Conclusions

Part one evaluates how we traditionally stage virtual representations of real-world performance in two-dimensions and provides the foundation of knowledge for informing music production staging practices. In part one I outlined the four dimensions of sound source placement as defined by Dockwray and Moore and discussed how these may apply to mono, stereo and surround sound staging practices, further noting which elements are illusory and how the presentation of these illusory elements are perceived to change across reproduction format. Additionally, part one concludes that non-realistic and musical meaning based staging approaches are often drawn upon to promote a sense of hyper-realism in stereo productions, and that this is in part due to the fronted and horizontal presentation of the phantom image requiring more creative mechanism for constructing the illusion of a multi-dimensional sound stage.

Part two builds upon and contextualises the concepts explored in part one by examining the systems that underpin our human perception. Part two provides an understanding of how we perceive real-world phenomena and highlights the importance of psychoacoustics and sonic embodiment in relation to music production and spatial audio technologies. I evaluate the nature of spatial audio, the current systems being used and how binaural phenomena may be employed in a creative way. Part three concludes that high-quality HRTFs are important in generating a coherent multi-dimensional staging illusion; that higher frequency content is fundamental to facilitating elevation perception, and that employing binaural DSP through object-based staging affords a reflexive approach unattainable using traditional binaural recording and rendering methods.

Part three builds on our understanding of perception in exploring how we make sense of real-world phenomena through cultural communication systems such as proxemics and memetics. Here I outline how these communications systems provide the culturally-informed cognitive mechanisms for constructing, interpreting and drawing meaning from experiences. What we can conclude from part three is that, in taking an ecological approach to sound staging, proxemics and memetics provide a vehicle to enhance the sense of surrealism in a spatial music production; surreally creating the impossible through metaphorical schematic representations of the possible. In part three I also introduce the concept of spectromorphologic analysis, concluding that it is a useful approach for the purpose of reverse-engineering sonified meaning in music composition and production contexts.

Part four provides an understanding of musical phenomena, further shaping our knowledge of the embodied Human perception. Here I examine musical cognition and explore how mechanisms of musical perception may be exploited in the construction and interpretation of metaphorical sonic schema. I noted how the brain works to differentiate particular musical events as being either significant or less significant and how this idea of segmentation can be applied to navigate the construction of large-scale spatial music schema. Part four further concludes that highly embodied musical attributes such as rhythm, melodic contour and pitch could help to create or enhance the perceived movement and elevation of sound sources within periphonic placements.

Holistically, these four parts interrogate the value systems underpinning the construction of spatial music production aesthetics. Together they provide the conceptual-musicological knowledge required to build a theoretical framework for constructing surrealist spatial music schema that metaphorically evoke concepts of real-world phenomena. This framework is important to the project's trajectory as it not only informs the conceptualisation of creative ideas but because it also influences the determination of appropriate strategies for thoroughly exploring and answering the research questions.

4 – Methodology

In order to answer the following primary research questions 'How can non-front orientated sound stages for music be approached and structured?' 'How does this new surround sound with height presentation affect mix practice?' and 'Does it correlate to or deviate from the sound box framework proposed by Dockwray and Moore?' a bricolage methodology was employed using a research as practice paradigm and experimental phenomenological methods. This approach affords investigating the research problem (1) and evaluating the results (2) within the context of which they exist; 1 - the problem concerns practice, 2 – judging the result concerns experience. In keeping with research as practice strategies and experimental phenomenological method, the method pertaining to quality judgements involves a tri-partite set of criteria defined through the research as practice paradigm; Artistic (evaluated through individual experience), Pragmatic (evaluated through communal experiences) and Activist (evaluated dependent on ethical/political agenda). A quantitative method involving A/B testing is employed in parallel to the ‘Pragmatic’ criteria outlined above. This mixed method approach aids in informing interpretation of the associated qualitative data and further reinforces its validity.

However, in order to answer the primary research questions above, the following preliminary research questions must first be addressed ‘What systems could be used to approach mixing for a periphonic-binaural format?’ and ‘How greatly would the variation of headphone affect the perceived musicality and periphonic translation of the music? A bricolage methodology was again employed but here we see a convergence in critical theory and research as practice paradigms with experimental phenomenological methods. The adoption of critical theory in this approach is necessary in informing the entire trajectory of the project. Through this paradigm I critically navigate the socio-historic impact spatial audio technology has on the development of practice. Further, this informs the measurement parameters surrounding the ‘Activist’ quality judgement criteria, mentioned above, which is employed in evaluating an appropriate toolkit for conducting the practice-led inquiries.

4.1 - Methodology and Research Phases

Methodology - Overview

To introduce the concept of research methods we can refer to Hughes and Sharrock (2007, pg. 35) who state 'method' and 'methodology' as being two distinct but relative elements of research practice. 'Method' defines those techniques employed to discern and collate data regarding the line of inquiry. Methodology, however, "examines the logic and rationale which underpins the use of particular methods" (Roberts, 2014). The function of methodology is therefore to critically enquire into the claims of the specific methods, with the methods lending credence to the often more abstract assertions of a methodology (Roberts, 2014; Ruane 2005, pg. 48-49).

Across most fields there are two standard ways of conducting research; qualitative and quantitative. Qualitative research, as opposed to quantitative, refers to subjective, non-statistical, non-numeric data that requires interpretation and in this project this specifically amounts to experiential phenomenological data gathered through the preliminary and primary practice as research and through the focus group questionnaires. Though the project mostly presents qualitative data, there is underpinning quantitative data (objective and quantifiable historical, numerical and statistical sets) in the form of historical evaluation, market analysis and the results of the HULTIGEN scaled A/B listening tests[18]. All of the above

[18] The HULTIGEN model, GUI and function is comprehensively detailed in the dedicated sub-section to follow.

strategies, both qualitative and quantitative, stand to represent differing types of 'method' in this respect and are applied appropriately as relative to the different phases and requirements of the research project.

Due to the interdisciplinary nature of the myriad topics underpinning the varied lines of enquiry it is necessary to adopt a more pragmatic bricolage approach; combining mixed methods in obtaining both qualitative and quantitative data sets. We will define the bricolage methodology framework toward the end of this section and discuss the importance of a mixed method approach in detail within the next sub-section of this chapter. As we will come to understand, a combination of methods is required in order to best triangulate the interdisciplinary nature of both the theoretical and practical research frameworks and to further validate the qualitative data sets. This ensures the most suitable and thorough approach in discerning, analysing and presenting conclusive research from a multitude of source and data types.

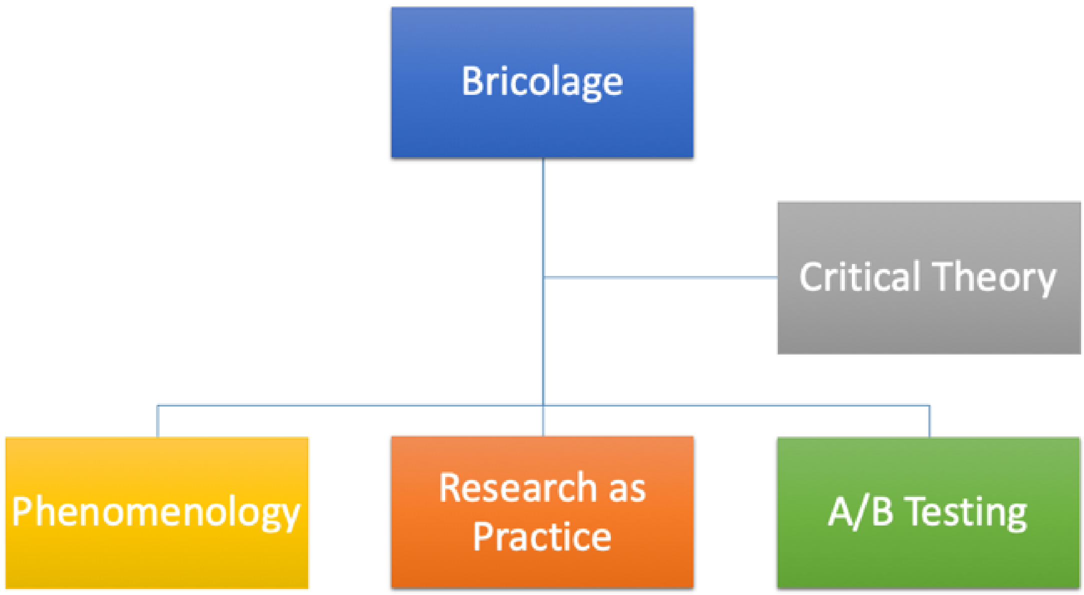


Figure 16: Methodological Framework – an overview.

The methodological framework diagram is representative of the method system hierarchy that constitutes the bricolage framework. In Figure 16 we can see how bricolage is presented as the presiding 'umbrella' methodology that all other methodological strategies branch off from. The first in the hierarchy is *critical theory*. Critical theory is not a methodology or a method but an adopted research paradigm and as such it is presented in series due to the common use of critical perspective employed throughout the study and the fundamental objective of informing practice. Critical theory is a way of questioning that is applied from the early stages of literary research, the hypothesising and synthesising of the methodological and conceptual frameworks through to the critical thinking and quality judgements employed in practice (see section 4.2.1 below).

On the framework diagram (figure 16) above we can see critical theory then secondarily expands in to *phenomenology*, *research as practice* and *A/B testing* which are framed in parallel to one another. In this project phenomenology acts as both a philosophical lens and as a methodology employed within the bricolage framework, with phenomenological methods being fundamental to the undertaking and evaluation of the preliminary and primary research practice (phases 1-2.5 & 3)[20], and in exploring the research problem within an applied context. Phenomenological methods also play a vital role in the phase 4 data collection, whereby focus group sessions are used as a means of conducting listening tests of the project material and

[20] See Figure 17 for a visual overview of the research phases.

extracting third-party experiential data from a small set of voluntary participants using questionnaire forms.

Research as practice, also referred to as practice-based research and practice as research, is an invaluable research strategy for this project as the research problem inherently concerns practice. Therefore, methods drawn from a practice-based approach provide an applied means of investigating the research problem within the context of which it exists (phases 1-3).

Finally, the third branch of the framework is A/B testing. This is not a methodology but a user experience research method for quantitative measurement of a comparison of two stimuli (A & B). This method is used in phase 4 as a parameter to comparatively measure a listener's experience of the new periphonic-binaural mixes produced through this project, against the original existing stereophonic mixes of the same musical content (the A & B stimuli). Utilising this quantitative method as an alternative means of measuring experience helps to reinforce the validity and meaning of the qualitative data extracted within the focus group questionnaires, consequently providing a triangulation of the third-party experiences as a numerical average which does not require interpretation but aids in informing interpretation of the associated qualitative data. We shall discuss the project's methodologies, methods and strategies in more theoretical detail in the section to follow but first we need to provide context to the methodological application by overviewing the four phases of research mentioned above.

Research Phases - Overview

In the previous section I introduced the concepts of method and methodology, and qualitative and quantitative research. I outlined the hierarchy of methodological strategies which comprise the bricolage framework and noted their significance relative to extracting certain types of data. This section details how the phases of research unfold and how the methodological strategies function to provide the differing data types required to answer the research questions posed through each phase.

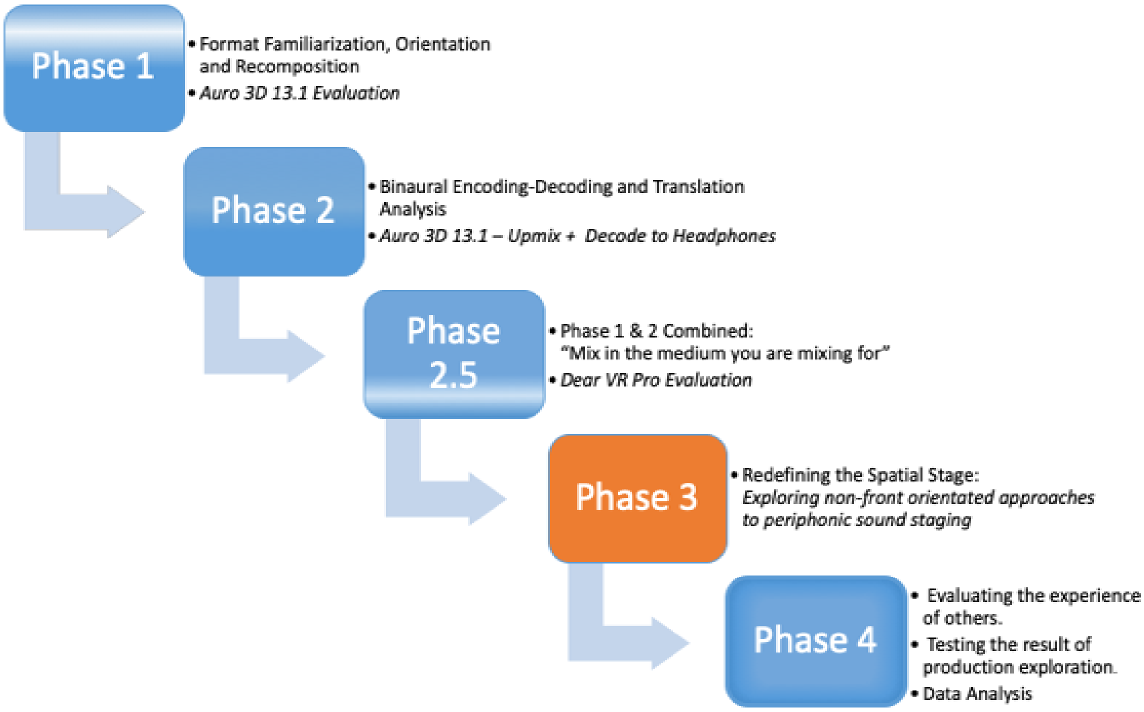


Figure 17: Research Phases – Overview

As detailed in figure 17 above, there are 4 distinct phases to this research project. Phases 1 - 2.5 are preliminary research phases addressing the following question: *'What systems could be used to approach mixing for a periphonic-binaural format?'* The aim of the preliminary research phases is to determine an appropriate method for creating headphone-based spatial mixes. The objective is to evaluate the quality and practicality of both speaker-to-binaural and direct-to-binaural formats.

Quality is assessed through the design and implementation of production techniques that examine problematic areas of binaural localisation and perceptual phenomena. Practicality is assessed through judgment based on the following criteria; accessibility, functionality and integration.

These phases use critical theory and phenomenological method employed through practice as research to explore the phase objectives. These were particularly important phases as only once a suitable toolkit had been defined could the project investigation then progress on to phase 3.

Phase 3 addresses the primary research question; *'How can non-front orientated sound stages for music be approached and structured?'* and is considered the main body of creative research practice.

The aim of phase 3 is to inform contemporary creative practice through defining new approaches to spatial music production. The objectives are to explore how we could construct cohesive and creative non-front orientated periphonic sound stages for music production, ultimately providing a taxonomy of approach to headphone-based spatial mixing through a portfolio of production work. Key areas of inquiry involve examining source audio types, performance versus production music, vocal staging, instrumental placement and the use of spatial effects. This phase employs critical theory and phenomenological method through practice as research to execute and explore the phase objectives.

Following the construction of the portfolio of periphonic-binaural production work in phase 3, the fourth and final phase evaluates the result of the audio work across a variety of listeners using A/B audio stimuli testing and phenomenological questioning. The aim of phase four is to generate an understanding of how others experience the non-front orientated spatial staging presented in the productions. The evaluation specifically questions their interpretation of the production aesthetic in an attempt to not only determine their overall preference (between fronted stereophonic and non-fronted periphonic-binaural mixes) but also to deduce whether their experience of the binaural presentation could be impacted by localisation or perceptual anomalies presented by the generic HRTFs used in the spatialisation process. This phase utilises phenomenological method through the design and implementation of a questionnaire and the use of focus groups, and employs an A/B listening test method for comparative analysis of the stereophonic and periphonic audio stimuli. Qualitative data collated through the questionnaire is analysed using thematic and interpretive phenomenological analyses and can be triangulated against the quantitative data from the A/B listening test to provide a more robust and conclusive meaning. The strategies for analysing the phase 4 data are discussed in the final sub-section of this chapter.

4.2 - Methodologies and Strategies of Research

In the two previous sections I have provided an overview of the project trajectory, detailing the hierarchal structure of the methodological framework and the progression of the research phases. This section follows the structural hierarchy governed by the methodological framework and focuses in on the processes of each of the methods, strategies and measurement parameters employed under the scope of the bricolage methodology. Here I unpick, describe and critically evaluate their specific function and deployment within each phase of research so that we may better understand the role each strategy plays in addressing the research problem.

4.2.1 - Methodologies

What is Bricolage and why is it important to this project?

Bricolage is a methodology that challenges the viewer in seeing the subject matter in an atypical manner, effectively it is the research that constructs the bricolage through means of ‘piecing together’ (Wibberley, 2017). Bricolage may include both existing (found) data (as in the traditional approach to bricolage) and new, or original data produced through, or for, the study specifically. This study uses both as a means to piece the research puzzle together, providing a more thorough and rigorous approach through an appropriately customised set of methods. This pragmatic mixed-method approach is shaped and driven by critical theory, which I use to set the activist goals and judgement criteria required to define the toolkit to progress the investigation. My approach to research as practice embeds the goal setting and judgement criteria informed by critical theory and actions these through experimental phenomenological method. It is this convergence that consequently constitutes the bricolage methodology, and determines the research journey and trajectory of approach to practice.

Above I delineate the bricolage data types; existing/found and new. The existing or found data in this project is initially that pertaining to the text analysis in the previous chapter, such as that critically examining the historical context of spatial music and theoretical context of fronted stereophonic music production. An example of such being the previous dissection of Dockwray and Moore's 'Sound box' article and Moylan's views in 'Considering Space in Recorded Music' contextualised against the historic creative and technological limitations of surround sound and spatial music production. Likewise, it is also presented in the form of the perceptual, cognitive and musicological texts underpinning the theoretical framework that provides governance to the philosophical conceptualisation of the hypotheses for praxis. All of the above literary examples are examined using critical theory and are reviewed in the previous chapter as a means for defining the theoretical framework and providing rationale for the study.

To prevent confusion, we will refer to this existing literary data as 'Secondary Research' in specificity because, as we will discuss later on in this chapter, the new data produced by this project can become existing data within the cyclical practice-based framework that reflexively informs and re-informs practice. Thus the term of existing data can incorporate both the secondary research literature and the older generations of new data produced through the project practice (see conceptual framework diagram below).

The new or original data produced specifically by and through this study is that pertaining to the research problem, as explored through the practical framework and hypotheses. This takes the form of new, experimental processes and approaches to staging and production as synthesised and developed through the project's practice. It also takes the phenomenological form of my experience of the work as gathered through the research as practice processes and of the listening participants' experiences as gathered through the focus group interviews – all of which are thematically interpreted, presented and discussed in the relative chapters further on in this thesis. This original data may also be referred to as 'primary research', and of course as mentioned above, the older generations of new data produced through this project remain within the primary research category, regardless of whether they transpire to become existing data in the re-information cycle.

It is in combining existing and original, and primary and secondary data sets that the project can holistically weave together answers to the research problem(s). Furthermore, when done with a judiciously democratic approach the output is both beneficial and appropriate to informing academic research and professional industry practice, alike. We discuss the democratic approach in detail further on in the critical theory section of this chapter.

Figure 18: Conceptual Framework

This custom methodological framework specifically works through the existing data informing practice, which then births new data that either presents possible solution to a research problem or continues to re-inform practice, regenerating and reusing new-from-existing data until an answer or solution to the research problem is found. This is where quantitative secondary research, critical theory and phenomenological method combine with primary qualitative research as practice in forming the bricolage methodology framework.

Practice is also informed through tacit knowledge, which is implicit knowledge, such as intuition or personal wisdom that may often be attained, or at least informed, through one’s experience. Experience also directly informs practice through acquired knowledge and in real-time through the production process and during the post-production analysis. Informed practice generates new knowledge which then re-informs the three key areas of knowledge input. See the graphic above for a visual diagram exampling this project's conceptual framework system.

However, although contemporarily bricolage is considered an acceptable approach to arts and humanities research, it should be noted that traditionally there have been critical voices from academic peers in regard to the validity of bricolage as a research methodology. This is mostly due to the spurious viewpoint of bricolage not being an actual methodology in its own right - although in fact it is a methodology that collages the use of many methods to piece together a framework - or indeed, it may be due to this contentious 'piecing together', often seen as being a lesser critical, less rigorous approach to research and data collection. Yardley (2008) details David Silverman's scathing critique of bricolage methodology as being that of lazy critical thinking and academic naval gazing. Silverman suggests, that Bricolage sets the stage "for a dialogue of the deaf between itself and the community" (Silverman, 1997, p.240). Subjective and unverifiable - Silverman beds in this view by describing attempts to engage directly (and interactively) with the narratives of research participants (fictional or otherwise) as a "romantic impulse" elevating "the experiential to the level of the authentic" (Silverman, 1997, p 248). In this instance Silverman is presenting a case for ensuring rigour in research, but what is it about the human 'experience' and the 'creative product as interpretive text' that lacks authenticity and rigour? Yardley suggests it is more the presentation of the 'experience' or 'creative product' that may be the issue, the language in which it is conveyed rather than the subject of the text itself (Yardley, 2008).

Yardley suggests the use of multi-layered texts as a means for presenting a case for narrative inquiry as a valid modality of ethical research activity (Yardley, 2008; Ledbetter, 1996; Nussbaum, 1990), importantly also acknowledging the notion that human cognitive networks process information and make meaning in their own way. Yardley refers to the terms narrative inquiry as broadly including visual, verbal, musical and written narrative texts (Yardley, 2008; Gardner, 1985) and further states that:

"To a certain extent the presence of a number of different kinds of texts within the meta-text provides a continual ground for self-critique not possible within a mono-text…. The literary, visual, and audio-visual texts are intended to speak directly to one another, to provide alternative interpretive angles to inform and enhance each other's meaning." (Yardley, 2008).

In respect to this project, the multi-layered texts can be found in the following guise, see Table 3 below.

	Text Layer Form	Data Type	Phase Collected
1-	The theoretical and historical background texts that inform the scope of research.	Quantitative	Pre-phase 1 (0) Phase 2 & 2.5
2-	The processes in generating the practical body of audio work, including the preliminary research and the audio work itself (artefact).	Qualitative	Phase 1 - 3
3-	The reflexive phenomenological discourse examining the audio work and one's practice.	Qualitative	Phase 1 - 4
4-	The phenomenological experiential data from the focus group questionnaires.	Qualitative	Phase 4
5-	The numerical HULTIGEN scaling data of the focus groups.	Quantitative	Phase 4

Table 3: Forms of the multi-layered text and their phases of generation

It is in this collation, interaction and presentation of multi-layered text that this project requires a mixed method approach to bridge the gap between the different types of data sets and philosophical schools of thought in connecting the varying bodies of knowledge. This allows us to widen the scope of investigation and extend the boundaries in exploring the hypothesis and philosophical research questions (Yardley, 2008).

The above assertions by Yardley strongly reinforce why a bricolage approach is explicit for an arts and humanities research project that is studying an interpretive product, such as art or music, in more than just a theoretical way.
The ‘creative product’ is that which is in question, yet the ‘creative product’ needs to be perceived to be evaluated. Therefore, the ‘creative product’ cannot be examined or verified without drawing upon the human experience as a means to generate the data required in evaluating the structural values, interpretation and efficacy of the ‘creative product’ itself. It would certainly lead to less authentic data if I were to evaluate solely my own experience of the work without due consideration of how other people may also perceive it. This is an especially important point where binaural audio and generic HRTF’s are involved because, as discussed in the previous chapter, we know that historically they may provide an irregular experience across users. This is what drives the rationale of the phase 4 listening tests, and the incorporation of phenomenological method – to explore the interpretation and presentation of the artistic work from a multiplicity of experience.

Therefore, to ensure rigour and truly elevate ‘the experiential to the authentic’ it is fundamental that a mixed-method approach oversees the multi-faceted aspects of study, incorporating and generating quantifiable and experiential data from several angles in ensuring the provision of a rich data set required to generate meaning from varied sources of knowledge and to provide a well-informed outcome.

To conclude this section and open the next, what we can understand from this is that when conducting practice-based research that uses and yields such varied and cyclical data types (music, experience, practice, literature), we may have to use a bricolage approach to custom build a set of methods that are specific to the lines of enquiry and that provide the most suitable means of data collection, interpretation, analysis and verification. It is important to understand that these artistic works (music productions) can be used as ‘interpretive texts’ as their original purpose is created for interpretation. They are designed to provide an experience for one from the creative process of another. Therefore, it seems logical that we study these forms in the way in which they exist or are intended to function. That is, if we intend to study the experience of music, then we do so through collating the musical experiences of real people, be it in the process of creation and or the experience of consumption. If we intend to inform industry approaches to record production and integrate new techniques into content creation, then we need to study the practical processes that underpin the content creation and ensure we do so in a democratic way that gives all creators the opportunity to explore and further develop the knowledge through practice.

Critical Theory as a Research Paradigm

In the chapter's introductory overview, we mentioned critical theory being employed as a research paradigm informing the breadth of this research project. It was used as a perspective in evaluating and analysing the literary secondary research to synthesise the hypotheses; conceptualise the research problem and most importantly, in formulating the democratic approach this thesis takes toward implementing and exploring the research practice. In essence, this demonstrates that it informs all research phases and the bricolage methodological design. The nature of a bricolage methodology is pragmatic in that it asks 'how can I conduct this research in a way that is available and accessible to me?' In this instance, critical theory is applied to further inform and expand on this question by asking 'how can I conduct this research in a *meaningful* way that *makes it* available and accessible *to all*?' It is in the reformulation of this methodological question that critical theory can be seen to inform the bricolage umbrella methodology and it is in answering this question through research as practice that we can see how it is so deeply embedded in informing how this project develops. Critical theory determines both the means of conducting the research; drawing upon set goals and experimental phenomenological method in informing the toolkit required in practice, and in evaluating it; through defining a set of judgement criteria and deploying the A/B testing method.

To introduce the concept, we can think of critical theory as being research that inquires 'against the grain'; research that challenges conventional methodologies and knowledge bases that make claim to scientific objectivity, regardless of whether quantitative or qualitative. This is in order to ask questions that go beyond prevailing assumptions and understandings, and to acknowledge the role of power and social position in phenomena (Intgrty, 2016). Given (2008) refers to critical theory as

> "A foundational perspective from which analysis of social action, politics, science, and other human endeavours can proceed. Research drawing from critical theory has critique at its centre (assessment of the current state and the requirements to reach a desired state)."

As simplistic as these definitions may seem, critical theory is a complex merger of two different schools of thought, both of which are based upon a critique of society and culture; Marxist theory and the ideas of the 'Frankfurt school' (Bohman, 2010). Belshaw (2011) states that the former has a more normative dimension (there is a way in which the world 'ought' to be) and the latter a more hermeneutic approach (gaining knowledge through interpretation of texts). It was the postmodern critical theorists such as Michael Foucault and Jean Baudrillard that initially fused the two streams in considering everything to be a 'text' and therefore open to multiple interpretations. The 1960's onwards saw the social sciences redefined by Saussure, Derrida, Chomsky and Barthes as dealing with symbolic representations of the world but it was not until the 1980's that the fusion of the two schools of thought became complete when Jurgen Habermas defined critical theory as being a theory of communication (Belshaw, 2011) and "the mode of inquiry which participants may adopt in their social relations to others" (Bohman, 2010). In the 1990's Horkheimer defined critical theory as being adequate only if it is simultaneously explanatory, practical and normative (Bohman, 2010; Belshaw, 2011). This means that it must explain the issues of social reality, identify the necessary actors to change it and provide both the clear norms for criticism and achievable goals for social transformation (Bohman, 2010; Belshaw, 2011).

Critical research attempts to reveal the socio-historic specificities of knowledge and how certain knowledges may reproduce structural relations of oppression and inequality and as such, is concerned with the critical meaning of experiences as they relate to varying systems of social oppression (race, gender, class etc.) (Intgrty, 2016). In research drawn from critical theory valid knowledge arises from the critique of the social structures and systems, as defined through the analysis of current discourse in society relative to the systems within which they operate. The aim of which is to reveal the power relationships within the system and its structures so that the oppressive nature of the system may be revealed (Intgrty, 2011).

How does this approach apply to the study of spatial music?

As detailed in both the introduction and literature review, and as further discussed within the preliminary research chapter to follow this, historically spatial *music* research (and practice) has long been a topic confined to the annals of academia and a speaker-based electro-acoustic autocracy. As we have learnt, speaker-based surround sound presents several difficulties as a music format and through its own construction and consequent lack of adaptation to societal behaviour, it presents a literal system of its own oppression established through the limitations it poses by default of design – that is to say that it were the specificities associated with the large and expensive multi-speaker configuration, the creative limitations that the system presents and the inability to accommodate the cultural-move to headphone-based listening that prevented it from assimilation as a popular music production format. There we have evidence of *oppressive system 1 – the lack of a democratic application in the production and delivery of recorded spatial music.*

It is from this system that further oppressive micro systems can be realised. It goes without saying that academia has long remained the realm of the middle-aged white man (Woolston, 2020; HESA, 2020), and as such, much of the historic research surrounding spatial audio and spatial music caters to the interests and needs of that demographic. Superficially there is nothing particularly wrong with this, however when we dig deeper into what this means systemically it - a) tends to be focussed on electro-acoustic / experimental / technical research with limited creative practice, thus reinforcing spatial audio as an academic niche, and b) up until recently the spatial audio systems used could not cater for a headphone-based (re)production and, as such, both of these things enforce a lack of application to the wider industry, do not correlate with the interests of popular culture and limit other demographics from accessing and exploring spatial audio technologies and practices. “One cannot expect positive results from an educational or political action programme which fails to respect the particular view of the world held by the people” (Freire, 2012, pg. 95).

In this we have covered step 1 of the critical theory approach – we have explained the problem with the social reality, and now step two requires us to define the actors for change and how can we implement it.

To take a democratic approach to spatial audio research would require input and exploration from a variety of researchers and practitioners whom do not fall solely into the above category as a start. Fundamentally to this, however, it also requires a low-cost or open source spatial system that can reflect and accommodate the current social behaviour of headphone-based listening whilst also being flexible enough to work within any standard DAW (digital audio workstation). It must also provide an accommodation for legacy audio formats for both the source and output delivery (mono / stereo wav files etc.). These criteria ensure flexible integration within past, present and future DAW projects using a typically normative approach to music production; a non-specialist software and standard audio formats.

Once the equipment and processes fulfil the above criteria and can be integrated into anybody's[21] workflow, then we have the beginnings of a democratic system for spatial production. This accessibility could then generate more interest in popular culture and consequently afford a broader creative input in application and a more varied research output as a result. *To achieve a democratic approach to spatial music production and consumption both the means of creation and delivery need to be accessible to everybody.* This idea of democratising spatial music making drives not only the prescription of technology used within this project but also how it is

[21] Academic, industry practitioner, student and enthusiast, alike.

used to inform practice in the creation of system agnostic *transferrable* production technique.

Though I do not purport to be the agent of change, nor do I think that a fully democratic approach will be possible in a capitalist society (the discussion of which is beyond the scope of this thesis), I do value that this research should aid as a step to bridging the gap between academia and industry in providing knowledge that informs both research and wider creative practice. “The starting point to organizing the programme content of education or political action must be the present, existential, concrete situation reflecting the aspirations of the people” (Freire, 2012, pg. 95)

Please see the *preliminary practice as research* chapter for more detail on the democratic toolkit prescribed for this work and the process of defining it (research phase 1 - 2.5). Please see the *redefining the spatial stage* chapter for a discussion regarding the transferrable techniques aforementioned (research phase 3).

Experimental Phenomenology: a governing philosophy and methodology for research praxis

As previously outlined, a phenomenological lens drives this entire investigation due to the study of the human experience that underpins the development, interpretation and evaluation of applied musicological practices in this context. Experimental phenomenological methods (doing phenomenology) work in conjunction with research as practice by providing an approach to unpick and analyse practice through the experience of it. This approach is applicable across all phases of research and is not confined to a solely creative or evaluative process, as these processes are regularly intertwined within the research practice where evaluation of technique is cyclical to the creation of it. Phenomenological method is informed by the goals set through the critical theory approach and is employed to execute the process of judgement in prescribing the toolkit to conduct the research practice. Further, experimental phenomenological method is integral in triangulating my experiences and evaluations of the practice against those of others. It is employed in parallel to the quantitative A/B listening test method to gather qualitative data on communal experience, and fundamentally provides a richer data set to draw meaning from.

Phenomenology is an umbrella term that encompasses both a philosophical school of thought and a range of approaches to research. The phenomenological movement was initiated by Edmund Husserl as a radical new approach to philosophy which focusses on consciousness and essences of phenomena and is known as transcendental phenomenology.

Martin Heidegger developed this further and moved away from the philosophical discipline of transcendental phenomenology toward elaborating existential and Hermeneutic (interpretive) dimensions (Finlay, 2009; Kafle, 2011). Paul Ricoeur's (1975) work states that this hermeneutic development can include but is not limited to; symbols, imagery, sonics and linguistics. Hermeneutic phenomenology is more concerned with historical meaning and experiences and their developmental and social effects on individuals.

As applied to research, phenomenology is the study of phenomena; their nature and meanings. Finlay (2009) states that the phenomenological focus is on the way in which things appear to us through our conscious experiences, where the researcher aims to provide a rich, textured description of lived experience. Langdridge (2007, pg.4) defines phenomenology as a discipline that "aims to focus on people's perceptions of the world in which they live in and what it means to them; a focus on people's lived experience". Langdridge further states that as a qualitative method phenomenology focuses on human experience as a topic, concerned with meaning and how meaning arises through experience. Lester (1999, pg. 1) states that "phenomenological methods are particularly effective at bringing to the fore the experiences and perceptions of individuals from their own perspectives, and therefore at challenging structural or normative assumptions". This supports the use of a mixed-method design, particularly the fusion of critical theory with phenomenological method and practice-based approach in examining the research problem through a multiplicity of lived experiences.

Merleau-Ponty (1962) writes that the aim of phenomenology is the description of phenomena, and identifies four strategic characteristics that are common to the varied branches of phenomenology; *description, reduction, essences* and *intentionality*. Description is just that, a description of the phenomena. Reduction is a process in which the phenomena is suspended so that the 'things themselves' may be returned to. Essences are the core meanings of an individual's experience that make it what it is and intentionality refers to consciousness; the total meaning of the object or idea.

As previously defined, transcendental and hermeneutic are two well-known 'classic' branches of the phenomenological discipline, another branch is that of experimental phenomenology. Don Ihde (1986, pgs.1-4) articulates that experimental phenomenology focuses on 'the doing' and that this is applied in research through experiments, observation, description and reflexive discourse. "Without doing phenomenology, it may be practically impossible to understand phenomenology. Phenomenology, in the first instance, is like an investigative science, an essential component of which is an experiment."

Heidegger claims that "Phenomenology is our way of access to what is to be the theme of ontology, and it is our way of giving it demonstrative precision. Only as phenomenology, is ontology possible" (Ihde, 1986, pg.6). The research ontology dictates the epistemology; the way in which we study the project subjects should reflect the way in which the subjects function (Wisker, 2008, pg.67-68). In this case specifically working with the phenomenological notion that though our environment may dictate particular sensory information and on a scientific level we know this to be true, our perception of the sensory material will be dependent on many internal and external factors and relationships that lead us to not necessarily all perceive the same material, individually, in quite exactly the same manner but that as an average, said experiences and perception can be generalised to generate objective data and apply meaning to knowledge. This is an important consideration when constructing musicological research methods and collecting and analysing data, especially subjective data such as music, or rather the experience of music, that then requires contextualisation and interpretation (Wisker, 2008, pg.66).

The research framework reflects an intrinsically phenomenological approach whereby the human experience underpins the topic and nature of study. This can be evidenced through the cataloguing of my own observations and 'first-hand experiences' within the research practice and the consumer-focussed, end-point evaluative reception of the resultant musical output – referred to in this project as 'second-hand experiences'. It is in this combination of both first-hand and second-hand experiences that a richer data set can be attained and the subjective experiences of all participants may be analysed, thus providing a more meaningful and more objective interpretation than any evaluation of a singular experience could alone.

This project lends itself more to the experimental and hermeneutic lines of phenomenological inquiry with both branches being directly relative to the conceptualisation, synthesis and practical exploration of the project hypotheses. The hermeneutic approach orientates more toward the philosophical, historical and theoretical contextualisation of the research problem and the conceptualisation of the creative inquiry. Whereas the experimental approach orientates more toward the exploration of the hypotheses in praxis. Both can be applied in the analysis and evaluation of the musical work, where the '*meaning'* of the sonic structures are assessed through *'doing'* phenomenology in practice.

By using the bricolage way of 'piecing together' we can see how the characteristics of these two branches fit well with critical theory and practice-based research methods (discussed in the next sub-section) to holistically fulfil the requirements of the theoretical framework in exploring spatial music staging through experiments in practice (the doing) and evaluating how people experience and interpret the creative staging product (the meaning).

We have already discussed the common criteria of description, reduction, essences and intentionality outlined by Merleau-Ponty as being fundamental to phenomenological approach. Further to this, typical methods of phenomenological research may include interviews, conversations, action research, focus group meetings and analysis of personal texts and that, as a principle, minimum structure and maximum depth, constrained by time and opportunities in practice, will aid to retain focus on the key aspects of research whilst avoiding undue influence from the researcher (Lester, 1999, pg. 2).

Research as Practice, Practice as Research – a practice-based approach to creative research.

This section outlines the importance of a practice-based approach in the undertaking of this research project, arguing that just as a phenomenological approach is important in the examination of experience, a practice-based approach is intrinsic to examining practice and the processes that constitute it. So where the two meet in study, as in the case of this project, it is just as important to both document, explain and evidence the creative process as it is in describing and interpreting the phenomena relating to the experience of it. The two ultimately go hand-in-hand in developing and informing practice using both the process and the experience of the practice as tools for examination. As we shall come to learn in this section, the judgement criteria defined through the research as practice paradigm works with the democratic agenda set out through the critical theory approach. This provides the measurement parameters needed to establish how the practice is conducted, which ultimately answers the methodological question presented earlier - 'how can I conduct this research in a *meaningful* way that *makes it* available and accessible *to all*?' This symbiosis of paradigm and method further examples the pragmatism of the bricolage methodology in informing the trajectory of approach to creating, collating and analysing varied types of data.

Practice as research, research as practice or research through practice – all interchangeable terms with the same meaning - is a newly emerging paradigm for creative research practitioners. Zagorski-Thomas (2015, pg. 28) observes that there is an amount of academic controversy pertaining to practice-as-research due to research and practice being generally regarded as distinct activities from one another and that it is seen as being impossible to engage in one ‘as’ the other. He references Ingold’s (2011, pg.419) likening of academic research to creative practice, stating both to involve “puzzle-solving… carried on within the context of involvement in a real world of persons, objects and relations” and argues that all knowledge is both embodied and related to the world one inhabits and that there is no such thing as disembodied or abstract knowledge. Zagorski-Thomas claims that although the written word (of research) provides a schematic representation of the rich communication process of speech and bodily experience, it can only be understood through reference to a lived experience (2015b, pg.28). This is where the phenomenological approach meets research as practice, in providing the reference to lived experience relating to both the process of undertaking practice and the interpretation of the audio work in production. To reiterate Finlay (2009); the phenomenological focus is on the way in which thing’s appear to us through our conscious experiences, where the researcher aims to provide a rich, textured description of lived experience.

When we document the acts or processes manifesting through the lived experience alongside the lived experience as experienced (i.e. recording the sessions and noting phenomenological description), we can then access the acts, processes and experiences from multiple angles, thus making the research richer. This is how we generate the multi-layered texts required of the bricolage methodological model.

Similar to the ontological and epistemological rationale made for phenomenology in the preceding chapter, Zagorski-Thomas (2015b, pg. 29) states that “If it is the ‘practice’ that is object of the research, then it is the practice that should be studied” and for practice to be constituted as research it needs to focus on the process of practice (not just the output) and provide a combination of method, theory and evidence to support it (Zagorski-Thomas, 2021, 07’:50”). He suggests that research outputs for practice-as-research should document, explain and evidence the creative process and/or the interpretation through which it may seek to embody knowledge. This can be validated further through the Research Excellence Framework (REF) stipulations on what constitutes research;

> “For the purposes of the REF, research is defined as a process of investigation leading to new insights, effectively shared... It includes the generation of ideas, images, performances, artefacts including design, where these lead to new or substantially improved insights.” (Higher Education Funding Council for England, 2012, pg.71).

What we can understand from this is that practice as research needs to intrinsically investigate and empirically explore the areas where practice is of key concern and the source of primary research data (Wisker, 2008, pg.67 and 74). Therefore, practice as research is vital to this framework because the research problem encapsulates and concerns practice - practice is inherent to the research problem – because the aim of the study is to inform approaches to spatial music production. Thus, informing practice through artistic practice-research is a fundamental step to producing the data required to inform the research problem.

What are the strategies involved with conducting practice as research and how do they apply to this work?

Zagorski-Thomas (2015b, pg. 30) articulates that the work should publish peer-reviewed outputs that are not specifically artworks but that may include audio, video or multi-media presentations that clearly convey the new and tacit knowledge involved in the process of practice as research. However, he clarifies that there should be a clearly defined research question to provide context to the artistic practice and an established set of criteria under which the researcher may assess the efficacy of their intentions and how the artwork affords or suggests a particular interpretation. In doing so, the artist/researcher can provide demonstrable examples of how a particular technique or approach may result in a particular set of physical attributes within the artwork. “The important point is not the specifics of the theoretical model but the fact that it should constitute a coherent and consistent basis for understanding how the artwork suggests potential for interpretation” (Zagorski-Thomas, 2015b, pg.30). He refers to these assessment criteria as ‘quality judgments’ and articulates that often in creative practice research these may be seen as being less objective to those established in science or social-science but that this is a key challenge of the practice research process: to not simply produce the research but to also establish the rigorous criteria to which it can be judged (Zagorski-Thomas, 2021, 04’:30”-05’:35”). Further to this, Zagorski-Thomas (2021) argues that how the quality judgements are set is dependent on the approaches to the practice research processes undertaken and provides the following tripartite criteria;

- Artistic: Quality judgments set by individual.
- Pragmatic: Quality judgments set by community.
- Activist: Quality judgments set by an ethical or political agenda.

	Quality Judgment Type	Research Goals	Methods	Reasoning
Autoethnography	Accuracy of interpretation (communal)	To understand	Qualitative	Inductive
Practice Research (artistic)	Aesthetic (individual)	To understand and improve (improvement criteria part of the research problem)	Qualitative	Inductive and deductive
Practice Research (pragmatic)	Aesthetic (communal)	To understand and improve (improvement criteria part of the research problem)	Qualitative	Inductive and deductive
Practice Research (activist)	Ideological	To understand and improve (improvement criteria set in the research question)	Qualitative	Inductive and deductive
Action Research	Measurable / ideological	To understand and improve (improvement criteria set in the research question)	Quantitative or qualitative	Inductive
Applied Research	Measurable / ideological	To improve (improvement criteria set in the research question)	Quantitative	Deductive

Table 4: Delineations of practice research paradigms against autoethnography, action and applied research models (Zagorski-Thomas, 2021)

Referring to Table 4 and the bullet points above, we can see that for artistic practice research, such as this, the quality judgments are set by the individual and are defined based on the aesthetics pertaining to the criteria relative to the research problem and the goal of the research practice. The quality judgment criteria for assessment of this work differs across the varying phases of research and are dependent on the goal of the phase and what it is that is to be assessed. The phases have been overviewed earlier in the chapter, however in this section we will draw upon them to show the quality judgement criteria they each require.

The first research phase requiring aesthetic quality assessment is phase 2, this is a preliminary research phase that defines the binaural spatialisation method to undertake the main inquiry. Phase 2 involves encoding and decoding the Auro 13.1 speaker-based mixes to binaural format to determine the best approach to achieving a headphone-based spatial delivery. This is referred to in this project as the ‘upmix-decode’ system and the quality judgments are designed relative to the goal of the binaural coding process; to establish the most appropriate method of binauralising the spatial speaker-based mixes. ‘Appropriate’ in this respect delineates that the mixes translate accurately, reflecting the original 13.1 mix structure, spectromorphology and sound source localisation (1 - aesthetic artistic judgement). There is also a secondary activist judgment in this that assesses the user-operability of the ‘upmix-decode’ system implemented. This focusses on evaluating the complexity and accessibility of the encoding and decoding processes in order to define whether it is in-keeping with the democratic ideology the project defends (2 – ideological activist judgment). The works will have passed the quality assessment threshold when both of these criteria can be fulfilled; the quality of translation reflects the original intention without the presence of any impeding anomalies and the process of en/de-coding is accessible, integratable and effective.

The method for judgment 1 involves listening to the binaural en/de-coded mixes across several headphone types (over ear, on ear and in ear) and phenomenologically describing and noting the resultant experience, paying particular attention to the translation of certain musical structures and sound sources, and offering reflexion on the following questions;

- Does this reproduction reflect the original staging intentions of the speaker-based mix?
    - Do I perceive sound sources from above?
    - Do I perceive sound sources from behind me?
    - Do the sound sources localise in reflection to their intended placement in the sound field?
- Are there any perceivable anomalies in the presentation? and if so,
- Where and when do they occur?

Using several headphone types helps in defining both how the differing qualities of the headphones may affect the binaural audio output (identifying variants) and as such provides a better understanding of how the coding process itself has affected the audio across reproductions (identifying invariants). The knowledge achieved through this process will then determine the efficacy of the coding system in question. For example, if ‘no’ is answered to the first question above then we know the system to immediately be ineffective, initiating further reflexive inquiry or the testing of a different system.

The method for judgment 2 involves evaluating the coding process and assessing how ‘easy’ it was to perform the coding. This is an evaluation formed on the following criteria; the accessibility in acquiring the tools and the ease of use in understanding and implementing the functions of the tools. This judgment is employed through reflexion on the following questions:

- Does the processing tool require specialist acquisition?
- Does the processing tool require specialist knowledge to implement?

- Did this process negatively impact the creative workflow?

If the answer is ‘no’ to each of these questions, then the second quality judgment has been passed.

This approach to judging practice-based research affords continued reflexivity and so if the practice work fails to meet either one or both of the judgments, then the knowledge attained through both the processes of practice and assessment may be used to revise the work and reformulate the methods and approaches (which led to phase 2.5).

The next phase requiring quality judgments is that of phase 2.5 - a reflexive off-shoot of phase 2 due to phase 2 not passing the quality judgments. This research phase again evaluates the aesthetics of a binaural output but does not concern itself with judging translation from a speaker-based format as phase 2 does, instead it focusses more on the periphonic qualities that the *direct-to-binaural* tool affords and whether the tool is in-keeping with the democratic ideology.

To assess the democratic ideology of the Dear VR processing tool the same method and criteria for judgment 2 above is applied. For the aesthetic assessment a slightly revised approach to judgment 1 above is required and is as follows; the revised method for judgment involves recomposing the previous 13.1 mixes directly to the binaural format using Dear VR Pro (a virtual reality audio spatialisation tool) and the 13.1 mixes as references in ensuring the structure, spectromorphology and sound source placement reflects that originally created upon the speaker-system. This is important as these elements have been designed through phase 1 to specifically assess the typically problematic localisation areas and anomalies common to binaural listening. The same approach then applies in terms of listening to the resultant output across the same variety of headphones as before (over ear, on ear and in ear) and judging the quality of the binaural aesthetic using phenomenological description, noting the variant and invariant properties of each experience and offering reflexion on the following questions;

- Does the production result reflect the original artistic intention?
    - Do the sound sources localise in reflection to their intended placement in the sound field?
    - Do I perceive sound sources from above?
    - Do I perceive sound sources from behind me?
- Are there any perceivable anomalies in the presentation? and if so,
- Where and when do they occur?

The works will have passed the quality assessment threshold when both of these criteria can be fulfilled; the quality of the *direct-to-binaural* production reflects the original artistic intention without the presence of any impeding anomalies, and the process of binaural spatialisation is accessible and effective. If the quality judgments have not been met then this provides another opportunity for reflexion, reformatting and recursion, or a new line of inquiry.

It is upon successful judgment of phases 2 or 2.5 that we can establish the technology required to move on from the preliminary research practice and fully explore the main research problem *'How can non-front orientated sound stages for music be approached and structured?' (Phase 3).* The aesthetic judgement criteria of which are based around the exploration of this question in practice and are wholly dependent on the hypotheses presupposing the individual structures created. Reflexive questioning takes the general form of; ‘what am I intending to achieve with this practice?’ and ‘how does the result achieve what was intended?’ We can use the above reflexive questioning to examine whether the structures fulfil the criteria as set out within the research question; how do they provide a non-front orientated approach to structuring a musical sound stage?

In respect to the criteria outlined by Zagorski-Thomas above, as applied to this research practice the preceding chapters provide the required context and a well-defined research question, and this chapter details the quality judgments, methods and methodology behind the research journey. Although the processes of artistic practice have been documented through audio or video recording and session notes, the publishable media presentations do come in the form of artworks (creative product / sonic artefact) and multi-layered text. The interaction of the multi-layered text and efficacy of the artworks and the associated practice is presented and evaluated through the research practice discourse. The discourse is organised within this thesis as a set of case studies examining practice. The case studies are formed through the phenomenological session notes and recordings, and the research as practice multi-media documentation data. These case studies draw upon the data from phase 4 to further examine the meaning generated through the work (the experiences of the focus group listening sample; their questionnaires and the A/B test data). The A/B method of phase 4 is discussed in the proceeding section.

Please see the appendices section for URL links to the published artefacts.

4.2.2 - A/B Listening Test: Method, Parameters for Measurement and Data Analysis

As with the above, these additional strategies are informed by methodology but are fundamentally constituted by method. Where phases 2-3 require the aforementioned individual aesthetic quality judgments necessitated by the research as practice framework, phase 4 involves implementing communal aesthetic judgments of the resultant creative work. This combination of individual judgment in practice and community judgement of result helps to formulate a deeper understanding of how the developed non-front orientated staging approaches may be perceived and interpreted by the wider community and as such it provides another lens through which the quality of work may be assessed and improved. This process intrinsically relates to the pragmatism of the bricolage methodology and the ideological impetus of the critical theory element of approach, both of which are further reinforced through the inclusion of this additional method of judgement, which too further increases the richness and validity of the data tapestry and aids extraction of meaning.

The aim of gathering this data is to find out how the quality and musicality of the spatial mixes compare to stereo, whether the perception and meanings of the staging concepts manifest as intended when using the same set of generic HRTFs applied across *different* listeners and to discern if there are any characteristics or anomalies that could lead to the binaural spatial presentation being less effective for music production.

The defined method for the listening tests directly relates to the phenomenological research methodology aforementioned. It involves collecting anonymised data through a series of one-to-one focus groups using a small pool of 7 volunteers. This data is then converted or interpreted – depending on data source - to comparatively assess the communal experience against the original intentionality and experience of the individual as judged and defined through phase 3.

Sample Background

The small sample size is typical of a phenomenological methodology where it affords more detailed and specific experiential data which can then be analysed and interpreted to generate meaning. A variety of musical backgrounds have been selected to provide a range of perceptions that reflect the real-world demographic of music consumers; people that listen to music, regardless of whether they are musically trained or specifically interested in it as a subject. The volunteers' musical backgrounds vary from none, untrained or self-taught through to formally trained or music industry professional. The majority of the sample do have some form of musical experience, be it self-taught musicianship or professional training, with the minority being without any musical experience or training. Though phenomenological studies often use a purposive sample - a group of people who have experienced the phenomena of interest - the focus groups are designed to provide the experience of the phenomena in study and, as such, the sample need not have any previous experience of spatial music beforehand, though of course this would be of benefit in providing a richer description of their experience (Ellis, 2016, pg. 129). Overall, the variety of sample helps to obtain a larger dynamic in data and to ensure variation and triangulation of critical and descriptive outputs in the questionnaires, allowing for a broader variation in the data pool without compromising on data quality. Consequently, this affords a more thorough analysis of the production aesthetic, providing opportunities for cross-correlation and corroboration of themes presented through a range of experiences.

Method Overview

The sample individually attend a one-to-one focus group session at a time and place most convenient to them. All tests were conducted outside of a laboratory setting at either the University of West London or my north London home-studio - both places provide a relaxed environment that is quiet enough to attend to the stimuli without distraction.

All tests used the same pair of over-ear headphones to reduce variables and provide consistency across the data sets. The headphones used throughout the tests were the over-ear model used in the previous quality judgments of phases 2 - 2.5, and these were also the same headphones used to create the spatial productions; AKG K271 MKII. This reduces variables across the judgments in each phase of research, as well as maintaining consistency in the playback across listening trials.

The focus group sessions are formed of three parts, which are repeated for each pair-wise comparison in question. There are 8 set pairs of A/B stimuli in total and therefore 8 rounds requiring the following:

- Listen to a randomised and normalised set pair of stimuli for A/B comparison (the stereo mix version against the spatial version of the same piece).
- Answer the accompanying questionnaire in their own words, answering each question for both stimuli where appropriate.
- Scale their overall aesthetic preference for each stimulus in each round using the customised HULTIGEN GUI system.

Quantitative data is collected through the trials using a scientifically recognised A/B scaling interface. The interface was created using a customisable, open-source Cycling ’74 Max-based listening test interface generator named HULTIGEN, created by Dr Hyunkook Lee and Dr Christopher Gribben of the University of Huddersfield, and is discussed in the sub-section below. The associated qualitative data is collected through a set of 21 open-ended questions pertaining to the experience of the stimuli in trial, the design of which is discussed in the relative sub-section later in this chapter. In keeping with phenomenological method, these questions request the participant to ‘describe’ their experience in their own words. A short and friendly discussion sometimes follows each focus group session, however this is not implicit to the method design and is more a supplementary and voluntary chat on-record initiated by the volunteer. A questionnaire was chosen over an interview so that the participant could, if required, attend to the stimuli and the questions in parallel and at their leisure.

For quality and rigour, the following must be ensured:

1. All users attend to both A/B variations of a musical piece, with the same task and questions to complete to avoid bias.

2. The perceived loudness of all audio samples has been normalised. This helps to ensure there is no preferential bias based upon perceived loudness levels -where the participant favours the louder version.

3. Ensure the order of stimuli and trial presentations are balanced. This is done by randomising the trial presentation and stimuli order. This helps to avoid any anchoring bias in participants (where they prefer a certain stimulus because they have experienced it first).

4. The questionnaire is presented in parallel to the audio stimuli. This helps to avoid the misinformation effect presented through memory-bias – where questions asked after an experience may alter a person’s memory of that experience.

5. The number of listens to each stimulus per trial is unlimited, however users are asked to note how many times they have listened to each stimulus within the questionnaire form for each trial. This allowance helps the listener acclimatise to the mix formats presented, prevents memory-bias in the questionnaire answers and consequently results in a richer data set.

Now we have outlined the design, structure and execution of the focus group listening trial method we can discuss the specificities of the measurement parameters used. The next section focusses in on the HULTIGEN GUI, providing functional description and detail as to how it was deployed with a questionnaire to collect and convert quantitative from qualitative data.

Huddersfield Universal Listening Test Interface Generator (HULTIGEN) – Overview and Functionality

The HULTIGEN GUI was developed to provide "a user-customisable environment, which takes user-defined parameters (e.g. the number of trials, stimuli and scale settings) and automatically constructs an interface for comparing auditory stimuli, whilst also randomising the stimuli and trial order" (Gribben and Lee, 2015). The tool templates are based on ITU-R recommendations[22] and provide scientifically and academically recognised methods of conducting listening test evaluations. The paper and Cycling '74 max-patch were published via the Journal of the Audio Engineering Society in 2015 and are now freely available via the University of Huddersfield research repository.

The HULTIGEN interface was selected as a method because it offers a versatile, reliable and accessible means of collating and transducing quantitative data from qualitative subjective experience by means of numerical scaling, and functions as both a playback system and a measurement tool. It affords customisable templates for several types of common listening test methods that may include multiple stimuli, a comparative scaling function and the option to included reference stimuli,

[22] BS.1116-3 (double-blind triple stimulus test - DBTS), BS.1534-2 (multiple stimulus test with hidden reference and anchor - MUSHRA)

hidden references and anchors depending on the requirements of the assessment. The listening test method templates include: ITU BS-1116-3 – double-blind triple-stimulus test (DBTS), MUSHRA (multiple stimuli hidden reference and anchor) and AB/X[23]. The difference between the three models is the depth of analysis they allow for.

The double-blind triple-stimulus test features a set of three audio stimuli per trial, two of which are graded against the third which is the reference. These judgments are made using a five-grade scale with descriptive anchors (e.g. imperceptible 5 – very annoying 1). This method is commonly used to detect small degradations of audio quality between high quality audio samples (Gribben and Lee, 2015 pg. 2)

In a MUSHRA test the user is presented with multiple stimuli to be compared against a reference, where one or more of the stimuli to be graded are anchors derived from the reference and one stimuli presents as a hidden reference. This is usually used to determine medium-large degradations of audio quality and was designed to test audio codec processing (Gribben and Lee, 2015, pg. 2).

AB/X equates a pair-wise test without grading, where the user must identify the hidden reference (X). This method is used for assessing smaller, less complex audio degradations (Gribben and Lee, 2015, pg. 2).

In this project we are simply looking for a preference grading of two different mixes of the same piece of music, rather than an identification of specific degradations between stimuli. Therefore, a simple pair-wise A/B comparison

[23] AB/X is pairwise comparison (AB) with a hidden reference (X).

using two stimuli without a reference was designed through combining elements of both AB/X and DBTS models. The preference judgment is based on subjective considerations of the overall aesthetic quality of each stimulus and are made using a five-grade scale with descriptive anchors (see figure 18). This five-grade reference scale is similar to the Mean Opinion Score (MOS) standard, which has been used to assess the transmission quality of audio (Gribben and Lee, 2015, pg.2). The stimuli were normalised as best as possible to prevent loudness being an overriding preferential factor or presenting a bias in the listeners' responses. However, given the obvious and intrinsic differences in the two format's aesthetics it was not possible, nor a requirement, to match them or make them undetectable from each other. This does not negatively impact the data collection as the judgment is designed to assess the listeners' subjective preference of the different aesthetic characteristics pertaining to each of the formats.

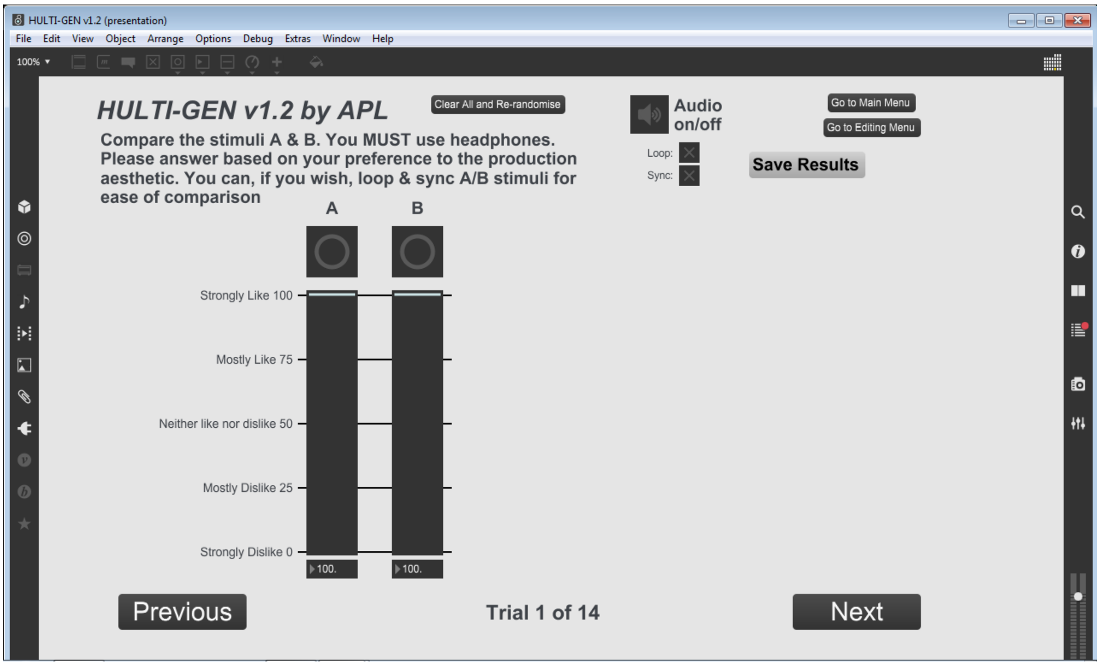


Figure 19: HULTIGEN GUI customised for the focus group listening tests.

Functionality

There are 8 rounds of pairwise trials with each trial presenting two versions of the same track to be judged by the participant – one spatial mix and one stereo mix. The GUI randomises the presentation order of each A/B trial and the order of the stimuli presented and asks the user to rate their preference of the stimulus using the associated sliders or by inputting a number into the value box.

Each slider has a five-grade value scale range of 0 – 100 associated with a semantic differential, whereby zero equals the most negative response (strongly dislike), 50 equates indifference and 100 equals the highest possible positive response (strongly like). All values between 0 and 100 are available to select, with the five-grades each presenting a judgment reference to a specific semantic differential (see figure 18). This type of continuous grading scale can present a lack of control over the way the participant grades stimuli; creating too much difference in the spread of scores between listeners. However, this can be addressed through normalisation (rounding up or down values) and should not present much issue over such a small pool of listeners (Gribben and Lee, 2015, pg.2).

The interface also offers user-activated synch and loop functions that allow for optional synchronous and continuous playback of each stimuli source in presentation, affording ease of comparison and allowing the participant to switch between the two stimuli at the same playback point.

When the participant has input their preference for each A and B stimuli presented they can press ‘next’ to move on to the next trial. When all 8 trials have been completed, the participants may ‘save results’ and export their data as a text file. The exported data sheet presents a log of the numerical values chosen for each A and B stimuli per round and details the presentation order of the stimuli and trials.

Although this scaling method gives numerical data on a person’s subjective preference for each stimulus, it does not give any further substantiating information as to why they have made such a decision. Therefore, a questionnaire accompanying each of the 8 trials was provided to garner a more in-depth understanding of the participant’s experience and to contextualise the quantitative HULTIGEN results across each trial round.

Questionnaire Design

The questions were designed to evaluate the reasoning behind the listener’s preference for each A/B comparison presented across the 8 trials. The questions examine the quality of the overall experience of each stimulus and gather data related to the listener’s perception of the periphony, sonic localisation, perceived movement (sonic kinesis) and their interpretation of the staging concepts and the meaning underpinning the production. These questions are open-ended and prompt the participant to describe what they are experiencing. This allows the participant to answer in their own words, preventing interference from the researcher, and consequently provides the descriptive, experiential data required for interpretive phenomenological analysis.

There are also preceding questions that collect data on the sample demographic; asking for the participants' background information pertaining to their musical experience/training, their musical preferences and their age. There are also questions regarding specific technical semantics relating to the project. This helped to determine the depth of relative knowledge the subjects had via judgement of their understanding of audio related terminology. The questions ask if the listeners had heard of the following terms; *Stereo*, *Binaural* and *Periphonic* and what these terms 'mean' to the listener. The expectation is for those with no musical background to have a basic understanding and experience of stereo as a musical format or sound system configuration, whereas perhaps those with more experience would have a deeper understanding of these terms and how they apply in the field. These were considered an important set of questions in determining the level of spatial audio knowledge or technical experience that the sample have. All of the background questions were asked to ascertain the dynamic variation across the sample; ensuring a broader audience range and to determine any possible communicative difficulties that may arise due to a lack of understanding of the field (more information on the sample diversity can be found in the preceding section entitled *Sample Background*). A copy of the evaluation questions can be found within the Appendices section of this document.

4.2.3 - Strategies for Data Analysis

There are two types of analytical strategy employed within phase 4 of this the project. Thematic analysis and Interpretive Phenomenological Analysis (IPA). In this section we will discuss how and why these two types are necessary in the analysis of the phase 4 data.

Thematic analysis is used to observe commonalities, patterns or ‘themes’ trending across the questionnaire and A/B test answers. Whereas IPA uses the researcher’s experience as a control reference in order to examine, interpret and draw meaning from the experiences of the listening test sample.

IPA is an associated form of phenomenological analysis “that attempts to work with the solipsistic (knowledge only of one’s existence) elements of phenomenological enquiry” (Ellis, 2016, pg. 128). Unlike transcendental and hermeneutic forms, IPA is concerned with accessing the emic (insider’s) perspective, acknowledging it as being an interpretation of the lived world. Consequently, IPA does not require the researcher to put aside their existing understandings of the phenomena in question but acknowledges that the researcher will use their existing understanding to apply an interpretation to the subject’s interpretation as a way to make sense of how the subject experiences the phenomena (Ellis, 2016, pg. 128; Smith and Osborn, 2004). Ellis (2016, pg. 129) states that “in all forms of phenomenology, the emphasis is to identify the important messages — the essence or kernel of the topic — so that the important aspects of the phenomenon are described.”

This is why a combination of IPA and thematic analysis is considered an important strategy in forming meaning from the experiential data analysis. IPA draws upon both my experience of the phenomena, as presented through the audio work and the process in generating it, as a point of reference in setting the criteria for data reduction and filtering, and the construction of meaning. It is through this reference to a lived experience that the essences obtained through the described experiences of the listening test subjects may be contextualised and a meaningful conclusive narrative established.

These two types of analysis are employed in the following four-stage analytical method:

1. The method of analysis begins by reading the transcripts of the questionnaires to ascertain broad observations, reducing the raw data to something more manageable.
2. The second stage involves filtering the important ideas from the less significant or irrelevant data (IPA). This is done through the identification of important phrases, keywords and messages within the transcripts (Thematic).
3. The third step identifies the emerging themes and trends from the important data and then groups the themes together in a meaningful way (Thematic).
4. The final step involves constructing a theory, hypothesis or providing a narrative to account for these commonalities in the data (IPA).

5 – Preliminary Practice as Research

Through the methodological design I holistically acknowledge that the research ontology dictates the epistemology – the way in which we study a subject should reflect the subject’s nature.

This statement should be applied not only in a philosophical sense when designing a research methodology but as we shall come to discover in this chapter, it should also be applied when defining the technology used to explore the research problem in practice. Though this was only partly actualised at the onset of the project due to the technological limitations presented at the time, the idea that *one should work in the format one is mixing for* gained clarity following the results of the phase 2 investigation and was further reinforced through the product release of the Dear VR Pro spatial panner, and the consequent undertaking of the preliminary research phase 2.5.

Ultimately, the aim of the preliminary phases of research is to answer the presiding questions; *‘What systems could be used to approach mixing for a periphonic-binaural format?’* and *‘How greatly would the variation of headphone affect the perceived musicality and periphonic translation of the music?’* Within the specific research phases there were a number of underpinning queries regarding the design of the staging approaches required to examine the problematic areas of binaural localisation, and the consideration of source audio and output formats, all of which are later revealed and discussed through this chapter’s discourse.

Phase 1:

Phase one explores the Auro 3D 13.1 multichannel speaker system as the initial tool for spatialising the musical work and evaluates the limitations and affordances of ambient recording versus the use of mono sound sources in constructing periphonic-binaural sound stages. The underpinning creative practice also presents experimental production approaches designed to examine the problematic areas of localisation typically associated with the binaural format and attempts to outline the affordances and limitations of a speaker-based approach to binaural mixing.

Phase 2:

Phase two examines the translation of the phase one speaker-based mixes across various headphone types when encoded to the binaural format and uses the quality judgments outlined in the preceding methodology chapter to assess the suitability of the encoding results. This phase implements experimental approaches to channel-based binaural encoding using two methods:

1. Upmixing to an Ambisonic format and decoding using third-party Blue Ripple Audio O3A encoders and decoders.
2. Re-amping the channel-based Auro 3D presentation to binaural using the Neumann KU100 dummy head.

As this chapter discusses, neither of these methods were considered to be aesthetically successful nor fully in-keeping with the project’s democratic requirements but at the time of undertaking these were the only DAW-based binaural rendering approaches available that could accommodate the specific and unique channel configuration of the tri-tiered Auro 13.1 system (7.1 + 5 + 1) (see figure 20).

Phase 2.5:

Later technological developments allowed for reflexivity and presented an opportunity to re-explore the queries posed and examined in phases one and two using a direct-to-binaural method - Dear VR Pro spatial emulation plug in - as the means for both spatialisation and binaural playback. This tool negated the need to encode and decode the mixes for a binaural format, offers the means for working with legacy source audio and offers a flexible variety of delivery formats. Fundamentally, the Dear VR Pro presented an opportunity to re-examine and reflect on the problematic areas of binaural perception in real-time within the binaural format during the re-composition and production process of phase 2.5. The importance of these features and how this benefits the exploration of the research problem are discussed and evaluated later in this chapter's phase investigations.

Out of the three methods explored it was decided that the Dear VR Pro spatial panner was the most suitable tool of choice due to fulfilling both the quality judgments and the democratic prerequisites defined in the preceding chapter. The aesthetic and ideological quality judgment criterion are reiterated within each phase conclusion as a means of evaluating the efficacy of each approach tested.

5.1 - Rationale and Context - Foreword

Before discussing the functionality and result of the periphonic-binaural methods examined in the preliminary investigations, this section first contextualises the importance of defining an appropriate spatialisation and binaural delivery method, and outlines key considerations of binaural localisation and spatial approaches as applied to music production.

Historic Approaches to Binaural Music Making

Binaural recording has been used in generating spatial content for a number of commercial music records, from Lou Reed to The Rolling Stones and Pink Floyd to Pearl Jam[24] (Hooke Audio, 2021; CBS Interactive, 2021). However, there are a number of limitations in the techniques traditionally used to construct binaural records (Ramsey, 2011; Hook Audio, 2021). One of these limitations has typically been the expensive cost of a professional quality binaural 'dummy head', with the Neumann KU100 retailing at circa £7000. However, due to technological developments, one can now purchase a set of in-ear binaural microphones for around £80, such as the Roland CS10em, or a Jecklin disc[25] could be bought for £120 - £200. Though the price may be more affordable, there is a compromise on the quality of HRTF filtering when using in-ear binaural microphones, as these rely on a single individual's

[24] One can find a shortlist of the top commercial music records that utilise binaural recording via Hooke Audio and Last.FM websites.

[25] A Jecklin disc is an acoustic panel that can be placed between two microphones to simulate the shadow of the head but lacks the spatial information that would be imprinted by the pinnae (outer ear). This can also affect the quality of the binaural capture and the reproduction across listeners.

HRTFs (the wearer) to generate the binaural imprint upon the capture. This may present a poor binaural perception of the recorded audio across other listeners who are fined tuned to their own individual HRTF response. Conversely, when using the Neumann KU100 dummy head - the design of which is based on an average measurement generated from a multitude of human HRTF data and affords a more consistent experience across listeners - you would then have to consider the size and weight of the microphone when looking at placement. For all these recording approaches, you would have to consider whether you were going to multi-track or record live. Individual instrument multi-tracking will sum the noise floor and binaural ambience considerably, which can reduce headroom and cause fidelity issues and phasing. This ultimately can result in a noisy mix and a distorted binaural image. The sonic information present in binaural recordings captures a singular listening position in time in space, and the musical content and characteristics of the microphones[26] used are as much a part of the spatial sonic imprint as the acoustic space and 'sonic object'. There is no altering any of this after the fact, the recorded sound and space are intrinsically interlinked and definitively captured within the recording print.

Zagorski-Thomas (2014, pg.75) states that "Even dummy head binaural recordings, the method that most accurately captures sound as we hear it, doesn't provide us with a completely accurate picture of an audio scene." Although it offers a good sense of realism, it represents a two-dimensional sonic schematic of one particular listening position. Stereo recording too is schematic but it works particularly well for music because techniques have

[26] 'Dummy head', Jecklin disc or 'ears' microphones (in-ear or modelled).

been developed that utilise the stereophonic sound field relative to our ‘natural’ binaural hearing and in ways that aren’t always based upon promoting realism. Much like stereo recordings, periphonic music production calls for a pragmatic compromise to sonic realism in presenting schematic representations of a metaphorical musical space[27]. One could assume that for periphonic music without a reinforcing visual stimulus, a schematic representation may be implicit, particularly given the perceptually possible but practically impossible staging schema and metaphor that could be afforded through a periphonic sound stage.

The Importance of a Headphone-based Approach to Spatial Music Making

“There's great attraction in being able to recreate spatial audio through any normal stereo speaker system” SOS Interview with Qsound Labs (White, 1995)

As previously outlined, a headphone-based spatial reproduction provides the closest possible way of reproducing 3D music in a manner that would successfully and conveniently work with the current headphone-based listening culture. This consideration provides an opportunity for this research project to both democratically inform industry approach and for a spatial

[27] This notion can be put into alternate terms; ‘realism’ versus ‘believability’. Stereophonic recording and production practices often present sonic schema that promote ‘believability’ over ‘realism’. These structures hint at realism through their association with real-world phenomena or experience, like a performance, but are often metaphorical sonic constructs that do not and could not exist within the real world; e.g. a voice having the acoustic properties of a concert hall whilst the drums (in the same perceived performance environment) have the reverberant properties of a small room.

format to achieve the music industry integration that traditional surround sound has so far failed to do. Without a binaural delivery we could explore the research problem in application to speaker-based spatial audio production systems. However, as history has dictated, the result of which would mostly remain confined to multi-channel speaker presentations and perhaps even limited to the particular format, tools and system on which the techniques were created.

Therefore, without a headphone-based delivery this research would likely not fulfil the overarching project aim of informing industry practice, at worst rendering it obsolete in commercial application or, at best, relegating it as another academic niche. It should be noted that there is no normative process or approach for creating headphone-based 3D audio production work and therefore it was necessary that the first stages of research explore, test and define a suitable method for binaural delivery.

As outlined previously in the literature review chapter, it is widely known in the field that binaural audio can be problematic in maintaining a consistent perceptual delivery across users. Some common and specific issues associated with binaural audio are those of inconsistent reproduction across differing headphone models and perceptual problems pertaining to the generic HRTFs not suiting the listener’s naturally fine-tuned binaural perception. Some common anomalies within binaural listening are those presented as front-back and elevation ambiguity; whereby the listener cannot accurately determine whether a sound is localised in front or behind, or above. This is especially common on the median plane where the sound source is situated centrally, equidistant to both ears and lacking the differential time-of-arrival and intensity cues. With this in mind, it is important to examine these areas of issue when defining a suitable method of binaural decoding. It is with this reasoning that during the spatial re-composition of phase one the staging techniques created were designed to explore these problematic areas of perception and act as anchored reference points to judge the presentation accuracy across the various encoding methods and headphone types tested. We discuss these techniques further on in this chapter, but for now we will continue contextualising the investigation and unravelling the rationale behind the choice of binaural rendering methods chosen for examination.

> *"When perception proceeds in an unproblematic way, we are usually unaware of the sensory aspects of the stimulus information, we attend only to the events specified by the stimulus structure but when that relationship is problematic, the stimulus structure itself can become more evident."* (Clarke, 2005, pg.32).

This statement from Clark asserts the motivation for examining the headphone translation and the presentation accuracy of the decoding methods used. It exemplifies the pre-requisite theory underpinning the preliminary investigation and asserts the notion that *with a coherent periphonic reproduction the listener will focus solely on the experience of the musical content and will not be preoccupied with the method and technicalities of which it is being delivered*. Clarke's statement and this hypothesis both reinforce and reflect on the necessity of an unobstructed binaural experience. This observation is further supported by proponents of the VR industry: 'We want a listener to be sucked into the acoustic world, we want the listener to believe the world he is hearing and not being distracted by it" (Boom Library, 2019). This consideration is imperative in providing appropriate means of exploring the research problem in phase 3 of the project and in ensuring the intended emotive interpretation and expected experience is delivered without the listener attending to anomalies in presentation or the structures that reproduce the spatial and musical values. With a problematic reproduction the effect and intention of the musical production may be dramatically lessened, the delivery system exposed and the perception disconnected - akin to how the 'magic' dissolves when the method behind a magician's trick is revealed. A problematic binaural presentation would not fairly represent the intended periphonic mix and would have considerable implications on the quality and substance of the data collected. As such, it is clear that defining a consistent and hi-fidelity method of binaural production and playback is fundamental in both exploring the research problem and in informing the entire project outcome; from investigating the creative possibilities of the research practice through to examining the result and providing meaningful data collection.

5.2 - Defining the Periphonic-Binaural Production Method

5.2.1 - Practice as Research Phase 1: Auro 13.1 Evaluation – Format familiarisation, orientation and recomposition.

At the time of the project commencement in the autumn of 2015 the University had installed an Auro 3D 13.1 system in Vestry Hall Studio 2. This system presented the opportunity to explore periphonic staging approaches within a familiar and industry-standard Protools DAW environment using three-layers of surround speakers[28] and the specific Auro channel-based panning tools. The Auro 3D toolkit also provided a binaural renderer that presented a culturally appropriate way of commercially delivering the spatial music mixes, suggestively addressing the democratic delivery problems that plagued traditional surround sound music making. Having seemingly found a 'perfect system' to use for the practical investigation, my attention turned to familiarising myself with the Auro 13.1 system and exploring the possibilities for periphonic sound staging using various types of source audio; mono stems, ambient recording techniques[29] and native Ambisonic recording formats. Not all of these formats were suited to this project's trajectory; binaural recording and Ambisonics recording approaches were particularly unnecessary given the multi-channel speaker configuration of the Auro system. Therefore, this section focuses on examining the use of mono sound sources against multi-microphone recording approaches as evidenced through research as practice case studies. However, this chapter does

[28] The three tier comprised a 13.1 configuration through a 7.1 lower layer, a 5.0 interim height layer and a single, central top layer speaker positioned overhead (see figure 20).
[29] Including both binaural and surround sound approaches.

provide a theoretic rationale for the consideration of Ambisonic formats for delivery which is discussed in Phase 2.

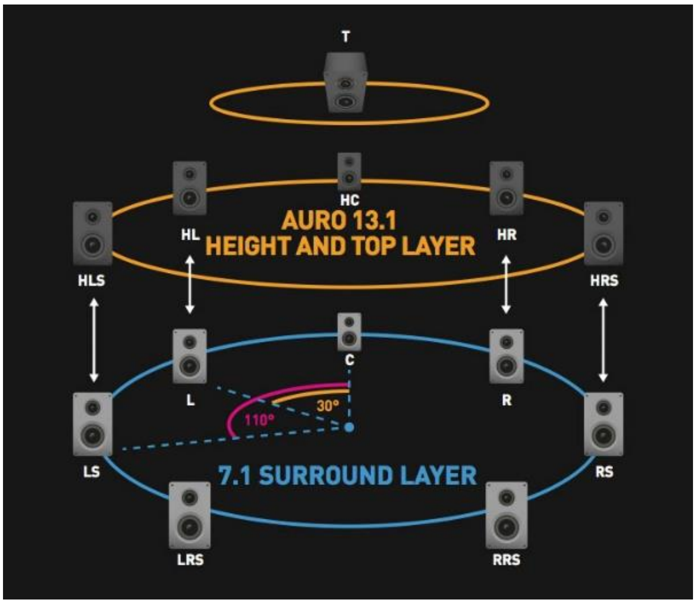


Figure 20: Auro 13.1 array configuration (Auro, [n.d]).

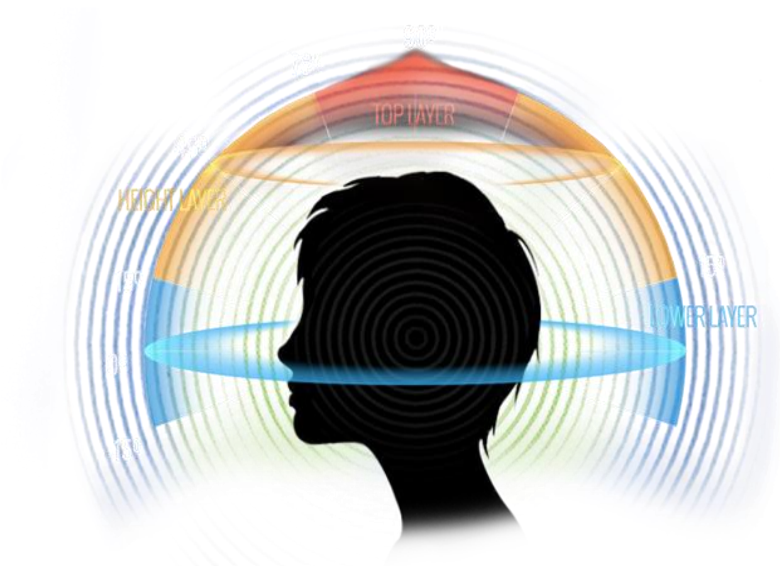


Figure 21: Schematised interpretation of effect of binaurally decoded audio relative to the system layers (Auro, [n.d]).

Spatial Recording versus Spatial Emulation case study

Part 1 - Ambient spatial recording approaches for periphonic sound staging

In phase one there was some preliminary experimentation with custom tiered multi-mic arrays that included the use of a Sound Field Ambisonic microphone, as well as integration of binaural ambience captured using the KU100 dummy head[30]. The multi-mic experimentation happened at 'Big Smoke Studios' in Wembley in 2016 with Mark Brocklesby (studio owner) leading the two performance music sessions; one session was allocated to record the Academy of Contemporary Music's Elektron Choir (directed by Kaya Herstad-Carney) and one session was allocated for recording Mosi Conde's EP. These two sessions presented an opportunity to collaborate and explore our ideas for experimental 3D multi-microphone techniques that would be compatible with the Auro 3D speaker configuration at the University of West London. For both sessions we were implementing 'a microphone per speaker channel' approach, minus the LFE which reproduces a sum of the low-frequency content of all channels.

For the Elektron choir recording we configured a 3D audio recording set-up comprising an INA-5[31] five-mic array for the lower tier and a Hamasaki quad array for the height layer. For the Mosi Conde recording we utilised a cross-adaptation on the 7-mic William's Star and INA-5 arrays for the lower surround layer, with the left and right surround capture provided by the

[30] Binaural recording and Ambisonic format considerations are discussed in their own research sections after the Phase 1 case studies.
[31] Ideale Nieren Anordnung (ideal cardioid arrangement).

KU100 binaural dummy head, instead of a matched pair of small diaphragm condensers[32] (see Appendix B for kit list and microphone plot). We also trialled the use of the Sound Field Ambisonic[33] microphone to capture the ceiling ambience for the top layer. However, we decided to use a Brauner Phantom large diaphragm condenser for this, instead[34].

To create a non-front orientated periphonic sound stage in these recording scenarios we had to place the sound sources around the microphone arrays. For the Elektron choir session there were enough voices to arrange a balanced vocal stage around the array and record each track live in a single take. However, it was slightly more difficult to generate a non-fronted stage for the soloist session with Mosi Conde. To create a non-fronted sound stage we had to ask Mosi to reposition himself and his kora around the rig numerous times and record multi-tracked layers for each piece[35]. However, to accommodate the size of Mosi's instrument and the correct spacing between sound source and the array within the live room, we also had to reposition the rig. This meant we had to re-measure the array and check that the microphone spacing and alignments were exactly the same as when we conducted the previous frontal takes. This ensured correct and consistent time-alignment and critical linking between channels, which is imperative to

[32] We had previously explored a similar technique together during our Masters degree course in Autumn 2013 but at the time there was no 3D system to facilitate the height playback – instead the channels were folded into a 5.1 mix-down.
[33] Ambisonic is an isotropic recording format from the 1970's. The functionality and considerations for Ambisonic formats is discussed in detail in the phase 2 study of this chapter to follow.
[34] This was decided as a means to retain phase coherence, image stability and format consistency between the array levels. There was also quite a lot happening with the production and decoding the Sound Field to print a legacy output added yet another task.
[35] This was not the case with the choir, as the nature of the repertoire did not require over-dubbed parts and layers, and there were enough voices to arrange the chorists around all sides of the microphone array.

preserve the phase relationship of the capture. We also set up a close vocal mic and a spot mic on the kora to retain the instrumental definition amongst the ambient array. These spot mics had to be phase inverted and panned appropriately to correspond to his position across the rear placement captures. It was very time consuming and precise in approach, and there were many audio channels and a lot of high-end equipment involved for what was a simple acoustic surround recording of one man and his kora. The session experimentation resulted in an excellent set of transparent surround recordings that were highly suited to the instrumentation and repertoire and presented a hyper-real capture of Mosi’s kora and vocal performances.

The ambient spatial approach to recording the ACM Elektron choir was also well suited to the performance application, the repertoire and the immersive staging of the choir voices in surround. Although in retrospect I realised that we should have explored switching the front-back positioning of the male and female voices in the choir, as the female voices presented as significantly higher in the frontal section of height array than the male voices did in the rear. It could have been an interesting experiment to explore what the inverse possibilities of this would have been and whether the male voices could have presented a better sense of height if positioned in the front. Unfortunately, given the nature of the ambient multi-microphone recording approaches there was no altering this after the capture; the stage had been set. This lack of reflexivity is one of the most pertinent limitations associated with this approach and one which is to define how this project moves forward.

Case study conclusions:

Although the captured ambience and the fixed stage positions were more than suitable for the type of recording tasks at hand, these approaches presented limitations pertaining to the level of creative agency and aesthetic reflexion achievable within a production. These recording approaches confine the production to a capture of a particular performance in a particular space at a particular time, with the intention that the percept is to be embodied and interpreted in a way that asserts realism through the re-creation of a performance in a specific environment. To record using tiered multi-mic arrays reflects and affords the same values as did past single-tier, traditional surround sound and binaural recording approaches - just with the inclusion of more microphones for more channels of immersive ambience. There was no affordance for exploring variations of the spatial stage after the capture nor for experimenting with abstract placements around the array (such as overhead voices and instrumentation).

This approach could be summarised as an exploration into the development of transparent surround sound recording techniques with captured height, and although it provided a sense of hyper-realism and increased immersion, fundamentally the result is not unfamiliar, and does not offer any new realm of creative possibility than previous surround sound and binaural recording techniques could.

Although both of these sessions were incredibly inventive and a great learning experience, the limitations to creative practice presented through the fixed sound staging could not comply with the investigative trajectory of this project (points 1 & 2). There were also conflicts pertaining to the quality judgments and democratic prerequisites set out by this project (points 1, 2 & 3).

1. The approaches to recording were suited more to presenting and enhancing realism through a performance capture than developing creative staging practice.
2. There was no means of staging anything overhead and fully exploiting the elevation capture. The height microphones added additional ambience through the capture of delayed direct sound and reflections.
3. The multi-mic arrays were costly, cumbersome and time consuming to rig and measure.

This research as practice study expanded my exploration into the importance of the sonic context and the sonic content underpinning a music production and initiated the ‘performance music versus production music’ argument (these topics are expanded upon in the next chapter of this thesis).
Through this case study we have concluded that the use of ambient multi-microphone recording approaches will depend on the appropriateness of the sonic context and sonic content of a production, as well as the intended interpretation; realism versus believability. The case study reinforces that multi-channel (and binaural) recordings are not the most appropriate source audio if you are looking for a reflexive and flexible approach to experimental and creative staging practice. However, what this case study does not tell us is what could fulfil this requirement for source audio that affords a reflexive creative approach to music production.

Part 2 –Spatial emulation through recomposition: using mono and stereo sound sources for periphonic staging practices.

In parallel to the above ambient recording investigation, phase one also explored spatial emulation using mono and split-stereo stems as source audio. This was explored in practice by spatially recomposing two production music pieces ‘Penny Drops’ and ‘Monomorphic’.

There were two aims of this practical study:

1. To define an approach that could afford reflexivity in exploring the creative staging research problem.
2. To create staging practices that could examine the typically problematic areas of binaural localisation for the binaural headphone translation evaluation of phase 2.

In this section I present the recomposed staging configurations that acted as reference anchor points in examining the binaural rendering result of phase 2. The initial work was produced on the Auro 3D speaker-system without any ability for binaural monitoring. Thus, this was an exercise in exploring the creation of periphonic staging technique with a binaural *potential*, rather than a reflexive exploration of technique within the binaural domain. The purpose of this study was to both to explore the possibilities of production music staging using mono and split-stereo sound sources and to generate techniques that would mainly focus on exposing the problematic areas of binaural localisation when the mixes were rendered for headphones in the next phase.

Piece 1: 'Penny Drops' – This electronic music piece can be described as a *percussive experimental audioscape.*

As 'Penny Drops' exists as the first periphonic piece explored it came as no surprise that some of the techniques involved were experimental prototypes that aesthetically did not lend themselves much to any real-world popular music application. However, the purpose of the re-composition was designed to primarily investigate periphonic image stability, source localisation and mix coherence when translated to headphones. There were various experimental staging aesthetics implemented to test the localisation and image stability across all azimuth dimensions (in front, behind, to the sides and above the listener), as well as techniques that examine the horizontal and vertical phantom imaging between the speaker layers[36].

Many of these techniques worked by generating static placements or automated movements across the known problematic areas of a binaural sound field. This was achieved by applying an Auro 3D panner to each source audio channel, judiciously positioning the sound sources in their considered placements and recording or drawing automation where appropriate. This approach, using static and kinetic sources, was considered as a way to not only explore the response of sound source movement in the binaural domain, but also as a means of examining whether moving sources would be better localised by the listener than static ones. The hypothesis underpinning this was based on the understanding that head movements help humans localise sounds in the real-world. This presents the notion that

[36] There were a few techniques implemented that explored the creative staging possibilities of the format relative to the research problem, although these are discussed in chapter 6.

movement of sources could help to resolve the localisation difficulties presented through binaural-reproduction, where the listener cannot use their head to naturally localise a source as they would in the real-world; *'if the listener cannot localise a sound source by moving their head, then the sound stage should move instead'*. Most of the techniques implemented used broad-band, percussive sound sources with a sizable level of higher frequency content[37]. The importance to use sources with higher frequency content was considered as another means to aid localisation. As previously outlined in the literature review, we know that higher frequency sound sources, typically of the range 6-8kHz, localise better and respond particularly well in elevated and periphonic positions. We can see from the associated spectrum analysis images that each sound source comprising these staging tests had a good frequency response in this range, aside from the source that made the 'crank' technique, which had more comparative spectral energy across the mid and low frequency ranges than it did the upper bands.

Follower and Chaser Technique – These techniques were created using the percussive audio sources titled 'creaking' and 'creaking 2'. Together they present an interactive sound stage that uses both stasis and automated kinesis to create illusions of exaggerated movement around the listener. Both the follower and the chaser techniques were implemented to explore how automated kinetic movement would respond across the front and back areas

[37] As relative to the rest of the spectrum. Sizable refers to either an increased or balanced amount comparative to the other bands. This judgement was not only deduced through spectrum analysis but also through listening to the sources and deciding if their spectral content sounds 'high' enough.

of the binaural sound field. This specifically looks to examine front-bank localisation ambiguities associated with the perception of binaural audio. The chaser technique was created to also utilise static sound source placement as a means for enhancing the perception of movement, which is discussed in chapter 6. The evaluation of these techniques when rendered binaurally are discussed in the phase 2 headphone analysis section of this chapter, and their use as creative *sonic cartoons* (Zagorski-Thomas, 2014, pgs. 49-69; 2018) is explored within the ‘Redefining the Spatial Stage’ case studies of chapter 6.

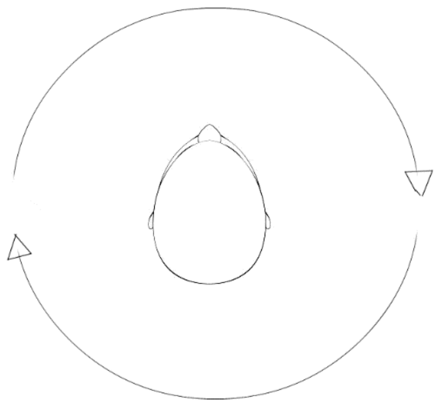


Figure 22: ‘Chaser’ automated panning technique (used in conjunction with judiciously placed static panned samples to create percussive stops).

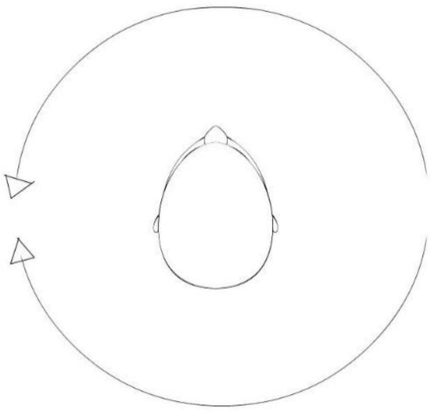


Figure 23: ‘Follower’ automated panning technique (used in conjunction with ‘Chaser’ and judiciously placed static panned samples (percussive stops).

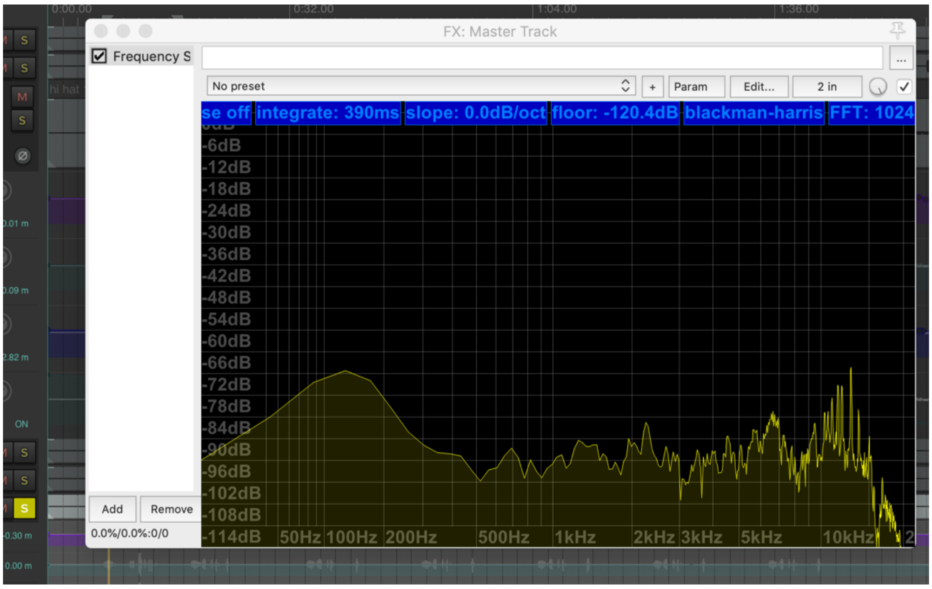


Figure 24: Frequency response of the 'creaking' audio source of the 'follower' technique.

Crank Technique – The 'Crank' was designed to test median front-back localisation over the Z & Y planes. This technique was implemented by centrally panning the percussive source element 'creaking 3' in a front-back, overhead trajectory. However, we can see from the spectrum analysis that this particular sound source lacked frequency content in the 6-8kHz range comparative to the balance across the rest of the spectrum. Though it looks to be a broad-band sound source, the graph shows a significant -20dB roll off after 4kHz. Effectively this acts to filter out much of the higher frequency content needed to help localise a sound source, especially in a non-frontal position relative to the listener. The impact of this is further discussed in the analysis in phase 2.

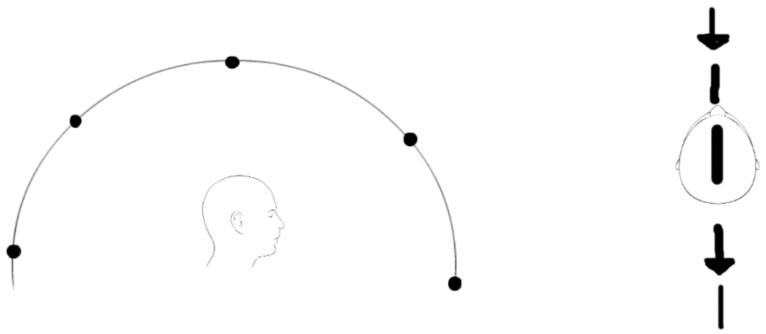


Figure 25 & 26: Automated panning trajectory of ‘Crank’ technique in profile and bird’s eye views. The dotted points in profile represent the static point of the sound source when audible.

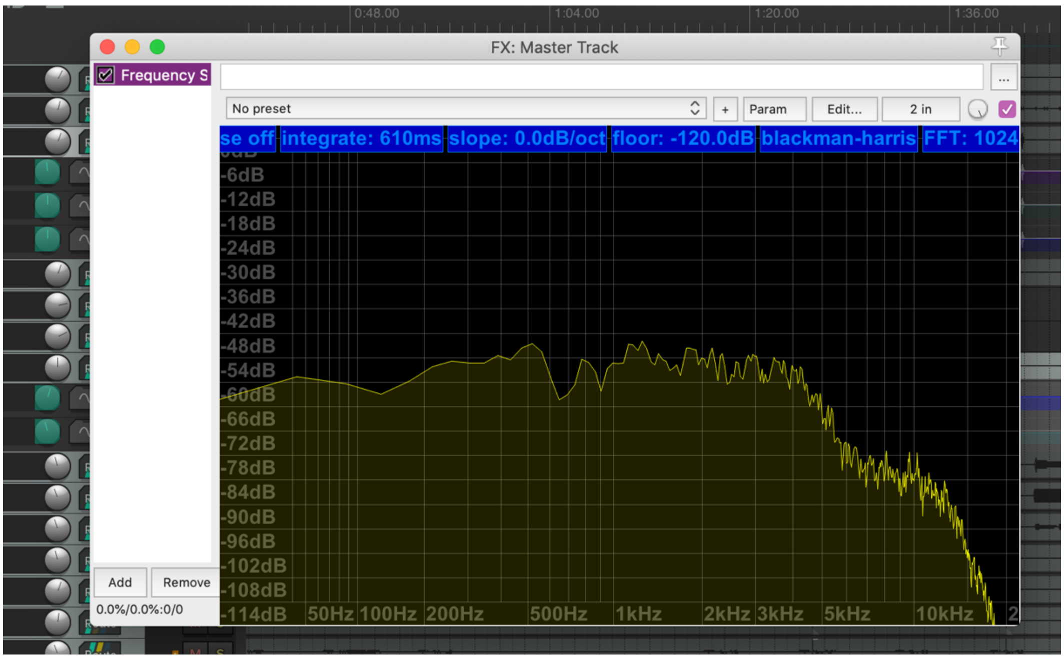


Figure 27: Frequency response of the ‘creaking 3’ audio source of the ‘crank’ technique.

Winch Technique (Spiralled kinetic hi-hat) – The winch technique automates an outward spiralling hi-hat around the listener that gradually moves further away from the central listening point. This technique was designed to test kinetic movement localisation of the XY planes and examines phantom image stability and depth perception across the left-right and front-back planes. This was also a useful way of exploring the vertical inter-layer linking between the speaker channels of the lower and height layers of the Auro system.

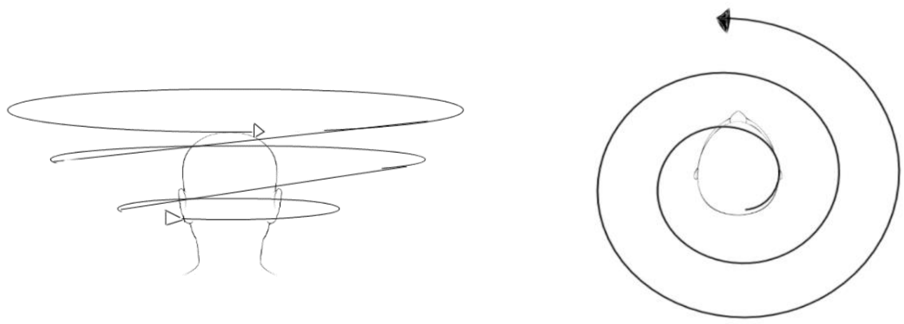


Figure 28 and 29: Automated panning trajectory of 'Winch' technique. In profile and bird's eye views.

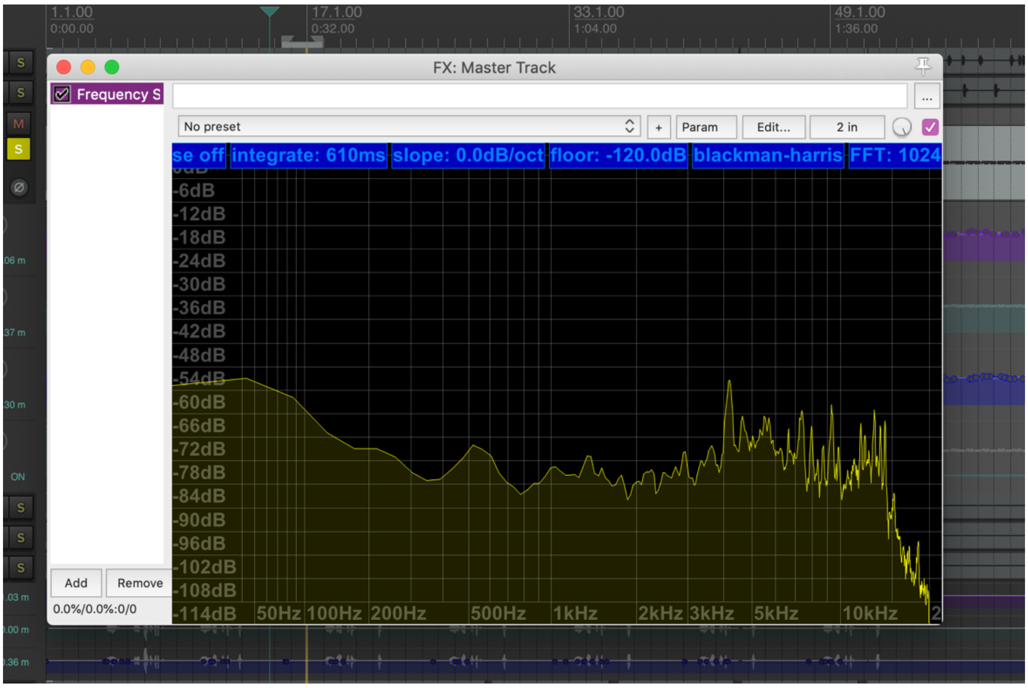


Figure 30: Frequency response of the 'hi hat 1' audio source of the 'winch' technique.

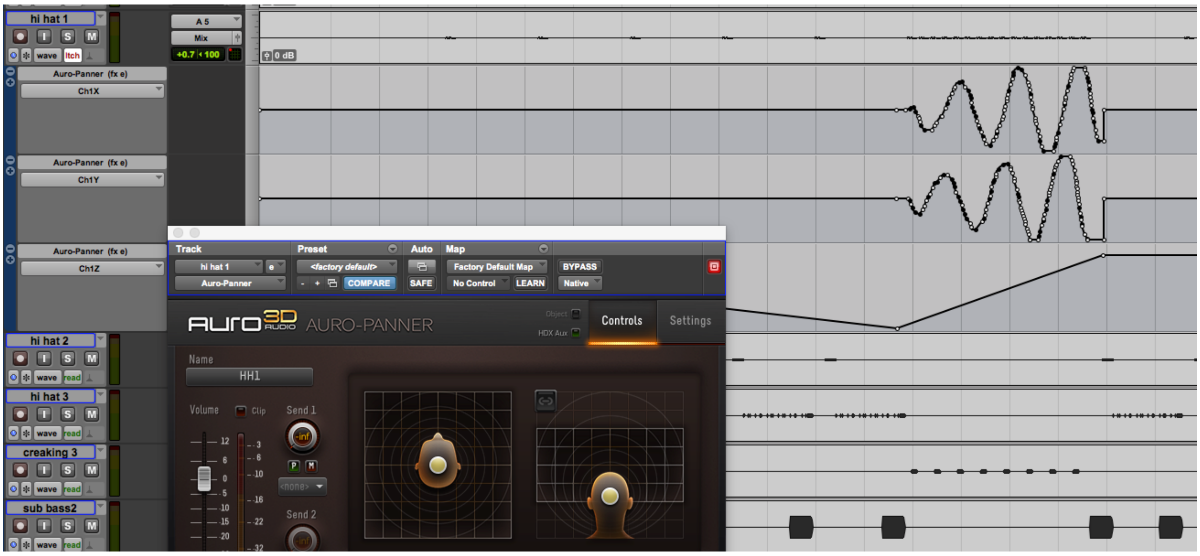


Figure 31: View of the 'Winch' hi-hat automation and designated Auro panner in Protools.

Penny Drop Sample – This aesthetic explores image generation through vertical panning, and was constructed to test the inter-layer linking between the height and lower layers on the Auro 3D speaker array. The aim was to examine the translation of the inter-linked aesthetic qualities when rendered binaurally. The technique was implemented by splitting the sample of a coin-toss landing and spinning to a stop into the two parts that individually comprise the main spectromorphology of the aesthetic; the 'clink' created by the landing of the coin (sample 1), and the 'rattle' that implies the coin spinning to a stop (sample 2). These spliced samples were then panned between the height and lower layers, respectively. Sample 1 was placed rear right in the height layer and sample 2 was placed low rear-centre. This implemented an ecological understanding of perception to create a schematic representation of the movement of a penny falling off a ledge and spinning to a stop. As applied to the Auro 3D system, the spacing between the speaker layers replicated the effect very well and presented an exaggerated percept of a giant coin falling behind the listener. This sample is what gave rise to the renaming of the piece from ‘Monomorphic Part 1’ to ‘Penny Drops’.

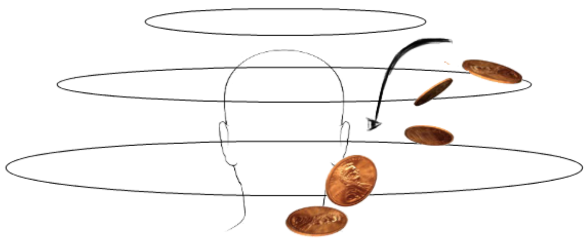


Figure 32: ‘Penny Drop’ sonic cartoon. Exampling the cartoon spectromorphology achieved through vertical stereophonic panning and inter-layer linking.

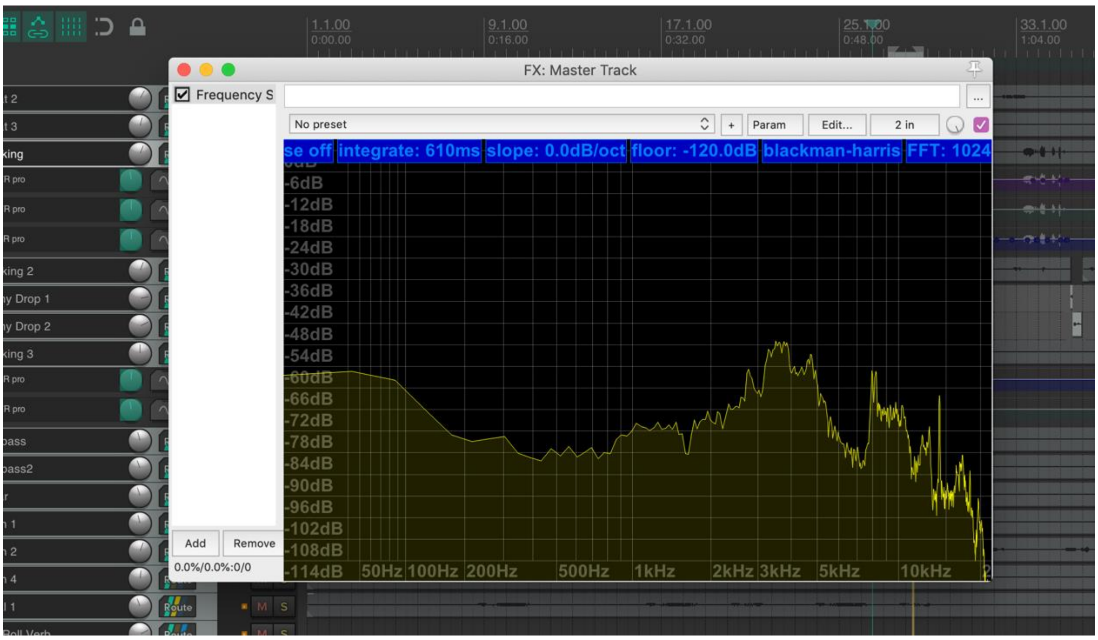


Figure 33: Frequency response of the ‘Penny Drop 1 & 2’ audio sources of the ‘Penny Drop’ technique.

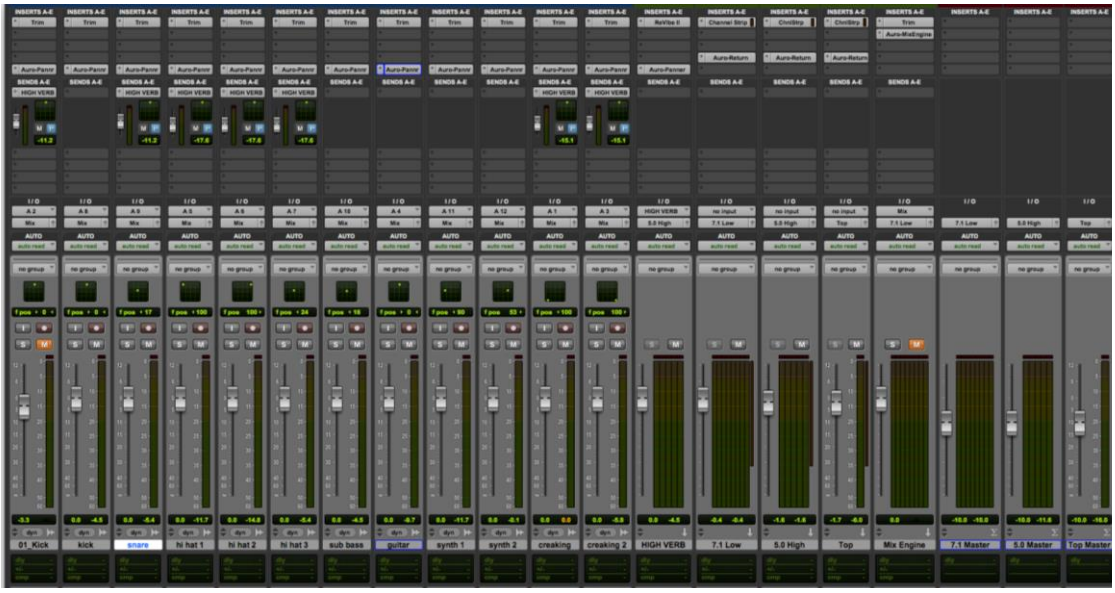


Figure 34: Protools mix window view of 'Penny Drop' channel configuration.

Piece 2 – Monomorphic

'Monomorphic' can be described as a more musical electronic music production piece than 'Penny Drops'. This piece experiments with vocal staging, sonic layering, rhythms and spatial effects. Although the staging constructs of this piece were created in the preliminary research phase and were also initially re-composed using mono and split-stereo stems on the Auro 3D speaker system, the focus of this piece was less about examining the problematic areas of binaural localisation and aimed toward exploring the possibilities for non-front orientated creative staging practice. For this reason, the techniques are discussed in the 'Redefining the Spatial Stage' case studies of chapter 6.

Case study conclusions:

Unlike the ambient recording techniques explored previously, direct mono source audio allows for the creation of spatial stages without fixing the production to a particular captured ambience or real-world staging arrangement. This approach afforded real-time reflexive experimentation with sound source positioning and allowed for the exploration of automated spatial staging constructs on the Auro speaker system. It should be reiterated that at this stage I was not working directly in a binaural format, but working on the speakers in a way that *considered* the binaural result as the intended output. Therefore, at the time there was no knowing how these techniques would respond in the binaural domain and as such there was no accommodation for reflexivity if they did not function well in a binaural placement.

To conclude part two of this case study we should recall the two aims set out in the phase introduction:

1. To define an approach that could afford reflexivity in exploring the creative staging research problem.
2. To create staging practices that could examine the typically problematic areas of binaural localisation for the binaural headphone translation evaluation of phase 2.

Although the efficacy of the second aim of this particular study can only be fully concluded following the binaural headphone analysis in the next sub-section, what we can conclude here is that using mono and split-stereo sound sources certainly affords a more creative and reflexive approach to spatial staging and production than those recording techniques previously outlined. As can be observed in the structure of the techniques detailed within the ‘Penny Drops’ staging practice[38], spatial emulation using mono and split-stereo sources provide a greater freedom in the construction of abstract and metaphorical sound stages and unusual sound source positioning that are otherwise unachievable through the examined methods of recording and the current methods of music production. This fulfils the first aim in providing a reflexive opportunity for exploring the research problem; investigating the possibilities of non-front orientated staging and production practice in a way that more comprehensively utilises the spatial sound stage.

This study fulfils part of the second aim through the creation of test staging structures that theoretically could be used to examine the problematic areas of binaural perception. However, because these techniques were constructed on a speaker system without any binaural output monitoring, there was no means for evaluating their efficacy until phase 2 was complete. Therefore, the next stage was to render the mixes for binaural reproduction and analyse their translation across varying headphone types. This evaluation then determined the success of the test structures in exposing any anomalies relating to the binaural format, and allowed for a full conclusion against the studies second aim.

[38] This can also be seen in the chapter 6 discussions regarding creative staging practice implemented in Penny Drops and Monomorphic, as well several other pieces to follow.

5.2.2 - Phase 2: Binaural Encoding – Decoding and Translation Analysis

Context and Rationale - Foreword

Shortly after embarking on the phase one investigation I discovered that the company behind the Auro system had recalled the binaural-rendering tool. This is why there was no means for binaural monitoring at the time of the preliminary mix creation. This came as a great inconvenience to the development and progression of this research project, as now it was no longer possible to move on to the translation examination of phase 2. Instead, this presented a new challenge and phase 2 became an investigation into how to get the 13.1 Auro mixes into the 2 channels of a pair of headphones.

'How can the Auro 13.1 channel mix be converted for a two-channel headphone-based delivery format?'

At the time of investigation in the spring of 2016, there were a number of DAW-based plug-in tools available that could 'binauralise' mono and stereo .wav signals, some even catered for traditional, single-tiered surround sound channel configurations (e.g. IRCam HEAR[39]). These methods work by applying a generic HRTF-based filtering to the master mix channel which manipulates the audio spectral content and L-R phase relationships to generate a binaural illusion. Although successful in converting mono, stereo

[39] This was the method I used for WhispersRed's sold-out 'ASMR Happens Live' – a binaural theatre show I worked on in 2016. I used IRCam HEAR to binaurally render the stereo sound cues and it worked well for the show requirements. However, had the DearVR Pro been released then, it would certainly have been the better tool for this performance as it affords individual sound source panning and automation within the binaural sound field.

or surround signals to binaural for basic applications, these binaural rendering tools were incompatible in this circumstance due to the channel count and the specific tri-tiered configuration of the Auro 3D system. There was a very limited choice of tools available at the time that could work with this format and those tools that could accommodate the number of channels could not always accommodate the specific three layers of surround the channels were spread across. Further to this, many of these tools were designed for Ambisonic formats and were Max MSP-based patches. Max is an excellent development environment that uses an object-based coding approach for the purpose of constructing digital signal processors, interactive audio-visual systems, gestural controls and interfacing tools[40]. The patches available are often open-source and promote a level of democracy across the audio coding and DSP field. However, the use of Max as a music production tool remains disconnected from traditional DAW-based music production workflows, outside of Ableton Live[41]. One could say its use tends to be confined to academic, coding and experimental audio circles[42]. For a creative record producer with little to no experience with Max, this format of programme is quite unfamiliar and can prove to not only be intimidating but also a hindrance; impacting one’s workflow, signal flow logic, creative agency and artistry. It can take a considerable amount of practice to become fluent with the functionality of a Max-based coding environment and, as such, this was considered a less than viable option, although a final resort. In accordance with the democratic ideology of this project, it is imperative that

[40] It is not within the scope of this project to code a GUI for the purpose of the binaural spatialisation.
[41] Ableton integrates a Max MSP workflow using the ‘Max for Live’ extension.
[42] Audio installations and interactive or avant-garde sound-art applications, particularly.

the binaural spatialisation or rendering tools exist to function in the environment a record producer or mix engineer would work in; as a VST or AU plugin accessible within a typically used digital audio workstation, such as; Protools, Reaper or Logic. Although Max can be used to build such a plug-in tool, this was definitely out of the scope of this project and my professional remit. Therefore, alternative options were first considered and two experimental DAW-based approaches were defined for testing; an Ambisonics-based VST option, and if that did not work, then there was the possibility of re-amping to binaural using the KU100 dummy head microphone[43].

Before we explore these approaches in practice, this section will first detail the functionality of Ambisonic technology and set the context for the use of Ambisonic formats within this project.

A theoretical consideration of Ambisonic formats.

Ambisonic is an isotropic recording and delivery format invented by Michael Gerzon in the 1970’s. Isotropic refers to uniformity in all directions; sound from any direction being treated equally, with isotropic radiation measuring the same intensity in any given direction (McWilliams, 2011). This fundamentally applies through the principle of spherical harmonics, which underpins the function of Ambisonic systems through both the physical

[43] In the past I have used binaural re-amping using the Roland CS10EM in –ear microphones to binaurally ‘print’ a 5.1 surround mix. The result was not perfect but it did provide a basic binauralisation without too many perceptual problems presented in the playback.

construction of the microphone's capsules and the processing algorithms. This affords the ability to reproduce the complete sphere of a sound using as few as four channels (first order); W – non-directional mono signal sum component and the X, Y, Z are the directional components of the sound in three dimensions. The various format orders run from first to sixth with the channel count requirements exponentially growing from 4 to 46, respectively. The higher the order, the higher the resolution (Waves, 2017). Higher orders also provide different polar patterns, and as such, provide more detailed spherical captures; third order is the minimum needed for capturing a complete sphere (lower orders than this do not reproduce height). The scientific principles[44] underpinning the functionality of Ambisonic systems are not common knowledge amongst most recording engineers or commercial music producers, and tend to be a specialist knowledge requirement of those in fields specifically seeking a more 'naturalistic' ambient spatial audio capture, such as; nature/wildlife/cityscape recordists, some AV industries – VR, film and television audio production, live ambient spatial audio capture (classical music performance etc.), spatial audio and psychoacoustic research. Traditionally, Ambisonic recording required an expensive proprietary rig comprising a large hardware decoder and a multi-coincident microphone, such as the Sound Field[45]. However, Ambisonic technology has now been integrated into more affordable consumer products, such as the Sennheiser Ambeo microphone, which uses a software app-based decoder to reformat the output audio. These democratic advances in Ambisonic

[44] This refers to spherical harmonics, the affordances of various levels of order and the function of the quad-coincident configuration of capsules etc.
[45] This was used in the aforementioned spatial recording case study in phase 1.

technology are driven by the increased popularity of VR, but research suggests that even proponents of VR are not keen on Ambisonic recording as a means to capture source audio (Boom Library, 2019). Research shows that, although Ambisonic captures rotate well and, in particular, that function suits VR audio, spaced microphone arrays and 3D microphones[46] offer more stable localisation and imagery than that offered through an Ambisonic microphone. This is due to the co-incident nature of the Ambisonic capsule configuration; at low resolution Ambisonic recordings can present a smaller image with a diffuse and difficult to localise phantom sound source, often with severe colouration (Boom Library, 2019; Politis et al., 2018). Contrary to this, Ambisonic imaging responds particularly well when used as a close microphone recording technique (Boom Library, 2019) (McWilliams, 2011) and it does present a convenient way to record in surround sound using a small footprint - which is especially useful in circumstances where a spaced array may be impractical. Ambisonic formats also upmix and downmix better than spaced arrays, and are *system agnostic* – this means they can be flexibly decoded for playback across a range of systems, from multi-channel speakers to headphones – this is certainly the biggest affordance of the Ambisonic format (McWilliams, 2011; Politis et al., 2018; Boom Library, 2019).

[46] Such as Shoeps ORTF3D spaced 3D microphone.

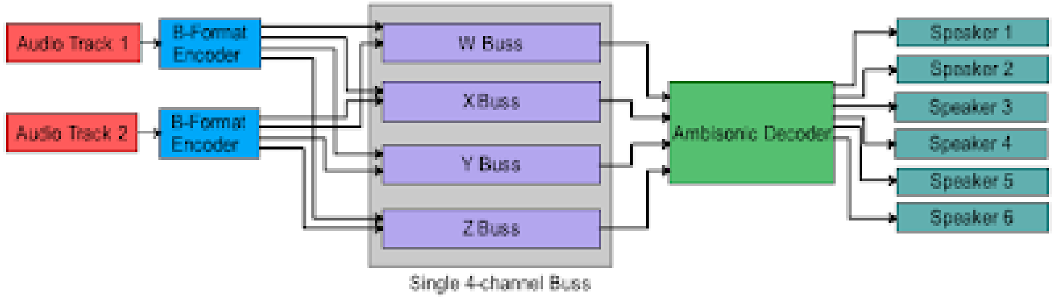


Figure 35: An example Ambisonic first order encoding-decoding flow (McWilliams, 2011)

What we can understand from this research is that an Ambisonic output is an incredibly useful consideration, although it is not a necessity of the project providing that, at a minimum, the rendering tool could output two-channel binaural in legacy .wav format. In reference to 3D audio systems using Ambisonic delivery formats, the Boom Library (2019) article states that there is no necessity for source audio to be native Ambisonic recordings when an Ambisonic output is an option for delivery. This statement implies that the legacy .wav output from the Auro system can be accommodated by an Ambisonic-based binaural rendering tool and suggests that the usefulness of an Ambisonic format lies not necessarily in the method of capture but in generating a headphone-based output. One company, Blue Ripple Sound, developed by Richard Furse, provides a possible DAW-based solution to the binaural delivery conundrum. Blue Ripple have a set of Ambisonic based plug-ins that can cater for the unique Auro 13.1 configuration, these are called ‘O3A encoders and binaural surround decoders’ and they are available in a VST format that can be used within a DAW environment (Blue Ripple Sound, 2015).

However, contrary to what the statement from the Boom Library article implies, the binaural surround decoder does require Ambisonic b-format audio as an input source[47] and this meant that I would first need to encode the project's discreet legacy audio[48] to Ambisonic b-format using the O3A encoder before I could use the binaural plug-in decoder to render the mixes. However, on the Blue Ripple Sound (2015) website there was a negative disclaimer that read "…these plugins are not generally the best way to make great-sounding binaural mixes with our tools!" and this left me anxious as to whether these would actually be the solution I sought. At the time of investigation (2016) to buy the O3A Upmixer and Decoder set the costing would have been circa £1000. Which was a substantially large financial investment for something that suggested such high risk. I contacted Richard Furse and explained my predicament. Richard kindly provided a one-month trial license of the plugins I required and so began the phase 2 O3A upmix/decode method exploration.

[47] B-format refers to the aforementioned WXYZ Ambisonic channels. See figure 35 above for an example of the encoding and decoding flow.
[48] Discreet legacy audio is an alternate format term for .wav output channels.

5.2.3 – Practice as Research Phase 2: Auro 13.1 x O3A Upmix/Decode to Headphones

Overview

The previous section outlined the binaural delivery issues pertaining to the multi-channel speaker-based mixes and concludes that select Ambisonic tools may provide a resolution. This section of the study attempts to address the following underpinning preliminary questions; *‘How can the Auro 13.1 channel mix be converted for a two-channel headphone-based delivery format?’, ‘What systems could be used?’* through exploring the ‘Upmix/Decode Method’: an experimental Ambisonic-based approach in rendering the channel-based Auro 3D mixes for a headphone-based delivery.

The sonic elements constructed in the preliminary mixes of ‘Penny Drops’ and ‘Monomorphic’ provide reference anchors to examine the presentation of the periphonic-binaural aesthetic and assess the mix translation from speaker array to headphones. The production aesthetics were designed with a particular focus on examining localisation and XYZ plane response in pursuit of identifying perceptual issues with the decoding that may be presented through front-back and up-down localisation anomalies.

The objective of this phase is to define a headphone-based spatialisation system on which to undertake the research as practice investigation of phase 3 and the data collection of phase 4. The aims of this study were to determine whether the O3A upmix/decode process is a suitable candidate for decoding the speaker-based Auro mixes, and whether the creative and musical aspects of the speaker-defined production techniques translate to headphones as intended.

O3A Upmix-Decode: key considerations

- Aim: Define whether the headphone decoding distorts the implied spatial aesthetic as designed on the speaker array.
- Research as practice: Implement the O3A upmix/decode method, as detailed below.
- Critical listening analysis: Check mix structure and image stability across vertical, overhead, rear & side placements using phenomenological method and the binaural test structures as anchors. Observe and describe how the staging test structures respond in these areas.
- Conclusion and Reflexion: Define adaptations and negotiations to production practice, if necessary.

This method uses the Blue Ripple Sound Upmixers toolkit and O3A plugin pack to first upmix the legacy audio to Ambisonics b-format. Once in b-format the Ambisonic content can be decoded to binaural or stereo C-format and printed to a 2-channel legacy format, such as .wav. Given the increased channel count per track required for higher-order Ambisonics it is necessary to change DAW as Protools cannot provide the number of output channels needed. Reaper can accommodate for the 16 channel output (per audio track) that the third order Ambisonics protocol requires [49].

O3A Ambisonics Upmixing and Decoding Method Process

Step 1 *– Export the 14 Auro mix output channels as discreet split multi-channel outputs.*

- Lower Layer - *Left, Centre, Right, Left Surround, Right Surround, LFE, Rear Left Surround & Rear Right Surround.*
- Height Layer - *Left, Centre, Right, Left Surround, Right Surround*
- Top Layer – *Top* channel.

Step 2 *- Configure the session and routing in Reaper and import the Auro output audio, as below.*

Reaper Configuration[50]**:**

It is very important to get the routing configuration correct, otherwise there will be issues in the processing that will deform the end result. Ensure that the channel and layer sequence in the correct order (L, C, R, Ls, Rs etc.), as well as the auxiliary routing for the upmix and decoding processes.

[49] I also had to use Reaper as the licenses were assigned to me and not to the University where I was working on the Protools and Auro system.
[50] A DAW switch to Reaper was necessary as the Blue Ripple O3A upmix and decoder tools required more output channels than Protools HD could offer at the time.

1. ***Channel creation***: Create 14 channels each with a 16 output count. Each individual channel track output must be set to carry a minimum of 16 channels to accommodate the requirements of third order Ambisonic b-format.
2. ***Import audio:*** Add the corresponding audio file to the configured channel and remove from main master mix (left and right) output routing.
3. ***Panner Assign:*** Each channel must then be assigned an O3A panner and each channel must be panned into the correct positon, as per the requirements of the speaker array and format you are working in (Auro has specific spaced measurements that differ to ITU 5.1 / 7.2 standard and this needs to be considered).
4. ***Aux Configuration:*** Create 2 auxiliary group channels (again carrying a minimum of 16 output channels each); one aux for the O3A upmixing plugin and the second aux for the binaural and or stereo decoder(s). Add the plugins as appropriate.
   Ensure that the correct upmixer is selected relative to the format being worked – O3A caters for many multi-channel formats, including Auro 13.1. Remove the main mix (left and right) output routing from the upmixer channel and route it to the decoder bus. Route the decoder bus to the main mix output (left and right).
5. ***Track to Aux routing:*** Route the audio track channels to the upmixer bus - the upmixer bus will convert the legacy audio into third order Ambisonic b-format (TOA) – there is no way to monitor this interim process without the decoder switched ‘in’. It is at this stage that the

desired settings can be made between the upmixer and decoder plugins in order to create the appropriate translation.

6. ***Render the master mix out:*** The binaural or stereo mix output may now be printed, as per the usual process.

Given the correct routing, the upmixer bus should send the TOA converted audio through to the decoder bus. This is then decoded to binaural by the decoder plugin and sent to the main mix output where it should appear as a widely playable binaural version of the 13.1 mixes and can be exported in legacy .wav audio format.

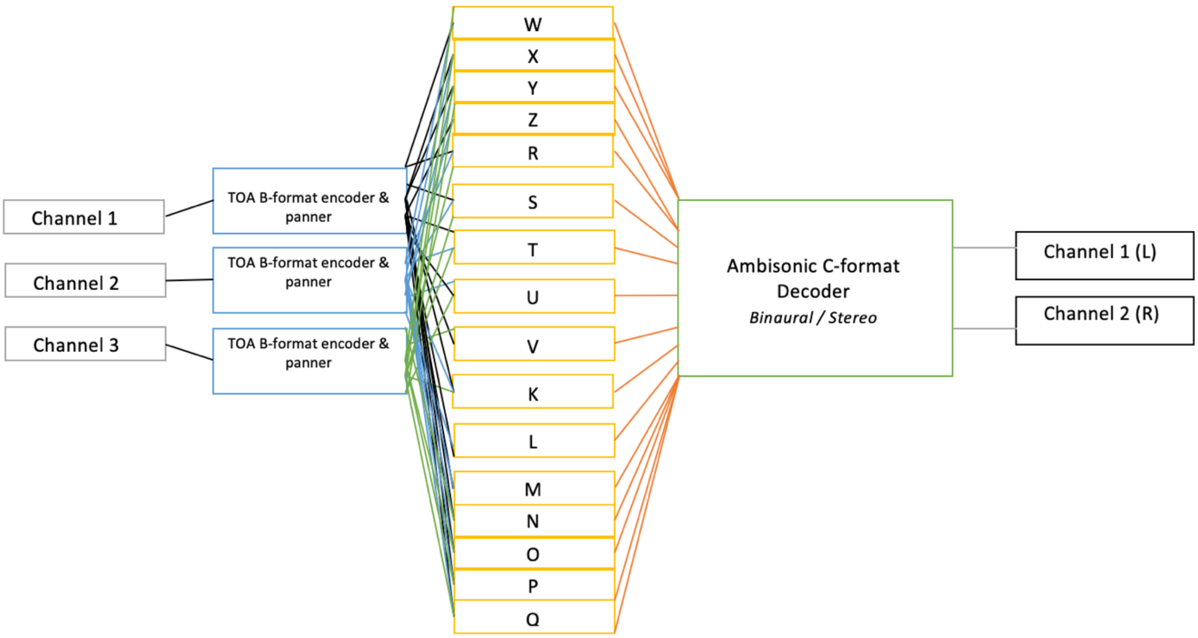


Figure 36: An example of the TOA O3A upmix/decode flow for 3 of the 14 channels.[51]

[51] This diagram depicts the signal flow and processes involved with upmixing legacy audio formats to third order Ambisonic b-format. It shows how the spatial information produced by the O3A panner & encoder is shared to all 16 TOA channels (K-Z) in their specific orders, and how this information is then passed to the c-format decoder to be rendered back to legacy audio format for binaural or stereo playback.

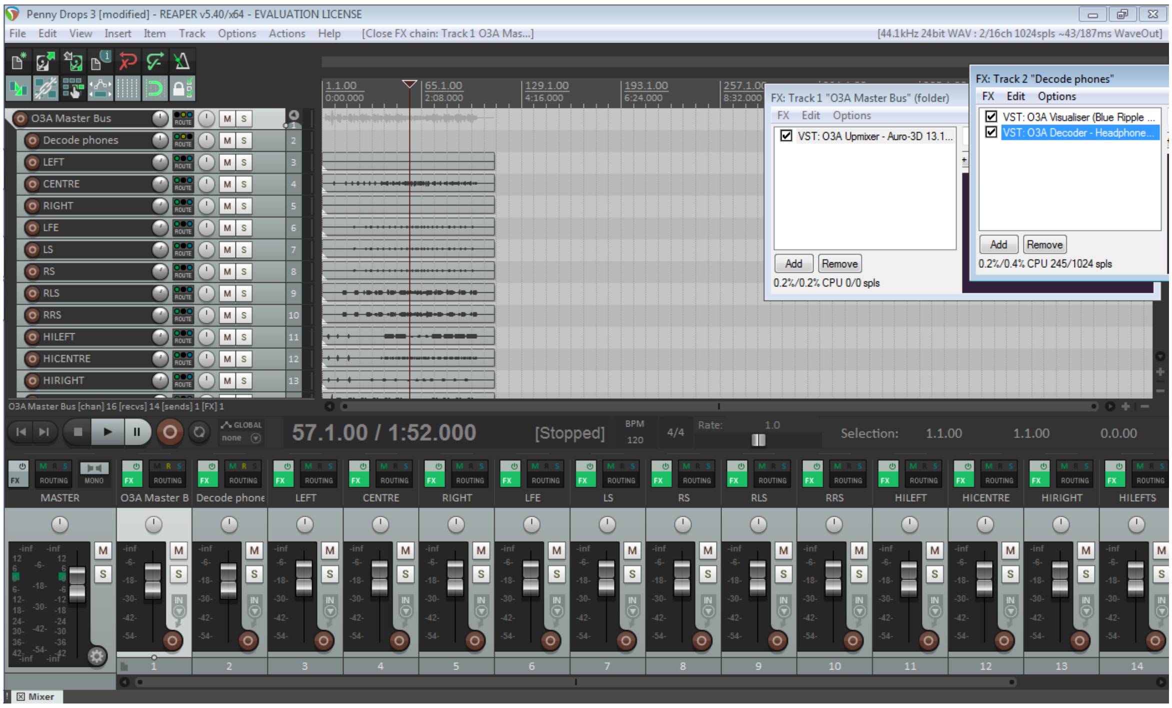


Figure 37: Overview of the upmixing and decoding process within Reaper.

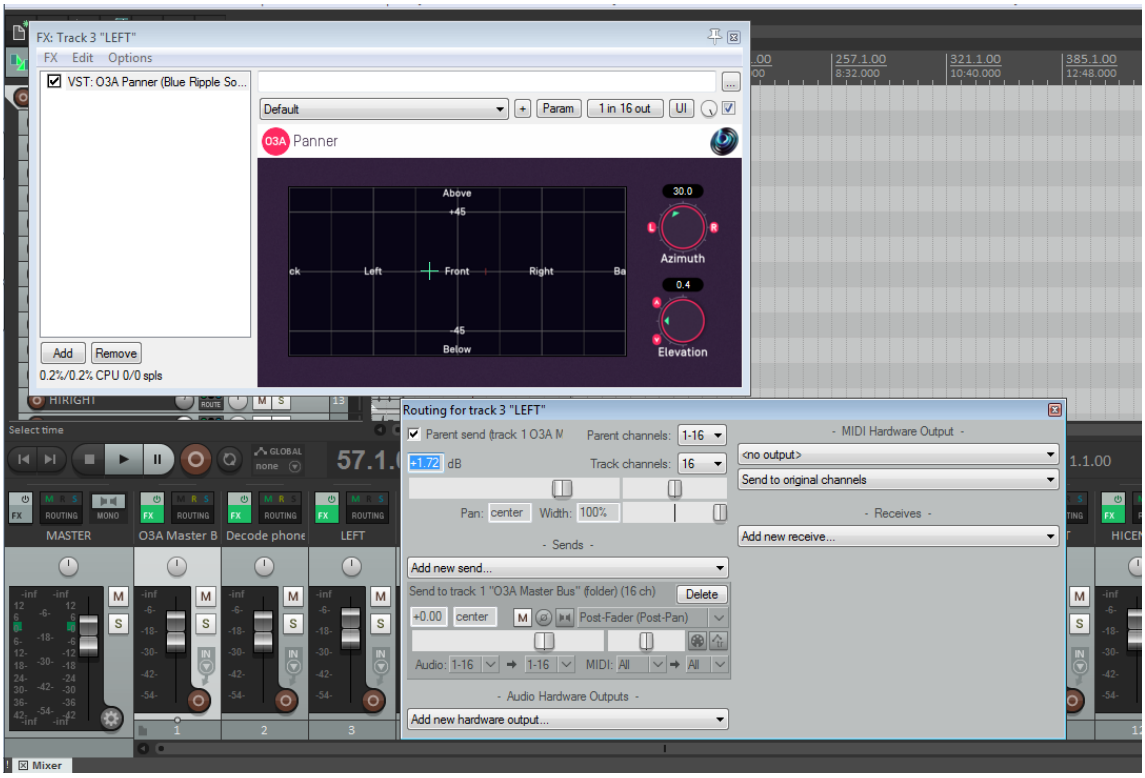


Figure 38: Example of audio channel panning and routing.

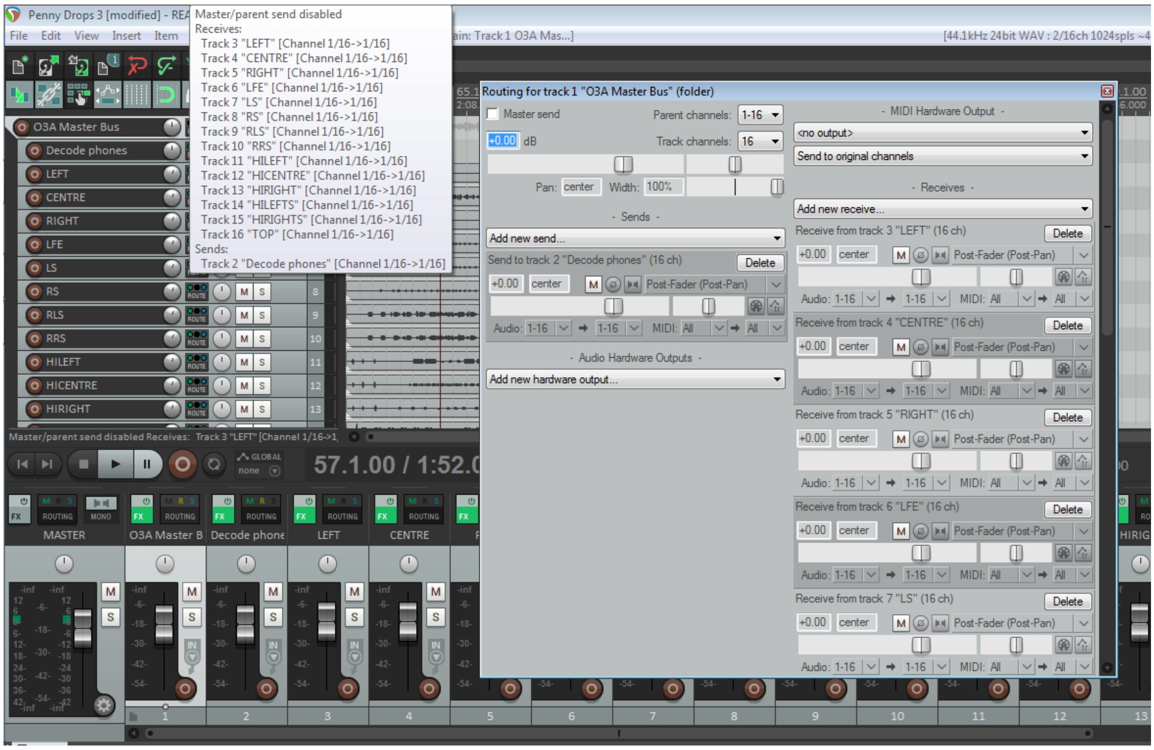


Figure 39: O3A Upmixer auxiliary channel send and receive routing information.

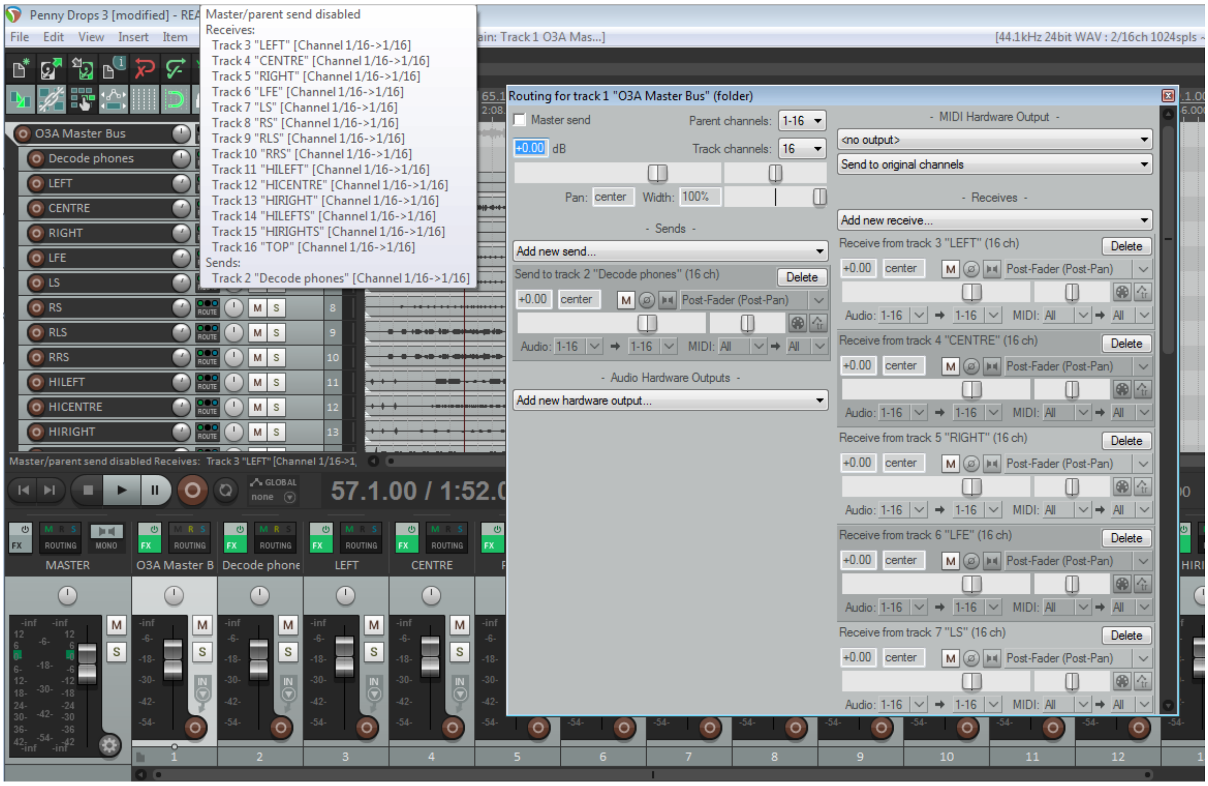


Figure 40: O3A Decoder auxiliary channel send and receive routing

5.2.4 – Summary of Findings: O3A Headphone Analysis

Although convoluted, the Ambisonic-based upmix-decode system worked. The 13.1 productions were reproduced binaurally, albeit with an interesting and surprising result. Some elements had translated well, several elements were exceptional, but there were obvious discrepancies between the speaker-based productions, the expectation and the resultant binaural output. The following reference headphone analysis was conducted with a pair of AKG K271 MKII over-ear, closed back headphones.

Piece 1 - 'Penny Drops' – O3A Binaural Headphone Analysis

1. The entire sound stage translated as narrowed and skewed, although some elements made this more noticeable than others.

This was particularly noticeable when observing the trajectory of the 'crank' technique in the 'Penny Drops' piece, which should have been in an arch shape, moving up and over through the centre of the sound stage. The intended trajectory as set out on the speakers was from front centre to top to rear centre. Instead, this technique appeared to traverse from front-left to top to rear right, although the localisation accuracy diminished considerably as it reached the top.

2. Sound source elements were not in the intended place.

This was particularly noticeable when observing the 'Penny Drop' sonic cartoon of a coin-toss. Which presented the first sample appearing as coming from the top of the head and the second sample appearing toward the cervical spine at the top of the neck. The intention was for sample 1 to be elevated rear-right and sample two to be low-rear-centre (directly behind the listener). The effect was also heavily internalised, which actually made it more cartoon-like. This presented the perception that the coin was falling into my head and rattled to a stop on the top of my spine, as if my skull was a piggy bank. Due to this exaggerated sonic-cartoon aesthetic, the misplacement of the sources was not severe enough to detract from the overall effect - in fact, it heightened it - and it's one of the more exceptional positive elements of this decoding result.

3. There were front-back discrepancies.

This was particularly noticeable when observing the 'follower & chaser' techniques. Although it was clear that they moved and the sense of movement overall was very good. It was not always distinguishable whether the left-to-right and right-to-left panning was happening in front or behind. It felt as if it vaguely went across and through the head. However, the combination of static and kinetic sources did enhance the sense of movement, making it feel like the percussion sources were flying around and bumping into boundaries (the theory, intention and aesthetics of this are discussed in the creative staging discourse of chapter 6).

4. There were spectral or perceptual masking issues.

There were seemingly masking or perceptual issues that prevented certain sources from cutting through at particular times and in particular areas. This was noticeable when observing the 'winch' spiralled hi-hat technique, in some places it would cut-through well and in others it would not – this problem would present quite consistently. Therefore, I first thought that perhaps this was to do with frequency-clashes presenting masking between sources such as the hi-hat and the 'crank'. However, over several listens the hi-hat also seemingly became less noticeable at other times and in other areas, unless I intently focussed in on it. Which leads to thinking that this could be due to an auditory scene overload presenting a sort of inattentional deafness (Bregman, 1990; Koreimann et al., 2014). Inattentional deafness is a perceptual masking phenomena whereby we do not attend to certain sources because they blend in to the background or because we are distracted by other, usually visual, phenomena. Upon further analysis it suggestively seemed as though too much simultaneous movement from multiple sources, along with an amount of frequency-masking, could possibly have led to distraction and the inability to properly localise and perceive the high-hat amongst the plethora of other kinetic sound sources and structures. I made a comment on this myself within one of my Auro mix session videos in 2017, where I make a state that *"Too much movement is not good – it sounds messy… simple patterns work best." – Jo Lord (2017)*[52].

Although there were other staging structures that both did and did not present well following the O3A binaural decoding, those reiterated above had the most significant impact on the binaural mix translation.

I checked the mix translation again across a further two variations of headphones; Sennheiser HD25 on-ear and Apple Earpods in-ear. Generally, each presentation was quite similar to the AKG reference analysis, with minor variations as detailed below.

[52] A quote of my own from production session video reviewing the percussive elements in 'Penny Drops'.

Sennheiser HD25 (on ear, closed back):

- Good movement and periphonic perception; a similar holistic response to the K271.
- Less accurate localisation than K271 that particularly effects the front-back and vertical perception of the ‘crank’ and ‘winch’ techniques.
- The overall instrumentation is more present than appears on the K271, presenting a minor disturbance to the depth and distance perception of sources.

Apple Earpods (in ear, open back):

- Heavily internalised reproduction.
- Good perceivable movement and periphony.
- Presenting similar localisation issues as common to the previous assessments (spectral masking, source position inaccuracies). The ‘crank’ technique responded better but it was still off-centre and the perception of the ‘winch’ hi-hat remained affected as previous assessments.
- The ‘penny drop’ coin-toss sonic cartoon was out of place but the overall internalised effect remained consistent.

Piece 1 – ‘Penny Drops’ O3A Stereo Decoding Headphone Analysis

As the O3A decoder plug-in also presented the ability to decode for speaker-based stereo I decided to implement this to see if the resultant output would respond better on headphones than the binaural output had, although I had no expectation for the periphonic sound stage to localise in surround. The source positioning was spread across an ultra-wide stereo image not too dissimilar to the binaural spread but lacking the spatial qualities of a binaural perception. Interestingly, most elements were in the correct position relative to a wide stereo sound field. However, the ‘crank’ positioning was still off-set to the left and moving in a rightwards trajectory. I double-checked the channel routing within the upmix/decode Reaper session, just in case the channel order was wrong but the order was correct. I can only assume this is an issue presented through the multiple format changes of the upmix/decode process itself. I used the same three pairs of headphones for the stereo decoding analysis and key elements can be found below.

AKG K271 (over ear, closed back):

- Responds particularly well, though left-right panning movements are exaggerated.
- Good localisation of sources with minor and sporadic issues.
- Presents the common image skewering pertaining to the ‘crank’.
- Penny Drop coin-toss cartoon remains out of place, though it is considerably more internalised in this stereo presentation.

Sennheiser HD25 (on ear, closed back):

- Good general sense of movement, though some movements appear skewered. This response to movement appears worse than in the binaural presentation on the same pair of headphones.
- There is still the apparent spectral and perceptual masking as previous assessments found. The hi-hat particularly suffers.
- The ‘crank’ technique responded well, although the trajectory remains off-set.
- Overall, the high-mid content of the presentation is very present. This is a characteristic of these headphones, which I assume is exacerbated by the increased amount of presence (and lack of spatiality) of the stereo render.

Apple Earpods (in ear, open back):

- Good perceivable movements and an exaggerated sense of width.
- A lot of spectral masking in this presentation causing a loss in perception of instrumentation, an unbalanced stage and a loss of definition.
- The same imaging and placement issues arise as perceived in the previous assessments; the hi-hat ‘winch’ and the percussive ‘crank’ remain problematic.
- ‘Penny drop’ coin-toss sonic cartoon presents as only slightly out of the intended placement, as defined on the speakers. However, the internalised effect remains unaffected.

Although it is obvious that the results are problematic, in order to determine the efficacy of the O3A upmix/decode approach we should revisit the reflective questioning set through the aesthetic judgement criteria.

- *Does this reproduction reflect the original staging intentions of the speaker-based mix?*

  No.
- *Do I perceive sound sources from above?*

  Yes.
- *Do I perceive sound sources from behind me?*

  Yes, although in some instances they are not where they should be.

- *Do the sound sources localise in reflection to their intended placement in the sound field?*

  No, many of them do not.
- *Are there any perceivable anomalies in the presentation*?

  Yes.
- *Where and when do they occur?*

  See the detailed analytical description above.

Although it is clear that aesthetically this approach is unsuitable, we should also refer back to the ideological judgement criteria outlined in the methodology chapter to fully assess whether this approach meets any of the standards set.

- *Does the processing tool require specialist acquisition?*

Yes. The upmix plugin pack was expensive, even though the binaural decoder is not and both were temporarily acquired through contacting the developer.

- *Does the processing tool require specialist knowledge to implement?*

Although the tools did not require a specific specialist knowledge of Ambisonics to implement, they did require a large amount of research to define and there were specific protocols required to ensure their correct functionality (such as specific channel ordering and signal flow).

- *Did this process negatively impact the creative workflow?*

No. These tools did not impact the creative workflow in generating the initial productions. However, they did impact the mix output and analysis workflows due to the convoluted approach requiring the use of multiple DAWs and format conversion.

5.2.5 - Conclusion - O3A upmix/decode study and headphone translation analysis

The O3A upmix/decode method presented a less than desirable workflow approach that produced a less than desirable binaural result. Although some elements of the reproduction were positive, the mistranslations that occurred were negatively impacting enough that they presented an unacceptable representation of the mix structures and intentions. The workflow involved lengthy exporting and importing processes into a secondary DAW[53] to convert the audio format, which possibly also abetted the channel-based mixes not presenting accurately. Due to the high cost, ineffective method and negative implication to rendering workflow we can fully conclude that the Blue Ripple O3A plugin set provided an ineffective means of producing a binaural render from the 13.1 mixes. Therefore, it was side-lined in favour of continued research in defining a more aesthetically and ideologically appropriate decoding solution.

However, perhaps the outcome would have been different if working on Reaper in the first instance with a full license for the O3A plug in's. This may have presented the opportunity to construct the mixes using a direct-to-binaural approach, bypassing the need for the Auro multi-channel speaker system. This would also have made it possible to monitor and analyse the binaural result of the decoder in real-time within the production process, rather than mixing 'binaurally blind' on the Auro system.

Secondarily to the outcome of the phase 2 study, what we can also conclude from these results is that the second aim of the phase 1 investigation - *to create staging practices that could examine the typically problematic areas of binaural localisation* - was successfully fulfilled. The test staging structures served well as reference points for analysing the translation of the mix content across a binaural reproduction and allowed for the examination of the particularly troublesome areas of perception often associated with

[53] Though in retrospect this could possibly have been avoided by working on headphones with the O3A tools in Reaper and negating the need for Auro, Protools and the speaker-array. I think it was the specific channel-based approach using the Auro system that presented the biggest problems for this project.

binaural HRTF filtering. Ultimately, they aided the synthesis of the phase 2 conclusion in defining that the Auro-O3A upmix/decode method was not the best approach to practice.

5.2.6 – Phase 2.5: Periphonic Binaural Re-composition using Deareality's Dear VR Pro

Context

In the Winter of 2017, a short while after the 'O3A' trials were undertaken and evaluated, a new spatial panner plug in was released by an unfamiliar VR company; the Dear VR Pro by Dearreality[54]. The Dear VR Pro provides a full 360° immersive panner, 46 acoustic (reverb) environments, the ability to generate real-time early reflections and occlusion modelling. The Dear VR Pro plugin allows for mono and stereo legacy source audio, and outputs in a number of formats, including; 26 multi-channel output formats (up to 9.1.6), as well as a binaural headphone output and third order Ambisonics (in ambiX and FuMa[55]). Not only does this provide the flexibility of being system agnostic, but this is also useful if you wish to use the Dear VR Pro as a panning tool and require, or wish to use, a different set of decoders for the delivery to those used in the Dear VR engine. This means that the Dear VR Pro offers the opportunity to explore a variation of HRTF algorithms in the binaural decoding, as well as providing ways to up-down mix for other formats. In summary, it is both functional and flexible, and at £280 for a full

[54] Acquired by Sennheiser Group as of 2019.
[55] These are different types of Ambisonic algorithm.

license it is far more affordable than the O3A tool kit, which definitely suggests it is the more democratic and more flexible tool of the two.

Alongside the Dear VR Pro, Dearreality have released the Dear VR Music - which contains the same features as the Pro minus the advanced acoustic synthesis tools for early reflection generation and occlusion – importantly, both are available as VST/AAX/AU supporting any DAW (dearVR.com, [n.d]; Noizefield, 2017). Further to this, the most advantageous development realised by Dearreality is that of a completely new workflow for constructing spatial audio. Dearreality have conceptualised and released a pioneering new system based within a VR environment called ‘Spatial Connect’. It is a virtual reality audio mixing interface that uses the HTC Vive system for “mixing VR, in VR” (Hills-Duty, 2017, Kuzminski, 2018). As per an Instagram positing on the 15th August 2019, Dearreality are also assimilating the Spatial Connect with Magic Leap headset technology to make their products and workflow available across a wider range of VR headsets (dearreality via Instagram, 2019), further adding to the level of democracy of the tool. Dearreality state that they conceptualised and built the Spatial Connect system to provide the user with a means to mix VR audio in the VR environment, providing a more relative epistemology for constructing VR auralisations (Hills-Duty, 2017, Kuzminski, 2018). This reflects and reinforces my conclusion following the phase 2 Auro x O3A experimentation, that mixing from a headphone-based beginning for a binaural output would provide a better workflow and epistemology in practice, and result in a better binaural result than mixing ‘blind’ using speakers and a detached decoding process. Here is where the '*mix in the medium you are mixing for'* ideology was both established and reinforced.

Practice as Research Phase 2.5 – Dear VR Pro Evaluation

- Aim: Define how the Dear VR Pro responds periphonically and whether the binaural decoding distorts the implied spatial aesthetic of each piece, as previously designed on the speaker array.
- Research as practice: Re-compose 'Penny Drops' and 'Monomorphic' using the same staging and sound source placement techniques as defined in Phases 1 and 2.
- Critical listening analysis: Check mix structure and image stability across vertical, overhead, rear & side placements using phenomenological method when placing and moving sound sources. Examine the output across the three variations of headphones. Observe and describe how the staging test structures respond in these areas.
- Conclusion and Reflexion: Define adaptations and negotiations to production practice, if necessary.

Step 1 – Re-compose 'Penny Drops' and 'Monomorphic' employing the same staging and sound source placement techniques as defined in Phases 1 and 2 (see below for workflow method).

Step 2 - Define adaptations and negotiations to production practice through real-time critical analysis and explore the variation of translation across headphones during practice.

Step 3 – Using the defined judgement criteria, outline whether the Dear VR Pro presents as a suitable aesthetic and ideological candidate for the phase 3 creative practice.

Step 4 – If the conclusions of step 3 are positive, continue on to the phase 3 investigation and re-compose 'Far From Here' using the Dear VR Pro.

Dear VR Pro Workflow Overview

The workflow when using the Dearreality Dear VR Pro is incredibly straight forward and easily integrated. You can simply use the plug in as you would any other VST plug in, you can load it directly to a channel as an insert or you could load it on an aux or bus and send audio to it, depending on your production aesthetic and work flow requirements. The output delivery format can be selected within the plug-in GUI (see bottom right of figure 41).

1. Import stemmed audio onto separate tracks.

2. Load the Dear VR Pro plug-in onto the desired audio channel, group or aux channel and route the channel output to the Master LR.

3. Configure output format in the plug-in, spatial parameters in use / not in use (occlusion, early reflections, reverb etc.), scale down to 3m (generally this is the best scaling for musical applications, although larger scaling can

be used it often presents sources as less defined). If using Ambisonics-based outputs, remember to ensure there are enough outputs per each DAW track and mix bus channels to accommodate the Ambisonic b-format output.

3. Spatialise sound sources, as appropriate.

4. Export final mix as; 2Ch Binaural, 2Ch Stereo or Ambisonic b-format (FuMa/AmbiX) for alternative later decoding. If using alternative Ambisonic decoding, ensure the TOA Ambisonic output of choice is selected on all plug in's used, bus the entire mix to the Master/Aux (with enough channels to accommodate the b-format output), load the desired Ambisonics decoder to the Master/Aux and then export/output the decoded mix as required.

Figure 41: Dearreality Dear VR Pro spatial emulator plug in (Audiofanzine, 2019)

Summary of Findings: Dear VR Pro Analysis

The Dear VR Pro plugin presented an opportunity to re-explore and reflect on the ‘Penny Drops’ and ‘Monomorphic’ staging approaches, re-examining the problematic areas of binaural perception in real-time within the binaural format. Consequently, this allowed for a greater focus on evaluating the periphonic qualities and the creative possibilities that the *direct-to-binaural* tool affords. The real-time monitoring meant the analysis of both source audio and headphone responses could happen during the re-composition production practice. This resulted in immediately reflexive practice achieved through a faster analysis process and quicker on-the-spot decision-making.

In terms of the re-composition practice, most of the binaural test structures could be replicated as originally intended using the Dear VR Pro spatial emulation plugin. The only production techniques that struggled to be represented were the specifically channel-based approaches, such as: the ‘Heavenly Voice’ reverb cascade; an automated and panned reverb output, presenting simultaneously across all horizontal channels and traversing through each of the 3 speaker layers of the Auro system. This was not possible to achieve using the Dear VR Pro approach. Although there were attempts to reformat it, these did not result in the same effect (this was a creative staging approach and as such this is discussed fully in chapter 6). The ‘penny drop’ coin-toss test structure also did not present the same effect as apparent through the Auro-O3A approach. Although the placement of sources could be replicated, the result of the binaural effect was less exaggerated than that presented through the Auro and O3A method, which resulted in the effect feeling less internalised (less like the coin falling inside of the listener's head). Overall, the other structures were recomposed in the same placements, which reproduced well using the Dear VR Pro plugin's binaural output format.

Although the Dear VR Pro plugin can output Ambisonics TOA b-format, the process associated with generating the production work shows us that the Boom Library (2019) article was correct in this case; native Ambisonics recording isn't a necessity and mono legacy source audio can be used in the construction of the productions. This affords a democratic and convenient approach to spatialising past and future projects, whilst allowing for a flexible Ambisonics decoded delivery, if required. This approach to spatial emulation using mono sources is further supported in an article discussing immersive sound design, where BBC sound engineer Tom Parnell refers to binaural DSP technology stating "This allows us to pan audio in any position around the listener's head when they are listening on headphones… this approach offers much more freedom when recording and more creative control when crafting immersive audio mixes" (Ramsey, 2017).

Dear VR Pro Headphone Analysis

The three pairs of headphones were evaluated before fully commencing the phase 2.5 staging practice so as to define the appropriate pair to use for monitoring when creating the Dear VR Pro versions of the ‘Penny Drops’ and ‘Monomorphic’ recompositions. It was decided that the AKG K271 generally presents a less colourful and characteristic reproduction across varied loudness levels, with a more balanced frequency response across the spectrum at lower mixing levels than either the HD25 or Apple Earpods afford (see figures 43, 45 and 48). The HD25’s have quite a lot of punch and presence and this does cause over-compensation in the mix process and a less ‘open’ binaural response. This could be due to the imbalance due to the increased bass response of the headphones presenting less ‘air’ and ‘sheen’ in the higher frequencies (see figures 42-46). As we know, higher frequencies are important for elevation perception. Although comparatively, the Apple Earpods have more perceived higher frequency content, the lack of low end and the U-curve present in their perceived responses does not present an accurate means for mixing. However, as many people wear this type of headphone it could be used as a comparative reference. The over-ear design of the K271 is more appropriate for precision mixing work than the on-ear or in-ear designs of the other two sets, with an overall sound quality that was more robust and reliable for defining the sound source placements.

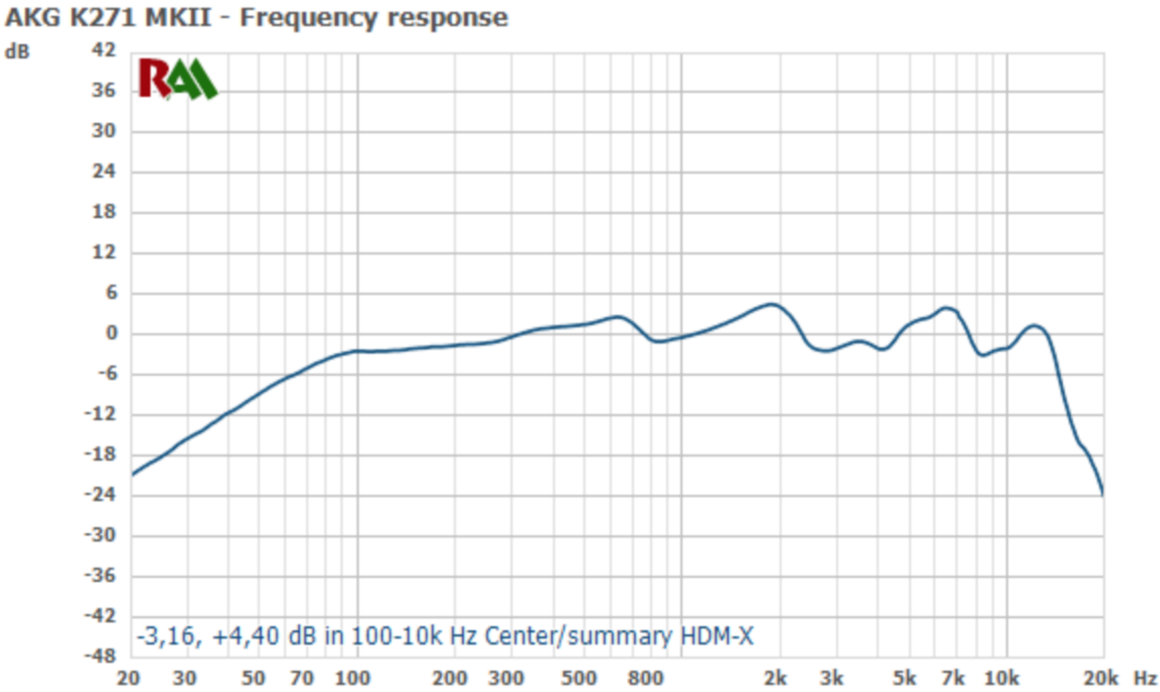


Figure 42: AKG K271 frequency response (RAA, n.da)

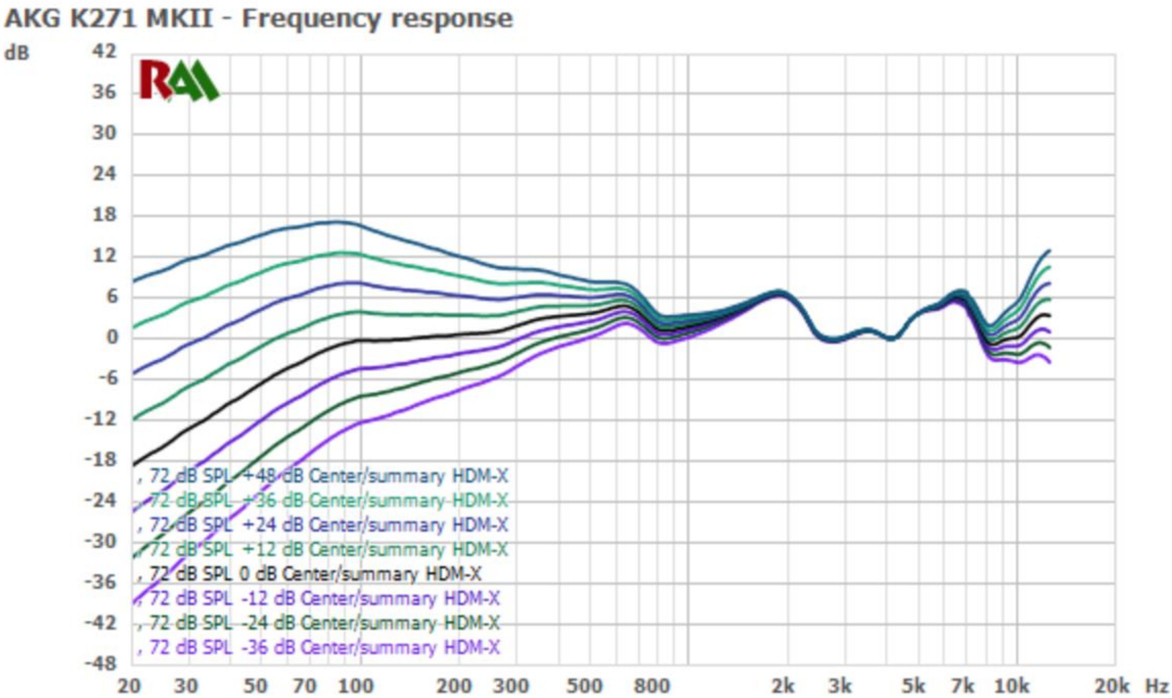


Figure 43: AKG K271 perceived frequency response (RAA, n.da)

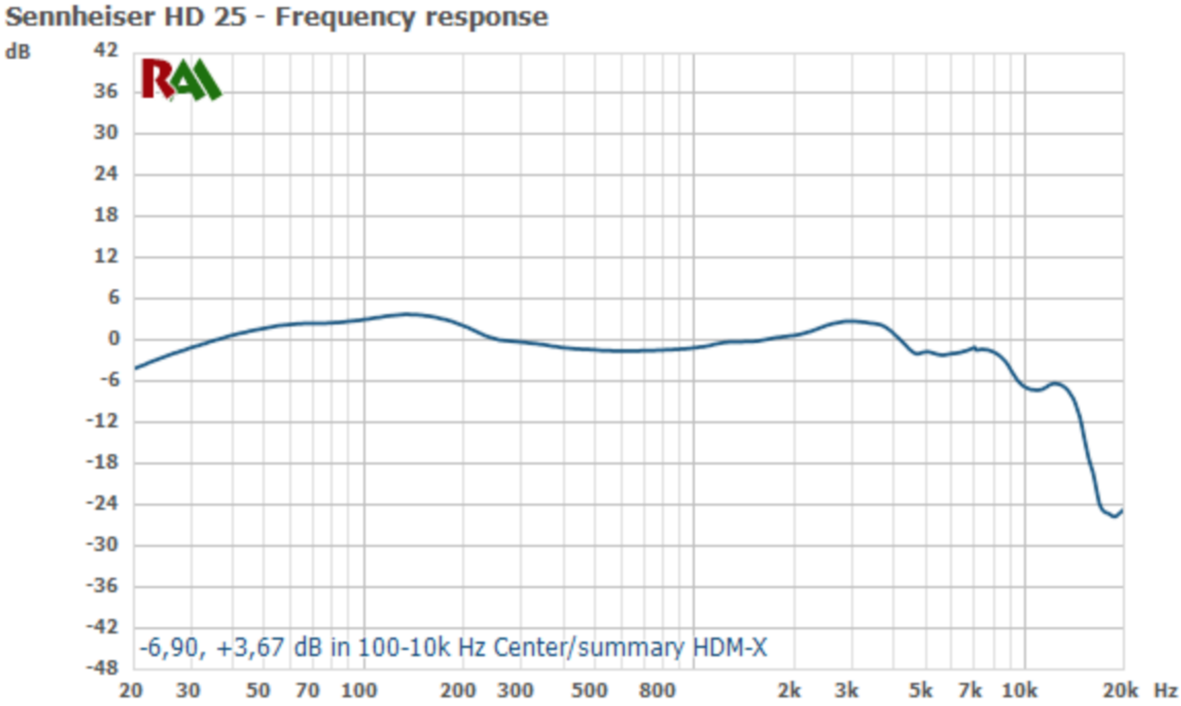


Figure 44: Sennheiser HD25 frequency response (RAA, n.db)

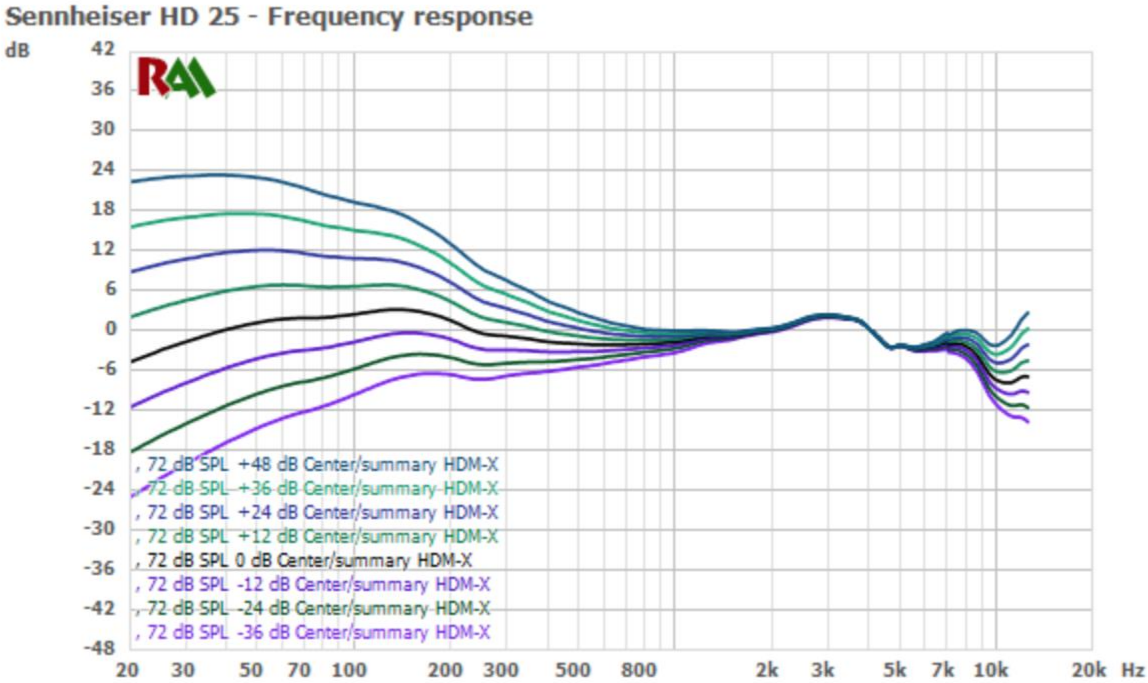


Figure 45: Sennheiser HD25 perceived frequency response (RAA, n.db)

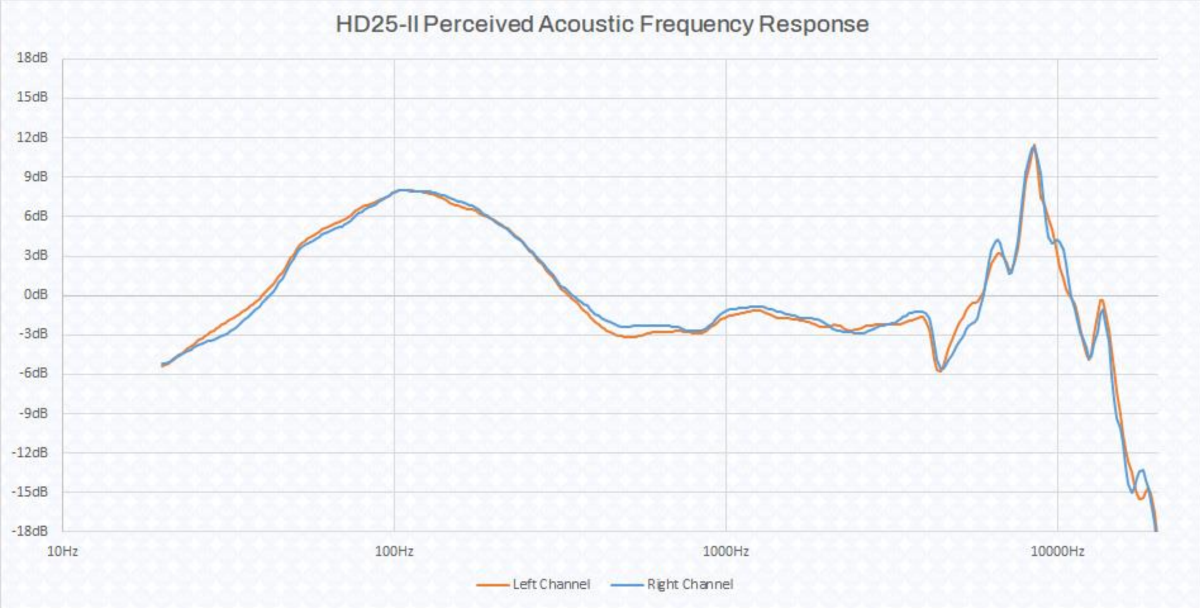


Figure 46: HD25 perceived frequency response showing U-curve 'smile'

(Sonarworks, 2019)

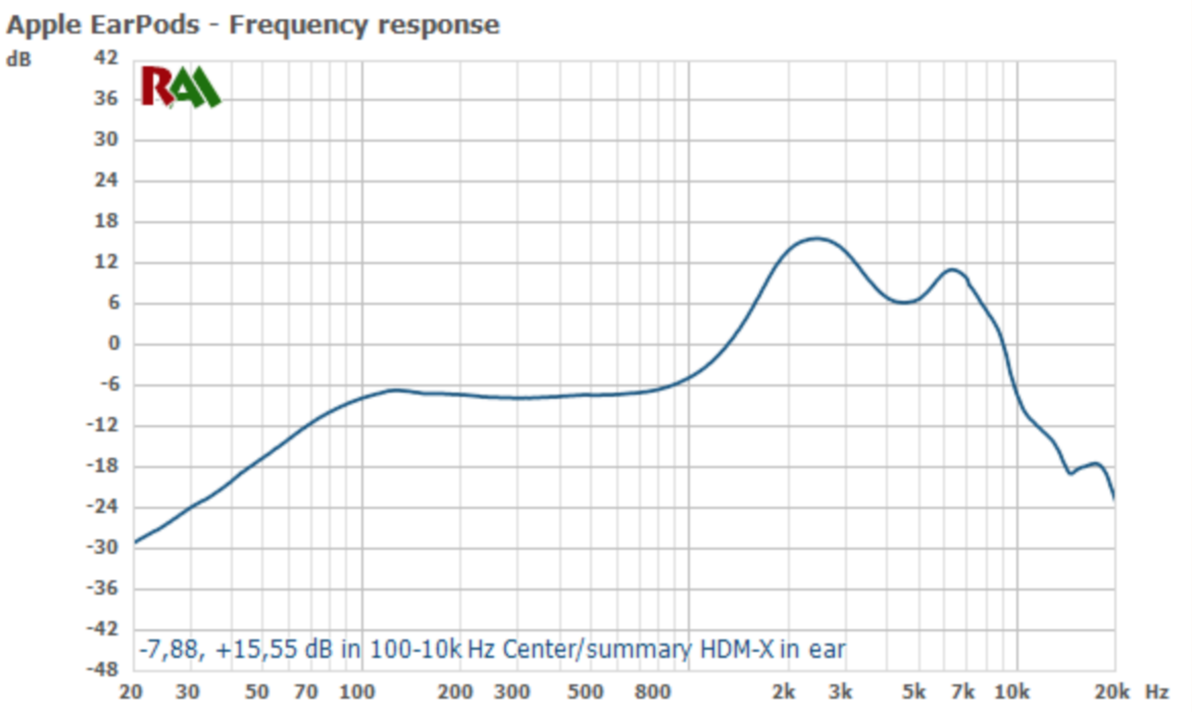


Figure 47: Apple Earpods frequency response (RAA, n.dc)

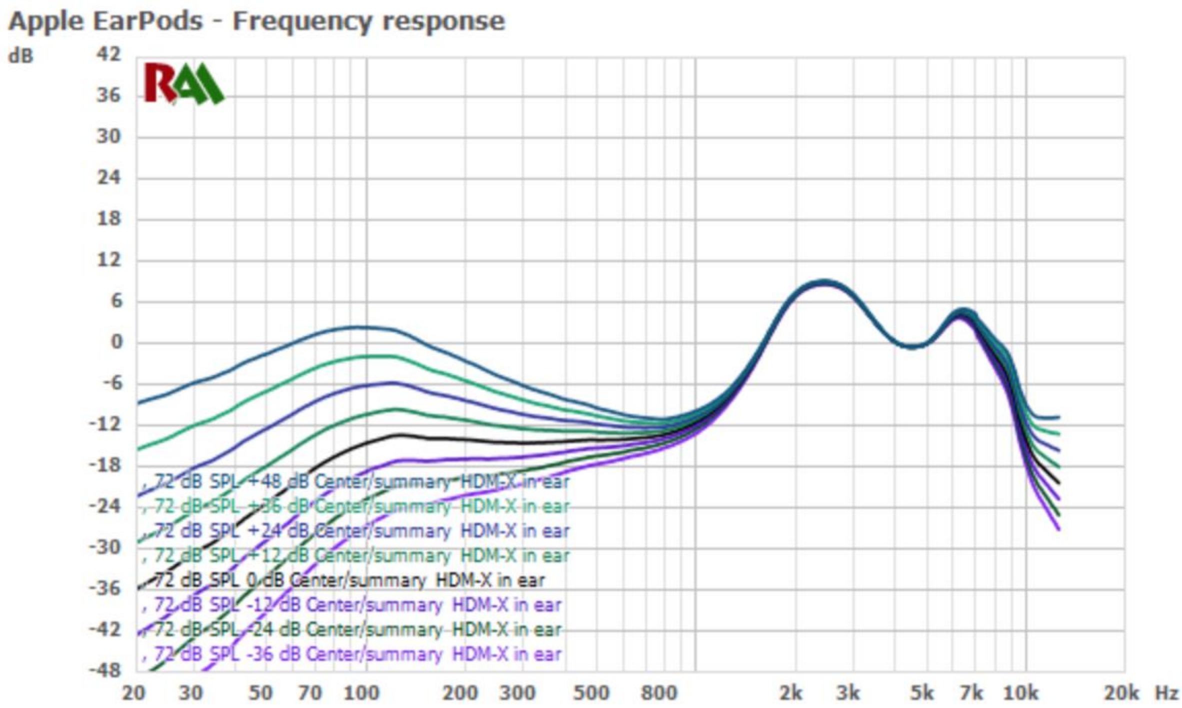


Figure 48: Apple Earpods perceived frequency response (RAA, n.dc)

5.2.7 – Conclusion: DearVR Pro Evaluation

To conclude the efficacy of this format we should revisit the aesthetic and ideological quality judgment criteria as set out in the methodology:

- *Does the production result reflect the original artistic intention?*

Yes. The staging practice reflects the original intentions and does so better than previous binaural presentations.

- *Do the sound sources localise in reflection to their intended placement in the sound field?*

Yes.

- *Do I perceive sound sources from above?*

Most definitely, yes.

- *Do I perceive sound sources from behind me?*

Most definitely, yes.

- *Are there any perceivable anomalies in the presentation? and if so,*
- *Where and when do they occur?*

Some sound sources when placed below the listener can be more difficult to discern than anywhere else in the sound field. However, none of the implemented staging concepts have so far utilised this placement as it was not offered by the original Auro system on which they were designed. This is more an observation made through exploration of the tool and the properties of the sound field generated.

- *Does the processing tool require specialist acquisition?*

The plugin costs £280 from the pluginalliance and Dearreality websites. However, a 14-day trial license is free and the tool can sometimes be found on sale for as little as £99. Therefore, the answer is no, there is no specialist acquisition required.

- *Does the processing tool require specialist knowledge to implement?*

No. The tool can be loaded onto an audio channel in your DAW as easy as you would load any other plugin. The GUI makes operation intuitive and user-friendly.

- *Did this process negatively impact the creative workflow?*

Not at all. In fact, I would say it enhanced it and expanded the possibilities of multi-dimensional staging and production practice.

Overall this tool was incredibly easy to use and integrate into my workflow. It provided many parameters of control, such as; gain, early reflection, reverb and scaling, as well as flexible output options. The binaural result was surprisingly good quality and the 360 degree panning ability made spatialising the pieces a highly creative and fun process, with immediately perceivable results.

5.3 Conclusion: Defining the Periphonic-Binaural Production Method

"What systems could be used to approach mixing for a periphonic-binaural format?' 'How greatly would the variation of headphone affect the perceived musicality and periphonic translation of the music?'

In seeking to answer these three questions I have considered and tested a variety of approaches to periphonic music making, including ambient spatial recording techniques and the spatial emulation of mono sound sources using both speaker-based and headphone-based approaches.

In these preliminary research as practice case studies I have outlined that ambient recording techniques do not present the most flexible, creative or even high-fidelity approach to a multi-dimensional spatialisation, and that the consideration for such approaches will be dependent on the sonic context and sonic concept of a particular musical work. This establishes the ‘performance music versus production music’ argument when considering source audio types and the most appropriate methods in generating the spatial sound stage.

The project opted to utilise binaural emulation rather than binaural recording because of the flexibility afforded in the production and delivery processes. Both the Auro channel-based spatial panner and the Dear VR Pro object-based binaural panner provide the ability to pan individual sound sources around the listener, including overhead, behind and, in the case of the Dear VR Pro, also below. This provides increased creative agency in the reflexive construction of ‘unnatural’[56] staging scenes and sound source placements that could otherwise not be attained through stereo, binaural or surround sound recording practice.

What we can understand from this is that binaural emulation achieved through a spatial panning plug-in (such as the Dear VR Pro) offers a more immersive and expansive experience than binaural recording or a binaurally rendered stereo or surround sound mix output. The addition of detailed elevation and rear-of-listener content adds artificial periphony to what would otherwise be a simple 2D horizontal sound field. This definition in height and surround, alongside reflexive functions for gain manipulation and make-up, depth panning, automation and reflection generation gives a more detailed, customisable and binaurally responsive result than static binaural or spatial

[56] ‘Unnatural’ in this context refers to surreal and metaphorical constructs, often physically impossible or, at the least, impractical to implement in recording scenarios.

recording techniques could. Fundamentally, this flexibility allows the user to make reflexive placement and production decisions in real-time, without commitment to captured ambience, or fixed source positions and staging arrangements. This method better benefits creative record production by affording a surrealist approach to periphonic staging practice, allowing the mix engineer the opportunity to utilise the full surround sound stage with mono and stereo sound sources in a way that traditional binaural recording and binaural master mix rendering cannot. Further to this, we can conclude that the variation in headphone types only affects the perceived binaural output to the level that they would affect any perceived stereo output – through the colouration and detail pertaining to their specific frequency and transient responses. As far as can be seen (or heard), the translation across the variation of headphone types assessed does not impact the interpretation, sonic localisation and sound source placement in such a severe way that would present as detrimental to the perceived musicality and periphonic translation of any given mix. However, it is imperative that an appropriate set of headphones be used in the initial production process, with alternatives types used as reference sets. It is the decoding process used that has the most impact on mix reproduction and translation. However, the affordance of working and monitoring within the binaural domain with alternative reference headphones means that any anomalies can be identified and avoided or resolved through actioned adaptations to practice. Further to this, using a spatial panning tool that allows for an Ambisonic output, such as the Dear VR Pro, provides the user with the means of reflexively addressing any resultant perceptual anomalies through the exploration of alternative decoding algorithms. We can conclude that as *one should work in the format one is mixing for,* the defined spatial tool for the phase 3 investigation is the Dear VR Pro method, as this offers a high-quality, flexible and affordable approach that fundamentally fulfils all the ideological and aesthetic judgements set out by this project.

6 – Phase 3 - Redefining the Spatial Stage: Non-front orientated approaches to periphonic sound staging for binaural reproduction.

6.1 - Overview

The previous chapter explored the technology through which we could approach investigating the main practice as research problem. We examined the pertinent and often problematic areas of binaural perceptual phenomena and we evaluated the importance of reflexive practice and in taking a democratic approach to spatial music making. Through this the previous chapter concluded that the DearVR Pro plugin would be the most suitable tool of choice in exploring the possible approaches for the creation of non-front orientated musical sound staging. Using the knowledge gained through the previous chapters, this chapter details the innovative creative research practice and discusses, demonstrates and evaluates the staging and production techniques realised through the phase 3 project praxis. The aim of this section is to answer the primary research question: *‘How can non-front orientated sound stages for music be approached and structured?’*

In an attempt to answer this question, and address the research problem posed through the thesis, this chapter investigates the musicology and musicality of non-front orientated record productions from a phenomenological viewpoint using practice as research strategies. The research focuses on the applied development of contemporary music production technique through creative practice, concurrently establishing a language for 3D music production techniques. The study collates and assesses periphonic sound staging approaches, outlining particular consideration when interpreting musical concept to sonic schema for a periphonic-binaural 3D audio arrangement. This chapter specifically discusses metaphor as a vehicle for enhancing the immersive musical experience and creative production practice. Hyper/surreal metaphorical interpretation is afforded through non-front orientated periphonic staging and is developed through an ecological approach to music perception, predominantly with a focus on embodiment and proxemics. Through case studies, this chapter evidences ways in which periphony may enhance musical staging beyond that which current industry record production practice affords, and explores concepts such as ‘omnimonophonic’ and ‘polyperiphonic’ vocal staging. The creative research informs the craft by defining new methods for periphonic sound staging through a non-traditional, democratic and unique approach to spatial music production.

The techniques discussed within this chapter focus on exploiting binaural perceptual phenomena as an aesthetic enhancer in music production, although it can be said that not all of the techniques defined herein are strictly reliant on a specific binaural synthesis tool (such as Dear VR Pro, used within this project) as a means of generating this perception. In many instances the same or a closely similar experience can be garnered through our natural binaural perception when listening upon a domed, tiered multi-speaker system (such as Auro-3D 13.1), thus these techniques are considered transferable - defined in this instance as being *system agnostic* and therefore not reliant upon or governed by any specific tool or playback system. There are, however, some specific techniques that are designed to employ proxemic reaction through binaural sensation (Hall, 1966), and thus are implicitly reliant on the binaural phenomena as presented over headphones in order to convey a more embodied experience, especially where the staging explores intimacy and in/externalisation.

This chapter explores the periphonic production framework (see table 5) and discusses several of the most pertinent, interesting or impacting instrumental stages synthesised through the project practice.

The periphonic framework presented in Table 5 implies a hierarchy governing approach. As the studies in this chapter explore, this is due to the top two tiers - ‘sonic context’ and ‘sonic content’ - fundamentally determining the level to which the other fields may be considered and applied in practice.

<table>
<tr><td colspan="2">Sonic Context: Performance vs Production Music</td><td>Sonic Content: Instrumentation. Spatial, Spectral and Lyrical content.</td></tr>
<tr><td colspan="3">Periphonic Staging: Non-front orientated, Focussed & Unfocussed.</td></tr>
<tr><td>Vocal Staging: Pitch-height periphonic placement. Omnimonophonia, Polyperiphonia or hybrid combination</td><td>Instrumental Staging: Pitch-height placement. Quadrants.Hybrid Mono/Stereo/Periphonic soundfield</td><td>Acoustic Effects: Non-traditional use of reverb, delays, doppler and Haas.</td></tr>
<tr><td colspan="3">Binaural Physiological Phenomena: EQ, Depth, Intensity and position as function of panning. Frequency Mapping. Externalisation.</td></tr>
<tr><td colspan="3">Proxemics and Embodiment: Sonic Cartoons and Metaphor</td></tr>
<tr><td colspan="2">Dynamic Staging: Percieved movement using static and kinetic sound-sources</td><td>Rhythm and Melodic Contour: Spectromorphologic staging approach; utilising temporal and tonal shape and movement</td></tr>
<tr><td colspan="3">Perceived space and depth: Conceptual blending, Internalisation & Externalisation, Acoustic Effects</td></tr>
</table>

Table 5: Periphonic production framework: Outlines a hierarchy of production considerations when approaching periphonic sound staging.

6.2 - Sonic Context

One of the key elements defined in this project is the influence of the 'sonic context' on the approach to practice, which establishes a delineation based on the different conceptual requirements of performance and production music. Although we touched upon this in the preceding preliminary research chapter through the exploration of the spatial recording versus spatial emulation query, this chapter continues the investigation with a further case study - which explores using mono and stereo live recordings as source audio - in a bid to further outline the importance of the contextual and conceptual requirements in informing creative practice.

In his book 'How Music Works' Byrne put forward that it is the context behind the composition, and the adaptations to new technologies, that subsequently determine the output of creativity (2012a, pgs. 15-33). In the case of this study, the same can be argued; *it is context that determines creativity*. Creative agency is reliant on not just the technology used, but also the context of the musical production being worked on; whether it be performance music or production music contexts governing a hyper-real or surrealist approach. By exploring periphonic approaches to music production we think beyond the confines of the stereo '(sound)box' and the front-respecting front-projecting triangular sound field that we are now so conditioned to, as defined by Dockwray and Moore (2010, pgs. 181-197). The proper holistic utilisation of the periphonic sound field, as governed by a non-fronted approach, presents new opportunities in creative production that could not have been achieved through traditional front-orientated approaches as employed in stereo and surround sound productions. However, the sonic context of the production plays an important role in informing the creative practice and the level to which the periphonic sound field can be utilised. Two conceptual approaches - performance versus production - narrow down the appropriate creative response through the contextual requirements relating to the retrospective sonic content and schema.

The characteristics of the two approaches can be defined as follows;

Performance music represents the capture and reproduction of a performance in a given space and time. However, Zagorski-Thomas (2015, p.76) states that in such music the boundaries between creations, performance and staging can become blurred through the creative and collaborative mediation processes. In an attempt to clarify such distinctions, Novotny (2019) states that the undertaking of performance music often involves minimal technological intervention and often does not involve virtual staging. Whereas, *production music* allows for the creative interpretation of a theoretical performance in imagined space(s), often undertaken through “technological means such as performance overdubbing, hyper-real microphone placement, midi & synthesizers, editing and processing, click tracks or prepared loops resulting in exaggerated sound and virtual soundscape”. As such, a study was conducted whereby the creative agency pertaining to each category was explored. The limitations and affordances are defined, collated and demonstrated through the following two-production case study.

Performance versus Production Music: Case Study

‘Joey Clarkson - Sort Yourself Out’ and ‘Beautiful Thing - Waiting’ were recorded pilots for a separate research project whereby unusual spaces and locations were utilised for a live audio-video performance series (see audio playlist in appendix A for examples). The performances took place at the Brentford Water and Steam Museum, a former Victorian steam-works in

London. Unlike the Mosi Conde and ACM Elektron choir recordings in the phase one investigation, the artists in this study were recorded live in differing acoustic spaces using common stereo approaches to live performance recording; a combination of direct sound and acoustic capture using close and ambient microphone techniques, and where possible, direct input source recording using DI boxes (although this could only be implemented on certain electric and electronic instruments).

The sonic content of these productions intrinsically had a space imprinted within them and so too an environment. The nature of live performance capture also presents a staging relationship within that environment. Moylan articulates the relationship between staging and performance across various texts (1992, pgs. 207-208; 2012, pgs.164-167; 2020). The ambient microphones captured the physical live stage at a fixed position along with the reverb generated by and containing the sonic properties of all the acoustic instruments and voices. Although the DI sources and close microphone techniques provided the opportunity to spatialise some of the individual instrument sources, the close microphones also captured an amount of the natural reverb, and also spill from other instruments nearby. Thereby limiting the available variation in their placements due to their relationships with one another (Moylan, 2012, pgs.164-167). In ‘Joey Clarkson – Sort Yourself Out’ the trumpet spills across most of the microphone channels, it being loud and the space particularly responsive to its energy. This presents much less autonomy over the placement of the trumpet within a virtualised stage, as the other tracks contain varying degrees of trumpet spill, with the violin presenting the most. It requires a balance between the direct sound capture, the spill and the captured ambience in order to achieve not only level balance in the mix but also the perceived position of a given instrument. For example; the trumpet could not be perceived as being high overhead if it could also be heard in the violin channel front right, unless it was much louder or arrives much sooner, as governed by the principles of the Haas effect (Mc Carthy Iatse, 2009; Everest and Pohlman, 2009, pg. 60-62). However, making the trumpet louder or arrive sooner are often not viable options when working with captures of physical live stages. Loudness increase affects our perception of sound source distance, it makes it feel closer, thus affecting perception of the intended position and the over all mix balance. Increasing the time of arrival of the trumpet, or delaying the other tracks that contained trumpet spill, would render the mix messy and the performance incoherent. Care must be taken with time delays, as the time of arrival relationship between stems can be quite inflexible due the live aspect of the content requiring the instruments, and the reverb, to be phase coherent and in time. It is truly, what it is; a performance captured in space and time and this presents staging limitations pertaining to the instrument sources, and acoustics, that one has to work with. Although it is possible that capturing the performance through a 3D mic-array[60] may have given a more coherent spatial aesthetic, there would always have been limitations pertaining to the possibilities of creative auralisation of a sound stage given the limitations in the physical positioning of instruments and the nature of ambient recordings of a performance dictating and fixing the stage position within the capture.

[60] Such like those used in the Mosi Conde and ACM Elektron choir recordings.

Conclusion:

This case study evidences that both the sonic context (live performance capture in a particular space) and the content (ambient recording, spill and acoustic imprint) limit the creative agency in approaching the periphonic staging for these pieces. Therefore, a more hyper-realistic production method was sought for these performances. A lightly enhanced spatial reproduction of the physical live stage was the best approach in presenting the most coherent possible sound stage whilst also affording some means of spatial enhancement. The production music pieces, such as: 'Hidden Behind Static – Penny Drops, Monomorphic & Far From Here' and 'Jerome Thomas – Late Nights'[61] present more flexibility and opportunity in conceptualising meta-realistic, virtual staging concepts and constructing a more creative, surrealist auralisation that fully makes use of the periphonic sound field (see audio link in appendix A for examples). This is due not only to the instrumentation and more complex textural layering featured through these studio productions (see audio link in appendix A for example), but fundamentally also due to no particular perceived environment or space dictating the staging relationship and the staging freedom that drier, more direct sound affords.

[61] These pieces and associated production approaches are discussed in more detail throughout this chapter.

6.3 - Sonic Content

As concluded in the previous section, the ‘sonic content’ also plays an important role in defining the spatial production. By utilising the lyrical, melodic and temporal content as a production narrative, the musical and lyrical content can be reinforced within the spatial production staging, and vice versa. The melodic and temporal techniques in the discussion to follow are featured and continually developed across several spatial productions and presented both through instrument and vocal staging arrangements. The lyrical production technique can be demonstrated to good effect throughout ‘Hidden Behind Static - Far From Here’ and is discussed in detail in the vocal staging case study to follow.

Research states that vertical perception is governed by spectral cues pertaining to higher frequency bands, with frequencies near 6-8 kHz being of particular importance for elevation decoding (Gibson, 2008, pgs. 24-26; Lee, 2017; Paterson, 2019, pgs. 169-170; Cheng and Wakefield, 2001, pg. 6). The pitch-height effect is a great example of spectral governance over vertical perception; whereby the higher the frequency, the higher the perceived height, relative to a single source loudspeaker (Cabrera and Tilley, 2003; Lee, 2017; Paterson, 2019, pgs. 169-170). Therefore, it seemed a logical approach to consider sound sources with higher frequency content as being particularly suited for a periphonic height placement, and therefore counter-defining sources with more lower frequency content as being best placed in the lower layers of periphony or presented in mono or stereo.

Having defined the differences in approaching a production based on the 'sonic context' and 'sonic content', there follows two further categories in informing production approach to non-front orientated periphonic sound staging.

6.4 - Staging (Re)focussed: unfocussed and focussed approaches.

These staging categories are presented as; unfocussed and focussed.

A *focussed* stage can be defined as having a main focal point situated within a given area - typically the phantom centre of stereo playback, or perhaps the median plane for surround sound with height.

Unfocussed staging can be defined as the holistic sound stage having no particular fixed focal position, even though the mini-stages that make up this holistic stage may be of a fixed-focus, or the point of focus may shift across different areas (such as above, behind, to the sides, or front).

To offer examples of each of these, we could say that a stereo or a traditional surround sound production has a typically front-*focussed* stage, whereas Hidden Behind Static's – 'Far From Here' presents a holistically *unfocussed* stage throughout (see audio link in appendix A for example), with the change in the focal area being dependent upon the spectromorphology of the composition and the lyrical narrative. Spectromorphology, as defined by Smalley, refers to the sound spectrum and its morphology – the ways in which they are shaped and change through time (1986, pgs. 61-63).

The following commentary of innovative research practice has been divided into two thematic sections addressing *Vocal Staging*, and *Perceptual Contrast, Movement, Space and Depth*. Although technically they do intertwine, this titular distinction has been made in order to discuss the vocal studies separately from the instrumental production work, as the vocal studies were defined to test and implement a fundamental theory, whereas the instrumental arrangement studies were mostly to explore the realms of aesthetic and creative possibility.

6.5 – (Re)staging the Voice: Non-front orientated approaches to periphonic vocal placement

One of the most effective areas of periphonic production practice within this project explores vocal staging and this is due to several factors. The first is that the voice is naturally easier for human beings to localise as we are physiologically attuned to it.[62] It has a familiar and recognisable sonic quality and a higher frequency range to which elevation perception can readily respond. Second, in popular music production, the voice is generally the principle point of attention in any stage, making it a particularly suitable candidate for staging experimentation (Lacasse, 2000, pgs. 9-11). Due to our familiarity with vocal placement most often being ‘upfront-centre’ in the stereo sound field, this presents an expectation, due to a lifetime of stereo listening, that the vocal should be front-centre. Therefore, the voice requires

[62] See the Literature Review chapter, section: ‘Part 2 - Lend me your ears: An overview of binaural phenomena and the embodied human perception’, for existing research and more detail on this topic.

careful consideration when placing within a non-front orientated periphonic sound stage.

A concern of a non-fronted lead vocal is that the mix would sound unbalanced, unusual or incoherent if the vocal was placed off to one side or less defined and impactful if placed to the rear of the listener. Although 'true' stereo recording and reproduction prevailed from the outset of classical-music stereo reproduction, one can often find examples in early stereo popular-music recordings where (say) the entire drum kit is in the left channel, the vocals in the right, or vice versa, and the rest of the instruments are similarly hard-panned – a two-channel-in-preference-to-stereo approach. This all-or-nothing approach to panning was a transitional phenomenon due to the technological limitations associated with the use of LCR switches in the 1960s and the creative limitations pertaining to the original four-track recording content (occasionally eight-track), which is supported by Dockwray and Moore's survey of recordings and their triangular mix taxonomy (2010, pgs. 181-197). Although this post-hoc stereo-from-mono mix approach was more mono compatible than some contemporary stereo mix approaches, the disjointed hard-panned placement present a less coherent soundstage with no interlinked stereo phantom imaging. As research, and familiarity, surrounding the 2-channel stereo system progressed, techniques were defined that presented a relationship between the two channels. This relationship better reflected our binaural means of listening (and Blumlein's intentions[63]) and subsequently real-life staging concepts were more

[63] Alan Blumlein is considered to be the 'forefather' of stereophonic recording. See Literature Review chapter, section 'Part 1 – Traditional Approaches to Record Production Sound Staging' for further information.

commonly accommodated within the schema of the stereo sound field. These staging concepts use phantom imaging techniques generated through the physical position and signal distribution of sound sources in relation to the two speakers. However, when looking at a periphonic-binaural sound field, the entire production is a “phantom image” that the user sits centrally within, as opposed to the stereo phantom image that is projected in front of the listener. This paradoxically presented creative opportunity in terms of staging practice, while also at the same time posing the problem of where to situate the lead vocal.

The vocal staging case study investigates the theory that the localisation and balance issue pertaining to non-fronted vocal staging could be rectified if the vocal cannot be localised as being in any one particular place but instead being perceived as emanating from everywhere; thus appearing as if the listener was inside the voice (although using binaural externalisation to consciously avoid the ‘in the head’ phantom-centre experience of normal stereo-headphone listening)[64]. Hidden Behind Static’s ‘Monomorphic’ seeks to address the aforementioned issue of vocal placement by implementing an ‘omnimonophonic’ aesthetic to the vocal stage. ‘Omnimonophonic’ is a term coined through this research practice to describe the phenomenon of the voice being perceived as ‘one voice from everywhere’. It can be considered as synonymous to ‘omnipresent’, but specifically relating to audio perception

[64] The phenomenon of stereo internalisation has been previously documented in the introduction section and during the headphone analysis of the O3A stereo vs binaural decoding effects in the previous chapter. It was noted that the stereo results commonly showed a more internalised reproduction across headphones than the binaural version. This juxtaposition of inter/externalisation is purposefully employed as a production aesthetic within ‘Far From Here’ which we discuss further on in this chapter.

and record production. The etymology can be broken down as; ‘omni’ meaning ‘in all places’, ‘mono’ meaning ‘one’, and ‘phonic’ meaning ‘voices or sounds’. Prior to establishing this, a previous experiment in ‘Penny Drops’ investigated an alternative approach to implementing such an omnimonophonic image. This involved positioning the lead vocal directly overhead and filling the lower layers with a cascade of vocal reverb, as a means for creating the surround[65]. Although this did not provide the desired omnimonophonic effect, it did however, provide a basis for the vocal-staging experiment that will be discussed in the following section (6.5.1), whereby pitch-height staging and the omnimonophonic vocal placement were successfully addressed.

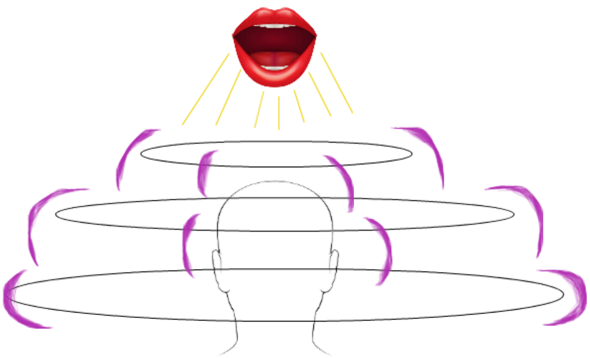


Figure 49: ‘Heavenly’ voice sonic cartoon schematic with depictions of automated vertical-panned reverb.

[65] This is referred to as the ‘Heavenly Voice’ vocal staging technique and was conducted during the preliminary practice as research in phase one using the Auro 3D system. This technique was difficult to replicate using the DearVR Pro as it was not possible to configure the reverb panning in the same way (i.e. sending to an entire isolated layer of the sound field at once. The dearVR presented it from a specific panned point-source in the field).

Figure 50: 'Heavenly voice' preliminary omnimonophony test. Vocal 1 positon using the Auro-panner.

Figure 51: 'Heavenly voice' preliminary omnimonophony test. Automated surround reverb cascade top positon using the Auro-panner.

Figure 52: 'Heavenly voice' preliminary omnimonophony test. Automated surround reverb cascade high positons using the Auro-panner.

Figure 53: ReVibe surround reverb setting configuration used for 'Heavenly Voice'.

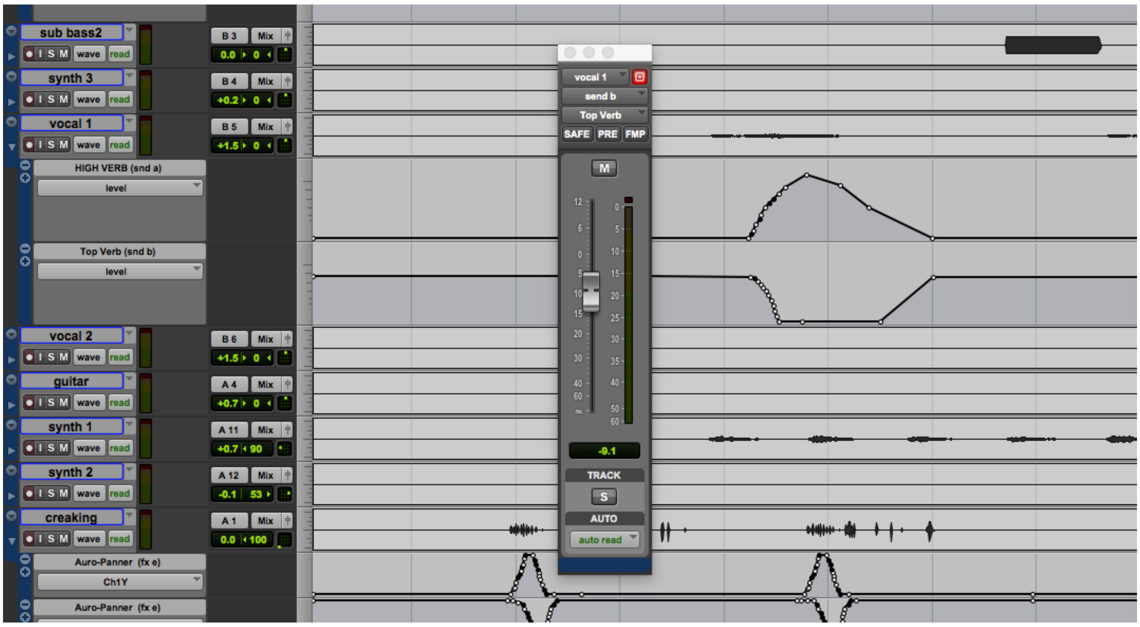


Figure 54: Rolling top > height reverb automation and send settings in Protools.

6.5.1 - Pitch-Height Vocal Staging Study – ‘Monomorphic’

Single tracked voice split into parts:	5 x vocal sound-sources
Pitch & Content - determines vertical position	Positioned in wide opposing pairs, lower pitching pair on the lower layers, a mid-pitching pair in height, highest pitched single voice top centre
Result: “Omni-monophonic” sonic cartoon based on a choral structure	Height separation can be perceived, plenty of upwards movement in melodic contour
	No specific single voice localisation, in unison they act as one voice surrounding the listener

Table 6: ‘Omnimonophonic’ Vocal Tree Staging Matrix. This table presents an overview summary of the staging decisions as outlined within the text below.

‘Monomorphic’ can be described as an electronic music production piece experimenting with vocal layering and rhythms, the salient aspects of which are summarised in the above table.

There is considerable understanding of the perceptual and cognitive boundaries that underpin multivoice composition. Individual voices can be heard singularly as melodic lines, whereas when in combination they create harmonic effect that must be balanced with melody, and although mutually intertwined, melody and harmony do not always work together in the same manner (Thompson, 2015, p.278). Such understanding of the different psychological effects melody and harmony can have on musical cognition helped to further inspire the concept underpinning the following vocal technique.

The concept behind the Monomorphic ‘Vocal Tree’ was to experiment with creating an immersive listener-centric vocal stage using an ecologic approach in the construction of a choral sonic cartoon (Clarke, 2005). The sonic context and sonic content of the vocal suggested an interpretation whereby the harmonic grouping of voices with a very similar timbre was theorised as a possible means to meld the voices into either a “single” omnimonophonic voice or ensemble. The vocal staging was defined based on phrase-matched pairs of single voices, vertically positioned to exploit pitch-height effects, and utilising the spectromorphology of the melodic contour as a further means to create upward movement and exaggerated height. A single voice recording was used to attain a user-centralised phantom image of the combined vocals in harmony, reinforcing the perception of one-voice ensemble.

The stage consisted of a single tracked vocal split into three parts over five vocal lines, delivered by the same vocalist and recorded in one single track on one take. The lower-layer voices were positioned on the listeners' shoulders, slightly to the rear and panned left-right, and voices in a height layer positioned front-back. This made the vocals feel more cohesive and the front-back height positioning was an aid to localising the voice at the rear.

This configuration of an immersive choir was based upon an ecological and schematic representation of how we would perceive the structure and auditory content of a choir in a performance space. Further, the voices were arranged by pitch - a very common approach to the physical positioning of a choir's parts – soprano, alto and so on. However, a choir comprising many singers with the same voice is not a real-world phenomenon and therefore could be thought of as a sonic cartoon – a metaphorical representation of real-world sonic structures (Zagorski-Thomas, 2018)[66].

Since all parts came from the same vocalist, they had an almost identical timbre, and this was most beneficial when the voices sang in both unison and in single layers. When singing in single-layer pairs, the wide pairing acted to define and extend the stereo image, and the similar timbre reinforced both unison and harmony between the voices, allowing the listener to perceive the layered pairings as each being one voice. The opposed pairing and group positioning in this fashion also provided vocal coverage around all areas of the listener's head and this was deemed an important consideration for an enveloping and immersive binaural

[66] See Literature Review chapter, section 'Part 2- Embodiment vs Memetics' for a theoretical background of sonic cartoonism.

experience. This technique was further supported by matching vocal register to elevation; the lowest register was positioned lowest, the middle register was positioned centrally and the highest register was positioned at the top (see figure 55). The voices were also grouped by phrasing on the XY plane in order to create a 'call and response' scenario and a wider image via the spaced pairings as indicated in figures 55 & 56.

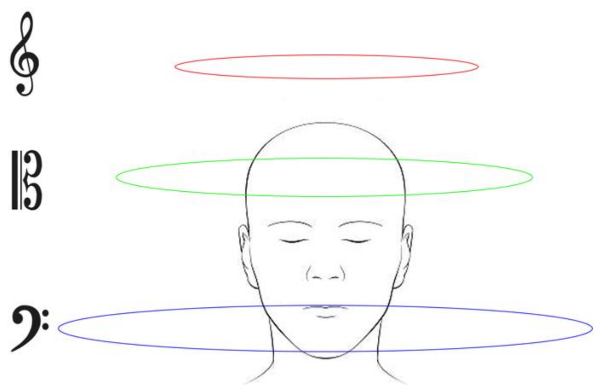


Figure 55: A basic structure representative of the vertical placement layers defined through pitch-height relationship.

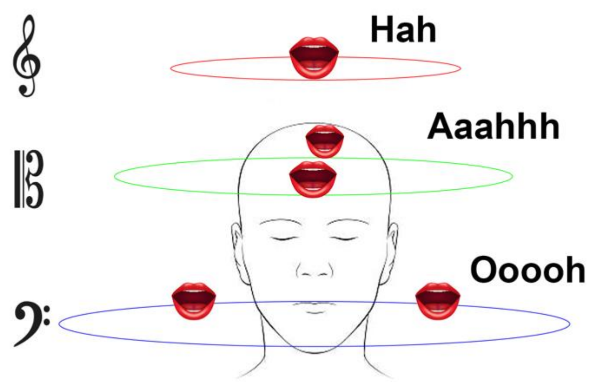


Figure 56: A basic structure representative of the Omnimonophonic vocal staging structure as presented within 'Monomorphic'. The vocal height positioning is separated and defined by phrase, and register or pitch.

As the melodic contour ascends, the voices perform from spatial bottom to top creating a further implied upward movement in the melody. This reinforces the elevation perception through both the upward movement of the melodic contour and the spectromorphology of the vocal sound stage. However, when all of the voice layers perform at once the voices and melodies meld into a unified harmonic structure, where the movement in the layers is identifiable but the individual voices and melodies are not. This technique utilises the phenomenon of inattentional deafness, whereby the harmonic vocal stream presents auditory-scene overload, and little differentiation can be made between the familiar timbres arriving simultaneously at the ears of the listener (Koreimann et al., 2014); thus creating a perceivable effect of one voice from everywhere, or omnimonophony.[67] This harmonic melding of periphonically arranged voices places the listener in the centre of a conically shaped phantom image, resulting in a fully encompassing, immersive sonic cartoon based on a metaphorical choral-meme configuration (see audio link in appendix A Sec. 02':19" for example). This staging structure exists as the first successful and coherent application of a non-fronted periphonic vocal arrangement that presented a single vocalist emanating from "everywhere". HRTF-matching permitting, there is clearly a perceivable upward vertical movement supported by the pitch-height based vocal placement and spectromorphology of the melodic contour.

[67] In the previous chapter inattentional deafness and auditory scene overload were observed to negatively impact the localisation perception of various sound sources within the periphonic sound field. As we can see through this example, these phenomena can be exploited to also positively benefit periphonic sound staging through purposeful implementation.

6.5.2 - Polyperiphonic Vocal Staging and Depth Study - ‘Far From Here’

<table>
<tr><td rowspan="10">Multi-tracked 'gang' style vocal technique</td><th>Female Voices</th><th>Male Voice</th></tr>
<tr><td>Not layered, multi-tracks periphonically placed.</td><td>Single tracked vocal</td></tr>
<tr><td>Female lead-vocal rear left-right (on shoulders)</td><td>Positioned top-front</td></tr>
<tr><td>Depth is dependent on lyrical context</td><td rowspan="2">Under-mixed</td></tr>
<tr><td>Rear vocal left-right call-and-response interaction</td></tr>
<tr><td>Contrasting intimate frontal accent</td><td rowspan="2">Lo-fi, telephone distortion effect</td></tr>
<tr><td>Mostly dry, with wet reverb rides for distance enhancement</td></tr>
<tr><td>Chorus female vocal top-centre</td><td rowspan="3">Drives and retains lyrical inteligibility.</td></tr>
<tr><td>Wet female vocal swells</td></tr>
<tr><td>Upward melodic contour</td></tr>
</table>

Table 7: Polyperiphonic Staging Matrix. This table presents an overview summary of the staging decisions as outlined within the text below.

The concept behind the track 'Far From Here' reflects mental anguish and escapism. The spatial environment was constructed to represent a 'void' or 'nothingness'. Proxemic theory[68] (Hall, 1966) was exploited via binaural synthesis to explore the possibilities pertaining to perceived depth and intimacy in the staging (Moylan, 2012, pgs.183-184). This piece also uses a combination of externalisation and internalisation phenomena - achieved through conceptual blending - combining stereo and periphonic sound fields, to imply a space existing outside of another space (see the 'Time' cue in the lyric-based production section further on).

The sociologist Edward T. Hall introduced the discipline of proxemics as being "the interrelated observations and theories of man's use of space as a specialised elaboration of culture." (Hall, 1966). Hall is responsible for the notion of so-called "personal space", or the invisible force field most people ensconce themselves in while moving through public places. A breach of implied boundaries (Hall suggested that the human ego extends about a foot and a half outside the body) is neither welcome nor tolerated (Petrusich, 2016). This notion of personal space was employed through the vocal staging constructs in *'Far From Here'* as a vehicle to enhance the depth and distanced implied by the lyrical concepts. Further, this created perceptual contrast across the differing sections of the song structure (verse, pre-chorus, chorus etc.) via the implied changes to the boundaries of the performance space relative to the centric listener position.[69]

[68] See Literature Review chapter, section 'Part 2 – Proxemics' for a theoretical background on this topic.
[69] This refers to the changes in the scale of the sound stage and source placements from small and close to expansive and far.

To extend the periphonic vocal work developed in ‘Monomorphic’, a multi-tracked lead vocal with an unfocussed (and coining the term) *polyperiphonic* stage placement is investigated to describe the attributes of the next development of the Vocal Tree. From its etymology, ‘polyperiphonic’ can be defined as meaning ‘many voices from everywhere’. The associated staging decisions are summarised in table 5, and the associated topology in figure 56.

This technique of multi-tracking the lead vocal provides the opportunity to split phrases into location based on timbral nuance, pitch/frequency content and lyrical content (see figures 57, 58 & 59). This technique utilises small changes in timbre and the performance interactions between voices to create the illusion of many voices from one voice. Typically, in stereo production the multi-tracked voices would be layered in the same position to create a thicker texture or panned left-centre-right (LCR) to enhance the width. This is commonly used in ‘gang’ style production and is often present through choruses. However, in this application the multiple voices are spaced, and present similar, but fundamentally more distinct timbres. The voices interact as one lead voice throughout the piece, and the stage focus and vocal tone changed depending on the spectromorphology of the musical constructs (verse, chorus etc.).

Lyrical production cues also informed the development of the vocal stage, and offered the opportunity to experiment with both depth and intimacy (again, see figures 57 & 58). However, the topography of the vocal stage remains focussed on utilising the pitch-height phenomena as a means for defining the tonal structure of the voices, with the exception of the spoken male voice, which was positioned high front-centre as a constant lyrical reinforcement.

The first cue point, *'Far'*, starts the female vocal in the mid-field – this vocal is made up of a double tracked pair of voices binaurally positioned directly to the left and right of the listener, in line with the ears. *'This'* then cues the automation to bring the voices into the closer mid-field, reducing the size of the perceived performance space. *'Go to anywhere'* takes the voices back to the farther mid-field and then further perceptual contrast is utilised when *'here'* draws them in very close to the listener where they can be perceived as almost being upon the listener's shoulders, singing into the ears. This feeling of intimacy is further enhanced using a multi-band compressor to apply compression to high frequencies (a familiar stereo technique) creating a sonic cartoon of closeness. This compression technique controls the dynamics of the high frequency content while allowing the lower spectrum to remain naturally dynamic, resulting in a heightened, hyper-real, breathiness to the vocal.

The next phrase begins with the lead pair of voices again in the mid-field. The cue *'slip away'* automates the voices out to the far field and *'lost'* takes the female voices far enough out that they are almost inaudible, effectively lost in the far reaches of the now expanded sound field. The position of the male voice is consistently high front centre. It is under mixed throughout (sitting low in the vocal blend) so that it acts as a supporting layer rather than a featured one. However, when the female voices are *'lost'*, the male voice then becomes more of a feature, helping to retain intelligibility of the lyrics when the lead voices are drawn away into the distance. To accentuate the feeling of the voices being lost, the reverb send was switched to pre-fader and the tail and output level increased. This reinforced the perceived *'vastness'* of the performance space, and made the voices appear as if they moved outside of their critical distance, whereby the reflections became far greater and more prominent in perception than their direct sound (White and Louie, 2005).

'Time' is an interesting cue that is unrelated to the vocal stage. It's a cue for exploring the conceptual blending of stereo and periphonic-binaural spaces in order to enhance the void/vacuum aesthetic that contextualises the production. The two bass synth pads had a similar tonality and timbre, such that when panned and layered together they created a thicker texture and a slight perception of musical movement through spectral and temporal flux. However, when experimenting with their purpose in the spatial sound stage, it was discovered that panning them in the stereo domain caused the perceived similarities and difference between the two sounds to come alive, with what appeared to be interference patterns that presented the synth pads in an alternately phasing-type manner. This created not only width but a directional and fronted sense of movement between the interactions of the two panned channels. When combined with the binauralised periphonic sound field, this action presented a very interesting spatial duality that lent a new aesthetic to the production. This combination of stereo and periphonic spaces presents a noticeable change in the character of the perceived sound field(s). The stereo sound field presents as internalised, as it tends to be on headphones, whereas the periphonic-binaural sound field externalises due to the effect of the HRTF (Xie, 2013) and the early reflections afforded by the

DearVR panner. It is this juxtaposition between the two spatial perceptions, along with the synth timbre and spectromorphology that exaggerates the sense of ‘void’ or ‘vacuum’. This adds to the perceived perceptual contrast within the staging schema, creating interest, intrigue and movement within the music, while presenting a new experience that is metaphorically representative of the sonic content and concepts.

Following this, the female voices remain in the near field but not overly close or as intimate as the *‘here’* cue had presented. They remain in this position until the cue *‘gone’* triggers their cut with the male voice again taking attention and retaining the intelligibility and meaning of the lyrical content.

To create further difference across the song structures the *‘oooooooh’* pre-chorus swells were positioned high overhead, and present movement through melodic contour and provide a larger sense of space - in contrast to the verses. The swells act to introduce the chorus, which utilise the rear lead-vocals in a close call-and-response fashion relative to the lyrical phrasing.

The lyrics *‘stuck between four walls’* cue a ping-pong call-and-response interaction between the voices, whereas *‘built by you’* follows this same pattern but *‘you’* then cues both voices to sing in unison - which doesn’t quite present the same definitive frontal phantom image we see with stereo, but the similar experience conveys a more fronted intimacy.

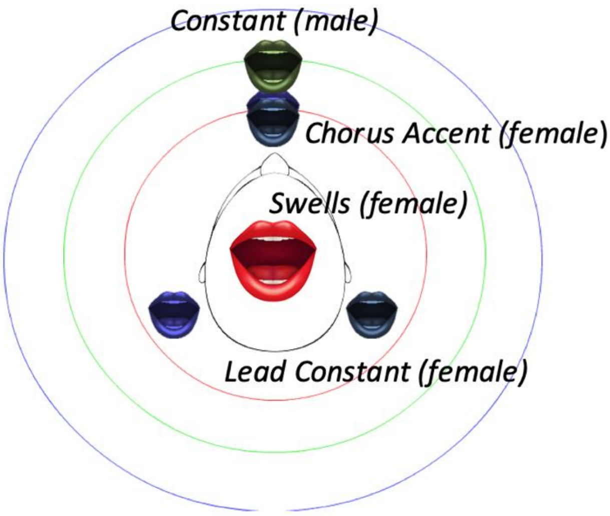


Figure 57: 'Far From Here' polyperiphonic vocal stage topography.

Far, far from ***this***,
go to ***anywhere***, than ***here.***
Dream, dreams you lost,
slip away from you.
Now you're ***lost.***

Time, time won't heal.
Give for anything, not to feel.

Everything you wanted, everything you needed.
Gone.
You're wounded, drag yourself through every day,
always numb.

Ohhhh Ohh Ahhhh Ohhhhh
Never can be found, when you're hiding in the darkness.
Stuck between four walls, built by **you.**

Ohhhh Ohh Ahhhh Ohhhhh
Ahuahuohhhh wahhh ohh ahhh ohhhh
Stuck between four walls, built by **you.**

Figure 58: Color-coded lyrical cue sheet presenting the lyrics and vocal staging cue points relative to Figure 5.

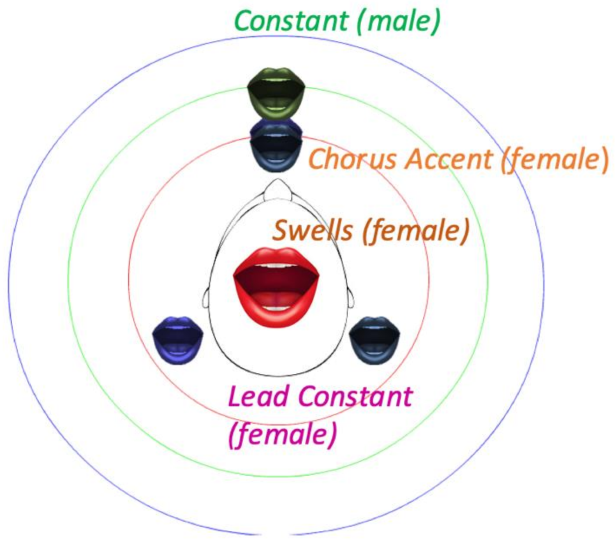


Figure 59: 'Far From Here' color-coded polyperiphonic vocal stage topography as related to the color-coded lyrical cue sheet.

6.5.3 - Variation of (re)development and (re)application

The techniques as aforementioned were (re)developed and (re)applied across the various pieces explored through the research project.

Jerome Thomas's – *'Late Nights'* (see audio link in appendix A for example) presents an amalgam of the 'omnimonophonic' and 'polyperiphonic' versions of the Vocal Tree, redeveloped and reapplied as relative to the sonic context and content of *'Late Nights'*. The combination of technique suited the vocal-work within the piece particularly well. Characteristic of the production practice associated with the work and style of Jerome Thomas, there are numerous vocal stems, each containing a specific harmony, backing vocal or lead vocal phrasing layer. In a typical stereo production, these vocals would be layered together 'gang style' to create texture and tonal harmony between the voices, and gently panned across the stereo image. In this instance, the periphonic sound field offers the opportunity to completely spread them out and make use of them as individual sound-sources in 3D. Accordingly, the backing vocals and vocal harmonies in this production were periphonically organised around the listener and arranged within a way that presents balance between the quadrants of the periphonic sound field. The lead vocals were arranged in a paired call-and-response fashion in the frontal left-right domain. One backing vocal was automated to sweep rear left-right on the backing-vocal-repeat of the lyrical cue *“Am I too late on arrival?”* (Sec. 0’:47”- 0’:53”) as a means of creating a movement-themed sonic cartoon influenced by the movement implied within the lyrical content (see audio link for example). The backing vocal placements work particularly well in the rear, as they are not a prominent feature in the mix and therefore the filtering and lack of presence is not a concern. The way they respond individually aids to define clarity to their placement, create movement through the vocal interactions and reinforce the immersion generated through their staging. This technique was an adaptation of the polyperiphonic technique, whereby the same voice was used throughout the recording with different timbres and phrases in order to construct the ‘many voices from everywhere’ aesthetic. Although there are points in this piece where all voices sing together, a different result is achieved from that in ‘Monomorphic’ – although the effect still lends itself to presenting an amount of perceived ‘omnimonophony’ through an immersive and enveloping vocal harmonisation.

The vocal staging work was also memetically redeveloped and reapplied to the two live pieces; Joey Clarkson – ‘Sort Yourself Out’ and Beautiful Thing –

'Waiting'. However, as previously discussed, creative periphonic staging approaches were much more difficult to implement due to the acoustic imprint present on the sonic content. 'Sort Yourself Out' was especially problematic and although the sound field was designed to replicate a periphonic version of the physical live sound stage, there was some experimentation with periphonic vocal production approaches. This was attempted through duplicating and splicing the pre-chorus lead vocal by phrase and panning the segments rear-left and rear-right in a call-and-response fashion.

The polyperiphonic technique was further adapted in the vocal staging work when redeveloped and reapplied to 'Waiting'. However, rather than present many of the same voice being perceived as one-voice from everywhere, it presents many voices being perceived from everywhere. This approach to vocal periphony was applied by using the backing vocals as the non-frontal enhancement, while the lead vocal remained low front centre to the listener. This technique was implemented to exploit the ascending melodic contour of the lead vocal in the break down, which was used as a cue point to pan the voice up to high front-centre in an approximate 45° angle upward relative to the listener's forehead. The automated panning was applied to reflect the *vocal lift* implied by the melodic contour and phrasing. This provided movement and used the melodic contour as a perceptual reinforcement for the vocal elevation.

Although the staging practice discussed throughout this chapter can be varied, redeveloped and reapplied in a manner dependent on the sonic content of a given production, the sound-stage boundaries remain consistent throughout several variations of concept. It is interesting that upon reflective analysis, it could be observed that periphonic sound staging often presents (though is not limited to) a three-dimensional cone shape around the listener, defining these as boundaries of the perceived performance environment (see figure 60). This could be considered an upright, multidimensional variation on the *triangular* and *diagonal* mixing taxonomies originally proposed by Dockwray and Moore (2010, pgs. 185-187), albeit with a new and contemporary framework defining the sound-source organisation therein.

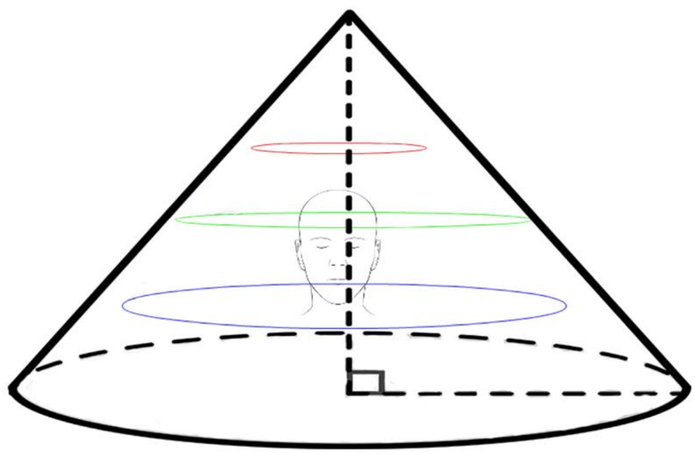


Figure 60: Periphonic Cone Staging Taxonomy.

6.5.4 - Conclusions

The omnimonophonic staging technique helps to resolve the placement and perceptual issues pertaining to non-front orientated periphonic vocal staging without compromising balance or the listener's experience. The polyperiphonic vocal staging technique provides the opportunity to creatively utilise multi-tracked voices in ways that not only provide textural and spatial enhancement, but that fundamentally also support and counter the loss in definition presented via farther depth placements. This offers the opportunity to explore the possibilities of an expanded performance area without compromising on intelligibility. Both approaches can be applied to create perceptual contrast across the different sections of a song's construction. Further, these techniques can be utilised to reinforce the conceptual, musical and lyrical narratives through implied meaning attained via metaphorical schema.

These techniques present suggestive solution to the issues associated with a non-front oriented vocal placement. They provide a unique, adaptable and contemporary approach to staging the voice, offering an enhancement to the musical production that cannot be achieved through traditional stereo or surround sound practice.

6.6 - Perceptual *Contrast, Movement, Space and Depth*: Instrumental staging and abstract spatial effects

6.6.1 - Movement enhancement through rhythmic cognition and embodiment

In an attempt to address the research question *'How can non-front orientated sound stages for music be approached and structured?'* kinetic and static approaches to dynamic staging were explored using rhythmic embodiment as a vehicle for movement enhancement. The dynamic stages in discussion relate to the *follower* and *chaser* techniques featuring in 'Penny Drops'. These techniques were partly outlined in the previous chapter as they were one of the binaural localisation test structures developed for the phase 2 study.

During synthesis of the techniques it was hypothesized that the perceived movement of the percussion could be enhanced via the metaphorical suggestion of implied boundary blocking their movements. Although the notion of creating movement by suggestively preventing it sounds counter-intuitive, the idea arose from an understanding that human beings are more sensitive to changes in their auditory perceptual stimuli than constants[70]. This can be understood through the mechanisms of auditory scene analysis (Bregman, 1990).

[70] I empirically evidence this in my lectures by playing a recording of an air conditioner whirring in the background as the students enter the classroom and take their seats. I then mute the recording at the point in my lecture where I begin discussing this sensitivity to change in perceptual input. It is only then the students become aware of the noise that has been whirring in the background the entire time – they only become aware of it once it has stopped.

Bregman states that the goal of auditory scene analysis is “the recovery of separate descriptions of each separate thing in the environment” (1990, pg. 9). Bregman notes that these separate things may be combined into a grouping of what he refers to as an ‘auditory stream’. To offer example of this he refers to a series of footsteps forming a single event, despite the fact that each footstep is a single sound[71] (Bregman, 1990, pg. 10). The auditory stream serves a purpose of clustering related qualities, grouping them as ‘belonging to an auditory object’ – the mechanism underpinning the function of the omnimonophonic vocal tree in the preceding section. Changes to these stimuli act as informative markers in the stream of auditory data, providing the ear with a more detailed picture of a given auditory scene. This can be exampled by referring back to the footsteps analogy.

If within the stream of footsteps the sound of crushed glass was heard at the same point in the grouping that a footstep simultaneously occurred, one could perceive these two sounds as being part of the same ‘happening’, consequently relaying the perception that the person walking had trod on and crushed the glass. However, if the sound of crushed glass was not heard simultaneously to a footstep or with the same rhythmic interval pattern of the footsteps, this auditory grouping would not occur in the same manner and the crushed glass may be perceived as being an event entirely distinct from the footsteps. There are other aspects that may effect these grouping associations, such as timbral qualities. To offer example, if the sound of the

[71] This example is perfectly analogous of the form of the dynamic percussive follower and chaser structures; which are a series of short percussive sounds grouped together through their related sonic qualities. It is this grouping that both suggests and reinforces the movement generated by the panning and static placements.

broken glass was more of a ‘smash’ than a ‘crush’, there is a possibility that, regardless of the coincidence in occurrence with the footsteps, it would be perceived as a separate event. This is due to sonic embodiment implying a pre-existing association with the sound of a glass ‘smashing’ – we associate this with a fall from height, rather than the ‘crush’ of being underfoot.

Employing this understanding via a judicious consideration of timing and timbre, it seemed plausible that a sonified blockade could be implied within the auditory stream of percussive sources, consequently enhancing the percept of motion by altering it. This idea was explored via the implementation of percussive *‘stops’* or *‘blockages’* that were suggestive of sonified *‘impacts’,* and a differing percussive sound source to imply a ‘deflection’ (see audio link in appendix A for example[72]).

We could think of this structure as being analogous to the motion observed during a table tennis match. The ping-pong ball is struck by a paddle, it flies across the table past the view of the observer (listener) and is then hit and deflected by the opponent’s paddle, which changes the ball’s direction. In this we can see that there are three-parts that could make up a sonified representation of this schema; the agonist (the first hit), the action of motion (flight) and the antagonist (the responsive second hit).

These three elements were applied as a rhythmic structure in ‘Penny Drops’ using a combination of kinetic and static sound source placements, and an understanding of the rhythmic cognition pertaining to the grouping of the percussive sources. As previously discussed in chapter 5, the *flight* staging

---

[72] This technique is demonstrated throughout the track but can be readily auditioned within the first 16 seconds.

elements were created using kinetic sources; applying panning automation to the series of rhythmic samples in the ‘creaking’ percussion track, generating the action of left-to-right motion in a semi-circle trajectory around the front of the listener (*follower*). This was also applied to the content of ‘creaking 2’ but in reverse (right-to-left) and to the rear of the listener (*chaser*).

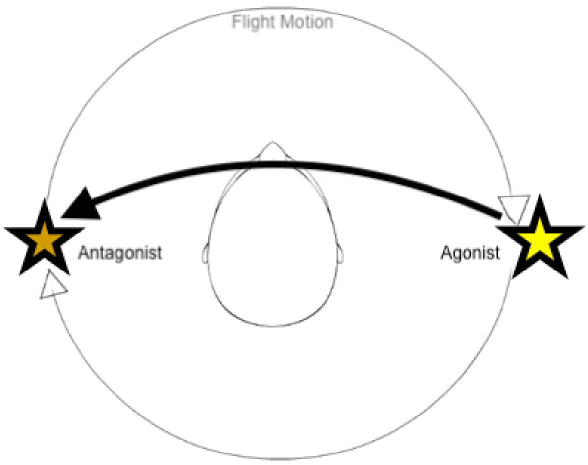


Figure 61: A schematic representation of the agonist and antagonist percussive staging.

A static percussive sound source was then strategically placed following the end of the two-part semi-circle sequence, creating the agonist, and another placed in the opposite sound field quadrant to this, creating the antagonist source. The agonist sample (‘dink’) implies the blockade and a change of direction to the motion of the kinetic percussive source and the antagonist sample (‘donk’) - statically panned to the opposite side of the sound field - implies the deflection (or rather, the landing after the deflection occurrence) (see figure 61).

Although the static sources were not of the same timbral quality as the percussive samples in motion[73], they were presented in a temporal and timbral manner which presented them as related to one another and to the kinetic percussive group. This was achieved via the consideration of short-spaced *'inter-onset intervals' (IOI)*[74] between the kinetic and static sound sources' perceived occurrence (Klein and Posner, 2019, Sec 04:00; Thompson, 2015, pgs. 96, 279). It is the timing of the IOI that aids to group these percussive elements to the related cluster of rhythmic sounds.

The timbral change of the antagonist sound presents a perceived drop in pitch, suggesting the phenomena of a Doppler effect[75], such that (in comparison to the higher pitch of the 'dink' sound) reinforces the percept of the sound having deflected off the boundary and flown quickly past the listener. Following these structures are a series of shorter and faster rhythmic samples that further enhance the ping-pong effect through a high velocity left-to-right to-and-fro action.

This utilisation of IOI to relate the rhythmic groupings, alongside the change in the timbral qualities of the sound sources, presents a spectromorphology that suggests implied motion through the metaphorical obstruction and deflection of a sound source, offering an enhancement to the movement through a ping-pong sonic cartoon.

---

[73] The elements in motion sound more 'clicky' whereas the static agonist/antagonist sound more like a 'dink' and 'donk', respectively

[74] IOI refers to the time between the onset of rhythmic elements that constructs the rhythmic pattern, rather than their tempo.

[75] The change in frequency of a sound wave in relation to an observer who is moving relative to the sound source. The Doppler phenomena presents a perceived upward shift in frequency for observers towards whom the source is approaching and a perceived downward shift in frequency for observers from whom the source is receding (The Physics Classroom, 1996).

6.6.2 - Abstract Spatial Effects: Creative reverb for movement and acoustic shapes

This section investigates the use of artificial reverb as an instrument in the creation of movement and spatial-shape projection, rather than a tool for emulating the spatial properties of an acoustic environment - as is typical with stereo and surround sound production. This technique can be demonstrated in Hidden Behind Static's - 'Monomorphic' and explores the construction of synthesised acoustic shapes within the sonic environment, offering depth enhancement via metaphorical spatial structures and an implied percept of movement.

This idea was synthesized through the hypothesis that if the panning tool can be used to construct a perceived performance environment via the periphonic placement of instrumental sound sources, then a related acoustic shape or environment may also be constructed via the placement of artificial reverb in relation to the instrumental sources.

As with the follower-chaser technique aforementioned, this approach explores IOI temporal intervals and sonic embodiment to group elements sharing similar qualities together. This was implemented in practice using the kick and snare drum sources which were placed in a wide opposing pair slightly rear-left and rear-right of the listener, respectively. Two auxiliary channels were configured, each containing a spatial panner which then fed the signal into a reverb plugin with a 100% wet output[76]. The original dry kick

[76] When using the DearVR Pro it is important to ensure the pan > reverb processing order, otherwise the reverb characteristic is lost through the HRTF processing. However, on the Auro 3D system the same rule does not apply, as the reverb auxiliary output can be panned

and snare signals were sent to the relative reverb auxiliary channels, of which the wet outputs were panned in an opposing pair, front-right and front-left, respectively.

The pre-delay/start position function on the reverb plug-in was switched in and manipulated to determine the temporal interval between the dry sound and the wet. This generated a perceived ‘echo’ reflection from the wet sources, exploiting the principles of the HAAS or precedence effect, through an IOI correlation relative to the dry signal occurrences. Care should be taken not to extend the pre-delay past the point affording temporal cohesion as this can lead to a decorrelation of the wet and dry sound sources, thus presenting them as being unrelated instances in the auditory stream. Although, this can be a welcomed aesthetic if the intention is to use the reverb as an instrument source in its own right. A pre-delay, or a time delay, may be used to create rhythmic patterns and repetitive sequences which can then be creatively panned around the periphonic sound field and mixed at different levels, offering a similar approach in constructing acoustic shapes but without the need for artificial reverb[77].

directly or discreetly allocated to any given channel and as there are no HTRFs involved in this process.

[77] This variation of technique is discussed in the sub-section of this chapter to follow; ‘*Acoustic Emulation through Temporal Delay and Level Difference’.*

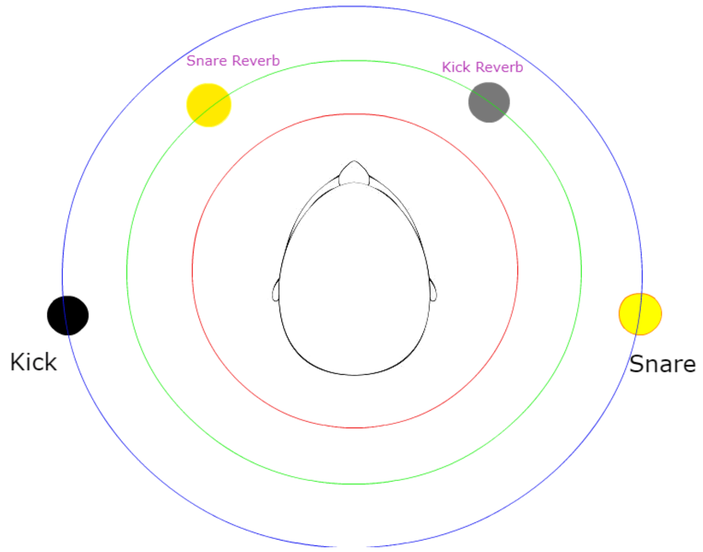
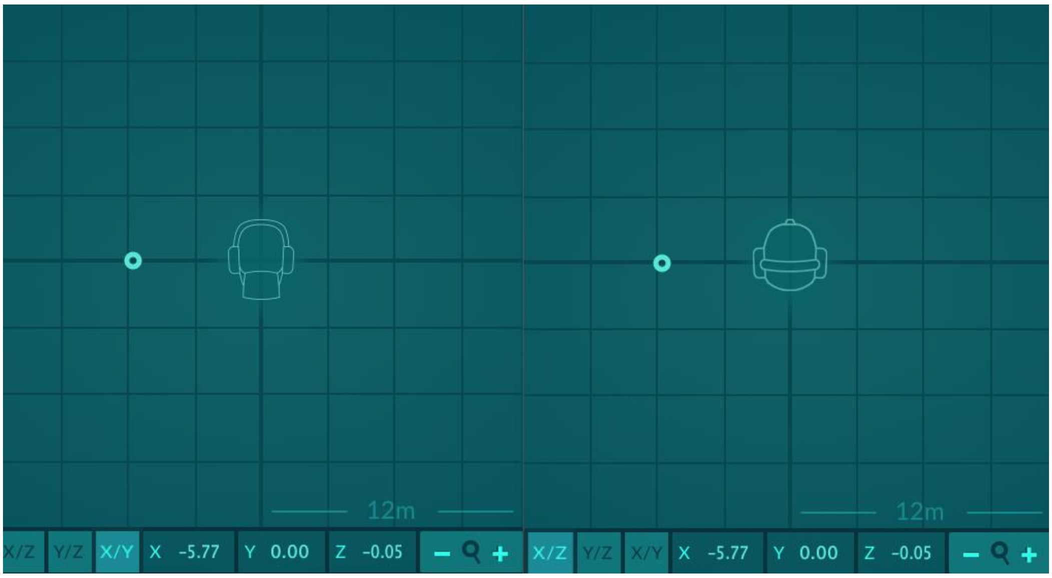
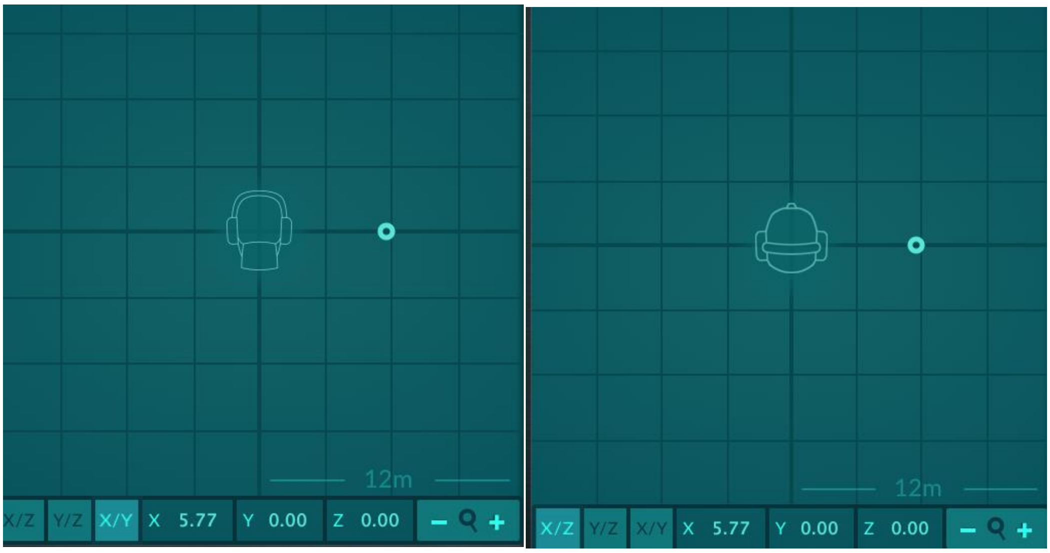


Figure 62: Location of wet and dry kick and snare

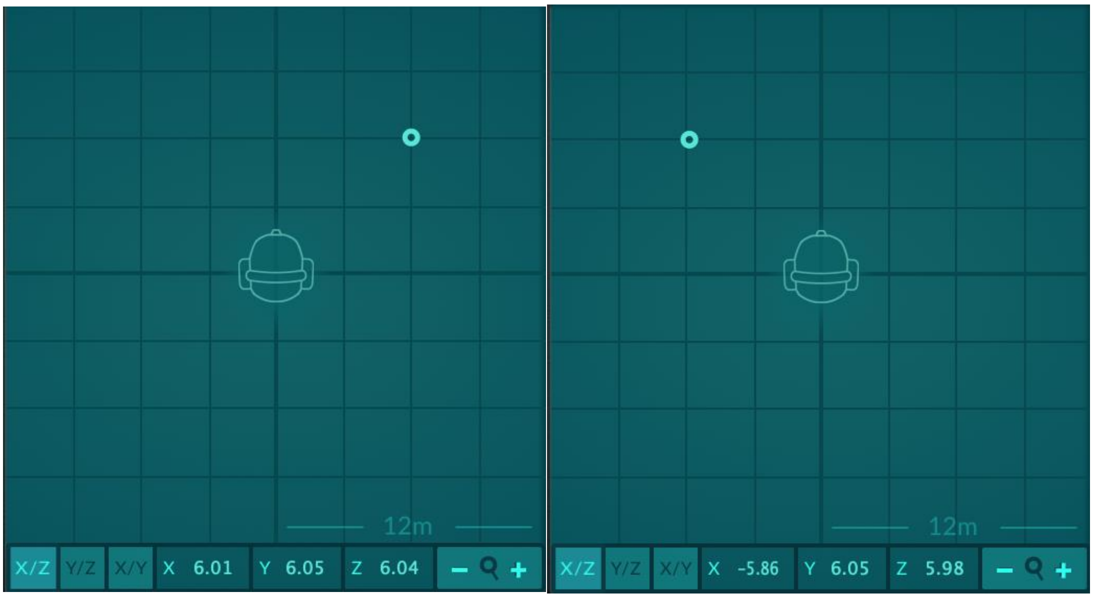
Initially, this technique was constructed on the Auro system which allowed the outputs of the reverb auxiliaries to be assigned directly to the stereo-configured front left and right channels of the speaker-array without the need for a panner. This is what created the perceptual off-set between the wide dry placements in the rear and the narrower wet placements in the front (see figure 62). However, there are no multi-channels to discreet output to when applying this process using the Dear VR Pro. Therefore, a close approximation of the same placements was made using the spatial panner. Panning the wet sources left and right in the stereo domain was explored - ranges between 20-30% either side were implemented – and although this gave a narrower image to the wet frontal sources, it removed much of the externalised directionality associated with the binaural HRTF filtering. Unfortunately, this did not result in exactly the same presentation as reproduced on the Auro 3D x O3A system and a slight adjustment to the dry source positions was implemented to achieve a similar effect using the Dear VR Pro (see figures 63 - 68).



Figures 63 & 64: Dry kick drum sound source placement in Dear VR in X/Y and X/Z views.



Figures 65 & 66: Dry snare drum sound source placement in Dear VR in X/Y and X/Z views.



Figures 67 & 68: Wet kick drum and wet snare drum sound source placement in Dear VR.

As apparent in both applications, the wet signals act as metaphorical “reflections” of the kick and snare drum sources. The opposed placement and pre-delay of the wet-dry sources adds to further enhance the reflection metaphor, suggesting a ‘slap-back’ reflection from a boundary, such as a surface or a wall. The ‘slap-back’ sonic cartoon provides an implied sense of movement across the sound field and adds depth to the frontal space. Further, the narrowing of the frontal reproduction, as presented across the Auro 3D system and O3A versions, presents a deeper triangular shape to the construct which enhances the perceived depth of the wet source placements, offering a ‘tunnel-like’ trapezoid shape. This shape reinforces both the ‘echo’ metaphor percept and further enhances the perception of depth to the frontal area of the sound stage.

6.6.3 - Abstract Spatial Effects: Acoustic emulation through temporal delay and level difference – 'Late Nights'

The technique previously discussed presents as the precursive experiment to that explored within this section. Although similar in principle and theory, this approach utilises time delay and level differences to emulate acoustic phenomena, creating perceived movement through implied reflection, or periphonic delays, without the need for artificial reverb or a spatial studio delay plugin.

This technique was implemented by firstly defining the number of different positons for repetition/reflection, these points account the number of delayed signals required (single or multiple). Following this, an auxiliary channel for each point was created and configured each with a sample/time delay and a Dear VR plugin. The original audio content was then bussed to each auxiliary via multiple sends and the appropriate time delay was implemented across each auxiliary channel. The relative IOI between the delays of each channel determines the rhythmic pattern structure of the emulated reflections (see figures 69 & 70).

Following this, the delayed sources were positioned within the periphonic sound field and the fader levels of each auxiliary output were set to progressively decrease (see figure 71). This was implemented to generate an illusion of diffusion, offering a metaphorical critical linking between each of the delayed source channels. The relative level control of each delayed source determines the critical distance as perceived between the emulated reflections and the original sound source, inferring an acoustic space and providing a manipulatable depth percept.

Figure 69: Reaper mix window showing the auxiliary channels, inserts and sends.

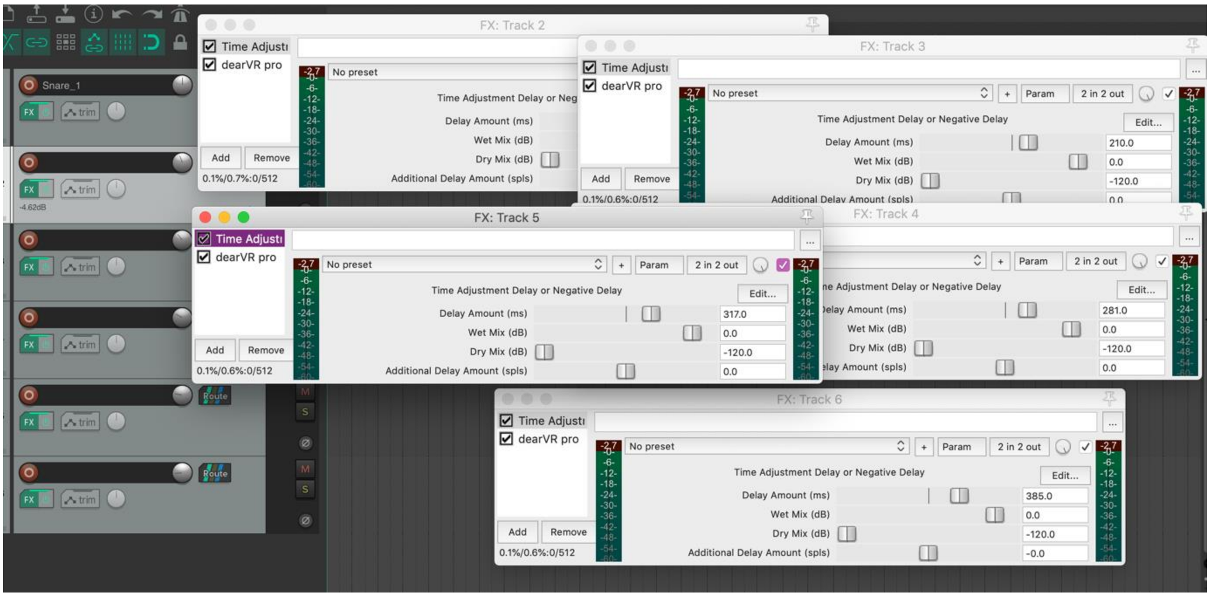


Figure 70: The different delay times implemented to create IOI linking between the delay sources.

If all levels are equal and the IOI is small enough, the result is an interesting sense of movement created through the panned and delayed source repetition, suggestive of a *'smear'* or *'transient blur'*. Whereas, if the amplitude levels exponentially decrease relative to the timeline progression of delays, the result is perceived similar to that of a studio echo delay, or as the dissipation of early reflections. This effect can be further enhanced by switching on the reverb emulation within the Dear VR plugin and setting the level gain and damping relative to the decrease in amplitude level of each delay source – the lower the amplitude level of the source, more damping and less gain the reverb should have (see figure 72).

Figure 71: Reaper edit window showing the decreases in mix level (amplitude) of each delayed source.

Figure 72: The decreases in reverb amplitude and damping parameters for each delayed source.

It is interesting to note that the whole structural percept of the repetition pattern can change by simply altering the time delay or the volume of one particular channel (see figure 73). This happens because of a change in how the grouping correlations between sources are perceived. When the IOI is reduced or increased this changes the temporal interval between the onset of each source, which alters the perception of the grouping pattern. The HAAS effect is exploited when a level difference is implemented, presenting the louder sound as being perceived to appear first. Consequently, this changes the start-point of the perceived grouping structure which alters the rhythmic perception of the grouping pattern. This can be utilised to produce creative rhythmic and spatial enhancement effects, especially when using a source with a fast transient, such as a snare drum.

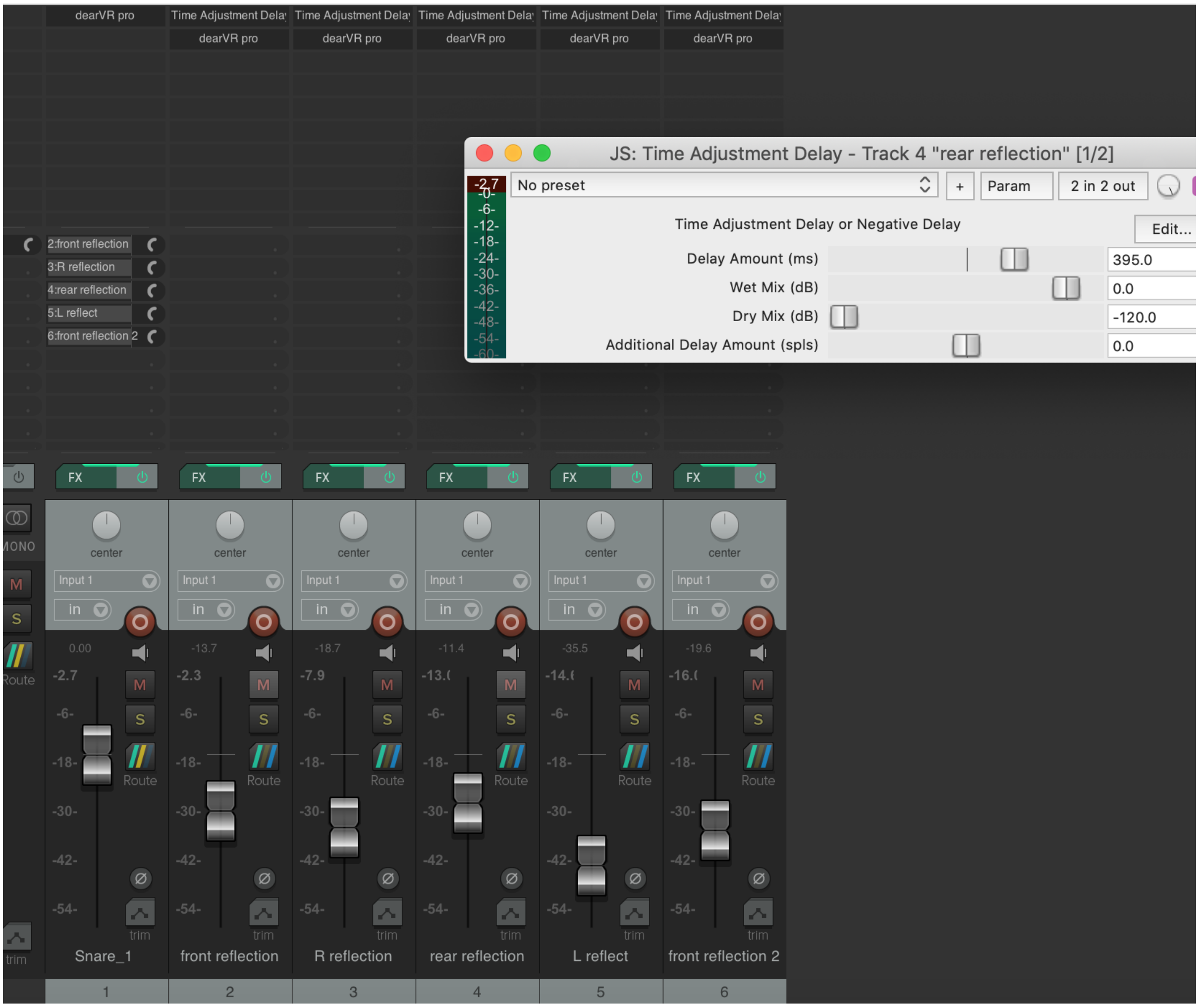


Figure 73: Reaper mix window showing irregular adjustment to mix level and delay time of a single source.

Alternative adaptations to method: Although using a single send to multiple auxiliaries is easier and provides quicker reflexion in practice, the delay effect may also be generated through duplicates of the original audio track; duplicate the track and content as many times as required and position the stems/samples in the edit timeline, as necessary, to generate the delay times required, the Dear VR plugin can then be directly inserted onto each of the duplicate channels and the process can continue proceeding from the point of source panning as in the aux-method describe above.

As applied to ‘Late Nights’, this technique demonstrates implied reflection using the snare and one auxiliary to create a short ‘one-shot’ slap-back reflection (see audio link in Appendix A for example). However, through the discussion above the technique may also result in implied movement and acoustic space, dependent on how it is configured and the repeated number of delayed sources. It may also be applied as a layer to enhance the musical instrumentation via the generation of rhythmic sequence.

Although developed via experimentation with the ‘Late Nights’ content, this technique was not necessarily appropriate for the musicality of the production. As such, two versions were created; one with an applied example of the technique and the other without it. This was undertaken to demonstrate what can be done with the technique, while providing an alternate mix that employs a more considered approach relative to the stylistic requirements of the sonic context and context[78].

[78] To aid demonstration of the approach, the version containing the ‘one-shot’ slap-back technique applied to the snare has been included within project playlist.

6.7 - General considerations for a balanced and defined approach in practice.

When using the Dear VR Pro to spatialise a musical production the following key considerations were outlined and provide a suggestive approach to practice.

As an aid to the construction of the mini-stages that comprise the large-scale schema of a periphonic-binaural production, it can be considered good practice to think of the spatial sound field as a hemisphere or a sphere surrounding the listeners head and upper torso. This hemisphere (or sphere) can then be divided into quadrants to help determine placement, interaction and balance. It is an important consideration to retain balance across quadrants in all directions. Imbalances are more noticeable between left-right and front-back relationships, than between the up-down quadrant relationships. However, this does not necessarily mean that similar sources must always features across opposing quadrants at once to retain balance. The balance can also be resolved through movement of sound sources through the quadrants. For example, a ‘quad-guitar’ technique has been implemented in ‘Monomorphic’ via splitting the phrasing of a guitar into parts and placing one phrasing splice per hemispherical quadrant from front left clockwise to rear left (L-C-R-Rr-Rl). The movement of the guitar phrasing between each of the quadrants helps to resolve any initial unbalance perceived from having only one sound present in any one given quadrant – it also helps to reinforce the balance resolution if the timing of the phrasing and the melodic contour resolves simultaneous to the quadrant movements. Opposing pairs also work well at retaining balance across the sound field, as can be exampled through the previously discussed vocal staging approaches of ‘Monomorphic’, ‘Far From Here’ and ‘Late Nights’. This paired approach can also be applied to instrumental sources, as evidenced through the paired opposing rear left-rear right-centre orchestral stabs featuring in ‘Monomorphic’.

The scaling system of the Dear VR Pro can range from 3m to 12m; at 12m the distance and lack of definition presented by the plug-in is vast. Although in some instances I have used a 6m and 12m scaling, I have found this distance between the end of the ‘panner boundary’ and the listener to be far too large for any meaningful musical definition. Therefore, scaling down to work within an implied 3m distance from the central-listener presents the audio sources with a workable definition and the listener with an appropriate association of depth relative to that perceived in a music record.

It is also advisable that upon first loading the plug-in that the scaling and output delivery format is configured and the early reflection and acoustic environment (reverb) generator functions are switched out until the staging configurations have been defined and acoustic effects are deemed necessary to implement.

6.8 - Applying the ‘soundbox’ framework: The four-dimensions of a virtual acoustic space as applied to a three-dimensional sound field.

In the literature review we referred to the *soundbox* article and discussed the four-dimensions required in the construction of a virtual acoustic space and the mixing taxonomies defined by Dockwray and Moore (2010). Here we posed the question of whether the soundbox framework could be expanded upon or redesigned within the context of this project, suggesting the mixing taxonomies and the four dimensions of placement and localisation as an adaptable starting point for the practical investigations.

In this section we focus the discussion on how these elements have been found to apply to the associated practical body of work and outline pertinent new findings and adaptations to practice to which they can be associated.

Dockwray and Moore (2010, pgs. 182-183) state that in the stereo domain “the use of a four-dimensional heuristic model consisting of the dimensions; laterality (width), register (height), prominence (depth) and temporal continuity (time)” is applied to construct and control the aesthetic mechanisms and functions pertaining to the stereo image and may be implemented using several production processes including; panning and width processing, frequency content manipulations, volume level adjustments, and time-based effects such as reverb, delay and HAAS effect – to name a few. The four-dimensions for placement and localisation, and the audio processing functions that are used to construct them, are comprehensively discussed across various texts (Moylan, 1992 and 2012; Zak, 2001; Dockwray and Moore, 2010; Frith and Zagorski-Thomas, 2012;

McLaughlin, 2014; Zagorski-Thomas, 2014; Gibson, 2018). They are also reiterated in the table taken from the literature review, as seen below (see table 8).

In the stereo domain these processes are important in constructing a perceived multi-dimensional aesthetic of the frontal illusory performance image. However, in a periphonic-binaural domain where the sound field and phantom image encompasses the listener, the approach to generating the illusory three-dimensions of the perceived performance are no longer solely attributed to distinct processes but are functions fundamentally entwined via the mechanisms of the periphonic panning process.

Lateral placement *– affecting perceived width*	***Foreground and background placement*** *- affecting perceived depth*	***Height*** *- affecting perceived verticality*	***Time*** *- affecting perceived spatiality, width and depth.*
L-C-R Panning	Volume	Spectral content of source	Time / Sample Delay
Stereophonic source recording techniques	Distortion	Equalisation adjustment	HAAS Effect (Volume/Time Delay & Panning)
M/S and Width Processing	Equalisation adjustment		Spatial Effects: True and artificial Reverberation and Echo Delay

Table 8 – The four dimensions of stereo imaging as defined by Dockwray and Moore (2010), and basic processes of affect.

Panning in the periphonic-binaural domain will affect perceived laterality (width), prominence (depth) and temporal continuity (time), as well as offering perceived elevation (height). With periphonic-binaural placement it is now not always necessary to use specific production processes in the construction of these four dimensions. This is due to the periphonic panning tool implementing these attributes via the placement of sonic sources within the spatial sound field as relative to the perception of the central listening position.

To create the feeling of depth; a source can be pulled away from the listener and the volume and frequency content will change in respect to the perceived placement manipulation. To create elevation; a source can be panned in the height layer or overhead, and the frequency content will be altered through HRTF filtering and volume level manipulation in a way that suggests the illusion of an elevated sound source in a given location relative to the central position of the listener. As such, EQ is therefore also a function of the panning process due to the direction dependent frequency changes. Further, this requires a negotiation toward the considered placement of sources relative to retaining their required timbral qualities, as further discussed below.

Although panning now presents an entwined affordance for imparting multi-dimensionality, the traditional stereo production processes associated with image manipulation (as outlined in table 8) also retain agency in generating the aesthetics of the four-dimensions, presenting a symbiotic and interlinked relationship between imaging processes, aesthetic characteristics and panning. This requires an amount of judicious attention to detail, consideration and negotiation in practice, as the periphonic-binaural framework provides multiple ways for controlling the dimensions of the sonic image and the aesthetics of the sound sources.

Care must be exercised when manipulating depth perception via panning that level differences (amplitude changes in the mix process, for example) do not then detract from the position and depth aesthetic achieved via source panning. Similar considerations need to be taken in approach to spectral content manipulations, as distortion and EQ also effect the perceived prominence attained through the panning process.

For example, it does not make sense to place a vocal behind the listener whereby the HRTF filtering will decrease the presence of the voice, and to then use EQ to reinstate the presence lost. This will result in a skewed perception of the rear-voice positioning, ultimately presenting poor binaural localisation and incoherent spectral quality. Further, the direction dependent frequency change pertaining to periphonic-binaural panning may be used '*as EQ'* to remove or enhance spectral properties of a sound source via *position-based frequency mapping*. For example, if the attack characteristic of a snare drum timbre is *'too snappy'*, rather than placing the source in a fronted position and using EQ to cut the offending 1-3kHz, the snare can be placed in a position relative to the user whereby the spectral adjustment is made as a function of the HRTF filtering (e.g. above the listener's head). This phenomenon was first observed when considering the snare staging for Hidden Behind Static's 'Penny Drops' using the Auro 3D system. The same phenomenon was presented when the technique was reproduced *direct-to-binaural* using the Dear VR Pro plugin.

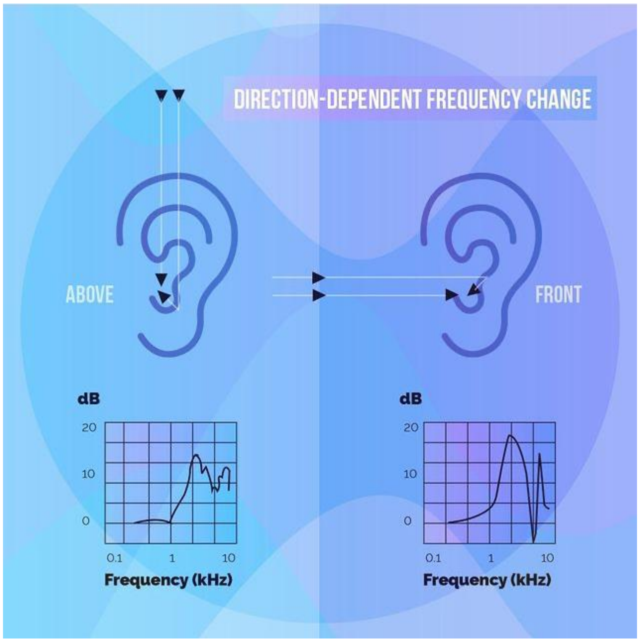


Figure 74: Direction dependent frequency change (Waves Audio, 2021)

However, pre-panning spectral manipulations achieved through a gentle application of harmonic distortion may be used to enhance the high-frequency content of a sound source (pre-pan), consequently offering an aid to localisation (post-pan).

This was explored through a study involving the 'crank' technique in Hidden Behind Static's - 'Penny Drops'. As outlined in Chapter 5, this technique was constructed as part of the binaural localisation test structures and when reproduced binaurally it presented as typically problematic to localise in the rear. Upon implementing a spectrum analysis, it was considered that this could be due to the lack of high frequency content present in the sound source comparative to the amount of low and mid-range spectra. Consequently, it was hypothesized that applying distortion to the source pre-panning, could increase the higher spectral content which may aid localisation.

This was tested in a study and although the result was subtle the ‘crank’ technique did present as easier to localise across all positions of the trajectory, including the rear which was previously problematic (see figure 75 below). However, when positioned in the front, the source presented as *‘too harsh’.* Such that, in retrospect, perhaps applying an automation to remove this distortion function at this point would help to smooth out the harsher timbre presented through the combination of frontal placement and the additional higher harmonic content.

The before and after spectral analysis graphs (taken at the rear position of the source trajectory) show that with the distortion there is a comparative dip around 5-6kHz with an extension increase in the 8-10kHz range, and a less smooth high-frequency roll-off over all than that presented in the ‘before distortion’ graph. Whether this is the reason for the increased localisation is yet to be concluded but the graph suggests this change in spectra as possibly being part of the explanation. A point of interest noted via this exploration is that the use of the Dear VR Pro integrated artificial reverberation on the ‘crank’ source also made it more difficult to localise, both with and without the distortion applied pre-panner.

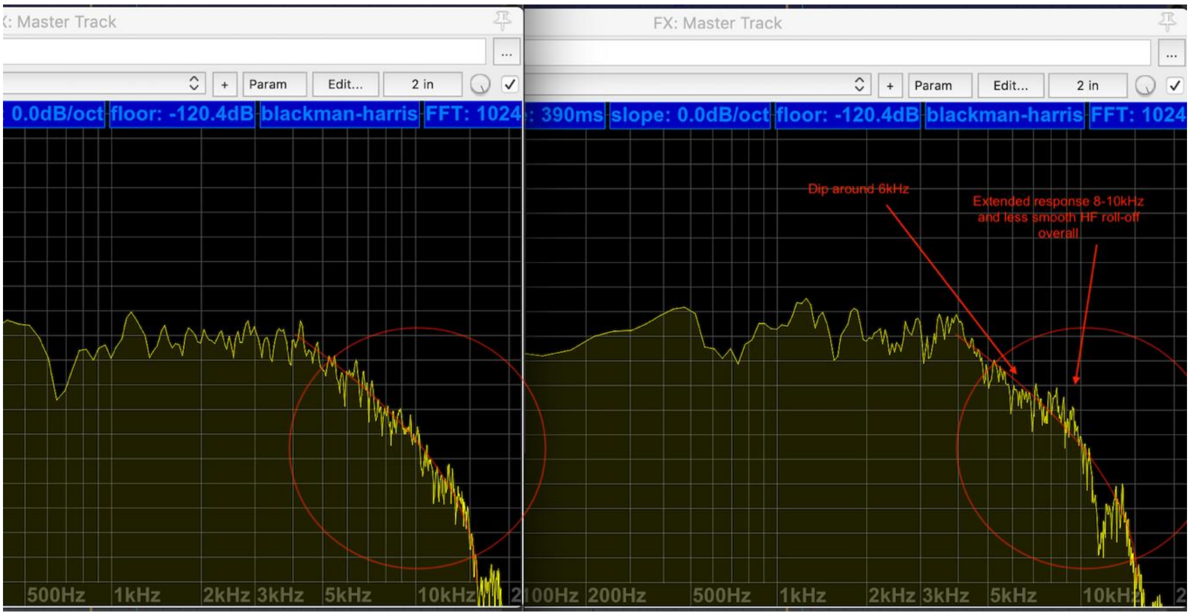


Figure 75: Spectrum analysis of the 'crank' in the rear. Before (left) and after (right) pre-pan distortion.

Soundbox study conclusions:

In terms of the taxonomies; it was theorised that the dynamic mix approach would seemingly be the most adaptable, offering the possibility for fully surrounding kinetic sound stages, and presenting opportunities for lateral, vertical, linear and diagonal movements through a non-front orientated spatial phantom image. It was also suggested that the idea of the diagonal mix could possibly be applied via vertical or horizontal diagonal placements, as a means of utilising the increased perceived performance environment (PPE).

As the practical studies in this chapter have outlined, panning can be used to generate movement and aid localisation via both automated-kinetic and suggestive static placements. Movement can also be implied through melodic contour, inter-externalisation created through conceptual blending of binaural and stereo spaces, and through constructed time-based abstract spatial effects (acoustic reflection and delay emulation). Therefore, it can be concluded that the dynamic mix approach outlined by Dockwray and Moore can be applied to the periphonic-binaural framework, either through actualised or implied movements.

Further, we can define specific musical attributes that may be exploited by the affordance of periphonic panning, and used to enhance the musicality of a production; melodic contour and register, rhythmic cognition and the lyrical narrative.

- Register (enhances elevation)
- Melodic Contour (enhances movement and elevation)
- Rhythm (enhances movement, prominence and perceived acoustic space through temporal effects)
- Lyrical Narrative (Influences source arrangement and staging structures, sonic cartoons and metaphorical concepts)

The studies discussed in this chapter have specifically outlined how the spectromorphology of melodic contour and source placement relative to register may enhance the perception of elevation and movement in a production. A judicious employment of rhythmic structure may enhance movement, prominence and laterality, as well offering the opportunity to construct abstract sonic shapes and spatial effects. The narrative associated with the lyrical content may also be employed in defining the concept of the virtual performance environment, influencing the arrangement and placement of sound sources which make up the spectromorphology of the individual 'mini stages' that comprise the holistic production.

Across the vocal staging case studies, we defined that the spatialised vocal stages often present in a conical or pyramid shape. This is true across the Hidden Behind Static and live recorded pieces, with the vocal staging in Jerome Thomas's 'Late Nights' presenting as more of a 'U' or inverted horse-shoe. However, though this presents as a possible thematic recurrence in regard to perceived staging boundaries and arrangements, there is not enough periphonic-binaural mix data to determine at this point if this is the case. Therefore, it is something to be considered in a future study, perhaps at a time when there is more periphonic-binaural content available for assessment.

In terms of the four-dimensions; what we have come to understand from the project's practical experimentation is that - within a periphonic-binaural framework - panning is a multi-dimensional master function that influences and controls all of the four-dimensions of a virtualised performance.

The effects of panning can be arranged into the following categories and sub-categories;

- Laterality
- Prominence
- Elevation
- Level (which affects prominence)
- EQ (which affects prominence, perceived elevation and source localisation)

- Movement (which affects prominence, laterality, perceived elevation)
- Time-based & Abstract Spatial Effects (which affects perceived acoustics/space, spatial shape, prominence, laterality, temporal continuity and movement).

It could be said that it is not necessarily *how* we create the four-dimensions constructing the 'sound box' that has the most significant impact on the development of periphonic staging practice, but the presentation of the sources in the virtual space relative to the listener's position. It is this encompassing central placement of the listener within the phantom image that has had the most impact on the development of the multi-dimensional non-fronted staging practice, rather than any inter-multi-dimensional or philosophical aspect of the 'sound box' itself.

Instead of having these four-dimensions of the perceived performance environment (PPE) projected in front of the listener, creating an *outside-looking-inside* perception of the *soundbox* betwixt the stereo triangle, the listener is now a part of the PPE and can experience the four dimensions from within the *soundbox* itself. It is this change in the presentation of the PPE phantom image that subsequently inverts the listener's perception to *inside-looking-outside* of the *soundbox*. This change in perception is fundamentally afforded through the periphonic-binaural panning, which explicitly provides the heightened multi-dimensional experience through periphonic sound source placement. Without a front-projecting *soundbox*, there is no need to utilise traditional processes to imply tri-dimensionality to front-respecting, two-dimensional sound stages. With the listener situated within a periphonic-binaural *soundbox*, the interactive multi-dimensionality afforded through a non-front orientated approach to sound staging seems to be implicit.

6.9 - Phase 3 Conclusion

In answering the question *‘How can non-front orientated sound stages for music be approached and structured?’* the techniques constructed within the scope of this study address key issues pertaining to periphonic sound staging and offer a new and contemporary approach to music production. This can be seen more specifically through the design and implementation of the vocal-staging practice created within this project, as well as via the philosophical, theoretical, and practical implementation of staging constructs throughout the associated practical work. In taking an ecological approach to staging through an applied understanding of embodiment theory, the study evidenced that non-front orientated techniques can be creatively utilised to reinforce the sonic narrative, content and metaphor of a production - beyond that which more traditional production approaches can usually afford. This increase in creative agency is demonstrated through the construction and redevelopment of periphonic memes, or sonic cartoons, that comprise the surreal, non-frontal sound stages.

When combined with the binaural phenomena, musical cognition and the integration of proxemics as a production theory, periphonic spatialisation allows for further creative agency in conveying emotion, musicality, depth and intimacy. Although the further work of phase four will evaluate this effectiveness for a range of listeners with differing HRTF compatibility.

Thus, the study presents a new perspective in approaching headphone-based spatial music production, offering a framework of technique that provides pathways for implementation across other musical contexts.

7 – Phase 4 Data Analysis: Focus Group Interviews & HULTIGEN Listening Tests

The objective of the phase 4 data analysis is to comparatively assess the communal experience of the spatial productions against the original artistic intentions[79]. This phase examines the meaning generated through the spatial production work, drawing upon the experiences of the focus group listening sample via the questionnaire answers and the HULTIGEN A/B test data[80].

There are three aims to gathering this data;

1- The first aim is to find out how the aesthetic quality and musicality of the spatial mixes compare to the original stereo arrangements to discern whether the sample prefer a particular format and why. Ultimately, the aim is to understand if there are any characteristics specific to the periphonic-binaural presentation that could lead to the format or the non-front orientated approaches being less effective for music production.

2- The second aim is to determine the efficacy of the non-front orientated staging approaches in conveying the metaphorical meanings and

[79] It is the developmental practice discussed in the previous chapter which forms the original intentionality.

[80] The anonymised raw data can be found in Appendix E, which presents both the questionnaire answers and the HULTIGEN ratings per participant and per each audio track in the listening trials.

production concepts implied through the sonic content and staging design of each piece.

3- The third is to examine the efficacy of the binaural emulation and determine whether the staging concepts manifest as intended when using the same set of generic HRTFs applied across *different* listeners.

The data required to explore these three aims is collected through open-ended, non-leading questions that examine how the stereo and spatial production phenomena are experienced and interpreted by the listeners. The questions request answers are formed as descriptive accounts of the listeners' experiences. The depth and detail of these descriptions varies between listeners, with some presenting more detail than others. As the answers are presented in the listeners' own words, they require interpretation before analysis. This phase employs interpretive phenomenological analysis (IPA) which aims to both reduce and interpret results while finding themes in the data.

> "Good studies will demonstrate how they derived their themes… In all forms of phenomenology, the emphasis is to identify the important messages — the essence or kernel of the topic — so that the important aspects of the phenomenon are described." (Ellis, 2016, pg. 129)

Unlike typical phenomenological approaches, IPA facilitates solipsistic phenomenological enquiry; it does not require that the researcher put aside their existing understanding or experience of the phenomenon in question.

Although the emphasis remains on accessing the *emic perspective*[81], IPA recognises that this is in itself an interpretation of the lived world, and that the researcher will apply an interpretation to the subject's interpretation (Smith and Osborn, 2004; Ellis, 2016). That is, I am attempting to understand the listeners' attempts to make sense of the phenomena, and to do this I may draw upon my pre-existing understanding and experience of the *'thing'* in question.

Due to the qualitative nature of the questionnaire data, the HULTIGEN scaling system was devised and employed as a way to transduce data types to aid interpretation and contain solipsistic influence, quantifying the listeners' format preferences for each production presentation into a value on a five-grade reference scale. This section details the anonymised and coded trial data in literary and graph form[82], providing both textual and visual formats for ease of observing, extracting and digesting meaning.

Ellis (2016, pg. 129) details the four steps to interpretive phenomenological analysis, as follows:

- Reduce the raw data to a manageable form
- Filter the important ideas from those less significant
- Identify important themes
- Construct a narrative account of the analysis.

The literary form has been evaluated using the thematic interpretive phenomenological analysis method above and the results of both data types

[81] Which in this instance would be the perspective of the music 'consumer'.
[82] The graphs are bar charts presenting various breakdowns of the quantitative HULTIGEN preference scoring.

have been summarised and organised into a narrative relative to the three aims outlined previously.

It is appropriate to reiterate that the same pair of headphones were used by each participant across each listening test (AKG K271 MKII). These were the exact pair of headphones that were judged in phase 2 and 2.5 as being the most appropriate for the spatialisation production processes. Consequently, they have been used in both constructing and evaluating the musical work throughout the previous practice as research phases (2 - 3), and as such, they provide a reference through which to deduce the comparative experiential findings of this phase.

This approach ensures consistency across all project quality judgements and between each of the listening trials, negating experiential variables arising from the different functional and technical characteristics presented across varied headphone types, such as; frequency response, in/on/over-ear (the utilisation, or not, of the effect of the pinnae), build quality, impedance and transient response etc. This should afford the Aim 3 determination that if perceptual anomalies arise, they are a result of HRTF mismatch and not an influence of the playback system.

7.1 - Aims, Analysis and Key Findings

<u>Aim 1:</u> *To find out how the aesthetic quality and musicality of the spatial mixes compare to the original stereo arrangements, and to discern whether the sample prefer a particular format and why. Ultimately, the aim is to understand if there are any characteristics specific to the periphonic-binaural format that could lead to the presentation being less effective for music productions.*

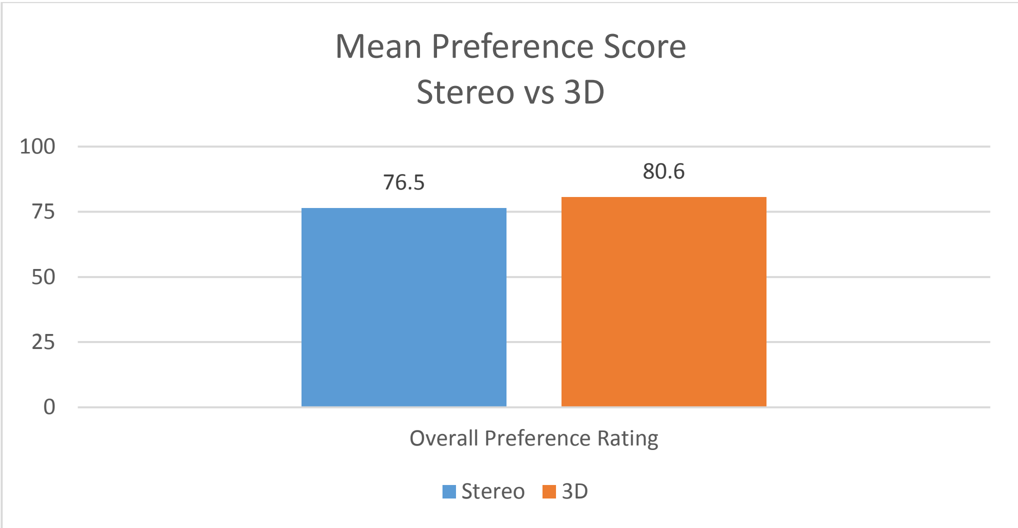


Figure 76: A bar chart showing mean scoring results collated from the HULTIGEN listening test data.

In reference to the above chart (figure 76), although the results are close, on average the HULTIGEN listening test data shows that the spatial productions are generally rated higher in preference to those presented in stereo. However, though this tells us an average of rated preference over all, it does not tell us why this is this case. To explore the underpinning reason for this result we can look to the preference results for each track to extract more detail (see figure 77 below). We can then draw upon the questionnaire data to determine a narrative as to what *aesthetic, quality or musicality* characteristics may inform this preference.

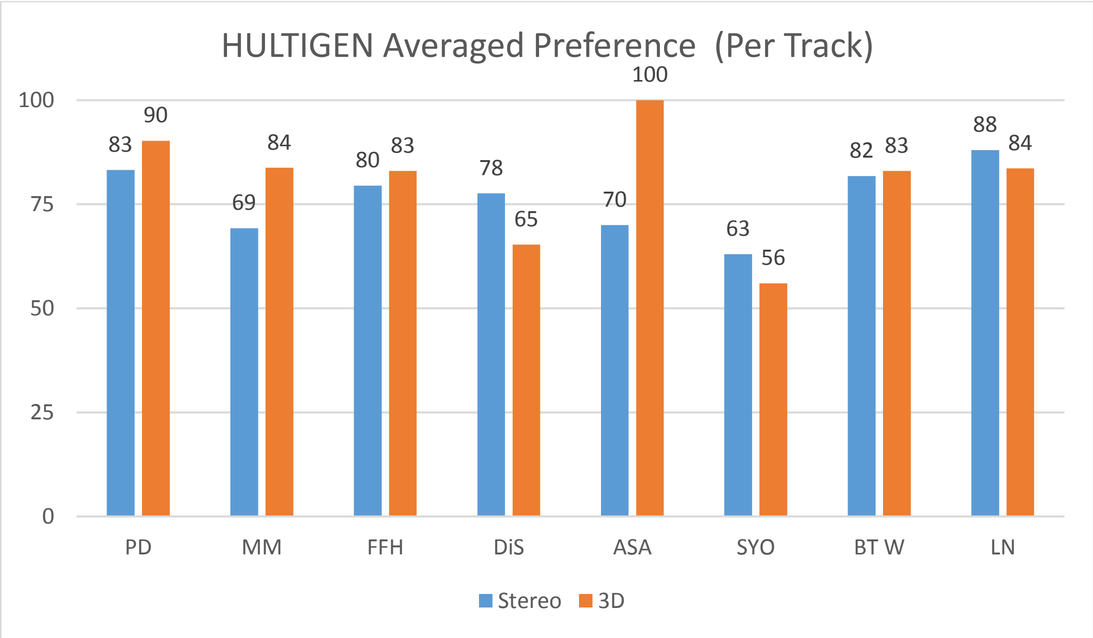


Figure 77: A bar chart showing the mean user preference score per stereo and 3D versions of each track collated through the HULTIGEN listening test data. The figures have been averaged and rounded to the nearest whole number.

Figure 77 provides more detail regarding the breakdown and distribution of preference across the two versions of each track presented. It shows that the preference scale between each format presentation is often very close. A minimum-maximum range of deviation between 1 – 30 scale points can be seen, presenting a mean deviation of 10 scale points. We can use the mean to delineate between a 'small difference' and a 'large difference' in preference, consequently setting a deviation threshold of 10. That is to say that anything 10 scale points and under in preferential deviation can be considered as a 'close' preference between the original and spatialised versions and any rating that deviates above this figure is considered as presenting a distinct difference in preference between the two versions.

Preferences for the stereo mix

Those that present a preference to the original stereo version were 'Sort Yourself Out' (SYO) and 'Late Nights' (LN). Joey Clarkson's - 'Sort Yourself Out' is one of the live pieces discussed in the sonic context and sonic content case studies detailed in the previous chapter, whereas 'Late Nights' is the Jerome Thomas production music piece that explored redevelopments of the vocal staging practice and abstract spatial effect studies.

In regard to SYO the themes found through the questionnaire data suggest that the stereo version presented as having perceivably better sound and performance quality with "more punch and presence" which made it "easier to listen to".

Although the spatial version was perceived as presenting a cleaner mix with easier to localise sources, it was perceived as lacking definition and that "the overall quality of performance and musicality was not very good". Some participants note the "unusual staging jumps" in the spatial version which negatively affected the enjoyment of the production. One participant comments "I think this sort of staging would benefit from better performances and better recorded sources" and another participant reflects this by explicitly stating that the band's performance was not that good and that this detracted from their enjoyment of both versions.

What we can conclude from this analysis reflects the arguments outlined in the previous chapter, reiterating that the sonic context and content of the music has a substantial impact on the sonic quality and the creative coherence of mix aesthetics of a spatialised production. In this case, the live performance context and the acoustic environment captured on the source recordings were not perceived as suitable for the spatial format, and the spatial techniques employed to enhance the production were perceived as more detrimental to the experience than the approaches presented in the typical stereo 'live stage' concept - which is what the recordings were originally indented for.

In terms of 'Late Nights' the commentary was very positive for both formats. No themes could be identified that would explicitly explain the preference result favouring the original stereo version. However, although there was much excitement about the immersion and creative production of the spatial mix, there was comment that the spatial version presents as being less comfortable to listen to comparative to the stereo version which presents a "typically professional 'industry' mix". Although the reason for this discomfort is not specified, it could be suggestively interpreted as being a result of the left-right call and response style lead vocal arrangement. This approached presented a more experimental vocal stage that deviates from the typically front-centre vocal placement associated with stereo listening, and which features on the stereo version.

Key findings

This data suggestively shows that the preference for the stereo versions in these two cases was due to the format presenting staging approaches that were more coherent and better suited to the sonic context and sonic content of each production. However, it is difficult to discern an exact reason behind this result for either track, as the negative feedback for SYO was orientated more toward the poor quality performance and musicianship than the specific qualities and attributes relating to the formats or practice. A similar indeterminate conclusion can be made for LN due to the lack of specifics in the data surrounding the ‘discomfort’ perceived in the spatial production. What we can determine is that the spatial presentations do not enhance the musicality sufficiently enough to warrant a preference to the spatial version, and that the stereo versions were likely chosen because they present a more suitable stage relative to the production context for each presentation.

Preferences for the spatial mix

The tracks that presented a higher preference rating to the spatial version were all three production music pieces from Hidden Behind Static[83] and surprisingly, Beautiful Thing’s - ‘Waiting’ (BTW) which was one of the live pieces discussed in the sonic context and sonic content case studies.

[83] ‘Penny Drops’ (PD), ‘Monomorphic’ (MM) and ‘Far From Here’ (FFH).

In reference to BTW it could be deduced from the questionnaire data that although all participants found both versions comfortable to listen to, the stereo version presented as “more muddy, mono and contained” and “less exciting” than the spatial version which presents “rich, defined space”. The questionnaire data suggests that the listeners enjoyed the enhanced live performance concept presented in the spatial version. A common theme is that the listener felt that they were in a performance space with the band around them and that they felt the production suited the music in the spatial presentation.

For Hidden Behind Static’s pieces a common theme throughout the questionnaire data is that the participants enjoyed the aesthetic qualities of the periphonic-binaural sound field, with statements such as “rich, exciting, immersive and expansive. More professional and more like a concert”. *Depth* was a recurring positive keyword found through the analysis of the questionnaire data of ‘Far From Here’. In regard to ‘Monomorphic’ the commonalities in data suggest that the spatial version was preferred to the stereo as it was perceived as being more “immersive”, “dream-like” and more “suited to the music”. One person commented that the stereo version of ‘Penny Drops’ felt lifeless or “dead” in comparison to the spatial production.

Key findings

What we can deduce from this data is that the spatial pieces were seemingly preferred to the stereo versions due to the exciting aesthetic qualities of the spatial sound field offering enhanced laterality and depth, and affording a richer and cleaner presentation befitting and enhancing the musicality of the production. The spatial enhancement has not only been found to benefit the production music pieces but does also positively benefit the live performance interpretation of BTW. The ‘real-world’ approach taken to the BTW spatial staging reflects the sonic context of a ‘live performance’, consequently enhancing the perceived immersion and providing a ‘hyper-realism’ to the performance piece. The sonic content of BTW contained both a backing vocal and a lead vocal, which presented the opportunity to construct a more coherent spatialised vocal stage than that featured in SYO[84], keeping the lead vocal front-centre, which is more representative of the ‘real-world’ performance context. The source performance and musicianship was better in this case and this too has had a positive impact on the perceived ‘quality’ of the BTW live performance production.

Large deviations to format preference

So far, the pieces and preferences discussed are all found to be under the mean HULTIGEN preference deviation of 10 scale-points. These could be considered as relatively close in terms of the preference rating for each version with most people positively commenting on various aspects of both format versions. However, there are two pieces that exceed the mean

[84] Where there was only one vocal, the lead, which was split into phrases to provide the ability to spatialise the vocal stage. This was then panned left-right in a call-and-response fashion.

threshold of deviance and show a distinctly large HULTIGEN score variation between each of the two versions.

The pieces that measure the largest preferential deviation are Drowned in Sound (DiS) and Art's Self Alteration (ASA). The data shows particularly interesting results in their preference deviation with DiS presenting a preference toward the original production by 13 scale points and ASA presenting a preference toward the spatialised production by 30 scale points.

Neither of these pieces were discussed in the previous chapter of research practice but they are included within the associated audio playlist. They were not included in the previous discussion as the production techniques involved do not explicitly explore the research problem and they were not constructed directly for the investigation in question. These pieces were both early spatial explorations in my back-catalogue of work that were originally constructed on surround and stereo systems with an *intention and consideration* for immersion and spatiality.

In the case of DiS, this track was a previous art-music piece for my 2013 BA dissertation project which explored the creative possibilities of binaural composition. Due to the technological limitations of the time, the project was undertaken using a 5.1 system for the composition and production processes and utilised re-amping with in-ear microphones to render the 5.1 mix to binaural for a headphone-based delivery.

In terms of ASA, this was an immersive composition/production project from my MA the following year. The concept was to present an immersive game-audio inspired musical-political soundscape. However, the production and staging concepts imagined for the piece were beyond what the technology at the time could afford, and as such, the implemented approach was a stereophonic actualisation of what was originally a spatial intention. The techniques employed were fundamentally designed and applied in the stereo domain but with a desire for the concepts and aesthetics to be immersive and spatial.

When the Dear VR Pro plugin was released it presented the opportunity to re-spatialise these pieces and explore implementing the creative ideas and concepts in a format and manner truer to the original visions. Therefore, it was decided that they should be included within the listening tests to determine whether the enhancement presented through the spatial recomposition has benefit the original production concepts or not.

When re-spatialising ASA using the Dear VR Pro, the staging concepts were adapted and applied in a spatial manner that reflected the original placement and conceptual approach as presented in the stereo version. Unfortunately, the DAW session had long been lost for DiS and therefore individual re-spatialisation of sources was not possible. However, the 5.1 channel outputs were located and re-spatialised with the Dear VR Pro plug in, instead[85].

Key findings

What we see from the phase 4 HULTIGEN data is that the original re-amped DiS binaural mix was preferable to the channel output re-spatialisation using

[85] All channels were positioned reflecting their ITU surround configuration.

the Dear VR Pro. However, given the questionnaire data it is difficult to determine why people prefer the original version. Commentary shows that all participants thought positively of both versions, noting the re-spatialised tracks to be more immersive and more ‘realistic’, suggesting heightened ‘real-world’ associations to the staging concepts in both tracks[86]. However, it is noted that in the re-spatialised DiS piece “the front placement feels inconsistent” and that the participants felt the original mix was “tighter and defined”. This statement presents a possible reason underpinning the overall preference to the original mix, even though most seem to enjoy the hyper-real production and ‘real-world association’ of the re-spatialised version. What this seemingly suggests is that although the presentations of the 5.1 channels using the Dear VR Pro worked to convey the concepts and present an immersive reproduction of the original piece, the spatialised channels presented a loss in the ‘glue’ and mix coherence that was apparent in the original re-amped piece. This caused the listener to attend to the inconsistencies in the frontal parts of the presentation, which detracted from the overall experience.

We can also see that the spatial ‘re-imagining’ of ASA presents a much more likeable or meaningful experience of the soundscape than the original stereo version. However, the questionnaire data shows experiences and opinions of each version as being heavily polarised. Data suggests that some people found the spatial version expansive with a lot of depth, whereas others felt it

[86] Most participants refer to the Dear VR presentation of the ‘water’ sounds featured in the piece as being related to ‘rapids’ or ‘rivers’, which is relative to the ‘Drowned in Sound’ concept underpinning the piece but with a more real-world association of the content. Interestingly, in the original mix they are referred to as ‘a running bath’ or ‘bathroom’ which is more representative of the source of the water audio content recordings.

to be ‘flat, harsh and lacking ‘punch’”. However, it also shows that the suggestive reason behind the preference result may be that the spatial version presented as ‘smoother’ to listen to, whereas the stereo piece was “unsettling or uncomfortable”. Seemingly, the participants liked that the spatial version was better at conveying the production concepts than the stereo piece was, with participants feeling as if the spatial mix is “evocative of war”, whereas the stereo version presented “no sense of space [environment] at all”.

This suggests that the definition and techniques of the original stereo mix production were more impacting than the spatial presentation but that the staging was less representative of the musical-political concepts and the imagined soundscape environment. This again suggests that the preference for the spatial piece was due to the periphonic-binaural sound field presenting an association with ‘real-world’ phenomena that better represented the implied spatial environment and the musical concepts of the production.

Aim 1: Conclusions

What does this data analysis tell us about the efficacy of the periphonic-binaural format for music productions?

The results of the thematic IPA analysis suggest that those productions where the techniques were specifically designed or considered relative to the sonic context, content and spatial format are the ones that generally present a more positive preference to the spatial version. This can be seen across the Hidden Behind Static pieces, the techniques of which were developed with the periphonic-binaural format, production concepts and lyrical narratives in mind. This reinforces the notion that the periphonic-binaural format requires new, specific and considered approaches to staging and production. Further, this supports the notion that the creative agency (and enjoyment) pertaining to each piece will be fundamentally reliant on the sonic content and contextual requirements of the musical work in determining how the format is utilised to best reflect the musicality of the piece.

We can see that, generally, the periphonic-binaural format for music is widely considered to offer an exciting experience that enhances the musicality of the productions across the trials. Even the tracks that show a higher rating to the stereo version often present positive experiences toward the spatial qualities and expansive sound field of the periphonic-binaural format. However, these tracks were let down by either the sonic content and or the spatial placement techniques employed, presenting a less coherent, problematic and disassociated sound stage which detracted from the overall positive experience.

We can see evidence referring to and supporting the importance of the hypothesis underpinning the vocal staging investigation of this project, reinforcing the necessity to present a vocal stage in a coherent and balanced manner that best reflects the concept and context of any given production. That is to say, if the piece suggests a live music performance, the voice(s) need to reflect the ‘real-world’ staging associated with the context of that performance. Likewise, where a single lead vocal source is concerned, it is evidently not the best approach to try and create the ‘correct’ sonic content for a vocal spatialisation where it does not exist. We can see the short-comings of this approach evidenced in the lead vocal staging of ‘Late Nights’ and ‘Sort Yourself Out’. In neither track did it benefit the periphonic production to present the lead vocal phrasing as split-panned-pairs. This resulted in a noticeably unbalanced stage and detracted from the overall enjoyment of the spatial pieces.

However, one theme that presents consistently across the data is that of a ‘lack of definition’ or ‘punch’ in the spatial productions. In some instances, the definition and detail presented through the spatial version is seemingly beneficial for clarifying the content (such as in BTW). However, there is frequent reference to the lack of these qualities as a negative aspect of the spatial production format, comparative to the definition and punch presented in the stereo versions. This could possibly be interpreted as a lack of ‘presence’ or ‘fullness’, resultant from the increased spatiality and distance presented through the periphonic-binaural sound field. Perhaps this could be rectified with a more blended approach that combines stereophonic elements to retain the definition and fullness against the spatialised stages. It could also be beneficial for future studies to explore spatial approaches to parallel processing as a possible way to help retain thickness and create variation in textures.

Aim 2: Key findings and conclusion

To determine the efficacy of the non-front orientated staging approaches in conveying the metaphorical meanings and production concepts implied through the sonic content and staging design of each piece.

The HULTIGEN scaling data cannot be used to help determine a conclusion to this particular aim and, as such, this section refers only to the results attained through the descriptive questionnaire answers reflecting the participants' experiences.

For each of the pieces the participants were asked the following questions relating to this aim;

1- What does the record as a whole invoke for you?
2- How would you explain the sonic space / environment?
3- Describe the placement of sounds you hear in each version.
4- Do you visualise what you hear when listening to the music?
5- Is there anything else you would like to add about the music, mixes or your experience?

There were also other questions asked that request descriptions of the placements of sounds heard in the presentations. Although these questions were more investigative of Aim 3, in some instances conceptual commentary can be found in those answers and therefore that data has also been used to help conclude Aim 2.

'Penny Drops' – *Interpretation of metaphorical meanings and production concepts.*

Although the majority of experiences were in agreement that the perceived meaning or intention underpinning the piece was about conveying movement, this piece presented unique variation in interpretations of concept across the sample of participants.

Some described their experience of the spatial presentation as conceptualising an 'abyss', while another felt as if they were inside a *ball* (and that the stereo presented a ball in front of them). Other experiences described "mechanical textures" and feeling as though they are "inside a machine" and some associated the piece with a "casino". Although these interpretations are all quite varied, the intention of conveying movement is completely accurate. The study was focussed on dynamic staging and specifically employed staging approaches that implied movement through the percussive sound source placements (see the *chaser and follower* techniques outlined in chapters 5 & 6).

What is very interesting is the ball interpretation, as I had previously likened the chaser and follower technique to the percussive schema reflecting a 'table-tennis' match. The 'casino' interpretation can also be readily explained by the 'Penny Drop' 'coin-toss' sonic cartoon construct which, when combined with the mechanical textures and rhythms of the percussive staging, could be interpreted as being inside of an environment suggestive of a 'fruit-machine'.

‘Monomorphic’ – *Interpretation of metaphorical meanings and production concepts.*

Some participants refer to their experience of the production concept as “dream-like or zen”, presenting an atmospheric ambience. Whereas, contrary to this, some describe it as “eerie or uneasy”, evoking a feeling of “death”.

One participant describes their interpretation of the sonic environment as being representative of a “horizontal, linear bar with the top and bottom cut away” and another reinforces this with a similar perception, describing an “open, tall box that has no top. The sound escapes from the top”. This interpretation of sound escaping from the top could be reflective of the upward movements of the vocal staging construct.

Again, showing an interpretation relative to the themes of the previous experiences outlined above, another participant refers to the environment as presenting a “wide open space, where the voices are high”. One participant makes the association of “a wide open space like that of a church” and refers to the voices as reminding them of a Gregorian chant. There were also suggestions made that the environment reflects an ‘empty chamber’ and ‘theatre’ and that the vocal staging was suggestive of “four nuns” or a “choir from above”.

The data shows several recurring themes across the collection of interpretations, with most stating the perception of a wide, open space, perhaps representative of a church containing voices like those of a choir. This fully aligns with the vocal production and staging concepts employed within the piece. ‘Monomorphic’ features the omnimonophonic vocal tree structure, where the voices were arranged by register in wide-opposing pairs, presenting a choral sonic cartoon as suggested by the sonic content of the vocal parts. What is even more interesting is that one participant refers directly to this vocal structure as a “plant or flower”, describing the leaves on the stem of plant as being representative of the voice positions and presenting a hand-drawn picture to represent their visualisation. This idea of a vocal “plant or flower” and the positions of the voices drawn by the participant was almost an exact representation of the ‘omnimonophonic vocal tree’ staging concept employed within the piece.

‘Far From Here’ – *Interpretation of metaphorical meanings and production concepts.*

Interpretations of the concepts underpinning this production were also slightly varied, 50% of participants expressed that the piece made them feel “lonely”, “sad” or “melancholy” whereas the other half describe “elation”, “dreamy”, “magical” and “mysterious” and like a “hallucination”. Some participants stated that the aesthetic “feels cold” and others reinforced this through visualising “a frozen lake” or “winter”. The concept underpinning the piece was supposed to convey mental anguish via the presentation of a ‘void’, ‘vacuum’ or ‘nothingness’. Although these keywords were not mentioned, many participants did interpret negative and upset emotions or feelings of ‘coldness’ that could be associated with the original conceptual intention.

As previously mentioned, *Depth* is a recurring keyword in this data. 100% of participants refer to the periphonic-binaural sound field and the placement of sounds as presenting perceivable depth. The production explored depth and intimacy as part of the lyrical staging study and, therefore, this observation provides a positive interpretation of the meaning behind the staging constructs.

Another interesting point found in the data states that a participant thought of the staging as having a changing focus, stating “the focal point was moving”. This positively reflects the unfocussed staging concepts explored within this piece. Several participants refer to the feeling of sources as being positioned in line with the forehead or crown of the head, and that other sources were floating around the head. Again, this is positively representative of the polyperiphonic vocal staging constructs employed in ‘Far From Here’.

‘Drowned in Sound’– *Interpretation of metaphorical meanings and production concepts.*

Two thirds of participants interpreted feelings of “anxiety and peril” from this piece, with others referring to “the unknown”, “undersea aliens, adventure and exploration”.

In terms of the sonic environment, most participants made the conceptual association with ‘water’, suggesting that “they were engulfed in water rapids”, or evoked visualisations of “running water”, “running a bath” or “a river or stream” and that the “sonic elements were less of an illusion and more a real sonic experience”. Two-thirds also found the experience to be uncomfortable due to the aesthetic of being underwater and the anxiety and panic induced through the sounds. Some describe the piece as “disorientating”, whereas another states it as “terrifying”.

Although these results are seemingly negative, they are a positive reflection that the original production intentions and concepts have been conveyed and interpreted correctly through the re-spatialised production. The associations with panic, anxiety and being underwater are fully representative of the metaphorical staging and production concept of ‘drowning in audio’ and the original compositional intention was to create feelings of discomfort and panic.

‘Art’s Self Alteration’– *Interpretation of metaphorical meanings and production concepts.*

100% of participants comment the term “soundscape” when describing the sonic environment and that it is evocative of “war”. One person states that the production sounds as if it belongs to a military game, such as ‘Call of Duty’ or ‘Army Commando’. 100% of participants refer to the presentation as feeling externalised and immersive. This positively reflects the intended concepts and metaphorical staging intentions of the piece and suggests that the re-spatialised version was better at conveying the production concepts than the stereo piece was. Data shows that participants correctly interpreted the political and war concepts through the re-spatialised piece and that the concepts were lost in the stereo version, which presented “no sense of space [environment] at all”.

‘Sort Yourself Out’ and ‘Waiting’ (live performance productions) – *Interpretation of metaphorical meanings and production concepts.*

In both pieces the participants stated that the productions conceptualised the feeling of being in the audience to a live performance in a space, presenting a periphonic sound stage with the feeling of “ambience around the listener” or the feeling of the band “floating around the listener”. This was the intended production concept and given the creative limitations of the sonic content there was not much else that could be done to present alternative concepts or metaphorical representations of schema within the live productions. Interestingly, in ‘Waiting’ (BTW), the concept of the production being associated with a “music video” was mentioned, which suggests an enhanced visual and embodied association between the performance and the space. In this piece one participant also referred to the lead vocal stage as being “high in front during the ad libs but generally in front for the verses”. This is a positive perception of the automated vocal staging approach which was employed as a metaphorical representation of the ascending melodic contour of the lead vocal in the break-down/bridge section.

‘Late Nights’– *Interpretation of metaphorical meanings and production concepts.*

Some participants refer to this piece as presenting the feeling of being in their mind or the mind of the artist, presenting an “immersive, empathic, first-person kind of feel, as if the listener is in the mind/thoughts of the artist”.

Others refer to it as a visualisation of “the [sonic] elements within their head” or that their “head is inside the realm of the music” and some mention that the space [or staging] reflects the state/feelings implied by the lyrics. Although there is no definitive holistic concept behind the spatial staging of this piece, there were lyrical production cues employed in the kinetic backing vocal placements that were supposed to imply concepts suggested through the narrative of the lyrics (example: “Am I too late on arrival?”). The spaced placement of the backing vocal arrangement was intended to explore immersive approaches to non-front orientated vocal staging and imply movement through the vocal placements.

Aim 2: Conclusions

What does this data analysis show in regard to the efficacy of the non-front orientated staging approaches in conveying the metaphorical meanings and production concepts implied through the sonic content and staging design of each piece?

Overall, the data shows that the non-front orientated and periphonic approaches to staging were successful at conveying the metaphorical meaning and production concepts of each spatial piece, even for the tracks that showed a preference to the original stereo presentation. The successful interpretations of metaphorical meaning can be readily seen through the participants’ acknowledgments of the staging constructs relative to each piece, and through their descriptions of the phenomena being suggestive of implied movement, depth and the lyrical narratives underpinning the conceptual and metaphorical meanings etc. This can be exampled across a variety of the productions but more specifically through the perception of the omnimonophonic vocal stage as suggestive of a choir, the ‘Penny Drops’ sonic cartoon coin-toss being associated with a ‘casino’, the “panic and anxiety” induced in ‘Drowned in Sound’, the negative emotions associated with the concept of ‘Far From Here’ and the enhanced immersive live performance association presented through the two spatialised live pieces.

The results consistently show that the spatial format does well to convey and enhance the metaphorical meanings and production concepts presented through the sonic content, context and staging structures in a way that the original stereo productions do not. This is evidenced across the perceived sense of ‘realism’ and the vivid auditory-related visualisations experienced through the spatialised pieces, offering the perception of a ‘hyper-real’ enhancement to the perceived sonic environment which reinforces the metaphorical and conceptual narratives underpinning the productions.

Aim 3: Key findings and conclusion

To examine the efficacy of the binaural emulation and determine whether the staging concepts manifest as intended when using the same set of generic HRTFs applied across different listeners.

The HULTIGEN scaling data cannot be used to help determine a conclusion to this particular aim and, as such, this section refers only to the results attained through the descriptive questionnaire answers reflecting the participants' experiences.

For each of the pieces the participants were asked the following questions relating to this aim;

1- How would you explain the sonic space / environment?
2- Describe the placement of sounds you hear in each version.
3- Do you perceive any sounds to be coming from above, below or either side of you?
4- Do you perceive any sounds to be moving?
5- Do you perceive the sounds to be coming from around you or coming from the front?
6- Is there anything else you would like to add about the music, mixes or your experience?

Generally, the descriptions of phenomena presented in the data suggest that the placements of sounds and the periphonic staging constructs manifest as intended for most participants across the majority of the spatial productions. This can be readily observed through the previous discourse in this chapter, specifically that attending to the descriptions of the sonic environment and perceived key staging constructs, such as; the omnimonophonic vocal tree in 'Monomorphic', the moving percussive sources and the 'coin' cartoon in 'Penny Drops', the unfocussed staging in 'Far From Here', and the vocal placements in 'Late Nights' and the two live pieces.

However, there are some noticeable statements that I wish to draw attention to regarding the experience of a fronted perception projection, and specific experiences that could suggest minor anomalies in perception.

The majority of participants across the listening trials stated that they perceive the original stereophonic versions as always being front-orientated and that the spatial versions were frequently perceived as non-fronted with sounds perceived from around the listener, or that the focus of the production changes. This is true across experiential descriptions of all tracks with the exception of DiS where the perception of a fronted sound stage was identified by two-thirds of participants across both format presentations. This does not seemingly suggest a perceptual anomaly as it can be explained by the binaural re-amping of the 5.1 reproduction in the original version. This approach presents more energy distributed in the front of the sound field due to the frontal L-C-R stereo configuration of the 5.1 surround system. This is further corroborated through the Dear VR re-spatialisation approach also reflecting the same L-C-R channel positioning across the front of the periphonic-binaural sound field.

The most interesting commentary regarding perceived frontal projection comes from one participant who consistently describes their experience as ‘projecting the spatial sound stage to the front as a listening habit’, further stating that as the tests progressed they are “learning” not to do this and that the spatial presentations are “educating me [them] differently”. This experience reflects and reinforces the hypothesis outlined in the introduction of this project, that the format will require a new way of listening to recorded audio and consequently this will also require the facilitation of new approaches to production aligned to this way of listening.

In ‘Late Nights’ (LN) most participants identify the backing voices as being placed in the rear. However, one person states that they perceived a voice as coming from below them. As there are no voices placed below the listener in this piece, this perception may be attributed to the HRTF filtering presenting an elevation anomaly, causing a mis-judgement in the perceived vertical placement of the voice.

The majority of participants also expressed that the stereophonic versions presented as more internalised than the spatial presentations, which appeared as externalised “creating an environment outside of your [their] head”. This is consistent with the well-documented presentation of the stereo sound field when perceived over headphones, which tends to be experienced as more internalised. Contrary to this, the function of HRTF filtering presents the binaural versions as being more externalised with sources perceived to emanate from outside of the head. In ‘Far From Here’ both internalisation and externalisation was felt in the spatial presentation, which is consistent with the conceptual blending of both sound fields in this production[87].

Two-thirds of participants felt that the original version of DiS could be perceived as externalised and that the spatial version presented as externalised also. Again, this consistently reflects the sound field similarities

[87] This was previously discussed in the lyrical production section of chapter 6, where the ‘time’ cue presented a combination of stereophonic and periphonic-binaural staging to create a duality in the perception metaphorically representative of the ‘void’ production concept.

of the two presentations, with binaural externalisation presented via the re-amping process of the original presentation and attained through the periphonic-binaural re-spatialisation associated with the Dear VR version.

Aim 3: Conclusions

The descriptions and statements of experience presented by the participants suggests that, in the majority of cases, the staging concepts manifest as originally intended and that there are few instances where generic HRTFs generate anomalies that detract from the overall experience or enjoyment of the audio. The analysis above details very few anomalies that could be attributed to the binaural filtering process and suggests that where fronted perceptions were projected or inconsistent, this was more suggestive of the staging placements reflecting a stereo arrangement, or the participants' listening behaviour being conditioned to the frontal stereo format. In both instances this is less to do with anomalies of the binaural emulation manipulating the perception of the presentation, and more to do with the perception being informed by stereophonic expectations. Interestingly, it seems as though spatial listening entrainment can be learnt quickly when stereophonic listening habits are acknowledged. This suggests that the listener has an active ability to change their perception of the sound field, attending more to the spatiality and periphonic placements as they become familiar with the attributes and affordances of the spatial format.

7.2 - Phase 4 conclusions

The phase four investigation concludes that the results of the thematic IPA suggest that spatial staging and production techniques function best to enhance the metaphorical concept, musicality and experience of a production when considered relative to the sonic context, content and spatial format in question. Further, the experiences documented through the listening tests support the Chapter 6 study findings that the creative agency pertaining to each piece will be fundamentally reliant on the sonic content and context requirements of the work in question, and these will determine the level of experiential and conceptual enhancement.

We can see that the questionnaire responses referred to and supported the importance of the hypothesis underpinning the vocal staging investigation of this project, reinforcing the necessity to present a vocal stage in a coherent and balanced manner that best reflects the concept and context of any given production.

The participant experiences also referred to and supported the hypothesis that the periphonic-binaural format requires a new way of listening and that this in turn requires specific and considered approaches to staging and production relative to the periphony of the format. The results of the listening tests suggest that stereophonic conditioning and expectations may have an impact on the spatial perception but that spatial listening entrainment is possible and can become more familiar with practice over time. We can see that in the majority of cases the staging concepts manifest as originally intended and that the musicality, concepts and experience of the spatial productions are not negatively impacted by perceptual anomalies or HRTF mis-matches.

Overall, the phase four listening tests have shown that the non-front orientated and periphonic approaches to staging were incredibly successful at conveying the original artistic intentions, enhancing the metaphorical meaning and musical concepts of each spatial production. The documented experiences consistently show that the spatial format does well to convey the metaphorical, musical and lyrical concepts in a way that goes beyond the ability of current practice, offering a ‘hyper-real’ enhancement to the perceived sonic environment which reinforces the conceptual narratives and generates increased immersion.

However, the questionnaire responses reveal that there is a distinct ‘lack of definition’ or ‘punch’ in the spatial productions, which in some cases detracts from the positive experience of the format. To overcome this, it may be beneficial for future studies to further explore conceptual blending of the stereo and periphonic sound fields, or to investigate spatial approaches to parallel processing and sonic layering as a way to help retain thickness and definition in the spatial mix.

8 – Conclusion

'Why continue to structure front-respecting sound stages when we have a much expanded and surrounding virtual performance area now available to us?'

With the on-going developments of technology record production historically has always continued to explore new means of creativity and ways of staging and presenting productions. Consequently, these developments in approach create different types of musical meaning. This thesis acknowledges and attempts to address the dichotomy relating to the historic use of binaural audio for realism and classical music recordings by taking the contemporary developments of binaural technology and applying it to a newer more popular music context. Similarly, this thesis also acknowledges the complex and fraught historical relationship between music production, listening culture and spatial audio technology. With the current culture of headphone-based listening firmly established, there seems to be a perfect storm of technological and cultural change. It is this that facilitates a timely binaural resurgence and provides the opportunity to explore the affordances of new spatial technologies in a culturally relative and creative contemporary context. It is through a novel non-front orientated approach to staging and production technique that this thesis addresses the historic stagnation of creative spatial music making and attempts to fill the subsequent gap in creative spatial music research.

The thesis evidences the importance of a flexible, creative and democratic approach to the production and consumption of spatial music. This is demonstrated through the practical and theoretical explorations of spatial music production systems, evaluating the possibilities they present for catering to the current culture of headphone-based listening.

The preliminary research phases 1-2.5 address the following key questions under pinning the practice-based research problem; *'What systems could be used to approach mixing for a periphonic-binaural format?' 'How can the Auro 13.1 channel mix be converted for a two-channel headphone-based delivery format?' 'How greatly would the variation of headphone affect the perceived musicality and periphonic translation of the music?'*

The preliminary studies comprehensively answer these key questions through research and practical experimentation, investigating and evaluating a variety of approaches, tools and formats in generating headphone-based spatial music productions. These studies determine that *one should work in the format one is mixing for* and conclude that *direct-to-binaural emulation* via an object-based spatial panner, such as the Dear VR Pro, offers a more integratable, reflexive and creative approach to spatial music production than spatial recording or binaural-rendered multi-channel mix output approaches. The study evidences that this method benefits creative record production in a manner that promotes democracy and increased creative agency in practice. This can be seen in the case studies comparatively evaluating the approaches underpinning production practice using the Auro 3D x O3A upmix-decode method, KU100 dummy head and the Dear VR Pro. Further, the study concludes that an Ambisonics mix output benefits a democratic approach to spatial music consumption, providing a variety of up-down mixable formats that offer a flexible delivery of spatial music. The preliminary studies define the importance of using a quality pair of headphones for the production process and suggest that the variation across headphone types in playback does not negatively detract from the immersive experience or impact the interpretation, conceptualisation or the perceived musicality of a spatial music production.

Through practice as research the study addressed the primary research question *‘How can non-front orientated sound stages for music be approached and structured?’* and offers a suggestive approach to sound staging that better utilises the periphony and multi-dimensionality afforded through the spatial sound field.

The techniques constructed within the scope of this study address key issues pertaining to periphonic sound staging and offer a new, contemporary and democratic approach to spatial music production. This is shown through the design and implementation of the instrumental and vocal-staging practice, as well as via the philosophical, theoretical, and practical implementation of the staging constructs throughout the associated production work. The vocal staging approaches address the issue of periphonic vocal placement through the development of concepts such as ‘omnimonophonic’ and ‘polyperiphonic’ vocal staging.

The ‘omnimonophonic’ vocal staging technique addressed this issue by creating an immersive listener-centric vocal stage which presents the voice as not being localised from any specific position but being perceived as ‘one voice from everywhere’. This was implemented through the grouping of voices with a very similar timbre to meld the voices into a “single” omnimonophonic voice or ensemble. The vocal staging was defined based on phrase-matched pairs of single voices, vertically positioned relative to register/pitch to exploit pitch-height effects, and utilising the spectromorphology of the melodic contour as a further means to create upward movement and exaggerated height (see figure 78).

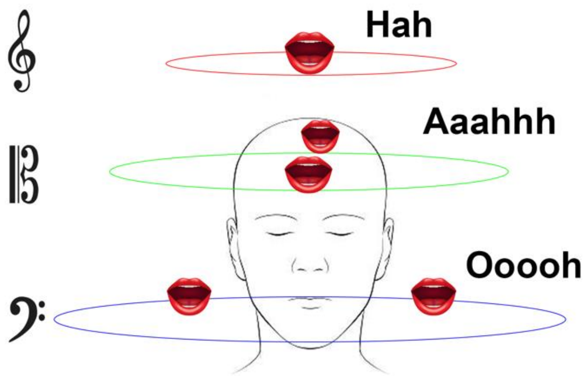


Figure 78: A schematic representation of the Omnimonophonic vocal staging structure as presented within 'Monomorphic'. The vocal height positioning is separated and defined by phrase, and register or pitch.

The 'polyperiphonic' vocal staging technique extends the 'omnimonophonic' vocal work using a multi-track approach to vocal recording to explore perceived depth, intimacy and metaphorical representations of the lyrical content. The technique of multi-tracking the lead vocal provides the opportunity to split phrases into location based on timbral nuance, pitch/frequency content and lyrical content (see figure 79). This technique utilises small changes in timbre and the performance interactions between voices to create the illusion of many voices from one voice.

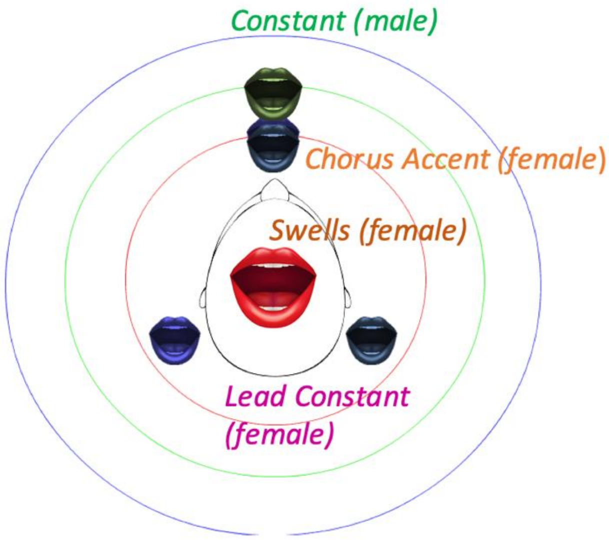


Figure 79: A schematic representation of a snapshot from a section of the 'Far From Here' polyperiphonic vocal stage topography.

The study evidences through both individual and communal judgements that non-front orientated production approaches can be successfully and creatively utilised to reinforce the sonic narrative, content and metaphor of a production - beyond that which more traditional production approaches can usually afford. This increase in creative agency is consistently demonstrated through the construction and redevelopment of periphonic sonic cartoons and the metaphorical staging schema that comprise the surreal, non-frontal productions. When combined with the binaural phenomena and the integration of proxemics, spatial memetics, sonic embodiment and musical cognition as production theories, periphonic spatialisation allows for further creative agency in conveying emotion, movement, depth and intimacy through representative staging 'memes'.

The study presents a new, contemporary framework for approaching virtualised performance and implying multi-dimensionality across spatial music productions, re-contextualising and re-developing the concepts and staging approaches of Dockwray and Moore's (2010) 'soundbox', as applied to the periphonic-binaural format. The research evidences that within a periphonic-binaural framework - panning is a multi-dimensional master function that influences and controls all of the four-dimensions of a virtualised performance, as well as effecting the processes that typically govern these dimensions in the stereo domain (EQ, level and temporal continuity). The study concludes that it is not *how* the dimensions of performance are created that has the most significant impact on the perceived multi-dimensionality and the immersive spatial experience, but how the presentation of the sources in the virtual space suggest metaphorical representations of 'real-world' schema. To provide example, this can readily be seen via the omnimonophonic vocal placements in 'Monomorphic', which uses a single voice, pitch-height placement and periphonic panning to suggest a metaphorical representation of a choir. This can also be evidenced through the implied intimacy and distance generated via the vocal production approach in 'Far From Here', which uses multi-voices and periphonic panning as well as traditional stereophonic processing approaches to create sonic cartoons of proximity.

The thesis presents the opportunity to inform industry and amateur practice, offering *transferrable* and *democratic* approaches to spatial music making through a radical approach to creative research. The collection of case studies offers a new perspective in approaching headphone-based spatial music production, presenting a framework of technique that provides pathways for implementation, expansion and future development across other musical contexts.

A suggested future context for research development should investigate the possibilities of practice as applied to a wider variety of musical genres. Further, and as outlined in the preceding chapter, future research should investigate the integration of spatial approaches to parallel processing as a way to help retain thickness and definition in the spatial presentations.

Another future context for development could explore the application of non-front orientated staging with head-tracking integration toward the construction of sonic content for interactive Virtual Reality music videos, or to investigate the application of non-front orientated approaches for live music performances, with or without head-tracking capabilities, Mixed Reality (MR) or Augmented Reality (AR) visuals. The notion of applying non-front orientated sound staging to a performance that may implicitly suggest a fronted sound stage presents as a particularly interesting study concept. This may provide the opportunity to develop and apply the concepts defined within this research to the wider field, presenting further reach and research impact and the possibility of integrating live and spatial music production research practices.

8.1 - Original Contributions to Knowledge

An original contribution to knowledge can be evidenced in the development of novel non-front orientated staging and production practice. The practice developed within this research resolves key historic issues relating to spatial music production and sound staging. This can be seen specifically in the ‘omnimonophonic’ and ‘polyperiphonic’ vocal staging concepts, as well as evidenced throughout the associated instrumental staging work.

An original contribution to knowledge can be evidenced through the creation of semantics used in describing the phenomena of the new practice. This can be evidenced in the original terms ‘Omnimonophony’, ‘Omnimonophonic’, ‘Polyperiphony’,’Polyperiphonic’ and ‘Periphonic-Binaural’. This presentation of semantics is specific and unique to the concepts developed within the research practice, and ultimately defines an original knowledge contribution through the development of field-specific terminology.

An original contribution to knowledge takes the form of the concepts, theories and philosophy underpinning the practical work investigations. This can be evidenced through the development and expansion upon pre-existing research frameworks in an application uniquely relative and beneficial to this study. Specifically, this is evidenced through the spatial re-contextualisation of Dockwray and Moore’s ‘soundbox’ concepts and mixing taxonomies (2010). Further, this can also be evidenced through the application and exploitation of ‘binaural phenomena’, ‘proxemics’, ‘embodiment theory’, ‘musical & auditory cognition’ and ‘memetic theory’ as vehicles to enhance the metaphorical, aesthetic and musical qualities of the spatial music productions. This presents a new and unique conceptual-theoretical framework informing spatial production practice.

An original contribution to knowledge can be found via the use of critical theory in defining a democratic approach to spatial music making. This presents an ideological approach to spatial production practice that is unique to this research project. Critically, this ideological approach promotes inclusion and affords the ability for the knowledge and practice defined herein to be experienced, explored and developed by a wide range of people.

An original contribution to knowledge can be found via the memetic design of key strategies for conveying metaphor, which provide a framework for understanding the technical and creative potential of this format. This promotes the notion that the embodiment of metaphorical spatial ideas requires a development of mixing and production/recording approaches, both in generating 'surrealist' and 'hyperrealist' presentations. Further, the performance and production music experimentation and the lessons that this project has set in motion may provide interesting models and directions for further study.

References

Books, Journals and Papers:

Bartlett, B. and Bartlett, J. (2002) *Practical Recording Techniques*, 3rd ed. London, Focal Press.

Begault, D. R. (Ed.) (2004) *Spatial Sound Techniques- Part 1: Virtual and Binaural Audio Technologies*. New York, USA: Audio Engineering Society.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound.* Cambridge, MA, US: The MIT Press.

Byrne, D. (2012a) *How Music Works*. Edinburgh, Scotland: Canongate Books

Clarke, E. (2005) *Ways of Listening*. New York, New York: Oxford University Press

Dawkins, R. (2016) *The Selfish Gene* [Kindle Edition]. Oxford, UK: Oxford University Press.

Dockwray, R. and Moore, A. F. (2010) Configuring the sound-box 1965–1972, in *Popular Music*, 29 (2), pp. 181–197. doi: 10.1017/S0261143010000024.

Everest, F. A. and Pohlmann, K. C (2009*) The Master Handbook of Acoustics*, 5th rev. ed. London: Mc Graw-Hill.

Freire, P. (2012) *Pedagogy of the Oppressed*. London, UK: Bloomsbury

Frith, S. and Zagorski-Thomas, S. (2012) *The Art of Record Production*. Ashgate Publishing: UK

Gardner, H. (1985). *The mind's new science: a history of the cognitive revolution*. New York: Basic Books.

Gerzon, M. A. (1973) Periphony: With-Height Sound Reproduction, in *Journal of the Audio Engineering Society*, 21 (1), pp. 2–10.

Hall, E. T. (1966) *The Hidden Dimension*. United States: Bantam Doubleday Dell Publishing Group.

E. C. Hamdan and F. M. Fazi (2021) A modal analysis of multichannel crosstalk cancellation systems and their relationship to amplitude panning, *Journal of Sound and Vibration*, Volume 490, 115743, ISSN 0022-460X, https://doi.org/10.1016/j.jsv.2020.115743.

Henry, M. J. and Grahn, J. A. (2017) Music, Brain and Movement; Time, Beat and Rhythm. In: Ashley, R. and Timmers, R. (Ed.) (2017) *The Routledge Companion to Music Cognition*. Routledge: Abingdon, Oxon, England.

Holman, T. (2007) *Surround Sound Up and Running*. Abingdon, Oxon: Focal Press

Howard, D. and Angus, J (2006) *Acoustics and Psychoacoustics*, 3rd rev. ed. Oxford: Focal Press.

Hughes, J. and Sharrock, W. W. (2007) *Theory and Methods in Sociology*, London: Palgrave.

Ihde, D. (1986) *Experimental phenomenology: an introduction*. Albany, State University of New York Press.

Ingold, T. (2011) *The Perception of the Environment: Essays on livelihood, dwelling and skill*. Abingdon, Oxon: Routledge.

L. Rayleigh (1907) On Our Perception of Sound Direction, in Begault, D. R. (Ed.) (2004) *Spatial Sound Techniques- Part 1: Virtual and Binaural Audio Technologies*. New York, USA: Audio Engineering Society.

Langdridge, D. (2007) *Phenomenological psychology: Theory, research and methods*. London: Pearson.

Ledbetter, M. (1996). *Victims and the postmodern narrative, or, doing violence to the body: An ethic of reading and writing.* New York: St. Martin's Press.

Lee, H. (2017) Sound source and loudspeaker base angle dependency of phantom image elevation effect. In: *Journal of the Audio Engineering Society*, vol. 65, no. 9, Sep. 2017, doi: 10.17743/ jaes.2017.0028.

Leman, M. (2008) *Embodied Music Cognition and Mediation Technology.* MIT Press: Cambridge, Massachusetts, USA.

Lord, J. (2021) Chapter 13 - Redefining the spatial stage: non-front orientated approaches to periphonic sound staging for binaural reproduction, in Lee, H. and Paterson, J., (eds.) *Perspectives on Music Production - 3D Audio*, London, England, Routledge, pgs. 256-273.

Martens, P. and Benedon, F. (2017) Musical Structure; Time and Rhythm. In: Ashley, R. and Timmers, R.(Ed.) (2017) *The Routledge Companion to Music Cognition*. Routledge: Abingdon, Oxon, England.

McLuhan, M. and Fiore, Q. (1967) *The Medium is The Massage*. London, England, UK: Penguin Books

Merleau-Ponty, M. (1962). *Phenomenology of perception*. London: Routledge & Kegan Paul.

Moylan, W. (1992) *The art of recording: the creative resources of music production and audio*. New York: Van Nostrand Reinhold.

Moylan, W. (2012) Considering space in recorded music, in Zagorski-Thomas, S. and Frith, S. (eds) *The Art of Record Production*. United Kingdom: Routledge (Ashgate Popular and Folk Music).

Nussbaum, M. C. (1990). *Love's knowledge: Essays on philosophy and literature*. New York: Oxford University Press.

Ruane, J. M. (2005) *Essentials of Research Methods*, Oxford: Blackwell.

Rumsey, F. (2001) *Spatial Audio*. Oxford, UK: Focal Press

Silverman, D. (1997). *Qualitative research: Theory, method, and practice*. London: Sage Publications.

Smalley, D. (1986) Spectromorphology and structuring processes. In S. Emmerson (ed.) *The Language of Electro-acoustic Music*, pp. 61–93. Basingstoke: Macmillan

Thompson, W.F. (2015) *Music, Thought and Feeling*. Oxford, UK: University Press

Torrick, E. (1998) Highlights in the History of Multichannel Sound. JAES, January-February, vol. 46, pg.27-31, in Rumsey, F. (Ed.) (2006) *AES: An Anthology of Spatial Sound Techniques Part 2: Multichannel Audio Technologies*. New York, USA: AES.

Vorländer, M. (2008) *Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*. Germany: RWTH Aachen University

White, G. D. and Louie, G. J. (2005) *The Audio Dictionary: Third Edition, Revised and Expanded*, REV- Revised, 3. University of Washington Press.

Wibberley, C. (2017) Bricolage Research Methods. In: *Nursing and Healthcare Research: at a glance*, Glasper, R. (ed.), ch 52, 106-107. Oxford: Wiley-Blackwell.

Wisker, G. (2008) *The Post-Graduate Research Handbook: succeed with your MA, MPhil, EdD and PhD*. 2nd rev. ed. Hampshire, UK: Palgrave Macmillan

Xie, B. (2013) *Head-related Transfer Function and Virtual Auditory Display*, 2nd ed. USA: J. Ross Publishing

Zagorski-Thomas, S. (2014) *The Musicology of Record Production*. Cambridge, United Kingdom: Cambridge University Press.

Zagorski-Thomas, S. (2015) Developing the formal structures of artistic practice-as-research, *New Vistas,* 1 (2), p.28-32. doi: https://doi.org//uwl.18 [Accessed: 24th August 2021]

Zagorski-Thomas, S. (2018) The Spectromorphology Of Recorded Popular Music: the shaping of sonic cartoons through record production, in Fink, R., Latour, M., and Wallmark, Z. (eds) *The Relentless Pursuit of Tone: Timbre in Popular Music*. Oxford, United States: Oxford University Press USA - OSO. Available at: http://ebookcentral.proquest.com/lib/uwestlon/detail.action?docID=5509530 [Accessed: 17 November 2020].

Zak, A. (2001) *The Poetics of Rock: Cutting Tracks, Making Records.* California, USA: University of California Press.

Internet Sources:

Belshaw, D. (2011) *Methodology section: Critical theory* [Online]. Available at: https://dougbelshaw.com/blog/2011/05/29/methodology-section-critical-theory/ [Accessed: 23rd August 2021]

Blue Ripple Sound (2015) *Pro Audio: Practical high resolution 3D audio tools for the studio* [Online]. Available at: https://www.blueripplesound.com/ [Accessed: June 2016]

Bohman, J. (2010) Critical Theory, in *The Stanford Encyclopedia of Philosophy (Spring 2010 Edition)*, Zalta, E. N., (ed.). Available at: http://plato.stanford.edu/archives/spr2010/entries/critical-theory/ [Accessed: 23rd August 2021]

Boom Library (2019) *Why we don't use Ambisonic microphones – even for VR* [Online] Available at: https://www.boomlibrary.com/blog/why-we-dont-use-ambisonic-microphones-even-for-vr/ [Accessed: May 2019]

Briscoe, R. (2015) *Action-based Theories of Perception* [Online]. Available at: http://plato.stanford.edu/entries/action-perception/ [Accessed: 26th October 2015]

Byrne, D. (2012b) *My Love Affair With Sound* [Online]. Available at: https://www.salon.com/2012/09/16/david_byrne_my_love_affair_with_sound/ [Accessed: June 2019]

CBS International (2021) *Binaural Music* [Online]. Available at: https://www.last.fm/tag/binaural/albums [Accessed: 25th September 2021]

Cheng, C. I. and Wakefield, G. (n.d) *Introduction to Head Related Transfer Functions: Representations of HRTFs in Time, Frequency and Space* [Online]. Available at:

http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.442.5500&rep=rep1&type=pdf [Accessed: June 2014]

Dearreality via Instagram (2019) *Our CEO working with the Magic Leap Headset* [Instagram post]. Available via Instagram at: http://www.instagram.com/p/B1L9nURApZj/?igshid=1m0jriyk63ile [Accessed: 15th August 2019]

dearVR.com (n.d) Ultimate tools for immersive 3D audio production [Online]. Available at: https://www.dearvr.com/ [Accessed: 19th August 2019]

Doctorlib.info (2019) *Medical Physiology, 3rd Edition, Temporal Representations: Time-Measuring Circuits* [Online]. Available at: https://doctorlib.info/physiology/medical/90.html [Accessed: 25th January 2022]

Ellis, P. (2016) The language of research (part 8): phenomenological research, in *Wounds UK*, 12 (1), pgs. 128-129. Available at: https://www.wounds-uk.com/journals/issue/46/article-details/the-language-of-research-part-8-phenomenological-research [Accessed: 31st August 2021]

Farina, A. et al. (2001) Ambiophonic Principles for the Recording and Reproduction of Surround Sound for Music, in *Audio Engineering Society Conference: 19th International Conference: Surround Sound - Techniques, Technology, and Perception*, Audio Engineering Society. Available at: https://www.aes.org/e-lib/browse.cfm?elib=10114 [Accessed: 17 November 2020].

Finlay, L. (2009) Debating phenomenological research, in *Phenomenology & Practice*, 3 (1), 6-25. Available at: https://doi.org/10.29173/pandpr19818 [Accessed: 23rd August 2021]

Gibson, D. (2018) *The Art of Mixing: A Visual Guide to Recording, Engineering, and Production.* Milton, United Kingdom: Taylor & Francis Group. Available at: http://ebookcentral.proquest.com/lib/uwestlon/detail.action?docID=5631496 [Accessed: 17 November 2020].

Given, L. M. (2008) *The sage encyclopedia of qualitative research methods*. SAGE Publications, Inc., Thousand Oaks, CA. [Accessed 17 August 2021], doi: 10.4135/9781412963909.

Gribben, C. and Lee, H. (2015) Towards the development of a universal listening test interface generator in Max, in: *138th Annual Audio Engineering Society AES Convention, 7th-10th May 2015*, Warsaw, Poland. Available at: http://eprints.hud.ac.uk/id/eprint/24809/ [Accessed: July 2018]

Guttenberg, S. (2009) *Whatever happened to 5.1-channel music?* [Online], Stereophile.com. Available at: https://www.stereophile.com/asweseeit/whatever_happened_to_51-channel_music/index.html [Accessed: 17 November 2020].

Guttenberg, S. (2014) Is surround sound for music and home theatre on its way out? *CNet - Audio*, 9th February 2014 [Online]. Available at: http://www.cnet.com/uk/news/is-surround-sound-for-music-and-home-theater-on-its-way-out/ [Accessed: 13th November 2015]

HESA (2020) *Who's working in HE? Personal characteristics* [Online]. Available at: https://www.hesa.ac.uk/data-and-analysis/staff/working-in-he/characteristics [Accessed: 24th August 2021]

Higher Education Funding Council for England (2012) *Assessment framework and guidance on submissions*, UK: REF 2014

Hills-Duty, R. (2017) *Dear reality release new workflow for spatial audio* [Online]. Available at: https://www.vrfocus.com/2017/07/dear-reality-release-new-workflow-for-spacial-audio/ [Accessed: November 2017]

Hooke Audio (2021) *The Best Albums Recorded in Binaural* [Online]. Available at: https://hookeaudio.com/blog/music/best-binaural-albums/#google_vignette [Accessed: 25th September 2021]

Hughes, J. (2012) *Meme Theory: Do we come up with ideas or do they, in fact, control us?* [Online]. Available at: https://www.independent.co.uk/life-style/gadgets-and-tech/features/meme-theory-do-we-come-up-with-ideas-or-do-they-in-fact-control-us-7939077.html [Accessed: 12th August 2018]

Intgrty (2016) *Research Paradigms: Critical Theory* [Online]. Available at: https://www.intgrty.co.za/2016/08/08/research-paradigms-critical-theory/#:~:text=Critical%20theory%20is%20any%20research,makes%20claim%20to%20scientific%20objectivity [Accessed: 23rd August 2021]

Kafle, N. (2011) Hermeneutic phenomenological research method simplified [PDF], in *Bohdi, an interdisciplinary journal,* 5, pgs. 181-200. Kathmandu University, Nepal. ISSN: 2091-0479. [Accessed: 24th August 2021].

Koreimann, S., Gula, B. and Vitouch, O. (2014). Inattentional deafness in music [PDF], in *Psychological Research. 10.1007/s00426-014-0552-x*. Available at: https://www.researchgate.net/publication/260998729_Inattentional_deafness_in_music [Accessed: August 2019)

Kuzminski, A. (2018) These fascinating new tools allow you to do 3D mixing – directly in VR [Online], *in A Sound Effect, Game Audio, August 15th 2018.* Available at: https://www.asoundeffect.com/vr-3d-sound-mixing/. [Accessed: December 2018]

Lacasse, S. (2000) *'Listen to my voice': The evocative power of vocal staging in recorded rock music and other forms of vocal expression* [PDF]. Available at: https://www.researchgate.net/publication/35721252_'Listen_to_my_voice'_the_evocative_power_of_vocal_staging_in_recorded_rock_music_and_other_forms_of_vocal_expression [Accessed: April 2018]

Lacasse, S. (2005) *Persona, emotions and technology: the phonographic staging of the popular music voice*. Available at: http://charm.cchcdn.net/redist/pdf/s2Lacasse.pdf [Accessed: 17 November 2020].

Lester, S. (1999) *An introduction to phenomenological research* [Online]. Available at: https://www.rgs.org/CMSPages/GetFile.aspx?nodeguid=7ad9b8d4-6a93-4269-94d2-585983364b51&lang=en-GB [Accessed: 31st August 2021]

Lovely Virus [n.d] *Binaural Barber Shop Experience* [YouTube]. Available at: https://www.youtube.com/watch?v=IUDTlvagjJA&feature=emb_title [Accessed: 2015]

Marsden, R. (2015) The Rise and Rise of Headphones. *The Independent*. 1st December 2015 [Online] Available at: https://www.independent.co.uk/arts-entertainment/music/features/the-rise-and-rise-of-headphones-why-the-set-you-buy-isn-t-just-a-question-of-sound-quality-but-a6756396.html [Accessed: June 2016]

Mc Carthy Iatse, S. (2009) Time Alignment & The Haas Effect: What is the Haas effect? *Audio, Computers and Everything Else.* Tuesday 2nd June 2009 [Online]. Available at: http://steve1mac.blogspot.co.uk/2005/03/time-alignment-haas-effect.html [Accessed: 7th July 2014].

McLaughlin, S. (2014) *The four essential elements of EQ and your audio mix* [Online] Available at: https://blog.discmakers.com/2014/04/essential-elements-of-eq-and-your-audio-mix/ [Accessed: August 2019]

McWilliams, A. (2011) *Introduction to Ambisonics* [Online] Available at: https://jahya.net/blog/introduction-to-ambisonics/ [Accessed: July 2017]

Moylan, W. (2009) *Considering Space in Recorded Music*. ARP journal https://www.arpjournal.com/asarpwp/considering-space-in-music/

Moylan, W. (2020) *Recording Analysis: How the Record Shapes the Song*. Milton, United Kingdom: Taylor & Francis Group. Available at: http://ebookcentral.proquest.com/lib/uwestlon/detail.action?docID=6109511 [Accessed: 17 November 2020].

Noizefield (2017) *Deareality release two 3D mixing plug-ins dear VR Music and dear VR Pro* [Online]. Available at: https://www.noizefield.com/news/dear-reality-releases-two-3d-mixing-plugins-dearvr-music-dearvr-pro [Accessed: November 2017]

Owsinski (2011) *Stereo Triangle Speaker Set Up* [Image] Available at: http://bobbyowsinski.blogspot.com/2011/08/6-steps-to-set-up-your-monitor-speakers.html#axzz6lLJQdkDP [Accessed: 1st February 2021]

Petrusich, A. (2016) Headphones Everywhere [Online]. The New Yorker, 12th July 2016. Available at: https://www.newyorker.com/culture/cultural-comment/headphones-everywhere [Accessed: 25th April 2017].

Politis, A., Tervo, S. and Pulkki, V. (2018) *COMPASS: Coding and Multidirectional Parameterization of Ambisonic Sound Scenes* [Online]. Available at: http://research.spa.aalto.fi/publications/papers/icassp18-compass/ [Accessed: June 2018]

Popova, M. (2011) *The medium is the massage; Shepard Fairey and Marshall Mcluhan* [Online]. Available at: https://www.brainpickings.org/2011/06/20/the-medium-is-the-massage-shepard-fairey-marshall-mcluhan/ [Accessed: June 2019]

Ramsey, C. (2017) *Spatial Awareness Inside the World of Immersive Sound Design* [Online]. Available at: https://audiomediainternational.com/spatial-awareness-inside-the-world-of-immersive-sound-design/ [Accessed: 17th November 2020]

Ricoeur, P. (1975) *Phenomenology and Hermeneutics [PDF]*. Noûs, 9 (1), Pgs. 85-102. *Available at: JSTOR*, www.jstor.org/stable/2214343. [Accessed: 27th July 2021]

Roberts, J. M. (2014) Critical Realism, Dialectics, and Qualitative Research Methods [word doc], in *Journal for the Theory of Social Behaviour* 2014, vol. 44, no. 1, pgs. 1-23. Available at: https://bura.brunel.ac.uk/bitstream/2438/11425/1/Critical%20Realism%20and%20Qualitative%20Research%20Methods4.doc [Accessed: May 2019].

Smalley, D. (1997) Spectromorphology: Explaining sound shapes [Online PDF], in *Organised Sound 2 (2)*, 26-107. Cambridge University Press: Cambridge, England. Available at: https://doi.org/10.1017/S1355771897009059 [Accessed: 26th February 2019]

Smith, J.A. and Osborn, M. (2004) Interpretative Phenomenological Analysis. in *Qualitative Psychology*, Smith, J.A. (ed.), 51–80. Sage Publications, London.

SSI (n.d) *Spatial Sound Institute – Spatial Memetics* [Online]. Available at: https://spatialsoundinstitute.com/spatial-memetics [Accessed: January 2018]

The Physics Classroom (1996) *The Doppler Effect* [Online]. Available at: https://www.physicsclassroom.com/class/waves/Lesson-3/The-Doppler-Effect [Accessed: 9th October 2021]

University of California (2011) *Psychoacoustics of Spatial Hearing* [Online]. Available at: http://interface.cipic.ucdavis.edu/sound/tutorial/psych.html [Accessed: 14th June 2014].

Waves (2017) *Ambisonics Explained: a guide for sound engineers* [Online]. Available at: https://www.waves.com/ambisonics-explained-guide-for-sound-engineers [Accessed: December 2017]

Webb, D. (2008) *The Virtual Barber Shop* [Online]. Available at: https://www.all-about-psychology.com/virtual-barber-shop.html [Accessed: 2015]

White, P. (1995) Qsound Labs: Right on Q*. Sound On Sound*. November 1995 [Online]. Available at: http://www.soundonsound.com/sos/1995_articles/nov95/qsound.html [Accessed: 20th April 2016]

White, P. (2000) ‘Improving your stereo mixing’ [Online]. *Sound on Sound*. October 2000 [Online]. Available at: http://www.soundonsound.com/sos/oct00/articles/stereomix.htm [Accessed 20th November 2020]

Woolston, C. (2020) *White men still dominate in UK academic science* [Online]. Career News, 12th March 2020. Available at: https://www.nature.com/articles/d41586-020-00759-1 [Accessed: 24th August 2021]

Wuttke, J. (2005) Surround Recording of Music: Problems and Solutions, in. *Audio Engineering Society Convention 119*, Audio Engineering Society. Available at: http://www.aes.org/e-lib/browse.cfm?elib=13356 [Accessed: 17 November 2020].

Yardley, A. (2008) *Piecing Together: A methodological bricolage* [Online]. Available at**:** http://www.qualitative-research.net/index.php/fqs/article/view/416/902 [Accessed: July 2018]

Conferences, Conventions and Proceedings:

Cabrera, D. and Tilley, S. (2003) Parameters for Auditory Display of Height and Size. In: *Proceedings of the 9th International Conference on Auditory Display*, July 2003, Boston, MA, USA.

Novotny, P. (2019) Performance music or production music? [Presentation] In: the *14th Art of Record Production Conference,* 17th –19th May 2019, Boston, MA. United States, May 2019. https:// arp19.sched.com/ event/ L4Dc/ performance-musicor-production- music [Accessed Nov. 17, 2020].

Audio Visual Media:

Audiofanzine (2019) *Dear VR Pro* [Image]. Available at: https://en.audiofanzine.com/surround-treatment/dear-reality/dearvr-pro/medias/videos/ [Accessed August 2019]

Auro Technologies (n.d) *Auro 3D System* [Online Image]. Available at: http://www.auro-3d.com/system [Accessed: 17th February 2016]

Beautiful Thing (2020) Waiting (Live at London Water & Steam Museum) [Periphonic Mix], *Soundcloud*, 17th November 2020. Available at: https://bit.ly/2MZpcWu [Accessed: 17th November 2020].

Doctorlib.info (2019) *Medical Physiology, 3rd Edition, Temporal Representations: Time-Measuring Circuits* [Online Image]. Available at: https://doctorlib.info/physiology/medical/90.html [Accessed: 25th January 2022]

Fielding, C. [n.d] *College of Santa Fe Auditory Theory: Lecture 008 Hearing III* [Online Image]. Available at: http://www.feilding.net/sfuad/musi3012-01/html/lectures/008_hearing_III.htm#ITD [Accessed: 14th June 2014].

Hidden Behind Static (2020) Far From Here, *Soundcloud*, 17th November 2020. Available at: https://bit.ly/3fugtI5 [Accessed: 17th November 2020].

Hidden Behind Static (2020) Monomorphic. *Soundcloud*, 17th November 2020. Available at: https://bit.ly/37sEhct [Accessed: 17th November 2020].

Hidden Behind Static (2020) Penny Drops. *Soundcloud*, 17th November 2020. Available at: https://bit.ly/2ZqYfol [Accessed: 17th November 2020].

IOSONO [n.d] *Spatial Audio Workstation 2* [Online Image]. Available at: http://www.iosonosound.com/spatial-audio-workstation.html [Accessed: 20th July 2014]

Jerome Thomas (2020) Late Nights [Periphonic Mix], *Soundcloud*, 17th November 2020. Available at: https://bit.ly/3e1okfV [Accessed: 17th November 2020].

Joey Clarkson (2020) Sort Yourself Out (Live at Brentford Steamworks) [Periphonic Mix], *Soundcloud*, 17th November 2020. Available at: https://bit.ly/30EX6HQ [Accessed: 17th November 2020].

Klein, E. and Posner, J. (2019) *Explained: Music* [Online]. Available at: https://www.netflix.com/watch/80243768?trackId=13752289&tctx=0%2C0%2C668ccef6-2602-44a8-861b-f848bbd97b1c-59293297%2C%2C

Lee, H. (2015) *Perceptually Motivated 3D Music Production* [Online Video]. Available at: https://www.youtube.com/watch?v=2zSnKUR_Gvo [Accessed: 2nd December 2015]

Lord, J. (2021) Art's Self Alternation [Periphonic Mix], *Soundcloud*, 19th October 2021. Available at: https://bit.ly/3pcSe9n [Accessed: 19th October 2021]

Lord, J. (2021) Drowned in Sound [Periphonic Mix], *Soundcloud*, 19th October 2021. Available at: https://bit.ly/3jdPJzy [Accessed: 19th October 2021]

Oodlepic.com (n.d) *An eye or a kitchen sink?* [Image]. Available at: https://www.oodlepic.com/weirdnutdaily/oodlebe/2u0/ [Accessed: August 2019]

RAA [n.da] *AKG K271MKII report for a pro* [Online]. Available at: https://reference-audio-analyzer.pro/en/report/hp/akg-k-271-mk2.php#gsc.tab=0 [Accessed: 25th September 2021]

RAA [n.db] *Sennheiser HD25 report for a pro* [Online]. Available at: https://reference-audio-analyzer.pro/en/report/hp/sennheiser-hd-25.php#gsc.tab=0 [Accessed: 25th September 2021]

RAA [n.dc] *Apple Earpods report for a pro* [Online]. Available at: https://reference-audio-analyzer.pro/en/report/hp/apple-earpods.php#gsc.tab=0 [Accessed: 25th September 2021]

Rodkimblestuntman (2014) *I Hope Your Day Is Legendairy Meme* [Image via IMGUR] Available at: https://imgur.com/gallery/F958y7a. [Accessed: January 2019]

Sonarworks (2019) *Sennheiser HD25II Headphone review* [Online]. Available at: https://www.sonarworks.com/soundid-reference/blog/reviews/sennheiser-hd25-ii-studio-headphone-review/#pros [Accessed: 17th November 2020]

Thornton, M. (2014) *Working In Immersive Audio with the Auro 3D Format in Music Recording & Production* [Online Video]. Available at: http://www.pro-tools-expert.com/home-page/2014/10/5/working-in-immersive-audio-with-the-auro-3d-format-in-music.html [Accessed: 18th November 2015]

Waves Audio (2021) *Direction dependent frequency change* [Image]. Available at: https://www.instagram.com/p/CSZe554N4D4/?utm_medium=copy_link [Accessed: 10th August 2021]

Zagorski-Thomas, S. (2021) *Approaches to Publishing Practice Research* [Video]. Available at: https://www.youtube.com/watch?v=EsCyxGJLO4Q&list=PLcNNsyIDf166mNKqRY1hExgM4MTPi_e1x&index=3 [Accessed: 31st August 2021]

Bibliography

4DSound (2019) *On Spatial Soun*d [Video]. Available at: https://vimeo.com/340861350?fbclid=IwAR1T_MefEpkHr8mwaVcNSlXJODc1J4N_wpok1xWzXJlXTaPczmaOHonHQIs [Accessed: July 2019]

Alba, A. (2017) Why 2017 could be the year of virtual reality. *The Week* [Online]. Available at: http://theweek.com/articles/671560/why-2017-could-year-virtual-reality [Accessed: 13th January 2017]

Alvarez, A. (2004*) Memetics: An evolutional theory of cultural transmission* [Online]. Available at: http://www.sorites.org/Issue_15/alvarez.htm [Accessed: March 2018]

Amatriain, X., Castellanos, J., Hollerer, T., Kuchera-Morin, JA., Pope, S., Wakefield, G. and Woolcott, W. (2007) Experiencing audio and music in a fully immsersive environment, in Kronland-Martinet, Ystad, S. and Jensen, K. (2008) *Computer Music Modeling and Retrieval. Sense of Sounds.* International Symposium, CMMR 2007, Copenhagen, Denmark, August 2007 [Online].Available at: https://books.google.co.uk/books?id=3XFqCQAAQBAJ&source=gbs_navlinks_s [Accessed: December 2017]

Anthony, B., Thompson, P. and Auvinen, T. (2019) Learning to be a 'tracker': A pedagogical case study of learning collaborative music production [Presentation] In: *The 14th Art of Record Production Conference*, 17th-19th May 2019, Boston, Massachusetts, USA.

Ashley, R. and Timmers, R. (Ed.) (2017) *The Routledge Companion to Music Cognition*. Routledge: Abingdon, Oxon, England.

Auro Technologies (n.db) *Press – Music* [Online]. Available at: https://www.auro-3d.com/press/category/music/ [Accessed: January 2018]

Barco (n.d) *Auro 11 or Object-based sound in 3D* [Online]. Available at: https://www.barco.com/secureddownloads/cd/MarketingKits/3d-sound/White%20papers/Auro%2011.1_versus_objectbased_sound_in_3D.pdf [Accessed: 2018].

BBC (2011) *Periphony for Broadcast* [Online]. Available at: https://www.bbc.co.uk/rd/projects/periphony-for-broadcast [Accessed: July 2015]

BBC (2019) *Binaural Broadcasting Research and Development* [Online]. Available at: https://www.bbc.co.uk/rd/projects/binaural-broadcasting [Accessed: July 2015]

Beard, D. and Gloag, K. (2005) *Musicology: The key concepts*. Great Britain: Routledge

Begault, D. R. (Ed.) (2004) *Spatial Sound Techniques- Part 1: Virtual and Binaural Audio Technologies.* New York, USA: Audio Engineering Society.

Born, G. (Ed.) (2013) *Music, Sound and Space. Cambridge*, Great Britain: Cambridge University Press

CIRT (n.d) *Phenomenology Research Overview* [Online]. Available at: https://cirt.gcu.edu/research/developmentresources/research_ready/phenomenology/phen_overview [Accessed: April 2018]

Clarke, E. [n.d] Music, Space and Subjectivity, in Born, G. (Ed.) (2013) *Music, Sound and Space*. Cambridge, Great Britain: Cambridge University Press

Clarke, E. and Cook, N. (2004) *Empirical Musicology: Aims, Methods, Prospects*. Oxford, Great Britain: Oxford University Press

Coleman, P. (2017) Recording the BBC Philharmonic Orchestra in 3D. *S3A Wordpress* [Online]. Available at: http://www.s3a-spatialaudio.org/wordpress/?p=518 [Accessed: March 2017]

Complicite (2019) *The Encounter* [Online]. Available at: http://www.complicite.org/productions/theencounter [Accessed: 21st July 2019].

Csikszentmihalyi, M. (1999) A Systems Perspective on Creativity, in Sternberg, R. (Ed) (1999) *Handbook of Creativity*, 313–35. Cambridge, England: Cambridge University Press

Dash, N. K., (2005) *Selection of the Research Paradigm and Methodology* [Online]. Available at: http://www.celt.mmu.ac.uk/researchmethods/Modules/Selection_of_methodology/ [Accessed: March 2017]

Deliege, I. and Sloboda, J. (Eds.) (1997) *Perception and Cognition of Music*. Psychology Press: East Sussex, England, UK.

Deutsh, D. (2013) *The Psychology of Music*. 3rd Rev. Ed. Academic Press: London, UK

EleVR (2014) *Audio for VR Film: Binaural, Ambisonic, 3D etc*. [Online]. Available at: http://elevr.com/audio-for-vr-film/ [Accessed: October 2016]

Emmerson, S. (2014) *Listening in time and over time – the construction of the electroacoustic music experience* [Downloaded PDF]. Proceedings of the Electroacoustic Music Studies Network Conference, Electroacoustic Music Beyond Performance. Berlin, June 2014.

EPSRC (2013) *S3A: Future Spatial Audio for an Immersive Listener Experience at Home* [Online]. Available at: http://gow.epsrc.ac.uk/NGBOViewGrant.aspx?GrantRef=EP/L000539/1 [Accessed: 21st November 2015]

Ernould, F. (2003) *Beyond the stereo sound Part 2 - Never 2 speakers without the 3rd dimension...* [Online]. Available at: http://www.macmusic.org/articles/view.php/lang/en/id/13/Beyond-the-stereo-sound-Part-2-Never-2-speakers-without-the-3rd-dimension [Accessed: 18th March 2016]

Facebook (2016) *Event Page: Odyssia Festival Launch Party: Francois K in Dolby Atmos* (6 hour set) [Online]. Available at: https://www.facebook.com/events/624207244415301/ [Accessed: June 2016]

Finklestein, R. (2008) *A Memetics Compendium* [Downloaded PDF].

Forsythe, G. (2018) *Audio Mixing for VR: The Beginners Guide to Spatial Audio 3D sound and Ambisonics* [Online]. Available at: https://sonicscoop.com/2018/02/05/audio-mixing-for-vr-the-beginners-guide-to-spatial-audio-3d-sound-and-ambisonics/ [Accessed: May 2018]

Fraser, L. (n.d) *Frequency response adaptation in binaural hearing* [Online]. Available at: https://slideplayer.com/slide/676877/ [Accessed: June 2016]

Henry, M. J. and Grahn, J. A. (2017) Music, Brain and Movement; Time, Beat and Rhythm, in Ashley, R. and Timmers, R. (Ed.) (2017) *The Routledge Companion to Music Cognition*. Routledge: Abingdon, Oxon, England.

HyperRadio (n.d) *Son 3D* [Online]. Available at: http://hyperradio.radiofrance.fr/son-3d/ [Accessed: July 2015]

IFPI (2019) *Global Music Report 2019* [Online]. Available at: https://ifpi.org/news/IFPI-GLOBAL-MUSIC-REPORT-2019 [Accessed: 29th April 2019]

Jan, S. (2016) *The Memetics of Music* [Kindle Edition]. Routledge: Abingdon, Oxon, UK.

Kirn, P. (2018) Mics that record in 3D and Ambisonics are the next big thing [Online], in *CDM, September 19th 2018*. Available at: https://cdm.link/2018/09/3d-ambisonic-microphones/ [Accessed: January 2019]

Lee, H., Gribben, C. and Wallis, R. (2014) Psychoacoustic Considerations in Surround Sound with Height, in *28th Tonmeistertagung: tmt 28*, 20th-23rd November 2014, Cologne, Germany

Linxia, C. and Ziran, H. (2006) *Analysis of memes I language* [Downloaded PDF]. Available at: http://en.cnki.com.cn/Article_en/CJFDTotal-WJYY200602004.htm [Accessed: March 2019]

Lipscomb, S. and Kendall, R. (1994) Perceptual judgment of the relationship between musical and visual components in film [Online]. *Psychomusicology, 13*(1), 60-98. Available at:

http://www.lipscomb.umn.edu/docs/LipscombKendall1994.pdf [Accessed: 11th April 2016]

Maes, P-J., Leman, M., Palmer, C. and Wanderley, M. M. (2014) *Action based effects on music perception* [Online]. Available at: https://www.frontiersin.org/articles/10.3389/fpsyg.2013.01008/full [Accessed: June 2016]

Moore, A. F. (2012) S*ong Means: Analysing and Interpreting Recorded Popular Song*. Routledge: Abingdon, Oxon, England, UK.

Moore, D. and King, A. J. (1999) *Auditory perception: the near and far of sound localization* [Online]. Available at: https://www.sciencedirect.com/science/article/pii/S0960982299802279 [Accessed: January 2017]

Moylan, W. (2015) *Understanding and crafting the mix* [Online]. Available at: https://www-dawsonera-com.ezproxy.uwl.ac.uk/readonline/9780203758410 [Accessed: February 2016]

Moylan, W. (n.d) *Considering Space In Recorded Music*. [Downloaded PDF], in Ashgate Popular and Folk Music Series: Art of Record Production: An introductory reader for a new academic field. Ashgate Publishing Group.

Palenchar, J. (2015) *Object-based surround sound coming to ATSC 3.0 standard* [Online]. Available at: https://search-proquest-com.ezproxy.uwl.ac.uk/docview/1667197761?pq-origsite=summon [Accessed: July 2015]

Parise, C., Knorre, K. and Ernst, M. (2013) *Natural Auditory Scene Statistics Shapes Human Spatial Hearing* [Online].Available at: https://www.pnas.org/content/pnas/111/16/6104.full.pdf [Accessed: September 2017]

Pike, A, (1967) *The Phenomenological Analysis and Description of Music* [Online]. Available at: https://www.jstor.org/stable/3343947?read-now=1&seq=1#page_scan_tab_contents [Accessed: Jan 2017]

Research and Markets (2019) *Earphones and Headphones Market – Global Outlook and Forecast 2019-2024* [Online]. Available at: https://www.researchandmarkets.com/research/xcdtqh/global_earphones?w=12 [Accessed: 16th June 2019]

Researchmethodology.net (2019) *Action-research* [Online]. Available at: https://research-methodology.net/research-methods/action-research/ [Accessed: January 2019]

RIAA (2019) *U.S. Sales Data Base* [Online]. Available at: https://www.riaa.com/u-s-sales-database/ [Accessed: 30th January 2019]

Robjohns, H. (2017) Dolby Atmos at Ministry of Sound. *Sound On Sound*. January 2017 [Online]. Available at:

https://www.soundonsound.com/techniques/dolby-atmos-ministry-sound [Accessed: January 2017]

Rumsey, F. (2006) *Spatial audio and sensory evaluation techniques – context, history and aims* [Online] Available at: http://epubs.surrey.ac.uk/530/1/fulltext.pdf [Accessed: 29th October 2013]

Rumsey, F. (Ed.) (2004) *AES: An Anthology of Spatial Sound Techniques Part 2: Multichannel Audio Technologies.* New York, USA: AES.

Sag, A. (2018) Virtual Reality in 2017: a year in review. *Forbes* [Online]. Available at: https://www.forbes.com/sites/moorinsights/2018/01/02/virtual-reality-in-2017-a-year-in-review/#3f1d33e835f4 [Accessed: 10th January 2018]

Sansom, M. (2005) Understanding Musical Meaning: Interpretive Phenomenological Analysis and Improvisation, in *British Forum for Ethnomusicology, 2005 Annual Conference – Music and Dance Performance: Cross-Cultural Approaches*. April, 2005. SOAS, London, UK [Downloaded PDF].

Savage, M. (2019) *Is this the end of owning music?* [Online]. Available at: https://www.bbc.co.uk/news/entertainment-arts-46735093 [Accessed: 15th July 2019]

Savov, V. (2016) Headphones are growing more expensive because we demand more of them. *The Verge.* 13th May 2016 [Online]. Available at: https://www.theverge.com/circuitbreaker/2016/5/13/11669906/headphones-market-price-worldwide-statistics [Accessed: June 2016]

Schachter, D. (n.d*) Acousmatic discourse and sound projection under the new multi-channel surround formats. Past, current, future* [Online]. Available at: http://www.ems-network.org/ems09/papers/schachter.pdf [Accessed: 2017]

Seamon, D. (n.d) *Interconnections, Relationships, and Environmental Wholes: a phenomenological ecology of natural and built worlds* [Online]. Available at: https://core.ac.uk/download/pdf/5165696.pdf [Accessed: January 2016]

Smalley, D. (2007) *Space-form and the acousmatic image.* Organised Sound, 12 (1), 35–58. Cambridge University Press. Printed in the United Kingdom

SSE Audio Group (2019) *L-isa's immersive experience with Mark Knopfler* [Online]. Available at: https://www.sseaudiogroup.com/Group-News/l-isas-immersive-experience-with-mark-knopfler?fbclid=IwAR0Of6rSz2atAjQ6TplBFWYAg7raURXfJ5aCLHgaK9D8M2ajq7hnQOaMxQk [Accessed: 30th July 2019]

Stanford.edu (2011) *Embodied Cognition* [Online]. Available at: https://plato.stanford.edu/entries/embodied-cognition/ [Accessed: July 2016]

Stanford.edu (2015) *Action based theories of perception* [Online]. Available at: https://plato.stanford.edu/entries/action-perception/ [Accessed: July 2016]

Statista (2019) *Headphones Market – United Kingdom* [Online]. Available at: https://www.statista.com/outlook/15010600/156/headphones/united-kingdom?currency=gbp [Accessed: February 2019]

Statista (2019) *What do you use your headphones for?* [Online]. Available at: https://www.statista.com/statistics/696862/uses-of-headphones-in-the-us/ [Accessed: June 2019]

Thornton, M. (2017) *Developing a 3D audio workflow for virtual reality in protools* [Online]. Available at: https://www.pro-tools-expert.com/home-page/2017/3/16/developing-a-3d-audio-workflow-for-virtual-reality-in-pro-tools#gsc.tab=0 [Accessed: April 2017]

Tribeca Film (2016) *Notes on Blindness: Into darkness* [Online]. Available at: https://tribecafilm.com/filmguide/archive/notes-on-blindness-into-darkness-2016 [Accessed: January 2017]

Walker, M. (2007) *Stereo image trickery with plug ins* [Online]. Available at: http://www.soundonsound.com/sos/jan07/articles/pcmusician_0107.htm [15th April 2016]

Wall, S. (2008) *Easier Said Than Done: Writing An Autoethnography* [Downloaded PDF]. Available at: https://journals.sagepub.com/doi/pdf/10.1177/160940690800700103 [Accessed: 17th July 2019]

Watercutter, A. (2013) Navigating Beck's performance of 'Sound and Vison' as a 360 degree video. *Wired*. 19th February 2013 [Online]. Available at: https://www.wired.com/2013/02/beck-360-degree-online-video/ [Accessed: August 2015]

Waves (2019) *Mixing on headphones: How to translate to speakers* [Online]. Available at: https://www.waves.com/mixing-on-headphones-how-to-translate-to-speakers?utm_source=wnletter&utm_medium=email&utm_content=inside-808-mixing-tips-step-inside-btn&utm_campaign=weekend-content-mixing-on-headphones-808-day-august-10 [Accessed: 10th August 2019]

WhispersRed (2017) *ASMR Happens: Live, 2017* [Online]. Available at: http://whispersredasmr.com/live-asmr-london-england/ [Accessed: May 2017]

Wishart, T. and Emmerson, S. (Ed.) (1996) *On Sonic Art*. Hardwood Academic Publishers: Netherlands.

World Atlas (2017) *Largest Music Industries in the World* [Online]. Available at: https://www.worldatlas.com/articles/largest-music-industries-in-the-world.html [Accessed: November 2017]

Yahoo Finance (2019) *Global Earphones and Headphones Report 2019-2024* [Online]. Available at: https://uk.finance.yahoo.com/news/global-earphones-headphones-market-report160000425.html?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAAFBRXVwyG9VI7rkHNLOTO8SO983JO48pSGTqamZfygiiFoIP9ebgzKYxk8iUj1NBe4R-DIhj8u1YtydkpaOaAQqZOShY7T2IyVeIABZTfEPYpg4yS5zJQtIUr0xhaT3cGgV3S7ZOTh8-AvNeEFYRDAfCBt9inzafNSi91tiW1bC [Accessed: 16th June 2019]

Appendices

Appendix A – Accompanying Audio Submission

Scan me

Please scan the QR code or click the link below to access the spatial audio playlist pertaining to this research project. This playlist includes all submitted periphonic productions that were recomposed using the Dear VR Pro.

Soundcloud Audio Playlist: https://soundcloud.com/dalis-deathmask/sets/the-art-of-periphonic-record-production?si=3998f3769ade4435a3df1cb729e76d0d

Accompanying Audio Submission Comprising:

10 Original Stereo Versions

10 Periphonic Audio Mixes (dear VR binaural)

10 PeriStereo Versions (dear VR speaker-based)

2 Auro x KU100 reamped

2 Auro 3D versions Upmixed and Decoded with Blue Ripple O3A (TOA > Stereo)

2 Auro 3D versions Upmixed and Decoded with Blue Ripple O3A (TOA > binaural)

3 Decoded with Kinicho Binaural Decoding

Total – c.39 tracks totalling approx. 170 minutes of audio

Sound Cloud Playlist - Track Listing

Hidden Behind Static

Piece 1 – ‘Penny Drops’ Available at: https://bit.ly/2ZqYfol

Piece 2 – ‘Monomorphic’ Available at: https://bit.ly/37sEhct

Piece 3 – ‘Far From Here’ Available at: https://bit.ly/3fugtI5

Piece 4 - ‘Dark Heart’ – (Not used in trial)

Dali’s Deathmask

Piece 5 - ‘Drowned in Sound’ Available at: https://bit.ly/3jdPJzy

Piece 6 - ‘Art’s Self Alteration’ Available at: https://bit.ly/3pcSe9n

Jimmy Logic

Piece 7 - ‘Sega’ – (Not used in trial)

Jerome Thomas

Piece 8 - ‘Late Nights’ Available at: https://bit.ly/3e1okfV

Joey Clarkson

Piece 9 - ‘Sort Yourself Out’ [Live] Available at: https://bit.ly/30EX6HQ

Beautiful Thing

Piece 10 – ‘Waiting’ [Live] Available at: https://bit.ly/2MZpcWu

Appendix B – Production & Technical Notes

Summer 2016 – Mosi Conde Recording Sessions with Mark Brocklesby and David Padilla

Sonic Ref: Dead Can Dance

Lower Array Comprising – 5.1 Array Adaptation

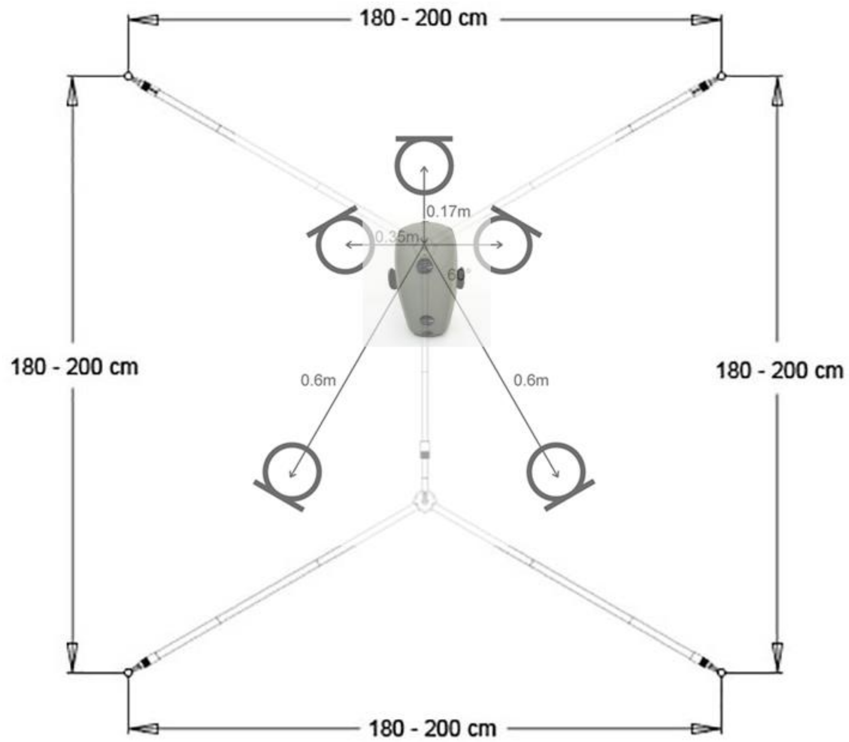
1. Williams Star x INA5 Array – C451B x 5
2. KU100 Binaural Dummy Head

Height Layer – Quad Array

3. Hamasaki Quad – Omni/QTC40 x 4

Overhead Layer – Ambisonics or Mono

4. Sound Field / Brauner – Ceiling Mic



Appendix C – Evaluation Questionnaire Templates

Contributor Interview

1. How old are you?

2. How would you describe your musical background?
   (Formal training (classical/university/conservatoire), informal tuition (family or friend influence/, self-taught, learnt an instrument at school, no musical background)

3. Do you listen to music? Active rather than passive

4. How do you usually listen to music? (Speakers, TV speakers, home theatre system / surround sound, laptop, car, personal radio, headphones, phone etc.)

5. What types/styles of music do you listen to most?

6. Do you create/record/produce music?

7. What type/style of music do you usually create/work with?

8. Are you familiar with the term ‘Stereo’? If yes, describe what it means to you.

9. Are you familiar with the term ‘Periphonic’? If yes, describe what it means to you.

10. Are you familiar with the term ‘Binaural’? If yes, describe what it means to you.

11. How would you explain the sonic space / environment created through the periphonic production?

12. What were the concepts, thoughts or influence behind the pieces submitted for recomposition?

13. Do you think they were represented well in the stereo production?

14. Do you think they were represented well in the periphonic production?

15. Which do you think had the best aesthetic?

16. Has your approach toward composition (theoretical/practical) been affected, changed or evolved since having experienced your work recomposed as periphonic?
How?

17. Do you feel your music has benefit from the periphonic production?
How?

18. Do you feel the periphonic production enhanced and expressed the music and concepts beyond that of the stereo version?

19. Do you see periphonic production as something you would continue working with?

20. Do you see a future in this method of production for your music?

21. Do you see a future in this method of production in the music industry?

Consumer Interview

1. How old are you?

2. How would you describe your musical experience?
   (Formal /, informal tuition, self-taught, learnt a un/graded instrument at school, no musical background etc.)

3. Do you listen to music?

4. How do you usually listen to music? (Speakers, TV speakers, home theatre system / surround sound, laptop, car, personal radio, headphones, phone etc.)

5. What types/styles of music do you listen to most?

6. Do you create/record/produce music?

7. What type/style of music do you usually create/work with?

8. Are you familiar with the term ‘Stereo’? If yes, describe what it means to you.

9. Are you familiar with the term ‘Periphonic’? If yes, describe what it means to you.

10. Are you familiar with the term ‘Binaural’? If yes, describe what it means to you.

Listening Questionnaire

(Please answer separately for Tracks A and B where specified)

1. What headphones are you listening on?

2. How many times did you listen to each track?

3. Which did you prefer, A or B, and why?

4. What does the record as a whole invoke for you?

5. How would you explain the sonic space / environment?

   Track A –

   Track B-

6. Describe the placement of sounds you hear in each version.

   Track A –

   Track B-

7. Did you perceive any sounds to be moving in either track? Please describe.

   Track A –

   Track B-

8. Did you perceive any sounds from above? Please describe.

   Track A-

Track B-

9. Did you perceive any sounds from behind? Please describe.
Track A-

Track B-

10. Did you perceive any sound coming directly from either side of you Left and or Right?
Please describe.

Track A-

Track B -

11. Did you perceive the sound to always be coming from in front of you? Please describe.

Track A -

Track B-

12. Do you imagine or visualise what you hear when listening to the music?

Track A-

Track B-

13. Is there anything you particularly liked/disliked about the production?

Track A-

Track B-

14. Do you feel immersed in the music?

Track A-

Track B-

15. Did you experience the music as if it was coming from all around you?

Track A -

Track B -

16. Did you feel as if the sound was coming from inside or outside of your head?

Track A-

Track B-

17. Were the tracks comfortable or uncomfortable to listen to? Why?

Track A-

Track B-

18. Which of the versions would you rather listen to, A or B?

19. Would you buy music in either of these formats?

Track A -

Track B-

20. Do you think production suits the type of music?

Track A -

Track B -

21. Is there anything else you would like to add about the music, mixes or your experience?

Appendix D – Interviews and Communications Transcripts

Appendix D1 - Email conversation with Auro Technologies RE: Binaural Rendering Tool

Query sent 24th May 2016
“Hello,
I am just emailing in request for some clarification on Auro over headphones. I'm a PhD student at London College of Music researching 3D record production. Your headphone format will feature as an important part of my production project and it’s fundamental to my practical work on the Auro system. A colleague of mine today however told me, quite certainly, that Auro is/has dropped this format.
Is this the case? As I cannot find any information to support this statement and I notice
The Auro website still advertises that format feature. I look forward to hearing from you and clearing the confusion.”

Response received 1st June 2016

“Dear J. Lord,
Thank you for contacting our support department regarding your questions.
In the Creative Tool Suite it was a business decision to remove the Auro Binaural Renderer
(AuroHP Plugin) considering it conflicts with other products in our portfolio. If you have any questions regarding this matter, please feel free to contact us.
Kind regards.”

Appendix D2 - Andy Caldecott Contributor Interview

1. How would you explain the sonic space / environment afforded through periphonic production?

Enhances the space, gives it a more defined and clearer audible space. New dimensions. The sounds aren't so restricted. They can move in more complex ways. Stereo could be considered to be a very static placement of sound and this opens up new ways of listening to music, it breaks tradition.

2. Has your approach toward composition (theoretical/practical) been affected, changed or evolved since integrating a periphonic production approach?

It has inspired me to look at how sounds fit, fill a space more. How composition of track can be altered. I haven't worked with the software, I've only worked in a passive way thinking about how or what sounds could be used, what sounds could provide movement and integrating this with traditional instruments and synthesisers.

3. Do you feel your music has benefit from the periphonic production?

Yeah. It has. It's definitely opened it up. Given it more spread within a sound field.

4. Do you feel the musical concepts of your compositions were appropriately conveyed through the periphonic production?

Yeah I believe they were, because they were designed in my mind to be used in that way. The outcome combined what I had made and took it to another level. So without using the software I had to just sit and imagine what it would be like to move the sounds around in a bigger field than just stereo. Thinking about what would be obvious to me as a sound effect that could be manipulated into a 3D field and combined with other instruments to make it more of a musical piece.

5. Do you see periphonic production as something you would continue working with?

Yes, I think it would be something that I would like to progress with, when the software is available to me I would definitely like to move onto that and maybe leave stereo behind.

6. Do you see a future in this method of production for your music?

I could see it as a future method of production for my music, but it could go the same way surround sound has. Enhancing it and giving more freedom and space for compositions but then again. Really depends on the market.

7. Do you see a future in this method of production in the music industry?

It could be, if it becomes a practical means of listening to music. Don’t need to buy a big system, if it is something accessible to everybody through headphones.

Appendix E – Focus Session Listening Trials: Raw Data

Listener 1

```
HULTIGEN DATA OUTPUT / TRIAL ORDER / AB ORDER

1 - A 75 "Art's Self Alteration VR.wav" B 100 "Art's Self
Alteration Redbook.wav";
2 - A 100 "Monomorphic VR.wav" B 60 Monomorphic.wav;
3 - A 50 "Sort Yourself Out Mstr1.wav" B 100 "Joey Clarkson –
Sort Yourself Out (Live at Brentford Steam Works) [Periphonic
Mix].wav";
4 - A 100 "Beautiful Thing - Waiting (Live at London Water &
Steam Museum) [Periphonic Mix].wav" B 75 "Beautiful Thing -
Waiting (Live at London Water & Steam Museum) [unmastered].wav";
5 - A 75 "Far From Here.wav"; B 100 "Far From Here VR.wav"
6 - A 51 "Penny Drops.wav" B 100 "Penny Drops VRv2.wav";

7 - HULTIGEN PROBLEM WITH PLAYBACK – TRIAL SKIPPED
"Drowned In Sound.wav" "Drowned In Sound VR.wav";
```

Raw Focus Session Questionnaire Data

Consumer Interview

1. How old are you? 42

2. How would you describe your musical background?
   (Formal training (classical/university/conservatoire), informal tuition (family or friend influence/, self-taught, no musical background)

   Mostly self-taught. Private guitar tuition, secondary school.

3. Do you / did you play an instrument? Yes

4. Do you listen to music? yes

5. How do you usually listen to music? (speakers, TV speakers, home theatre system / surround sound, laptop, car, personal radio, headphones, phone etc.)

Headphones, Laptop, Phone.

6. Do you create/record/produce music?

Yes

7. Are you familiar with the term ‘Stereo’? If yes, describe what it means to you.

Yes, split space left and right, not just centred.

8. Are you familiar with the term ‘Periphonic’? If yes, describe what it means to you.

No

1 **- A 3D 75 and B ST 100:**
(Reversed order in HULTIGEN A-3D B-ST not A ST-ASA- B 3D)

1. What headphones are you listening on? K271

2. How many times did you listen to each track? 1

3. Which did you prefer? B (see hultigen data TEST 0)

4. What does the record as a whole evoke for you?
Soundscape

5. How would you explain the sonic space / environment in A/B?

A- A lot of the sound seems to be coming from one place, the centre, with a few sounds in stereo effect.

B- All the sounds aside from bass, seems to spread around and sound like they create a space around you with distance. Depth.

6. Describe and/or draw the placement of sounds in each version:

A – A is more like a pipe

B- is more like a ball

7. Did you perceive any sounds to be moving in either track? Please describe.

There are in both to an extent but definitely more in B. B is more spatial things move around you and spread across.

8. Did you perceive any sounds from above you in either track?

Bomb moving from the left to the right, it made it sound as if the bomb goes off on the L and moves over you in arc to the right – Track B.

9. Did you perceive any sounds from behind you in either track?

No. Not at all. From L to front to R but not the back.

10. Did you perceive any sound coming directly from either side of you Left and or Right in either track?

A definitely had stereo effect on it, yes I could feel sounds coming from the sides. B felt that sounds were coming very clearly from the left and right, top left and right there was height. It was more enhanced.

11. Did you perceive the sound to always be coming from in front of you in A or B?

A always kind of feel like it was always from the front. B always from around you.

12. When listening to track A/B. Do you imagine or visualise what you hear?

Yes. A not so much. It felt like a soundscape that didn't lend itself to any space. Whereas B made it feel like you were in the space, in a warzone. It engulfed you more.

13. Is there anything you particularly liked/disliked about the production in A/B?
Didn't really like or dislike anything but I would prefer to listen to B it sounds more interesting.

14. Do you feel immersed in the music in A/B?
More so in B.

15. Did you experience the music as if it was coming from all around you?
In B, more so yes.

16. Did you feel as if the sound was coming from inside or outside of your head for either track?
A felt inside whereas B felt more outside.

17. Were the tracks comfortable or uncomfortable to listen to (A/B)? Why? (Explain if you wish).
A uncomfortable as it was louder but there was a quality to it that was different. A sounded slightly bit crushed, distorted, whereas B sounded more smooth.

18. Is there anything else you would like to add?

2 - **Listening Questionnaire (A 3D – MM- B ST / 100A and B60):**

1. What headphones are you listening on? K271

2. How many times did you listen to each track?

3. Which did you prefer (see hultigen data TRIAL1)? A

4. What does record as a whole evoke for you?

Dreamy shoegazey kind of track.

5. How would you explain the sonic space / environment in A/B?

A has a sphere to it whereas B was a horizontal linear bar where the top and bottom are cut away

6. Describe the placement of sounds in each version:

A – Sound from left right above and below.

B- Pretty much middle, left and right.

7. Did you perceive any sounds to be moving in either track? Please describe.

In A you could hear the guitar was moving. Started slightly panned out and then moved even further out. Percussive sounds were moving, jumping around from left to right.

Subtle movement left right.

8. Did you perceive any sounds from above you in either track?

9. Did you perceive any sounds from behind you in either track?

Still not from behind me

10. Did you perceive any sound coming directly from either side of you Left and or Right in A/B?

Most definitely in A yes.

11. Did you perceive the sound to always be coming from in front of you (A/B)?

B is very static, mostly in front. Some sounds in A were static but mostly around you.

12. When listening to track A/B. Do you imagine or visualise what you hear?

More so in A. You could be in a theatre with singers on podiums and machinery moving in front of you, felt like the percussion. Guitarists moving from left to right.

13. Is there anything you particularly liked/disliked about the production in A/B?

I really liked the production in A. it gave it more space and made it more interesting. B quite static.

14. Do you feel immersed in the music in A/B?

Definitely more so in A.

15. Did you experience the music as if it was coming from all around you?

In A I did, not so much from behind but definitely above below left and right.

16. Did you feel as if the sound was coming from inside or outside of your head?

Outside in A. Created an environment outside you head. B didn't feel like this as dramatically as A.

17. Were the tracks comfortable or uncomfortable to listen to (A/B)? Why? (Explain if you wish).

I didn't find either uncomfortable to listen to.

18. Is there anything else you would like to add?

3 - **Listening Questionnaire (A ST – SYO – B VR both 50A and 100B):**

1. What headphones are you listening on? K241

2. How many times did you listen to each track? 1

3. Which did you prefer (see hultigen data Trial 2)? B

4. What does record as a whole evoke for you?
Sounds like you're an audience to a band in a room.

5. How would you explain the sonic space / environment in A/B?

There are stereo sounds in A. the Trumpet on right Violin on left. Aspect of the drums but B is definitely more spread out and you can hear everything more clearly because of it.

6. Describe the placement of sounds in each version:

A – Vocals in the middle, trumpet goes from middle to right, violin on the left. Bass and drums in the middle.

B- Vocals move from middle to left and right.

7. Did you perceive any sounds to be moving in either track? Please describe.

Only in B, the vocals.

8. Did you perceive any sounds from above you in A/B?

Not really not in either.

9. Did you perceive any sounds from behind you in A/B?

Not in either of them.

10. Did you perceive any sound coming directly from either side of you Left and or Right in A/B?

In B, her vocals were coming from left and right.

11. Did you perceive the sound to always be coming from in front of you (A/B)?

In A yes it was like a wall of sound. B was in front of you but more spread out. It had more space to it. Like you were in the room watching them.

12. When listening to track A/B. Do you imagine or visualise what you hear?

Only in B does it sound like you're actually sat in the room with them.

13. Is there anything you particularly liked/disliked about the production in A/B?

Didn't like the production in A. It sounded too squashed into a mush of sound whereas B really spread everything out and you could hear more clearly.

14. Do you feel immersed in the music in A/B?

Not at all in A. B felt like you were sat in a chair with a band playing at you.

15. Did you experience the music as if it was coming from all around you?

No, you were sat as the audience with them in front. It was and oval of sound.

16. Did you feel as if the sound was coming from inside or outside of your head?

B outside. A just felt like inside your head in your face. Close.

17. Were the tracks comfortable or uncomfortable to listen to (A/B)? Why? (Explain if you wish).

I found A uncomfortable because of the drums they were loud but squashed and in your face.

18. Is there anything else you would like to add?

3-Listening Questionnaire (A 3D – BT – B ST both 100A and 75B):

1. What headphones are you listening on? K271

2. How many times did you listen to each track? 1

3. Which did you prefer?

A

4. What does record evoke for you?

Similar to the last, it felt like you were watching a band. Part of an audience.

5. How would you explain the sonic space / environment in A/B?

A was almost like you were in space and the band were floating around you. Whereas B was more grounded and very much in front of you.

6. Describe the placement of sounds in each version:

A – In A there is an organ on right, guitar left and right, violin on left. Bass mid below left. Vocals in front and sort of above around.

B- B felt like everything was in front of you.

7. Did you perceive any sounds to be moving in either track? Please describe.

I wouldn't say they were moving but they had a place that spread around you.

B was very narrow and in front of you.

8. Did you perceive any sounds from above you in A/B?

Vocals in A – slightly above lead vocal during the adlib fill seems high but for verses generally he in front, for both. No height in B.

9. Did you perceive any sounds from behind you in A/B?

No.

10. Did you perceive any sound coming directly from either side of you Left and or Right in A/B?

In A yes, guitars left right, violin left, and organ right.

11. Did you perceive the sound to always be coming from in front of you (A/B)?

In B yes.

12. When listening to track A/B. Do you imagine or visualise what you hear?

In A it felt like I was in a void in space with the band floating in front of me. In B felt like standing in a narrow hall with the band in front.

13. Is there anything you particularly liked/disliked about the production in A/B?

I wouldn't say I dislike it [B] but I preferred A.

14. Do you feel immersed in the music in A/B?

In A yes.

15. Did you experience the music as if it was coming from all around you?

In A it felt like that yes.

16. Did you feel as if the sound was coming from inside or outside of your head?

In A outside, slightly outside in B but not a big spread.

17. Were the tracks comfortable or uncomfortable to listen to (A/B)? Why? (Explain if you wish).

I found both quite comfortable to listen to.

18. Is there anything else you would like to add?

They're a bad ass band, I really like them.

4- **Listening Questionnaire (A ST- FFH – B 3D / 75A and B100):**

1. What headphones are you listening on?
   K271

2. How many times did you listen to each track?
   1

3. Which did you prefer (see hultigen trial 4 data)? B

4. What does record invoke for you?

It's a hard one to picture in my mind. Something futuristic. No particular setting it just puts the listener in a lonely place.

5. How would you explain the sonic space / environment in A/B?

A elements of stereo. Moving from the centre outwards, vocals are also stereo. B a dramatic difference. There is more movement, especially with the vocals. They are disappearing into the horizon and then coming very close – more depth. Slightly above. Music fills up the whole space, percussion quite central.

6. Describe the placement of sounds in each version:

See above.

7. Did you perceive any sounds to be moving in either track? Please describe.

Yes, more so in B. Voices moving around, being dragged away, up close.

8. Did you perceive any sounds from above you in A/B?

Female vocals left and right perceived as being positioned higher, with the male vocal in the front further away but even higher up than those two female vocals.

9. Did you perceive any sounds from behind you in A/B?

No

10. Did you perceive any sound coming directly from either side of you Left and or Right in A/B?

Slightly for A but predominantly in B.

11. Did you perceive the sound to always be coming from in front of you (A/B)?

In A yes, to some degree, B no it was moving. Focal point was moving.

12. When listening to track A/B. Do you imagine or visualise what you hear?

Not in this track, it was quite cold. There was no visual reference to the sound.

13. Is there anything you particularly liked/disliked about the production in A/B?

No dislikes, just preferred B.

14. Do you feel immersed in the music in A/B?

Definitely in B.

15. Did you experience the music as if it was coming from all around you?

To a certain degree in B.

16. Did you feel as if the sound was coming from inside or outside of your head?

Outside of the head in both.

17. Were the tracks comfortable or uncomfortable to listen to (A/B)? Why? (Explain if you wish).

Yeah both comfortable to listen to. Non-intrusive, nothing harsh. Both mellow.

18. Is there anything else you would like to add?

5- **Listening Questionnaire (A ST - PD – B 3D both 51A and B100):**

1. What headphones are you listening on? K271

2. How many times did you listen to each track? 1

3. Which did you prefer? B

4. What does record invoke for you?
   Sounds like the listener in inside a machine, inside something mechanical.

5. How would you explain the sonic space / environment in A/B?

In A it sounds like the sounds are coming from a ball shape in front of you emanating outwards whereas B sounds like you are inside a ball.

6. Describe the placement of sounds in each version:

A – Slight stereo but generally quite central. Certain percussive clicky sounds are left and right.

B- In B the sounds are coming from all over, above, left right. Presence (bass etc.) from below perhaps.

7. Did you perceive any sounds to be moving in either track? Please describe.

In B it is very apparent, not so much in A

8. Did you perceive any sounds from above you in A/B?

In A not so much, in B yes. HH are top of my head, centre left and right above my head. Vocal seemed further out but still elevated.

9. Did you perceive any sounds from behind you in A/B?

In this one yes, in B. Mechanical clank that went around from right to left, it gave the sensation it was behind you and then catching into your left ear

10. Did you perceive any sound coming directly from either side of you Left and or Right in A/B?

B yes. Mechanical sounds, dropping left and right. hihat sounds above left and right.

11. Did you perceive the sound to always be coming from in front of you (A/B)?

In A yes it was generally in front.

12. When listening to track A/B. Do you imagine or visualise what you hear?

In B I visualised more that I was inside something mechanical or a machine, things moving around me

13. Is there anything you particularly liked/disliked about the production in A/B?

A in comparison to B is far less in terms of quality. B lends itself more to sounds that have movement to them whereas A is static.

14. Do you feel immersed in the music in A/B?

In B you are immersed, it does feel like you are taken inside of something.

15. Did you experience the music as if it was coming from all around you?

In B yes, A not really.

16. Did you feel as if the sound was coming from inside or outside of your head?

Outside in both.

17. Were the tracks comfortable or uncomfortable to listen to (A/B)? Why? (Explain if you wish).

I didn’t find them uncomfortable to listen to.

18. Is there anything else you would like to add?

No

6- **Listening Questionnaire (DiS / A and B):**

PROBLEM WITH HULTIGEN AUDIO PLAYBACK – SKIPPED TRIAL

Listener 2

No HULTIGEN Trial Undertaken– 2 x Single 3D Record Evaluation

Raw Focus Session Questionnaire Data

Consumer Interview

1. How old are you?
   32

2. How would you describe your musical experience?
   (Formal /, informal tuition, self-taught, learnt a un/graded instrument at school, no musical background etc.)

Formal Training: HND Music Production Played Bass. Self-Taught

3. Do you listen to music?   Yes, all different genres

4. How do you usually listen to music? (Speakers, TV speakers, home theatre system / surround sound, laptop, car, personal radio, headphones, phone etc.)

Headphones on phone and club sound system

5. What types/styles of music do you listen to most?

Dance, Rock, Hip Hop, Trip Hop, Electronica

6. Do you create/record/produce music?

I used to but not currently.

7. What type/style of music do you usually create/work with?

Dance Music or Electronica

8. Are you familiar with the term ‘Stereo’? If yes, describe what it means to you.

The format we listen to audio in. One side for each ear.

9. Are you familiar with the term ‘Periphonic’? If yes, describe what it means to you.

No I’ve never heard it before. Phonic is to do with sound but I don’t know what periphonic means.

10. Are you familiar with the term ‘Binaural’? If yes, describe what it means to you.

I understand it’s like stereo but more complex, I’m not exactly sure how.

Listening Questionnaire

(FFH Periphonic Mix)

1. What headphones are you listening on?
AKH K271 MK2

2. How many times did you listen to the track?

3. What does the production make you feel or think of?

Dreamy.

4. How would you explain the sonic space / environment?

More full [than JT]. Sounds fuller. Part of the dreamy thing I guess.

5. Describe the placement of sounds you hear.

Similar to JT but wider – not all the time. Seemed quite stereo. Lead vocals very wide backing vocal narrower.

See drawings.

6. Did you perceive any sounds to be moving in the track? Please describe.

No

7. Did you perceive any sounds from above? Please describe.

No

8. Did you perceive any sounds from behind? Please describe.

No

9. Did you perceive any sound coming directly from either side of you Left and or Right?

   Please describe.

Yes, vocals.

10. Did you perceive any sound from below? Please describe.

No

11. Did you perceive the sound to always be coming from in front of you? Please describe.

No. Sides and middle.

12. Do you imagine or visualise what you hear when listening to the music?

Yeah

13. Is there anything you particularly liked/disliked about the production?

Very vocal based, full and big sounding. Sparse. Dreamy, fantasy and kind of surreal.

14. Do you feel immersed in the music?

Yes

15. Did you experience the music as if it was coming from all around you?

No sides and middle

16. Did you feel as if the sound was coming from inside or outside of your head?

Not really inside the head but near and yes definitely outside

17. Was the track comfortable or uncomfortable to listen to? Why?

No it was comfortable.

18. Would you buy music in that format?

Yeah.

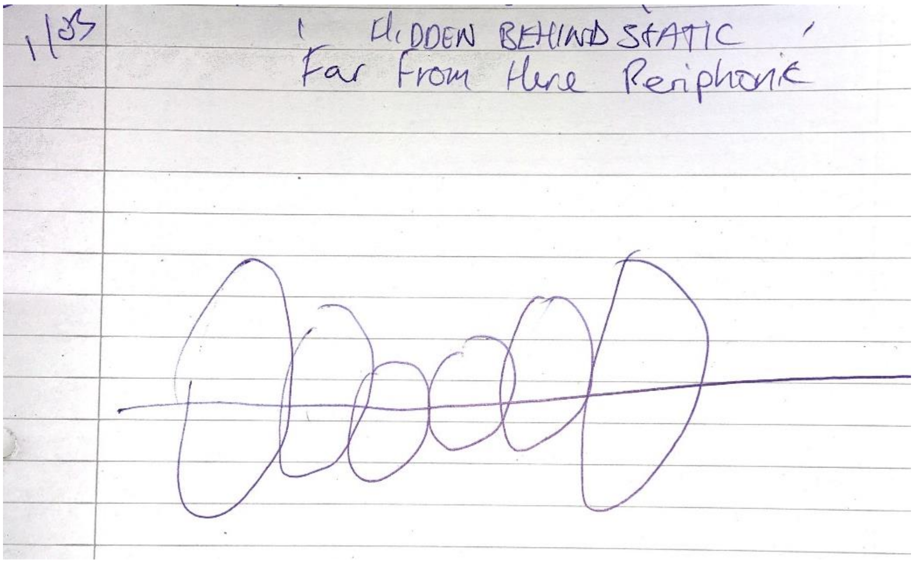
19. Do you think the production suits the music?

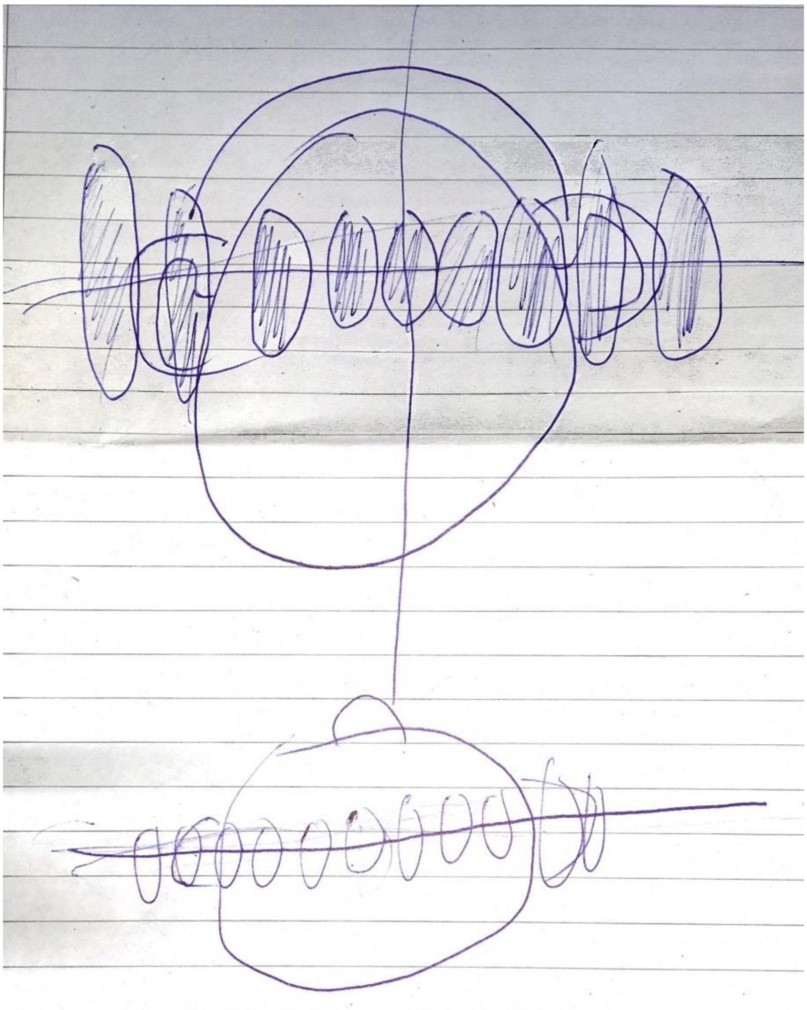
Yeah.

20. Is there anything else you would like to add about the music, mix or your experience?

No.

Listener 2 – Drawings of the perceived sonic space and sound source placements for 'Far From Here'





Listening Questionnaire

(JT Late Nights Periphonic Mix)

1. What headphones are you listening on?

AKH K271 MK2

2. How many times did you listen to each track?

5

3. What does the record as a whole evoke for you?

I visualise it as space.

4. How would you explain the sonic space / environment?

Spatial. Vocals very detailed. Simple and minimal instrumental. Subtle tones.

5. Describe the placement of sounds you hear.

   See recording and drawings.

6. Did you perceive any sounds to be moving in the track? Please describe.

Snare and syncopated verb ‘lurch’ movement. The synth. Maybe the vocals but it was harder to tell.

7. Did you perceive any sounds from above? Please describe.

   No

8. Did you perceive any sounds from behind? Please describe.

   No

9. Did you perceive any sound coming directly from either side of you Left and or Right?
   Please describe.

Yes.

10. Did you perceive any sound coming from below? Please describe.

Yes. I think I heard vocals slightly below.

11. Did you perceive the sound to always be coming from in front of you? Please describe.

No. I'd say it was from the middle of my head and the sides.

12. Do you imagine or visualise what you hear when listening to the music?

Yes.

13. Is there anything you particularly liked/disliked about the production?

See recording.

14. Do you feel immersed in the music?

Yes.

15. Did you experience the music as if it was coming from all around you?

No, from the sides and the middles of my head, a little bit below I feel but not massively.

16. Did you feel as if the sound was coming from inside or outside of your head?

Both.

17. Was the track comfortable or uncomfortable to listen to? Why?

Very comfortable to listen to

18. Would you buy music in that format?

Yes.

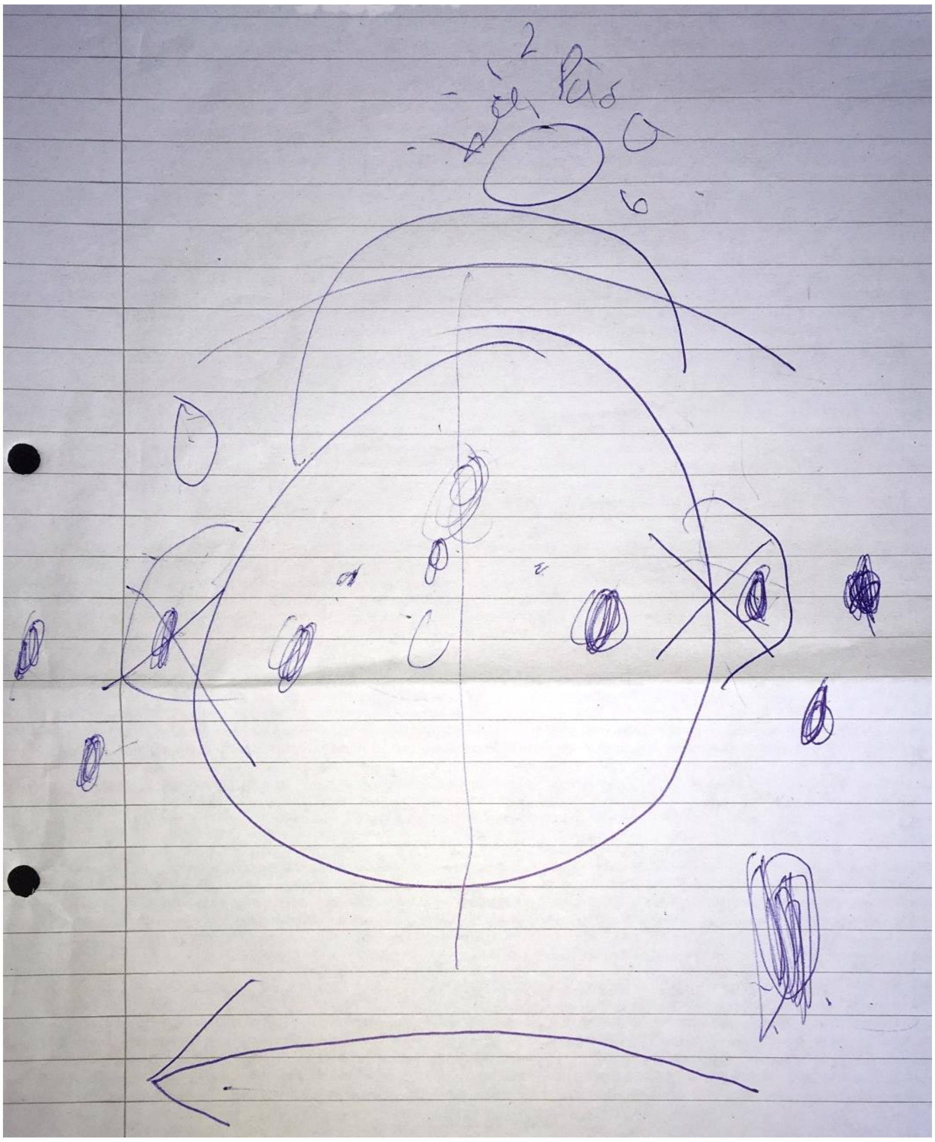
19. Do you think the music suits the type of production?

Yes definitely

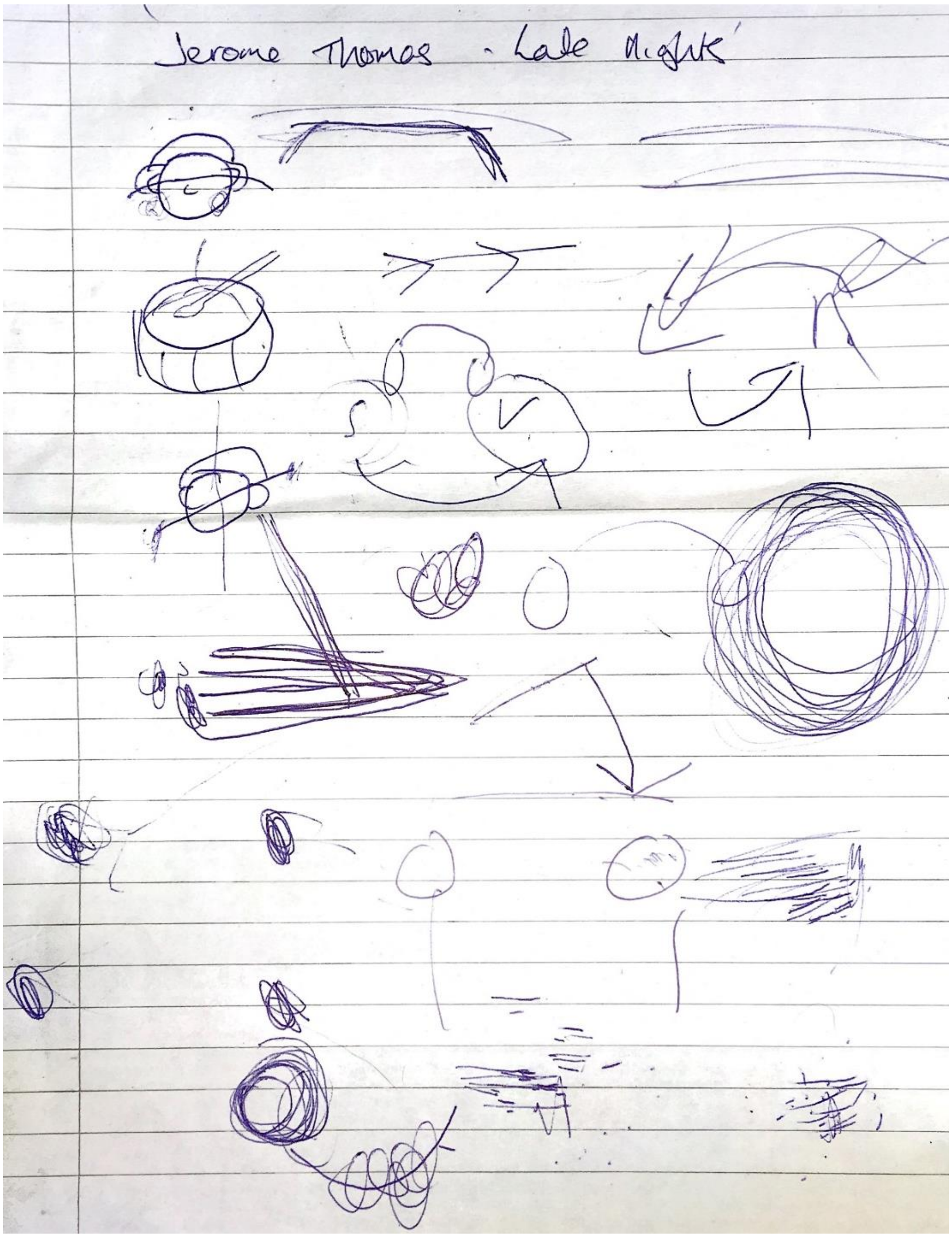
20. Is there anything else you would like to add about the music, mix or your experience?

I enjoyed it.

Listener 2 – Drawings of the perceived sonic space and sound source placements for ‘Late Nights’



Listener 2 – Drawings of the perceived sonic space and sound source placements for ‘Late Nights’


Jerome Thomas - 'Late Night'

Listener 3

No HULTIGEN Trial Undertaken – Single 3D Record Evaluation

Raw Focus Session Questionnaire Data

Consumer Interview

1. How old are you? 30

2. How would you describe your musical experience? *(Formal /, informal tuition, self-taught, learnt a un/graded instrument at school, no musical background etc.)*

I have 15 years of playing as a guitarist and singer in a range of bands of different styles of music, I attended Guitar – X when I was 16 which eventually joined with the instate and now is known as ICMP. I also trained in studio and live engineering.

3. Do you listen to music? Everyday

4. How do you usually listen to music? (speakers, TV speakers, home theatre system / surround sound, laptop, car, personal radio, headphones, phone etc.)

I listen to music through my studio monitors, my beyerdynamic headphones, my shitty headphones that I found on the train, my hi-fi system, my laptop and my TV.

5. What types/styles of music do you listen to most?

Rock, Blues, Metal, Jazz, Grunge, Folk and Classical

6. Do you create/record/produce music?

I do

7. What type/style of music do you usually create/work with?

Metal, Rock, Psychadelic and folk

8. Are you familiar with the term ‘Stereo’? If yes, describe what it means to you.

Yes, for me it means not mono

9. Are you familiar with the term ‘Periphonic’? If yes, describe what it means to you.

No.

10. Are you familiar with the term ‘Binaural’? If yes, describe what it means to you.

Yes, it means hearing sound as your ears hear it… basically.

Listening Questionnaire (FFH)

(Please answer separately for Tracks A and B where specified)

1. What headphones are you listening on?
   AKG K27 MII

2. How many times did you listen to each track?

   2

3. Which did you prefer, A or B, and why?

   A

4. What does the record as a whole invoke for you?

Elated, familiarity, Final Fantasy, Bubbles, Friends, Experimental, popping, Rhythmical

5. How would you explain the sonic space / environment?

   Track A – See Diagram

   Track B-

6. Describe the placement of sounds you hear in each version.

   Track a – See Diagram

   Track B-

7. Did you perceive any sounds to be moving in either track? Please describe.

   Track A – I felt the bass and the small ambient sounds to be swirling from place to place in time with the music

   Track B-

8. Did you perceive any sounds from above? Please describe.

   Track A- I perceived sounds to be from the centre of my forehead

   Track B-

9. Did you perceive any sounds from behind? Please describe.

Track A- I felt that there were sound from either side of the back of my head forming almost a triangle from the front to the back as I've drawn on the diagram.

Track B-

10. Did you perceive any sound coming directly from either side of you Left and or Right?
    Please describe.

    Track A- Yes from both sides.

    Track B -

11. Did you perceive the sound to always be coming from in front of you?
    Please describe.

    Track A – I felt the bass changed from being in front of me to the sides of my head

    Track B-

12. Do you imagine or visualise what you hear when listening to the music?

    Track A- I imagined and visualised the sounds as swirling masses of colour on a back

    Track B-

13. Is there anything you particularly liked/disliked about the production?

    Track A- I loved the production in its creativity and its experimental side.

    Track B-

14. Do you feel immersed in the music?

    Track A- I did feel immersed in the music

Track B-

15. Did you experience the music as if it was coming from all around you?

    Track A – Yes

    Track B -

16. Did you feel as if the sound was coming from inside or outside of your head?

    Track A- I felt some sounds were in my forehead whilst others were floating around my head.

    Track B-

17. Were the tracks comfortable or uncomfortable to listen to? Why?

    Track A- They were comfortable, I found the tones to be soothing.

    Track B-

18. Which of the versions would you rather listen to, A or B?

    A

19. Would you buy music in either of these formats?

    Track A – Because this is the only one I heard

    Track B-

20. Do you think music suits the type of production?

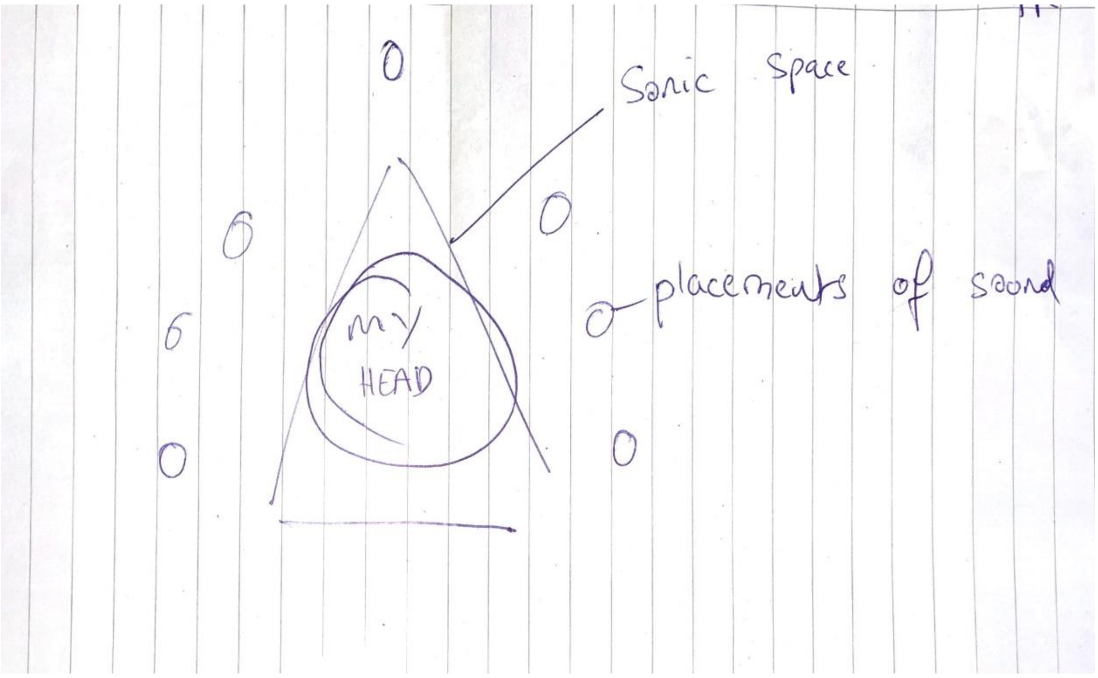
    Track A – I think this type of production fits perfectly to the genre.

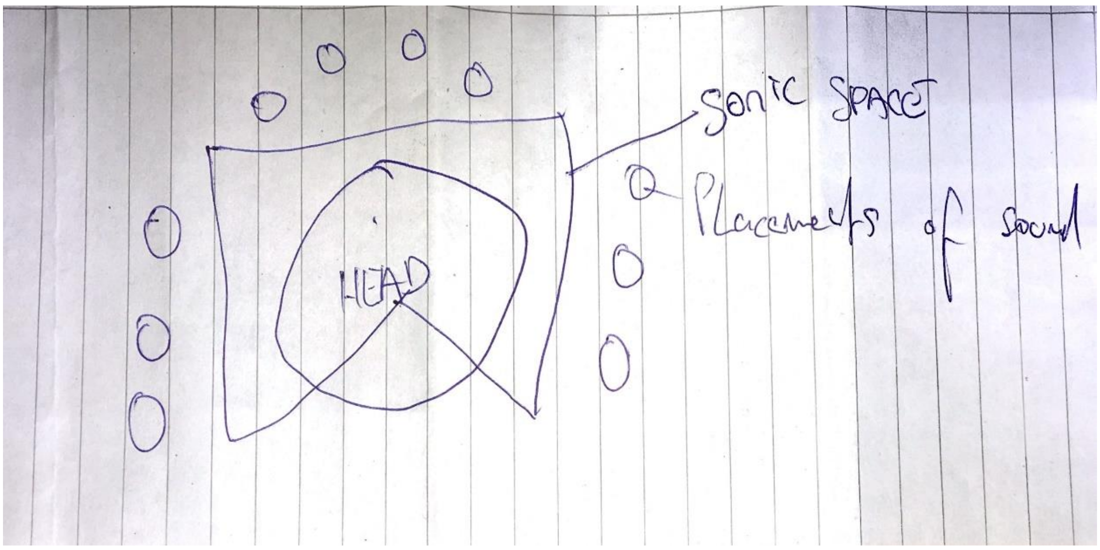
    Track B -

21. Is there anything else you would like to add about the music, mixes or your experience?

    I don't know what to say.

Listener 3 – Drawings of the perceived sonic space and sound source placements for 'Far From Here'





Listener 4

No HULTIGEN Trial Undertaken – Single 3D Record Evaluation

Raw Focus Session Questionnaire Data

Consumer Interview

1. How old are you? 27

2. How would you describe your musical experience, if any?
   (Formal /, informal tuition, self-taught, learnt a un/graded instrument at school, no musical background etc.)

   No musical background

3. Do you have experience with music production?

I've edited sound for film – not trained in it.

4. Do you listen to music?

Yes

5. How do you usually listen to music? (Speakers, TV speakers, home theatre system / surround sound, laptop, car, personal radio, headphones, phone etc.)

   Laptop, Headphones

6. What types/styles of music do you listen to most?

Mostly rock and alternative

7. Do you create/record/produce music?
   No

8. What type/style of music do you usually create/work with?

NA

9. Are you familiar with the term ‘Stereo’? If yes, describe what it means to you.

Yes, comes out of both speakers

10. Are you familiar with the term ‘Periphonic’? If yes, describe what it means to you.

No

11. Are you familiar with the term ‘Binaural’? If yes, describe what it means to you.

No.

Listening Questionnaire

(Please answer separately for Tracks A (FFH ST) and B (FFH VR) where specified)

1. What headphones are you listening on? AKG K271

2. How many times did you listen to each track? 1 each

3. Which did you prefer, A or B, and why?
   A – Preferred A, felt more experimental, a bit like the instruments were just put all over the place.
   B – felt more professional, and more like a concert. More structured.

4. What does the record as a whole invoke for you?

Tim Burton vibes. I didn't understand the lyrics but it felt kind of like 'Sally's Song' – melancholy.

5. How would you explain the sonic space / environment?

   Track A – more like a T shape, across the front and narrowly behind

   Track B- more like the 'power on' sign circular wide around but with a narrow line behind.

6. Describe the placement of sounds you hear in each version.

   Track A –

   Male vocals are forward facing, female vocals are above, behind and to either side intermittently, and there is an orbital use of microphone on instrumentals which creates a pendulum effect. Background effect noises are softer and seem to come from below.

   Track B-

   Male vocals are forward facing but feel as if they drop to behind, female vocals seem stronger forward as well as behind, the female vocals on either side are less intermittent and the instrumental sound is more static/still, the background effect noises are louder and come from above.

7. Did you perceive any sounds to be moving in either track? Please describe.

   Track A – bubbly sound that rises, the microphone on guitar was moving in a swung motion

   Track B- no movement, the voices appeared spontaneously on either side

8. Did you perceive any sounds from above? Please describe.

Track A- female voice but sounded more behind in A

Track B- female vocals above and more on the crown on the head

9. Did you perceive any sounds from behind? Please describe.
Track A- female vocal shoulder width behind

Track B- female vocal in level with the head behind

10. Did you perceive any sound coming directly from either side of you Left and or Right?
Please describe.

Track A- spontaneous vocals and the pendulum guitar

Track B – female vocals fading in and out, guitar on both sides but more static

11. Did you perceive the sound to always be coming from in front of you? Please describe.

Track A – man vocal

Track B- man vocal and female vocal before moving behind

12. Do you imagine or visualise what you hear when listening to the music?

Track A- kind of – more complex, felt more like just swinging a mic, more uncomfortable to listen to –made me feel a bit sick.

Track B- felt more like a studio setup – circular set up, more all around me

13. Is there anything you particularly liked/disliked about the production?

Track A- nothing particularly

Track B- nothing particularly

14. Do you feel immersed in the music?

    Track A- felt like I was there in the music but it all felt a bit hazy and made me feel sick

    Track B- more immersed in B but didn't like B that much

15. Did you experience the music as if it was coming from all around you?

    Track A – just felt more in the front and to the side, like more of a line in front

    Track B – to a degree yes – B felt more surround

16. Did you feel as if the sound was coming from inside or outside of your head?

    Track A- inside

    Track B- outside

17. Were the tracks comfortable or uncomfortable to listen to? Why?

    Track A- not so comfy, a bit nauseous

    Track B- comfortable, felt like I was sat listening in the middle of a performance

18. Which of the versions would you rather listen to, A or B?
    Prefer A because it feels more experimental but I don't like the nauseous feeling, B felt a bit boring in comparison.

19. Would you buy music in either of these formats?

    Track A – more inclined to A

    Track B- but yes to both

20. Do you think the production suits the music?

    Track A - yes

    Track B - no

21. Is there anything else you would like to add about the music, mixes or your experience?

A felt very Danny Elfman Tim Burton-esque (Nightmare Before Xmas), B felt more like Danny Elfman in Batman.

Listener 5

HULTIGEN Data Output

Stimuli Set.

1, "Far From Here VR.wav" "Far From Here.wav";
2, "Beautiful Thing - Waiting (Live at London Water & Steam Museum) [Periphonic Mix].wav" "Beautiful Thing - Waiting (Live at London Water & Steam Museum) [unmastered].wav";
3, "Penny Drops.wav""Penny Drops VRv2.wav";
4, "Jerome Thomas - Late Nights (Periphonic Mix Logic).wav" "Late Nights Demo Vocal Up.mp3";
5, Monomorphic.wav "Monomorphic VR.wav";
6, "Drowned In Sound.wav" "Drowned In Sound VR.wav";
7, "Joey Clarkson - Sort Yourself Out (Live at Brentford Steam Works) [Periphonic Mix].wav";"Sort Yourself Out Mstr1.wav"

100, Trial Order 7 6 4 5 1 3 2;

1, "Trial 7" 76. 52.; "Sort Yourself Out Mstr1.wav" "Joey Clarkson - Sort Yourself Out (Live at Brentford Steam Works) [Periphonic Mix].wav";

2, "Trial 6" 82. 96.; "Drowned In Sound.wav""Drowned In Sound VR.wav";

3, "Trial 4" 100. 72.; "Jerome Thomas - Late Nights (Periphonic Mix Logic).wav" "Late Nights Demo Vocal Up.mp3";

4, "Trial 5" 85. 68.; "Monomorphic VR.wav" Monomorphic.wav;

5, "Trial 1" 84. 89.; "Far From Here.wav" "Far From Here VR.wav";

6, "Trial 3" 89. 82.; Penny Drops.wav" "Penny Drops VRv2.wav"";

7, "Trial 2" 100. 78.; "Beautiful Thing - Waiting (Live at London Water & Steam Museum) [Periphonic Mix].wav" "Beautiful Thing - Waiting (Live at London Water & Steam Museum) [unmastered].wav";

Stimuli AB Order / Scale Correlation

A/B
101, 2 1; A "Sort Yourself Out Mstr1.wav" B "Joey Clarkson - Sort Yourself Out (Live at Brentford Steam Works) [Periphonic Mix].wav";
102, 2 1; A "Drowned In Sound.wav" B "Drowned In Sound VR.wav";
103, 1 2; A "Jerome Thomas - Late Nights (Periphonic Mix Logic).wav" B "Late Nights Demo Vocal Up.mp3";
104, 2 1; A "Monomorphic VR.wav" B Monomorphic.wav;

```
105, 2 1; A "Far From Here.wav"; B "Far From Here VR.wav"
106, 1 2; A ""Penny Drops.wav" B "Penny Drops VRv2.wav;
107, 1 2; "Beautiful Thing - Waiting (Live at London Water &
Steam Museum) [Periphonic Mix].wav" "Beautiful Thing - Waiting
(Live at London Water & Steam Museum) [unmastered].wav";
```

Raw Focus Session Questionnaire Data

Consumer Interview

1. How old are you?

40.

2. How would you describe your musical experience?
   (Formal /, informal tuition, self-taught, learnt a un/graded instrument at school, no musical background etc.)

Graded instrument at school and self-taught

3. Do you listen to music?

Yes, a lot.

4. How do you usually listen to music? (speakers, TV speakers, home theatre system / surround sound, laptop, car, personal radio, headphones, phone etc.)

Speakers and headphones.

5. What types/styles of music do you listen to most?

Hip-hop, electronica, rock and metal.

6. Do you create/record/produce music?

Yes.

7. What type/style of music do you usually create/work with?

Hip-hop, electronic, rock/industrial.

8. Are you familiar with the term ‘Stereo’? If yes, describe what it means to you.

(Hell yeah. I am Stereo!) Yes, stereo is an illusion (width), I only get to it after mono-staging (depth and height).

9. Are you familiar with the term ‘Periphonic’? If yes, describe what it means to you.

Yes, it is the sonic periphery, can be illusory (implied with a single or two speakers) or actual, created with multiple speakers.

10. Are you familiar with the term ‘Binaural’? If yes, describe what it means to you.

Yes, to me it stands for a recreation of recorded (or constructed) sonic phenomena that have been captured/made in the ‘image’ of - and for the consumption of – (two) human ears.

Listening Questionnaire

(Please answer separately for Tracks A and B where specified)

Far from here

1. What headphones are you listening on?
AKG K271 mkII

2. How many times did you listen to each track?
4-5 times.

3. Which did you prefer, A or B, and why?
A – Exciting, immersive, rich, expansive

4. What does the record as a whole invoke for you?
Magical, mystical.

5. How would you explain the sonic space / environment?

Track A –

Wow, this is all around me, above me, large and epic, suiting the music. As a result, sometimes, less defined.

Track B-

B is more controlled with definitive L-R image movements.

6. Describe the placement of sounds you hear in each version.
Track A –

As per 5 above, very immersed by the sounds. It’s almost not a space anymore, but a hallucination.

Track B-

Width and defined stereo image placements.

7. Did you perceive any sounds to be moving in either track? Please describe.
Track A –

Track B-

Yes, in the textures and voices of B.

8. Did you perceive any sounds from above? Please describe.

Track A- Certainly on A. Synth strings, voices and space on voices.

Track B-

9. Did you perceive any sounds from behind? Please describe.
Track A-

Yes, learning to on A.

Track B-

Not so much on B.

10. Did you perceive any sound coming directly from either side of you Left and or Right?
Please describe.

Track A-

Track B – Very much on B. Particularly the voices.

11. Did you perceive the sound to always be coming from in front of you?
Please describe.

Track A -

Track B -

I tend to, as a habit, but A, in this case, is educating me differently.

12. Do you imagine or visualise what you hear when listening to the music?

Track A-

Track B-

Yes.

13. Is there anything you particularly liked/disliked about the production?

Track A-

Track B-

Love both, and their differences are also enjoyable.

14. Do you feel immersed in the music?

Track A-

Track B-

Much more in A.

15. Did you experience the music as if it was coming from all around you?

Track A -

Track B -

Much more on A overall.

16. Did you feel as if the sound was coming from inside or outside of your head?

Track A – outside, but potentially more inside than B, if I consciously change my perspective.

Track B - outside

17. Were the tracks comfortable or uncomfortable to listen to? Why?

Track A- apart from a little less definition due to the huge spatialization on A, both comfortable.

Track B- safer and more defined but still comfortable/enjoyable, and in fact growing on me with multiple listens.

18. Which of the versions would you rather listen to, A or B?

A for excitement, B for clarity/control.

19. Would you buy music in either of these formats?

Track A - yes

Track B- yes

20. Do you think production suits the type of music?

Track A - yes

Track B - yes

21. Is there anything else you would like to add about the music, mixes or your experience?

Listening Questionnaire

(Please answer separately for Tracks A and B where specified)

Waiting

1. What headphones are you listening on?
AKG K271 mkII

2. How many times did you listen to each track?

3-4 times.

3. Which did you prefer, A or B, and why?

B – because of the impressive stereo image and definition. But A is cool as well if less exciting after the comparison.

4. What does the record as a whole invoke for you?

Coolness, groove.

5. How would you explain the sonic space / environment?

Track A –

This feels like a good pop mix, some good aux automations and spatial changes between sections.

Track B-

B is unique, exciting, changes the listening focus – rich, defined spaces and elements on the L-R axis.

6. Describe the placement of sounds you hear in each version.

Track A –

A classic pop stage, stylistically relevant.

Track B-

As per 5. above, width and defined stereo image placements.

7. Did you perceive any sounds to be moving in either track? Please describe.

A lot of width rather than movement everywhere, but more so on the L-R on B.

8. Did you perceive any sounds from above? Please describe.

Track B- Yes on B, the spaces above elements, e.g. guitars.

9. Did you perceive any sounds from behind? Please describe.

Track A-

Not really

Track B-

Yes on B, more the reflections (sometimes even the combined effect with the lead voice), and the BVs.

10. Did you perceive any sound coming directly from either side of you Left and or Right ? Please describe.

Track A-

Track B –

Very much on B, guitar riffs on L, electric keys on R, guitar mutes on R, drum overheads L-R, delays a bit on vox

11. Did you perceive the sound to always be coming from in front of you? Please describe.

Track A – very much on A

Track B – comparatively not on B

12. Do you imagine or visualise what you hear when listening to the music?

Track A-

Track B-

Yes.

13. Is there anything you particularly liked/disliked about the production?

Track A-

Track B-

Love both; prefer B!

14. Do you feel immersed in the music?

Track A-

Track B-

Yes, in both, but much more in B. I think the experiment has informed me in what can be done – commercially – beyond stereo!

15. Did you experience the music as if it was coming from all around you?

Track A -

Track B -

Much more on B overall.

16. Did you feel as if the sound was coming from inside or outside of your head?

Track A – inside on A.

Track B – outside on B.

17. Were the tracks comfortable or uncomfortable to listen to? Why?

Both comfortable, but A feels more muddy, mono and ‘contained’ after the comparison with B.

18. Which of the versions would you rather listen to, A or B?

B. It also retains a mean space for the bassline (which is always a great thing).

19. Would you buy music in either of these formats?

Track A - yes

Track B- yes!

20. Do you think production suits the type of music?

Track A - yes

Track B – yes!

21. Is there anything else you would like to add about the music, mixes or your experience?

Listening Questionnaire

(Please answer separately for Tracks A and B where specified)

Penny Drops

1. What headphones are you listening on?
AKG K271 mkII

2. How many times did you listen to each track?
3-4 times.

3. Which did you prefer, A or B, and why?

A – Expansive space, quite epic.

4. What does the record as a whole invoke for you?

Power.

5. How would you explain the sonic space / environment?

Track A –

This is all around me, above me, and – strangely – also inside me. Large and epic, suiting the music.

Track B-

B is more controlled with definitive L-R image movements again, less exciting than A.

6. Describe the placement of sounds you hear in each version.
Track A –

As per 5 above, very immersed by the sounds. It's almost not a space anymore, but a hallucination.

Track B-

Width and defined stereo image placements, but less expanded height and size.

7. Did you perceive any sounds to be moving in either track? Please describe.
Track A –

Track B-

A lot of movement everywhere, but more restricted to the L-R on B.

8. Did you perceive any sounds from above? Please describe.

Track A- Certainly on A. The epic spaces on voices, guitars, and the contained elements themselves.

Track B-

9. Did you perceive any sounds from behind? Please describe.
Track A-

Not really

Track B-

Yes on B, the drony textures feel behind me.

10. Did you perceive any sound coming directly from either side of you Left and or Right?
Please describe.

Track A-

Track B –
Yes on both tracks, granular bits, Foley.

11. Did you perceive the sound to always be coming from in front of you? Please describe.

Track A -

Track B -

I tend to, as a habit, but B, in this case, is educating me differently.

12. Do you imagine or visualise what you hear when listening to the music?

Track A-

Track B-

Yes.

13. Is there anything you particularly liked/disliked about the production?

Track A-

Track B-

Love both; the stuttered acoustic guitar textures feels a little ‘library’

14. Do you feel immersed in the music?

Track A-

Track B-

Yes, in both, in A in a more reverb-ed sense, in B in a more clearly defined – around my ears – sense.

15. Did you experience the music as if it was coming from all around you?

Track A -

Track B -

More on B overall.

16. Did you feel as if the sound was coming from inside or outside of your head?

Track A – outside, but potentially more inside than B.

Track B – outside

17. Were the tracks comfortable or uncomfortable to listen to? Why?

More interesting discomfort on B!

18. Which of the versions would you rather listen to, A or B?

A for epicness, B for surprising imaging movements.

19. Would you buy music in either of these formats?

Track A - yes

Track B- yes

20. Do you think production suits the type of music?

Track A - yes

Track B - yes

21. Is there anything else you would like to add about the music, mixes or your experience?

Yes, the experiment would work even better, if A-B tracks were sync'd and mastered to a similar loudness throughout (not on this track necessarily).

Listening Questionnaire

(Please answer separately for Tracks A and B where specified)

Late Nights

1. What headphones are you listening on?
AKG K271 mkII

2. How many times did you listen to each track?

4-5 times.

3. Which did you prefer, A or B, and why?
B. Very unique - and new to me - sonic experience.

4. What does the record as a whole invoke for you?
Mystery, sexuality.

5. How would you explain the sonic space / environment?

Track A –

Quite a professional ‘industry’ mix, but after hearing B, I’m not as excited about its professionalism. Some nice long, progressive R&B tails on the lead vox.

Track B-

Not sure how to describe B, apart from really immersive. I am one with the experience. The space could be my mind/emotions.

6. Describe the placement of sounds you hear in each version.
Track A –

As per 5. above, quite a professional stereo, wide, depth-implied-through reverbs placement.

Track B-

In B, the placement feels 3D, immersive, inside my head.

7. Did you perceive any sounds to be moving in either track? Please describe.
Track A –

Track B-

Much more in B, in an emotive rather than physically/naturally placed/moved fashion. Love the synth treatment.

8. Did you perceive any sounds from above? Please describe.

Track A-

Track B-

Yes, in B, as it feels *inside* my head.

9. Did you perceive any sounds from behind? Please describe.
Track A-

Not so much on A.

Track B-

Yes, again, after reading the question, on B. I think it is a habit, of projecting the 3D picture in front of me.

10. Did you perceive any sound coming directly from either side of you Left and or Right?
Please describe.

Track A – on A there is traditional stereo placement.

Track B – on B the stereo feels more like a constant experience.

11. Did you perceive the sound to always be coming from in front of you? Please describe.

Track A – quite a bit on A.

Track B – not on B.

12. Do you imagine or visualise what you hear when listening to the music?

Track A-

More in a traditional sense of instruments, singer

Track B-

More in an 'elements' in my head experience.

13. Is there anything you particularly liked/disliked about the production?

Track A-

Track B-

Great production, takes a fantastic new realm in B.

14. Do you feel immersed in the music?

Track A- A bit, in a classic sense

Track B- Very much so.

15. Did you experience the music as if it was coming from all around you?

Track A - No

Track B - Yes
More on B overall.

16. Did you feel as if the sound was coming from inside or outside of your head?

Track A – outside / in front

Track B – inside

17. Were the tracks comfortable or uncomfortable to listen to? Why?

Track A- comfortable in both cases

Track B- less, but more exciting, on B

18. Which of the versions would you rather listen to, A or B?

B.

19. Would you buy music in either of these formats?

Track A - yes

Track B- yes!

20. Do you think production suits the type of music?

Track A - yes

Track B – yes!

21. Is there anything else you would like to add about the music, mixes or your experience?

Great exploration of a great, more standard, pop production through the spatialisations on B.

Listening Questionnaire

(Please answer separately for Tracks A and B where specified)

Monomorphic

1. What headphones are you listening on?
AKG K271 mkII

2. How many times did you listen to each track?
4-5 times.

3. Which did you prefer, A or B, and why?

B – Both good, but there was more ‘space’ for elements here. The loudness of A is quite impressive, but if they were matched, I’d go for B.

4. What does the record as a whole invoke for you?
Zen epicness.

5. How would you explain the sonic space / environment?

Track A –

Quite a concentrated in the middle energy

Track B-

An open space, where the voices are in a high, physical space, while the electronics are in constructed places within what the voices have defined.

6. Describe the placement of sounds you hear in each version.
Track A –

Some width and depth, but once compared to B, quite mono in comparison.

Track B-

Width and immersion.

7. Did you perceive any sounds to be moving in either track? Please describe.

Track A –

Track B-

Much in the electronic percussion of B.

8. Did you perceive any sounds from above? Please describe.

Track A-

Track B-

Choirs on B.

9. Did you perceive any sounds from behind? Please describe.

Track A-

Not so much on A.

Track B-

Yes, on B.

10. Did you perceive any sound coming directly from either side of you Left and or Right?
Please describe.

Track A-

Track B – Very much on B. Electronic percussion and particular choir voices/layers.

11. Did you perceive the sound to always be coming from in front of you? Please describe.

Track A -

Track B -

I tend to, as a habit, but B is educating me differently.

12. Do you imagine or visualise what you hear when listening to the music?

Track A-

Track B-

Yes.

13. Is there anything you particularly liked/disliked about the production?

Track A-

Track B-

Love it, and B is very immersive (sometimes even distractingly so, but in a good way).

14. Do you feel immersed in the music?

Track A-

Track B-
Absolutely, in both cases, more in B.

15. Did you experience the music as if it was coming from all around you?

Track A -

Track B –

Much more on B overall.

16. Did you feel as if the sound was coming from inside or outside of your head?

Track A - inside

Track B - outside

17. Were the tracks comfortable or uncomfortable to listen to? Why?

Track A- a bit over compressed for me, but in a way many commercial formats/outputs would require
Track B- open and enjoyable.

18. Which of the versions would you rather listen to, A or B?

B.

19. Would you buy music in either of these formats?

Track A - maybe

Track B- yes

20. Do you think production suits the type of music?

Track A - somewhat

Track B - yes

21. Is there anything else you would like to add about the music, mixes or your experience?

Listening Questionnaire

(Please answer separately for Tracks A and B where specified)

Drowned in sound

1. What headphones are you listening on?
AKG K271 mkII

2. How many times did you listen to each track?
4-5 times.

3. Which did you prefer, A or B, and why?

B in this case – the material really lends itself to it. Feels like the elements take a life of their own, a more ‘expansive’ experience.

4. What does the record as a whole invoke for you?

Adventure, exploration, ‘brave unknown’.

5. How would you explain the sonic space / environment?

Track A –

An expanded stereo width, with good height courtesy of EQ processing, but more of a ‘constructed’ space.

Track B-

In B, building from 2. Above, it feels as if the sources/elements are ‘alive’, almost like this is not a sonic illusion anymore, but a real sonic phenomenon I am experiencing (even if there are synthetic sources and electronic processing/manipulation). So the space feels like a real-life experience of a synthetic landscape.

6. Describe the placement of sounds you hear in each version.

Track A –

In A there is a very wide stereo placement and movement, quite rich in the low mids / drone-like textures.

Track B-

In B, the placement feels 3D, more immersive.

7. Did you perceive any sounds to be moving in either track? Please describe.

Track A –

There is a lot of movement in width, depth and height on both tracks, but I feel the width and height is pronounced in A, while…

Track B-

…the above and the *depth* are really working in tandem on B.

8. Did you perceive any sounds from above? Please describe.

Track A-

Yes, particularly the early bell-like timbres on A.

Track B-

The strings on A feel much more ‘on the horizon’ while on B they are ‘higher’. What’s interesting here, is that the abstraction of ‘height’ as EQ is extended by ‘height’ as placement ‘from above’.

9. Did you perceive any sounds from behind? Please describe.
Track A-

Not so much on A.

Track B-

A bit on B, but I felt immersed in an ‘everything is happening in front of me’ here. Upon reading the question, I feel some of the wave-y/noise-y stuff is behind be. So, I may be breaking ‘stereo’ habits.

10. Did you perceive any sound coming directly from either side of you Left and or Right?
Please describe.

Yes, throughout A and B.

Track A-

Track B –

11. Did you perceive the sound to always be coming from in front of you? Please describe.

Track A – more on A

Track B- as per 9. above.

12. Do you imagine or visualise what you hear when listening to the music?

Track A-

Track B-

Yes, I’m visualising both the space(s) and a more synesthetic movement of the electronic textures as shapes in an unknown environment.

13. Is there anything you particularly liked/disliked about the production?

Track A-

Track B-

I love this production, a more electronica, contemporary take on avant-garde/electroacoustic textures. The strings feel a bit ‘library’. Love the rest.

14. Do you feel immersed in the music?

Track A-

Track B-

Absolutely, in both cases, more in B.

15. Did you experience the music as if it was coming from all around you?

Track A -

Track B –

More on B overall.

16. Did you feel as if the sound was coming from inside or outside of your head?

Track A - outside

Track B - inside

17. Were the tracks comfortable or uncomfortable to listen to? Why?

Track A- comfortable in both cases, but

Track B- more on B

18. Which of the versions would you rather listen to, A or B?

I like both, but I’d choose B if I had to.

19. Would you buy music in either of these formats?

Track A - yes

Track B- yes

20. Do you think production suits the type of music?

Track A - yes

Track B - yes

21. Is there anything else you would like to add about the music, mixes or your experience?

I love these compositionally and in terms of mix / post-production. I feel the B mix is excellent, tight, defined and still explorative. A is brave, perhaps more disorientating in places.

Listening Questionnaire

(Please answer separately for Tracks A and B where specified)

Sort yourself out

1. What headphones are you listening on?
AKG K271 mkII

2. How many times did you listen to each track?
4-5 times.

3. Which did you prefer, A or B, and why?

A, even though it was crunchier/saturated, because of clarity, punch and presence.

B sounds further away, less present and less defined (muddier). I also find some sudden, if striking, staging jumps on B (before the chorus).

4. What does the record as a whole invoke for you?
Relaxation, thought-provoking, laid-back.

5. How would you explain the sonic space / environment?

Track A –

There is a striking, large, reflective space around the drums and acoustic guitar intro that kick off the tune. The spaces on the horn and vocal feel 'added' in post, while the bass quite dry (and DI'd).

Track B-

In B, I am appreciating a tiled room quality over the whole record (particularly the instrumental), while the horn and lead vocal feature longer spatial tails. The latter feel like processing, while the former, 'captured'.

6. Describe the placement of sounds you hear in each version.
Track A –

An upfront lead vocal, surrounded by the band, further back horn, bass/bottom end low (on the height axis) and front (on the depth). Guitar strumming also quite front. The saturation/crunch is bringing everything quite forward.

Track B-

In B, the horn feels more direct (upfront) comparatively, while the staging set architecturally by the drums gives more of a tiled room impression. Once, I get used to it (after A), the lead vocal in B feels forward. The bass here is more muddy / less defined and even if not spatially processed, it blends more into the percussive space described above.

7. Did you perceive any sounds to be moving in either track? Please describe.

Track A –

It could be a result of post-fader sends, but I feel the spaces are responding to the instrument/vocal dynamics, creating interesting depth dynamics. These could also be the results of rides in the mixing stage.

Track B-

Yes, as mentioned above, vocals before the chorus in the stereo domain.

8. Did you perceive any sounds from above? Please describe.

Track A-

It's a strange impression when flipping from B to A, where A feels like I am more surrounded by – not so much the band but – the reflections the sources are causing to the space.

Track B-

9. Did you perceive any sounds from behind? Please describe.

Track A-

Track B-

Conversely, B feels much more like I'm in the middle of the tiled, reflective room, and there's much more happening around me and behind me, which explains the lesser definition.

10. Did you perceive any sound coming directly from either side of you Left and or Right?
Please describe.

Track A-

Track B –

Yes, as mentioned above but also on the drum positions, particularly noticeable on fills.

11. Did you perceive the sound to always be coming from in front of you? Please describe.

Track A –

More on A, and with the elements described above as frontal and present.

Track B-

Not really with B, I feel I'm listening to the space in which a band is performing.

12. Do you imagine or visualise what you hear when listening to the music?

Track A-

Yes, I'm visualising both the space(s) and the instrumental positions.

Track B-

13. Is there anything you particularly liked/disliked about the production?

Track A-

Love the drum compression, very Ringo. Emotive horn flourishes and ambience.

Track B-

14. Do you feel immersed in the music?

Track A-

Track B-

Yes, in both cases, but more on B.

15. Did you experience the music as if it was coming from all around you?

Track A -

Track B –

Much more on B.

16. Did you feel as if the sound was coming from inside or outside of your head?

Track A - inside

Track B - outside

17. Were the tracks comfortable or uncomfortable to listen to? Why?

Track A- comfortable in both cases, but

Track B- less on B

18. Which of the versions would you rather listen to, A or B?

A.

19. Would you buy music in either of these formats?

Track A - yes

Track B- no

20. Do you think production suits the type of music?

Track A - yes

Track B - potentially

21. Is there anything else you would like to add about the music, mixes or your experience?

I think this sort of staging would benefit from better performances and better recorded sources. But the spatial exploitation is inspiring and demonstrates what can be done.

Listener 6

Stimuli Set

1, "Far From Here VR.wav" "Far From Here.wav";
2, "Beautiful Thing - Waiting (Live at London Water & Steam Museum) [Periphonic Mix].wav" "Beautiful Thing - Waiting (Live at London Water & Steam Museum) [unmastered].wav";
3, "Penny Drops.wav" "Penny Drops VRv2.wav";
4, "Jerome Thomas - Late Nights (Periphonic Mix Logic).wav" "Late Nights Demo Vocal Up.mp3";
5, "Monomorphic VR.wav" Monomorphic.wav;
6, "Drowned In Sound.wav" "Drowned In Sound VR.wav";
7, "Sort Yourself Out Mstr1.wav" "Joey Clarkson - Sort Yourself Out (Live at Brentford Steam Works) [Periphonic Mix].wav";

100, Trial Order 7 3 1 4 5 6 2;

1, "Trial 7" 74. 24.; "Sort Yourself Out Mstr1.wav" "Joey Clarkson - Sort Yourself Out (Live at Brentford Steam Works) [Periphonic Mix].wav";

2, "Trial 3" 100. 75.; "Penny Drops.wav" "Penny Drops VRv2.wav";

3, "Trial 1" 49. 75.; "Far From Here VR.wav" "Far From Here.wav";

4, "Trial 4" 51. 100.; "Jerome Thomas - Late Nights (Periphonic Mix Logic).wav" "Late Nights Demo Vocal Up.mp3";

5, "Trial 5" 75. 49.; Monomorphic VR.wav"; Monomorphic.wav"

6, "Trial 6" 49. 100.; "Drowned In Sound.wav";"Drowned In Sound VR.wav"

7, "Trial 2" 75. 100.; "Beautiful Thing - Waiting (Live at London Water & Steam Museum) [Periphonic Mix].wav" "Beautiful Thing - Waiting (Live at London Water & Steam Museum) [unmastered].wav";

Stimuli AB Play Back Order / Scale Correlation

NOTE: When AB stimuli samples playback is switched (2, 1 – A, B)the order of the AB result above is also presented reversed (B, A). This must be due to a HULTIGEN glitch, the project has experienced intermitted anomalies with the software but usually in playback, not so far presented in data until now. However, the results can be correlated with the Spatial Producers handwritten notes of the scores for the session as seen below.

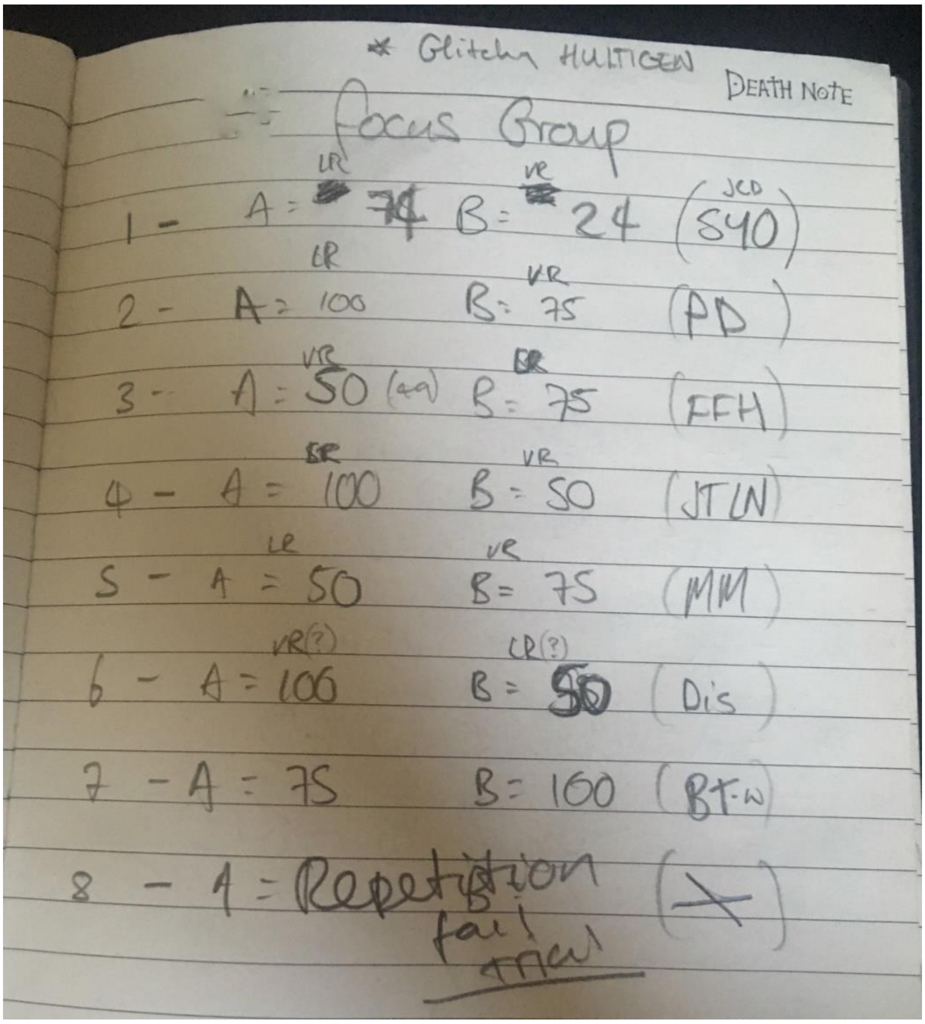
A/B
101, 1 2; A "Sort Yourself Out Mstr1.wav" B "Joey Clarkson - Sort Yourself Out (Live at Brentford Steam Works) [Periphonic Mix].wav";
102, 1 2; "Penny Drops.wav" "Penny Drops VRv2.wav";

103, 1 2; A "Far From Here VR.wav" B "Far From Here.wav";
104, 2 1; A "Late Nights Demo Vocal Up.mp3" B Jerome Thomas - Late Nights (Periphonic Mix Logic).wav" ;"

105, 2 1; A Monomorphic.wav" B Monomorphic VR.wav";
106, 2 1; A "Drowned In Sound VR.wav" B "Drowned In Sound.wav";

107, 1 2; A "Beautiful Thing - Waiting (Live at London Water & Steam Museum) [Periphonic Mix].wav" B "Beautiful Thing - Waiting (Live at London Water & Steam Museum) [unmastered].wav";

Researcher's handwritten HULTIGEN scoring sheet



Raw Focus Session Questionnaire Data

Consumer Interview

1. How old are you?

24

2. How would you describe your musical experience?
   (Formal /, informal tuition, self-taught, learnt a un/graded instrument at school, no musical background etc.)

Musical Theatre – diploma (sing, dance, act) I can read music, learnt theory in secondary school

3. Do you listen to music?

Yes

4. How do you usually listen to music? (speakers, TV speakers, home theatre system / surround sound, laptop, car, personal radio, headphones, phone etc.)

Headphones and laptop

5. What types/styles of music do you listen to most?

Pop or country

6. Do you create/record/produce music?

No

7. What type/style of music do you usually create/work with?

None

8. Are you familiar with the term ‘Stereo’? If yes, describe what it means to you.

Yes, it’s when you have two speakers.

9. Are you familiar with the term ‘Periphonic’? If yes, describe what it means to you.

No.

10. Are you familiar with the term ‘Binaural’? If yes, describe what it means to you.

Binaural.

Listening Questionnaire

(Please answer separately for Tracks A and B where specified)

Trial 1

1. What headphones are you listening on?

K271

2. How many times did you listen to each track?

Track a – 3 times, track b four times

3. Which did you prefer, A or B, and why?

Track a had a better sound overall and was easier to listen to and immerse yourself in the music

4. What does the record as a whole invoke for you?

Not sure

5. How would you explain the sonic space / environment?

   Track A –Wide area, echoing

   Track B-smaller area, sounds closer

6. Describe the placement of sounds you hear in each version.

   Track A –voice sounds like it's coming from directly in front of you, band sounds like it's behind you

   Track B-guitar to the right, drums to the left, sax and vocals coming from the centre

7. Did you perceive any sounds to be moving in either track? Please describe.

   Track A – brass instruments sounded like they were in motion

   Track B- the vocals seemed to move midway through the track

8. Did you perceive any sounds from above? Please describe.

   Track A-no

   Track B-no

9. Did you perceive any sounds from behind? Please describe.

   Track A- instruments from the band e.g. guitar and drums

   Track B- vocals occasionally sounded like they were coming from behind, sax consistently sounds like it's behind the listener

10. Did you perceive any sound coming directly from either side of you Left and or Right?
    Please describe.

    Track A-music sounded like it was coming from all around you

    Track B –guitar to the left, drums to the right

11. Did you perceive the sound to always be coming from in front of you? Please describe.

    Track A – the vocal track sounded like it was coming from in front

    Track B- no, the track was not always coming from in front

12. Do you imagine or visualise what you hear when listening to the music?

    Track A- imagined a band in a garage

    Track B-no

13. Is there anything you particularly liked/disliked about the production?

    Track A-quite a lot of echoing like it was produced in a wide open space, no cohesion with the band, sounded like they could have been recorded separately

    Track B-didn't like the production of the vocal track, drops in volume at times

14. Do you feel immersed in the music?

    Track A-yes

    Track B-no

15. Did you experience the music as if it was coming from all around you?

    Track A – there was a definite perception that the music was coming from multiple places

    Track B –at times but was inconsistent

16. Did you feel as if the sound was coming from inside or outside of your head?

    Track A-outside

    Track B-inside at moments

17. Were the tracks comfortable or uncomfortable to listen to? Why?

    Track A- more comfortable, more cohesive sound in terms of the mix between the vocals and band

    Track B-less comfortable, vocal track seemed to drop in quality at moments

18. Which of the versions would you rather listen to, A or B?

    Track A seemed to have a better mix and sound quality overall, track b wasn't as full sounded and had drops in volume and sound quality at moments

19. Would you buy music in either of these formats?

    Track A – more likely to buy music in this format

    Track B-unlikely to buy music in this format

20. Do you think production suits the type of music?

    Track A – more suitable, suits the acoustic nature of the music

    Track B – less suitable, feels disjointed, highlights differences between vocals and band

21. Is there anything else you would like to add about the music, mixes or your experience?

No

Listening Questionnaire

(Please answer separately for Tracks A and B where specified)

Trial 2

1. What headphones are you listening on? K271

2. How many times did you listen to each track?

Track A twice, track B 3 times

3. Which did you prefer, A or B, and why?

Track A was more immersive and felt more cohesive in terms of the mix

4. What does the record as a whole invoke for you?

Intrigue, tells an interesting story

5. How would you explain the sonic space / environment?

   Track A –wide space

   Track B-more intimate but still a large space

6. Describe the placement of sounds you hear in each version.

   Track A –tinkling sounds and clicks coming from the right, vocals from behind, drops coming from all around

   Track B-clicking and tinkling sounds left to right, vocals from behind and drops from behind

7. Did you perceive any sounds to be moving in either track? Please describe.

   Track A – continual rise and fall in the sounds

   Track B- felt like things were moving left to right and right to left throughout the track

8. Did you perceive any sounds from above? Please describe.

   Track A-occasional sounds from above, sounds coming from all around

   Track B-yes, some of the clicks sounded like they were coming from above

9. Did you perceive any sounds from behind? Please describe.
   Track A- occasional sounds from behind sounds coming from all around

   Track B-vocals and the drops sounded like they were coming from behind

10. Did you perceive any sound coming directly from either side of you Left and or Right?
    Please describe.

    Track A- clicks and tinny noises coming from the right

    Track B – clicks and tinny sounds coming from the left

11. Did you perceive the sound to always be coming from in front of you? Please describe.

    Track A – sound constantly felt like it was all around including in front

    Track B-no felt more like it was coming from above and behind me

12. Do you imagine or visualise what you hear when listening to the music?

    Track A- imagined the intro to a movie like a thriller or noir murder mystery

    Track B-a bar or casino

13. Is there anything you particularly liked/disliked about the production?

    Track A-liked the mix and how full the sound was

    Track B-liked the production

14. Do you feel immersed in the music?

    Track A-yes

    Track B-yes

15. Did you experience the music as if it was coming from all around you?

    Track A - yes

    Track B – not consistently

16. Did you feel as if the sound was coming from inside or outside of your head?

    Track A-inside

    Track B-outside

17. Were the tracks comfortable or uncomfortable to listen to? Why?

    Track A- comfortable, felt immersive

    Track B-comfortable, not as immersive as A though

18. Which of the versions would you rather listen to, A or B?

    Track A felt more immersive

19. Would you buy music in either of these formats?

    Track A – more likely

    Track B-less likely

20. Do you think production suits the type of music?

    Track A - yes

    Track B - yes

21. Is there anything else you would like to add about the music, mixes or your experience?

No

Listening Questionnaire

(Please answer separately for Tracks A and B where specified)

Trial 3

1. What headphones are you listening on? K271

2. How many times did you listen to each track?

Twice

3. Which did you prefer, A or B, and why?

Track B felt more personal and intimate

4. What does the record as a whole invoke for you?

Makes me feel sad and cold

5. How would you explain the sonic space / environment?

    Track A –big space like an empty room

    Track B-more intimate, smaller sounding

6. Describe the placement of sounds you hear in each version.

   Track A –male voice coming from the left, female voice from the right, the band sounds like it's coming from below

   Track B- vocals from left and right, piano sounds like it's above

7. Did you perceive any sounds to be moving in either track? Please describe.

   Track A – no

   Track B- no

8. Did you perceive any sounds from above? Please describe.

   Track A-no

   Track B-piano track

9. Did you perceive any sounds from behind? Please describe.

   Track A-the piano at points

   Track B-no

10. Did you perceive any sound coming directly from either side of you Left and or Right?
    Please describe.

Track A- male voice coming from the left, female voice from the right

Track B –both vocals were coming from left and right

11. Did you perceive the sound to always be coming from in front of you? Please describe.

    Track A - no

    Track B-no

12. Do you imagine or visualise what you hear when listening to the music?

    Track A- I imagined winter, frozen lake

    Track B-an empty house

13. Is there anything you particularly liked/disliked about the production?

    Track A-occasionally vocals felt disconnected

    Track B-it felt more personal

14. Do you feel immersed in the music?

    Track A-no

    Track B-yes

15. Did you experience the music as if it was coming from all around you?

Track A – at points but inconsistently

Track B –more consistently than track a

16. Did you feel as if the sound was coming from inside or outside of your head?

    Track A-outside

    Track B-inside

17. Were the tracks comfortable or uncomfortable to listen to? Why?

    Track A- comfortable but less personal

    Track B-comfortable, felt more intimate

18. Which of the versions would you rather listen to, A or B?

    Track B

19. Would you buy music in either of these formats?

    Track A – possibly depending on the track

    Track B-yes

20. Do you think production suits the type of music?

    Track A – no, doesn't suit the style of song as much

Track B – yes, accentuates the melancholy of the song

21. Is there anything else you would like to add about the music, mixes or your experience?

I liked both and felt like they both worked in different ways but much preferred track b overall

Listening Questionnaire

(Please answer separately for Tracks A and B where specified)

Trial 4

1. What headphones are you listening on?

2. How many times did you listen to each track?

Track A twice, Track B three times

3. Which did you prefer, A or B, and why?

Track a felt more cohesive and had a fuller sound, track b felt disconnected

4. What does the record as a whole invoke for you?

Has a nice beat, makes me tap my foot

5. How would you explain the sonic space / environment?

Track A –intimate space like a bar

Track B-larger area, occasional echoing

6. Describe the placement of sounds you hear in each version.

   Track A –VOCALS FROM THE FRONT, DRUMS TO THE LEFT

   Track B-piano from above, vocals from behind

7. Did you perceive any sounds to be moving in either track? Please describe.

   Track A – no

   Track B- no

8. Did you perceive any sounds from above? Please describe.

   Track A-no

   Track B-piano

9. Did you perceive any sounds from behind? Please describe.
   Track A-backing vocals

   Track B-lead and backing vocals

10. Did you perceive any sound coming directly from either side of you Left and or Right?
    Please describe.

Track A-drum track from the left

Track B –lead vocals from the left, backing from the right

11. Did you perceive the sound to always be coming from in front of you? Please describe.

    Track A – main vocals consistently felt like they were in front

    Track B-no

12. Do you imagine or visualise what you hear when listening to the music?

    Track A- smoky bar

    Track B-not sure

13. Is there anything you particularly liked/disliked about the production?

    Track A-full sound, good mix of lead and backing vocals with the band

    Track B-vocals sounded far away

14. Do you feel immersed in the music?

    Track A-yes

    Track B-no

15. Did you experience the music as if it was coming from all around you?

Track A - yes

Track B -no

16. Did you feel as if the sound was coming from inside or outside of your head?

    Track A-inside

    Track B-outside

17. Were the tracks comfortable or uncomfortable to listen to? Why?

    Track A- comfortable, warm sound

    Track B-less comfortable, didn't feel like it was all recorded at the same time

18. Which of the versions would you rather listen to, A or B?

    track a

19. Would you buy music in either of these formats?

    Track A - yes

    Track B-no

20. Do you think production suits the type of music?

Track A - yes

Track B - no

21. Is there anything else you would like to add about the music, mixes or your experience?

No

Listening Questionnaire

(Please answer separately for Tracks A and B where specified)

Trial 5

1. What headphones are you listening on? K271

2. How many times did you listen to each track?

Track A three times, track B twice

3. Which did you prefer, A or B, and why?

Track B, easier to distinguish different instruments and vocal lines

4. What does the record as a whole invoke for you?

Uneasy

5. How would you explain the sonic space / environment?

   Track A –wide open space like a church

   Track B-same as track a

6. Describe the placement of sounds you hear in each version.

Track A –vocals from behind, drums from behind but moving forward

Track B- vocals sound separate, some from in front and some from behind

7. Did you perceive any sounds to be moving in either track? Please describe.

   Track A – drum beats sounded like they were moving from behind to in front of me

   Track B- certain vocal lines sound like they are getting further away

8. Did you perceive any sounds from above? Please describe.

   Track A-no

   Track B-high female voice

9. Did you perceive any sounds from behind? Please describe.
   Track A-drums and vocals

   Track B-certain vocal lines

10. Did you perceive any sound coming directly from either side of you Left and or Right?

Please describe.

Track A-no

Track B –drums and guitar are coming from left and right

11. Did you perceive the sound to always be coming from in front of you? Please describe.

    Track A - no

    Track B-no

12. Do you imagine or visualise what you hear when listening to the music?

    Track A- empty chamber

    Track B-four nuns in a room getting down

13. Is there anything you particularly liked/disliked about the production?

    Track A-liked the beat and the instrumental line, didn't like how all vocals melded together

    Track B-liked being able to distinguish different vocal lines

14. Do you feel immersed in the music?

    Track A-yes

Track B-yes

15. Did you experience the music as if it was coming from all around you?

    Track A - yes

    Track B -yes

16. Did you feel as if the sound was coming from inside or outside of your head?

    Track A-outside

    Track B-outside

17. Were the tracks comfortable or uncomfortable to listen to? Why?

    Track A- comfortable when the beat was dropping, less comfortable with the vocal lines on their own

    Track B-more comfortable, preferred the mix of the vocal line with the beat

18. Which of the versions would you rather listen to, A or B?

    Track B

19. Would you buy music in either of these formats?

    Track A - yes

Track B-yes

20. Do you think production suits the type of music?

    Track A - yes

    Track B - yes

21. Is there anything else you would like to add about the music, mixes or your experience?

Vocal line reminded me of a Gregorian chant

Listening Questionnaire

(Please answer separately for Tracks A and B where specified)

Trial 6

1. What headphones are you listening on? K271

2. How many times did you listen to each track?

Twice each

3. Which did you prefer, A or B, and why?

Track A was very immersive

4. What does the record as a whole invoke for you?

Anxious

5. How would you explain the sonic space / environment?

    Track A –sounds like it’s outside, in a wide area

Track B-sounds like a bathroom

6. Describe the placement of sounds you hear in each version.

   Track A –running water from behind, squeaking noise to the right, heavy breathing noises to the left

   Track B-squeaks and clicks from left and right, draining water sound from above, strings from behind

7. Did you perceive any sounds to be moving in either track? Please describe.

   Track A – sound of water running downwards

   Track B- running water from above moving downwards

8. Did you perceive any sounds from above? Please describe.

   Track A-no

   Track B-water running down a drain

9. Did you perceive any sounds from behind? Please describe.
   Track A-running water

   Track B-strings

10. Did you perceive any sound coming directly from either side of you Left and or Right?
    Please describe.

    Track A-squeaking noise and heavy breathing to the left

    Track B –clicks and squeaks

11. Did you perceive the sound to always be coming from in front of you? Please describe.

    Track A - no

    Track B-at moments but not consistently

12. Do you imagine or visualise what you hear when listening to the music?

    Track A- someone running near water like a river or stream

    Track B-bath running and overflowing

13. Is there anything you particularly liked/disliked about the production?

    Track A-liked the suffocating nature of production, suited the track

    Track B-disliked that some elements got lost in this mix

14. Do you feel immersed in the music?

    Track A-yes

    Track B-yes

15. Did you experience the music as if it was coming from all around you?

    Track A - yes

    Track B -yes

16. Did you feel as if the sound was coming from inside or outside of your head?

    Track A-outside

    Track B-inside

17. Were the tracks comfortable or uncomfortable to listen to? Why?

    Track A- uncomfortable, moving sounds started to feel suffocating as track went on

    Track B-uncomfortable, felt like I was listening underwater

18. Which of the versions would you rather listen to, A or B?

    Track A

19. Would you buy music in either of these formats?

    Track A - yes

    Track B-no

20. Do you think production suits the type of music?

    Track A - yes

    Track B - no

21. Is there anything else you would like to add about the music, mixes or your experience?

Terrifying

Listening Questionnaire

(Please answer separately for Tracks A and B where specified)

Trial 7

1. What headphones are you listening on?

K271

2. How many times did you listen to each track?

Twice

3. Which did you prefer, A or B, and why?

Track B had a nicer blend and mix of vocals and band

4. What does the record as a whole invoke for you?

Happiness

5. How would you explain the sonic space / environment?

    Track A –sounds like a studio space

    Track B-very similar to track a but smaller

6. Describe the placement of sounds you hear in each version.

Track A –vocals from all around, drums to the left, bass to the right, backing vocals from behind

Track B-violin to the right, vocals from in front and behind, drums to the left

7. Did you perceive any sounds to be moving in either track? Please describe.

   Track A – vocals move from left to right

   Track B- no

8. Did you perceive any sounds from above? Please describe.

   Track A-no

   Track B-no

9. Did you perceive any sounds from behind? Please describe.

   Track A-backing vocals

   Track B-backing vocals

10. Did you perceive any sound coming directly from either side of you Left and or Right?
    Please describe.

Track A-lead vocals, drums

Track B –lead vocals, drums, violin

11. Did you perceive the sound to always be coming from in front of you? Please describe.

    Track A - no

    Track B-yes

12. Do you imagine or visualise what you hear when listening to the music?

    Track A- music studio

    Track B-professional music video

13. Is there anything you particularly liked/disliked about the production?

    Track A-liked the fullness of the vocal track, felt prevalent

    Track B-liked the cohesiveness of production, vocals blended well with band

14. Do you feel immersed in the music?

    Track A-yes

    Track B-yes

15. Did you experience the music as if it was coming from all around you?

Track A - yes

Track B -yes

16. Did you feel as if the sound was coming from inside or outside of your head?

    Track A-outside

    Track B-inside

17. Were the tracks comfortable or uncomfortable to listen to? Why?

    Track A- comfortable, nice full sound

    Track B-comfortable, nice blended mix

18. Which of the versions would you rather listen to, A or B?

    Track B

19. Would you buy music in either of these formats?

    Track A - yes

    Track B-yes

20. Do you think production suits the type of music?

    Track A - yes

Track B - yes

21. Is there anything else you would like to add about the music, mixes or your experience?

No

Listener 7

HULTIGEN Data Output

```
Stimuli Set

1, "Far From Here VR.wav" "Far From Here.wav";
2, "Beautiful Thing - Waiting (Live at London Water & Steam
Museum) [Periphonic Mix].wav" "Beautiful Thing - Waiting (Live at
London Water & Steam Museum) [unmastered].wav";
3, "Penny Drops.wav" "Penny Drops VRv2.wav";
4, "Jerome Thomas - Late Nights (Periphonic Mix Logic).wav" "Late
Nights Demo Vocal Up.mp3";
5, "Monomorphic VR.wav" Monomorphic.wav;
6, "Drowned In Sound.wav" "Drowned In Sound VR.wav";
7, "Sort Yourself Out Mstr1.wav" "Joey Clarkson - Sort Yourself
Out (Live at Brentford Steam Works) [Periphonic Mix].wav";
8, "Art's Self Alteration VR.wav" "Art's Self Alteration
Redbook.wav";


100, Trial Order 2 5 3 7 4 1 8 6;


1, "Trial 2" 81. 100.; "Beautiful Thing - Waiting (Live at London
Water & Steam Museum) [Periphonic Mix].wav" "Beautiful Thing -
Waiting (Live at London Water & Steam Museum) [unmastered].wav";

2, "Trial 5" 75. 100.; , "Monomorphic VR.wav" Monomorphic.wav;

3, "Trial 3" 100. 90.; , "Penny Drops.wav" "Penny Drops
VRv2.wav";

4, "Trial 7" 52. 48.;  "Sort Yourself Out Mstr1.wav" "Joey
Clarkson - Sort Yourself Out (Live at Brentford Steam Works)
[Periphonic Mix].wav";

5, "Trial 4" 92. 100.; "Late Nights Demo Vocal Up.mp3";"Jerome
Thomas - Late Nights (Periphonic Mix Logic).wav"
```

6, "Trial 1" 98. 100.; "Far From Here VR.wav" "Far From Here.wav";

7, "Trial 8" 65. 100.; "Art's Self Alteration Redbook.wav";"Art's Self Alteration VR.wav"

8, "Trial 6" 51. 51.; "Drowned In Sound VR.wav";"Drowned In Sound.wav"

Stimuli AB Order / Scale Correlation

A/B

101, 1 2; "Beautiful Thing - Waiting (Live at London Water & Steam Museum) [Periphonic Mix].wav" "Beautiful Thing - Waiting (Live at London Water & Steam Museum) [unmastered].wav";

102, 2 1; A Monomorphic.wav ; B "Monomorphic VR.wav"

103, 1 2; "Penny Drops.wav" "Penny Drops VRv2.wav";

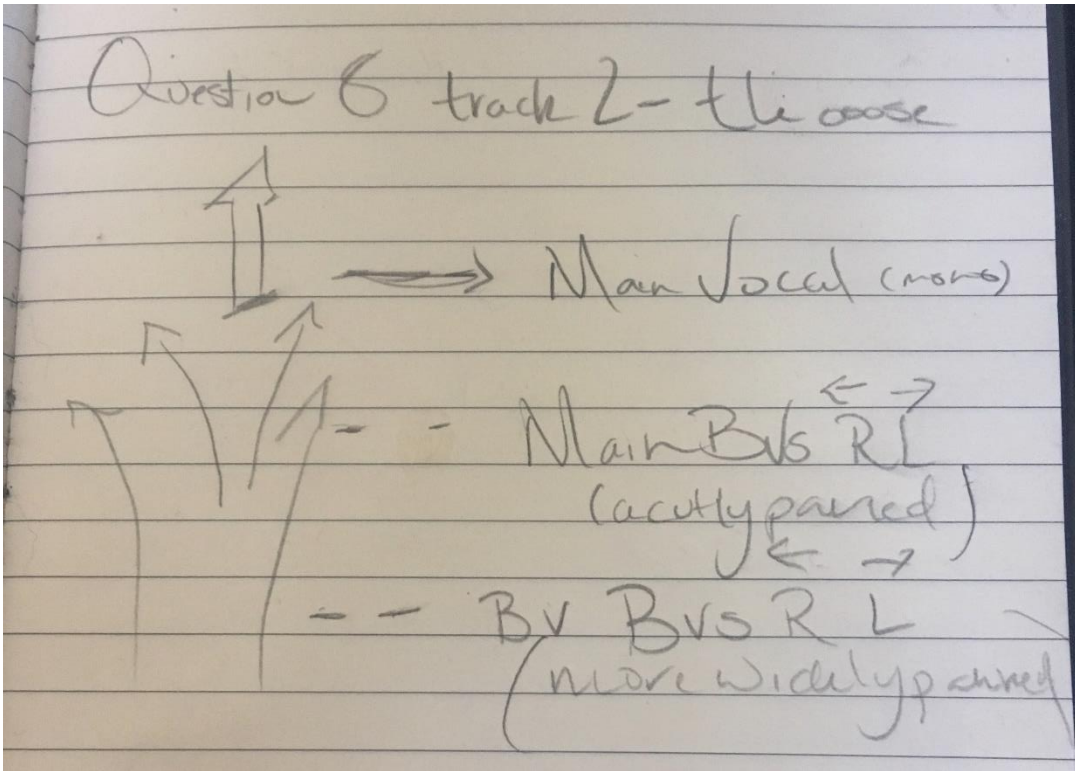
104, 1 2; A "Sort Yourself Out Mstr1.wav" B "Joey Clarkson - Sort Yourself Out (Live at Brentford Steam Works) [Periphonic Mix].wav";

105, 2 1; A "Late Nights Demo Vocal Up.mp3";B "Jerome Thomas - Late Nights (Periphonic Mix Logic).wav"

106, 1 2; A "Far From Here VR.wav" B "Far From Here.wav";
107, 2 1; A "Art's Self Alteration Redbook.wav"; B "Art's Self Alteration VR.wav"

108, 2 1; A "Drowned In Sound VR.wav" B "Drowned In Sound.wav"

Hand drawn vocal placements in Monomorphic in support of answer to Q8 in evaluation.



Raw Focus Session Questionnaire Data

Consumer Interview

1. How old are you?

29

2. How would you describe your musical experience?
   (Formal /, informal tuition, self-taught, learnt a un/graded instrument at school, no musical background etc.)

Self-taught, FE in Music Production, Rapper

3. Do you listen to music?

Yes. All the time

4. How do you usually listen to music? (speakers, TV speakers, home theatre system / surround sound, laptop, car, personal radio, headphones, phone etc.)

Headphones

5. What types/styles of music do you listen to most?

Rap, Trap, Funk, Instrumental, G Funk, Jazz, Swing, Rhythm n Blues,

6. Do you create/record/produce music?

Yes, to all.

7. What type/style of music do you usually create/work with?

8. Are you familiar with the term ‘Stereo’? If yes, describe what it means to you.

Yes, left and right. 180 degrees. Listening in headphones.

9. Are you familiar with the term ‘Periphonic’? If yes, describe what it means to you.

No, I’m not.

10. Are you familiar with the term ‘Binaural’? If yes, describe what it means to you.

Yes, listening with your ears and your head

Listening Questionnaire

(Please answer separately for Tracks A and B where specified)

Trial 1

1. What headphones are you listening on?

AKGk271

2. How many times did you listen to each track?

3. Which did you prefer, A or B, and why?

4. What does the record as a whole invoke for you?

Get up and keep keeping it gangster!

‘

5. How would you explain the sonic space / environment?

Track A –excellent use of space. Vibrant.

Crisp and precise use of eq compression and scarce used fx give what I felt was a very direct connection to temporal and melodic exciters.

Track B-

A somewhat more immersive experience. However, the spaced style doesn't immediately catch my ear.

That being said, halfway the track comes together with the sum of all it parts and the grander perceived space does then have its advantages.

Overall

Big! Talented musicians making well executed strokes and grand design shows development though t an overall vision!

6. Describe the placement of sounds you hear in each version.

Track A –upfront

Bam

Tight

But very individual

Track B-almost a curry bean formation? With the outer most apex of the bean pointing the same direction as I'm looking

The vocalist may be behind me even but more li his head is directly above mine singing out to the semi circle from right to left of drums. Bass on to the keys. Etc. but all of the instrumentation sounds as if in the floor?? That's just what I hear.

7. Did you perceive any sounds to be moving in either track? Please describe.

Track A – no

Track B- vocals but only through the perceived space not actual motion /position

8. Did you perceive any sounds from above? Please describe.

Track A-// kind of vocals but maybe just upfront.

Track B- vocals

9. Did you perceive any sounds from behind? Please describe.

Track A-no

Track B- there very beginning the clonking glass may just be behind me

10. Did you perceive any sound coming directly from either side of you Left and or Right?

Please describe.

Track A-no

Track B -no

11. Did you perceive the sound to always be coming from in front of you? Please describe.

Track A – no the space was used exquisitely not just every sound cutting through but also having its own possession in the place and you notice the moment they add an element or take one away and you miss it..

Track B- ‘’

12. Do you imagine or visualise what you hear when listening to the music?

Track A- white people music videos of passionate drummers in the sunset.

Track B-‘’

13. Is there anything you particularly liked/disliked about the production?

Track A-the mix was very bright and jumped out at me grabbed my attention and kept it

Track B- harder to get into and took time before you were able to see why things were done the way the where done

14. Do you feel immersed in the music?

Track A- very much so

Track B- intriguing worth the wait to the end but less likely to be played that far.

15. Did you experience the music as if it was coming from all around you?

Track A – no stereo

Track B – in front of me and far at points

16. Did you feel as if the sound was coming from inside or outside of your head?

Track A- out

Track B- out

17. Were the tracks comfortable or uncomfortable to listen to? Why?

Track A- no

Track B- no

18. Which of the versions would you rather listen to, A or B? A

A

19. Would you buy music in either of these formats?

Would need more data to choose between them but yes I would pay for this quality for sure.

Track A n/a

Track B-n/a

20. Do you think production suits the type of music?

Track A - yes

Track B - yes

21. Is there anything else you would like to add about the music, mixes or your experience?

banger

Listening Questionnaire

(Please answer separately for Tracks A and B where specified)

Trial 2

1. What headphones are you listening on?

AKG K271

2. How many times did you listen to each track?

3. Which did you prefer, A or B, and why?

4. What does the record as a whole invoke for you?

A feeling of eerie precision with sinister intent.

Siren song

5. How would you explain the sonic space / environment?

Track A –up and down. Percussion acts like piston on the floor of an open tall box the has no top and the rest of the noise escapes in different for me out the top

Track B-still epic yes definitely flatter rhythm section loses drive and no longer is the hard driving force muddled at parts

6. Describe the placement of sounds you hear in each version.

Track A –rhythm bottom

All other sound rising for the most part.

Track B-vox down the centre

Guitar slight right

Rhythm off to the left

Underlying bass pad

7. Did you perceive any sounds to be moving in either track? Please describe.

Track A – vox.. up and away

Track B- no.. filling space gradually

8. Did you perceive any sounds from above? Please describe.

Track A-

Definitely sound seems to travel upward into the wind if one would.

Track B- higher pitched vocal harmonies however, again, my just be an exceptional use of the mono channel and subsequent acute panning to create an upfront effect and backups either side like peddles of a blooming flower

E.g.

I – vox1

V - bv1

V- bv2

Consult diagram from journal at the end of the questionnaire.

9. Did you perceive any sounds from behind? Please describe.

Track A-no

Track B-no

10. Did you perceive any sound coming directly from either side of you Left and or Right?

Please describe.

Track A-

Track B -

11. Did you perceive the sound to always be coming from in front of you? Please describe.

Track A - yes

Track B-yes

12. Do you imagine or visualise what you hear when listening to the music?

Track A- machine of marvels inner workings demonstrating the precision and design

Track B- atmospheric ambiance maybe in death or a dream

13. Is there anything you particularly liked/disliked about the production?

Track A-the drum machine pattern & paunchiness. Brightness. Choice of enfaces instrument wise.

Track B- overall concept and mix are evocative and enable the listener to wonder freely

14. Do you feel immersed in the music?

Track A-very

Track B-able to drift away to but does not grasp me instantly like a

15. Did you experience the music as if it was coming from all around you?

Track A - yes filled space – captivated

Track B – yea sparsely spaced and distant

16. Did you feel as if the sound was coming from inside or outside of your head?

Track A- all in front

Track B-

17. Were the tracks comfortable or uncomfortable to listen to? Why?

Track A- no

Track B- no

18. Which of the versions would you rather listen to, A or B?

A

19. Would you buy music in either of these formats?

Track A – Most definitely.

Track B-need more data. I would like to listen out of other transducers to see how things translate

20. Do you think production suits the type of music?

Track A – yes

Track B - somewhat

21. Is there anything else you would like to add about the music, mixes or your experience?

Listening Questionnaire

(Please answer separately for Tracks A and B where specified)

Trial 3

1. What headphones are you listening on?

AKG K271

2. How many times did you listen to each track?

Several

3. Which did you prefer, A or B, and why?

A - What can I say seems every time the A mix is far more bright!! And attention to detail made easier by what seems to be proximity to each individual element down to the lifespan from attack to the release of every single sample

4. What does the record as a whole invoke for you?

Aww of the recordings that have been sampled to create the tapestry. The detail is hyper presentational... magnifying... and illuminating.

5. How would you explain the sonic space / environment?

Track A – dead nothing but the intended sound…. I like that

Track B-abyss

6. Describe the placement of sounds you hear in each version.

Track A –moving almost to a visual level along each sound presented... close up sound to accentuate the relationship between whatever two object interact to create their resulting sample.

Track B- more in front but up and down... in front of me

7. Did you perceive any sounds to be moving in either track? Please describe.

Track A – the whole point of the soundscape to me seems to be a show of the production team's ability to convey movement within sound.

Track B- same effect yet not as strong

8. Did you perceive any sounds from above? Please describe.

Track A- in front and inside

Track B- yes in fact most of the mix is above me t as effective

9. Did you perceive any sounds from behind? Please describe.

Track A-no but I do feel like you are somewhat able to hear the back of some of the sounds... sounds and on surround NOT SURROUND SOUND... 3D

Track B-same effect but not a strong

10. Did you perceive any sound coming directly from either side of you Left and or Right?

Please describe.

Track A-again... seems as if the sound is moving around objects as you observe their existence with your ears... rather than where you are. It's about what/where is the emitting sound source

Track B – same effect not as strong

11. Did you perceive the sound to always be coming from in front of you? Please describe.

Track A - moving with the centred sample and rest of space used well

Track B-same effect not as strong

12. Do you imagine or visualise what you hear when listening to the music?

Track A- more than any other song so far... the imagery is so vivid, that the fact my visual cortex is not too being stimulated is a bit difficult to reconcile at times

Track B- same FX not as evident

13. Is there anything you particularly liked/disliked about the production?

Track A-the amazing recording technique used when gathering these samples

Track B- ‘’

14. Do you feel immersed in the music?

Track A-mesmerised

Track B- ‘’

15. Did you experience the music as if it was coming from all around you?

Track A - I more moved to its whim

Track B - ‘’

16. Did you feel as if the sound was coming from inside or outside of your head?

Track A- no like was in the space of a 3 dimensional sound examiner chamber of sorts

Track B- same FX not a strong

17. Were the tracks comfortable or uncomfortable to listen to? Why?

Track A-

Track B-

18. Which of the versions would you rather listen to, A or B?

19. Would you buy music in either of these formats?

Track A -

Track B-

20. Do you think production suits the type of music?

Track A - amazing impressive and perception exciting

Track B – amazing

21. Is there anything else you would like to add about the music, mixes or your experience?

Do it again please

Listening Questionnaire

(Please answer separately for Tracks A and B where specified)

Trial 4

1. What headphones are you listening on?

AKG K271

2. How many times did you listen to each track?

3. Which did you prefer, A or B, and why?

4. What does the record as a whole invoke for you?

n/a

5. How would you explain the sonic space / environment?

   Track A – live

Track B- multi tracked

6. Did you perceive any sounds to be moving in either track? Please describe.

   Track A – no

   Track B- no

7. Did you perceive any sounds from above? Please describe.

   Track A- no

   Track B-no

8. Did you perceive any sounds from behind? Please describe.
   Track A-no well maybe the drum kit

   Track B-no

9. Did you perceive any sound coming directly from either side of you Left and or Right?
   Please describe.

   Track A-no

   Track B –vox seem to spread the whole spectrum like maybe they have been spread or they are very upfront

10. Did you perceive the sound to always be coming from in front of you? Please describe.

    Track A - ye

Track B- ye

11. Do you imagine or visualise what you hear when listening to the music?

    Track A- not really live vibes

    Track B- no

12. Is there anything you particularly liked/disliked about the production?

    Track A- just not feeling the track too much

    Track B- brighter better polish but still not the bomb

13. Do you feel immersed in the music?

    Track A- no

    Track B-no

14. Did you experience the music as if it was coming from all around you?

    Track A - no

    Track B - no

15. Did you feel as if the sound was coming from inside or outside of your head?

    Track A-outside

    Track B- outside

16. Were the tracks comfortable or uncomfortable to listen to? Why?

Track A- kind of uncomfortable, vocals are whack (sorry), overall instrumentation is boring played out and lacking imagination and soul

Track B- ‘’

17. Which of the versions would you rather listen to, A or B?

N/A

18. Would you buy music in either of these formats?

Track A – 50/50

Track B- yes sound is fair enough

19. Do you think production suits the type of music?

Track A - more than b

Track B - less than a

20. Is there anything else you would like to add about the music, mixes or your experience?

Na I’m cool

Listening Questionnaire

(Please answer separately for Tracks A and B where specified)

Trial 5

1. What headphones are you listening on?

AKG K271

2. How many times did you listen to each track?

3. Which did you prefer, A or B, and why?

B

4. What does the record as a whole invoke for you?

Spoke to my soul as result of the vocalist's performance... he seems to bear his soul honestly

5. How would you explain the sonic space / environment?

   Track A –chamber... small space enclosed

   Track B- spacious, immersive, and hard to determine the boundaries however, enclosed to ensure nothing but desired sound

6. Describe the placement of sounds you hear in each version.

   Track A – close together. Mix is good but not much to say for position of elements

   Track B- layered, beautify mixed , instrumentation scattered around the space each section with its own space no crowding perfect for the style of genre. Immersive effect on the listener is created by the placement of bvs has an empathetic, first person, kind of feel. As if the listener is in the artists mind/ thought

7. Did you perceive any sounds to be moving in either track? Please describe.

   Track A – synth bass percussive boom in the drum pattern has a downward effect but most likely created by some sort of pitch shift tool rather than making use of space manipulation

   Track B- less movement more set positions yet the sounds come from all over making the attention of the listener move around the audible scape

8. Did you perceive any sounds from above? Please describe.

   Track A- no

   Track B- vox main & bv and keys somewhat

9. Did you perceive any sounds from behind? Please describe.
   Track A-//

   Track B- bvs

10. Did you perceive any sound coming directly from either side of you Left and or Right?
    Please describe.

    Track A- no

    Track B – click seams to move from side to side. Bvs also coming in from all sides but off side

11. Did you perceive the sound to always be coming from in front of you? Please describe.

    Track A - yes

Track B- no all over immersive

12. Do you imagine or visualise what you hear when listening to the music?

    Track A-

    Track B-space the actual feeling that you are lost in the state on uncertainty the lyrics elude to

13. Is there anything you particularly liked/disliked about the production?

    Track A- great

    Track B-touched my soul... couldn't imagine a more effective execution... the song is a gift

14. Do you feel immersed in the music?

    Track A- // more so captivating

    Track B- totally

15. Did you experience the music as if it was coming from all around you?

    Track A – no

    Track B – exceptionally

16. Did you feel as if the sound was coming from inside or outside of your head?

    Track A-

    Track B-more like my head is inside the realm of the music than the other way

17. Were the tracks comfortable or uncomfortable to listen to? Why?

    Track A- comfy

    Track B- sole soothin.... live a mobo lozenge

18. Which of the versions would you rather listen to, A or B?

B

19. Would you buy music in either of these formats?

    Track A – might do

    Track B-exclusively

20. Do you think production suits the type of music?

    Track A – enough to get by

    Track B - more than I have heard in most other attempts at this particular sound

21. Is there anything else you would like to add about the music, mixes or your experience?

Hot shit

Listening Questionnaire

(Please answer separately for Tracks A and B where specified)

Trial 6

1. What headphones are you listening on?

AKG K271

2. How many times did you listen to each track?

Several

3. Which did you prefer, A or B, and why?

Hard to say both have their merits and pitfalls

4. What does the record as a whole invoke for you?

5. How would you explain the sonic space / environment?

   Track A – soft yet reflective, not solid/permeable, much of the sounds become/emerge from depths rather than get introduced

   Track B- ‘’ brighter

6. Describe the placement of sounds you hear in each version.

   Track A – same as b but softer overall

   Track B- layered – or- opposing partnerships rather provide the textures... Elements work in pairs. All attributes seem to be in fact teams of two elements opposing at all times. either above & below one another or coming in form opposing sides, or rhythmically one sounds then leaves space for the other (EVEN WHEN THE ELEMENTS ARE MALODIC THEY STILL MAY USE THIS RYTHMIC TECHNIQUE) extremely innovative

7. Did you perceive any sounds to be moving in either track? Please describe.

Track A – same as b but softer overall

Track B- underlying bass pad gives an effect of forward movement big time and where it is placed at the bottom of the perceivable space and just off centre, it's like an airport walkway conveyor belt that brings the whole song along with it .. Then the rhythm following segment containing the soft beaten drums have a definite motion feel... locomotion like

8. Did you perceive any sounds from above? Please describe.

   Track A- yes but dynamically this affect seems more to do with tonality and pitch than space

   Track B- definite emission of sound on a latitudinal scale

9. Did you perceive any sounds from behind? Please describe.
   Track A- no

   Track B- many sounds seem to start from behind and move to the forefront however, the song has a forward motion progression vibe that may be the main contributor to this particular effect

10. - Did you perceive any sound coming directly from either side of you Left and or Right?
Please describe.

Track A - same as b but softer overall

Track B- soft hydraulic pump sounding synth & keyboards seem to be quite directional

11. Did you perceive the sound to always be coming from in front of you? Please describe.

    Track A - yes

    Track B- immersive, enveloping

12. Do you imagine or visualise what you hear when listening to the music?

Track A - same as b but softer overall

Track b - the abys, becoming of things, growth, journey, serenity, acceptance of undesired truth and the ‘carry on with it’ beginning of living with the fact

13. Is there anything you particularly liked/disliked about the production?

    Track A- all

    Track B- everything its pure beauty

14. Do you feel immersed in the music?

Track A - same as b but softer overall

Track B -bigtime and moved by the time you come out of the experience

15. Did you experience the music as if it was coming from all around you?

    Track A - same as b but softer overall

Track B -yes

16. Did you feel as if the sound was coming from inside or outside of your head?

    Track A-yes

    Track B- yes

17. Were the tracks comfortable or uncomfortable to listen to? Why?

    Track A- don't any me... it sounds like God made it

    Track B- ‘’

18. Which of the versions would you rather listen to, A or B?

    b

19. Would you buy music in either of these formats?

    Track A – ye

    Track B-100000001%

20. Do you think production suits the type of music?

    Track A - same as b but softer overall

    Track B - unrivalled

21. Is there anything else you would like to add about the music, mixes or your experience?

Listening Questionnaire

(Please answer separately for Tracks A and B where specified)

Trial 7

1. What headphones are you listening on?

AKG K271

2. How many times did you listen to each track?

Several

3. Which did you prefer, A or B, and why?

A of course

4. What does the record as a whole invoke for you?

5. How would you explain the sonic space / environment?

   Track A – ever morphing (realistic adjacent) a genuine soundscape. A plane of existence illustrated with audio... exceptional example of sound movement. Places listener in a first person plane where you are no longer objective observer, but part and parcel of an audibly manifested ‘moment of happening’ in perceivable time and space... pretty fuckin ace

   Track B- somewhat flat... Dynamic, however, does seem as if its aim is to convey imagery that warrants at least more punch. Maybe more abrupt transitions and harsher louder versions would jar the listener a bit more

6. Describe the placement of sounds you hear in each version.

Track A – aw mate... literally all perceivable directions... It's an audio encoded ecosystem immersing the participant in the epicentre of its real-time reality

Track B-somewhat flat. Good concept for a soundscape could use expanding

7. Did you perceive any sounds to be moving in either track? Please describe.

   Track A – big time. And with exceptional realism

   Track B- somewhat

8. Did you perceive any sounds from above? Please describe.

   Track A-above and beyond

   Track B-

9. Did you perceive any sounds from behind? Please describe.

   Track A- behind and the continuous movement over listener to the other side with an expert realness factor. Even so far as to include the Doppler Effect in relationship to the 'observers' positioning

   Track B-

10. Did you perceive any sound coming directly from either side of you Left and or Right?
    Please describe.

    Track A- yes

    Track B -

11. Did you perceive the sound to always be coming from in front of you? Please describe.

    Track A – yes

    Track B-yes

12. Do you imagine or visualise what you hear when listening to the music?

    Track A- entirely evocative of the evils of war made like a montage

    Track B- ‘’

13. Is there anything you particularly liked/disliked about the production?

    Track A- the realistic movement of the overhead missile

    Track B- concept

14. Do you feel immersed in the music?

    Track A-yes

    Track B-less so

15. Did you experience the music as if it was coming from all around you?

    Track A – the most

    Track B -no

16. Did you feel as if the sound was coming from inside or outside of your head?

    Track A- out

    Track B-out

17. Were the tracks comfortable or uncomfortable to listen to? Why?

    Track A- I'm cool however, probably intended to unsettle

    Track B- ''

18. Which of the versions would you rather listen to, A or B?

A

19. Would you buy music in either of these formats?

    Track A – and pay good money

    Track B- 50/50

20. Do you think production suits the type of music?

    Track A - yes

    Track B – less so

21. Is there anything else you would like to add about the music, mixes or your experience?

Good work... this shit need to be in games like call of duty and army commando

Listening Questionnaire

(Please answer separately for Tracks A and B where specified)

Trial 8

1. What headphones are you listening on?

AKG K271

2. How many times did you listen to each track?

Several

3. Which did you prefer, A or B, and why?

B

4. What does the record as a whole invoke for you?

Time running out imminent peril... and maybe some undersee aliens

5. How would you explain the sonic space / environment?

   Track A – completely engulfed water rapids

   Track B- upfront. Bathroom, contradicting (light-hearted tonality yet conveying uncertainty and build-up)

6. Describe the placement of sounds you hear in each version.

   Track A - tones upfront running water focal

   Track B- tones/ glitches focal running water background but rushing & gushing

7. Did you perceive any sounds to be moving in either track? Please describe.

   Track A – very little

   Track B- very little

8. Did you perceive any sounds from above? Please describe.

   Track A-no

   Track B- no

9. Did you perceive any sounds from behind? Please describe.
   Track A-no

   Track B-no

10. Did you perceive any sound coming directly from either side of you Left and or Right?
    Please describe.

    Track A- no

    Track B –no

11. Did you perceive the sound to always be coming from in front of you? Please describe.

    Track A - ye

    Track B-ye

12. Do you imagine or visualise what you hear when listening to the music?

    Track A- running a bath

    Track B- creepy kid running a bath… or even a paedophile running as bath as the tonality is innocent but played sinisterly and only at the end do we realise we are the ones under water and out of breath

13. Is there anything you particularly liked/disliked about the production?

    Track A-no

    Track B-no

14. Do you feel immersed in the music?

    Track A-no

    Track B-no

15. Did you experience the music as if it was coming from all around you?

    Track A - no

    Track B –no

16. Did you feel as if the sound was coming from inside or outside of your head?

    Track A-out

    Track B-out

17. Were the tracks comfortable or uncomfortable to listen to? Why?

    Track A- not the most pleasant couple of minutes, continuous noise + creepy music little to no variation or development

    Track B- ‘’

18. Which of the versions would you rather listen to, A or B?

B

19. Would you buy music in either of these formats?

    Track A – this example would not make me feel I compelled to rush out and purchase it no... Does not show off the features being offered by the format...

    Track B- ‘’

20. Do you think production suits the type of music?

    Track A - 50/50

    Track B - ‘’

21. Is there anything else you would like to add about the music, mixes or your experience?

Appendix F – Music Permission Informed Consent

'The Phenomenal Rise of Periphonic Record Production'

Principle Researcher: Jocelyne Lord
Jo.Lord@uwl.ac.uk
+44 (0)7753 421 228

PhD Music - Research Ethics Consent Form

This research study has been ethically reviewed and approved by the URDSC in compliance with section 1.10 (g) of the University's Research Degree Regulations;

> *1.10 - in approving an application for registration, the University Research Degrees Sub-Committee, on behalf of the Academic Board, shall satisfy itself that: (g) ethical considerations have been addressed*

The purpose of this study is to investigate whether there is a future for audio-only 3D music in the record industry and could ultimately lead to development of novel composing and production technique for 3D music and the discovery of new knowledge in the field of 3D record production. The research will be available in publication through the University or by email request to the Principle Researcher.
- The practical workshops are designed to investigate and demonstrate how we might use our understanding of acoustics and the human perception system to enrich music through 3D production techniques.
- The listening surveys are used to collect anonymous data about the consenting participant's audio-only or audio-visual experience of 3D music.

Please tick box to confirm you have read and accept the statement:

1. ☒ *I agree to my music being used in this research project.*

2. ☒ *I understand that my paternal and Intellectual Property rights are not affected.*

3. ☒ *I understand that the music I have lent the project will not be used for anything other than the research project. It will not be released commercially or made available publically without my express permission first.*

4. ☒ *I agree that the music I have submitted may be used for data collection purposes and I understand I will be informed first and asked permission before use in any data collection.*

5. ☒ *I confirm that I have been explained the details of the above study and have had the opportunity to ask questions. I understand that if I have any problems or questions I can contact the principle researcher using the details specified above.*

6. ☒ *I understand that my participation is voluntary and that I am free to withdraw from the study at any time, without giving reason or receiving penalty. I understand that upon withdrawing, any unprocessed data previously supplied will also be withdrawn from the study.*

7. ☒ *I understand that I am not requested to provide any personal or sensitive data for this study. All research data collected will be anonymous. Data will be securely and privately stored for up to 3 years after the study concludes.*

8. ☒ *I voluntarily consent to take part in the above study.*

Participant Signature:………… …… Date:……… …..

*** I have reviewed the information detailed above and confirm I have explained the purpose of the study* **Principle Researcher Signature: …… ….. Date:…… …………

All participants who have submitted music for use within this study were required to sign a copy of the above consent form. They were briefed on the purpose of the study and how their music would be used within it. They have been made aware that any Intellectual Property rights associated with the borrowed music are not affected and that the music will not be released commercially, without first attaining their express permission. Participants have also been made aware that they are free to withdraw from the study, including their music and any data associated with it, at any time.

The researcher holds secure copies of the signed documentation, however given the requirements for anonymity these documents will remain held in secure encrypted storage and will not be submitted within this document. If for any reason these documents need to be viewed by the participant or degree examiners then a request to do so should be made to the researcher using the contact details specified.

Appendix G – Participants Informed Consent

'The Phenomenal Rise of Periphonic Record Production'

Principle Researcher: Jocelyne Lord
Jo.Lord@uwl.ac.uk
+44 (0)7753 421 228

PhD Music - Research Ethics Consent Form

This research study has been ethically reviewed and approved by the URDSC in compliance with section 1.10 (g) of the University's Research Degree Regulations;

> *1.10 - in approving an application for registration, the University Research Degrees Sub-Committee, on behalf of the Academic Board, shall satisfy itself that: (g) ethical considerations have been addressed*

The purpose of this study is to investigate whether there is a future for audio-only 3D music in the record industry and could ultimately lead to development of novel composing and production technique for 3D music and the discovery of new knowledge in the field of 3D record production. The research will be available in publication through the University or by email request to the Principle Researcher.
- The practical workshops are designed to investigate and demonstrate how we might use our understanding of acoustics and the human perception system to enrich music through 3D production techniques.
- The listening surveys are used to collect anonymous data about the consenting participant's audio-only or audio-visual experience of 3D music.

Please tick as applicable:

1. ☒ *I agree to the interview / practical workshop / survey / submission of stems.*

3. ☒ *I agree to the interview / practical workshop being audio recorded*

3. ☒ *I agree to the interview / practical workshop being video recorded*

4. ☒ *I agree to the use of anonymised data and quotation in publications*

Please tick box to confirm you have read and accept the statement:

9. ☒ *I confirm that I have been explained the details of the above study and have had the opportunity to ask questions. I understand that if I have any problems or questions I can contact the principle researcher using the details specified above.*

10. ☒ *I understand that my participation is voluntary and that I am free to withdraw from the study at any time, without giving reason or receiving penalty. I understand that upon withdrawing, any unprocessed data previously supplied will also be withdrawn from the study.*

11. ☒ *I understand that I am not requested to provide any personal or sensitive data for this study. All research data collected will be anonymous. Data will be securely and privately stored for up to 3 years after the study concludes.*

12. ☒ *I voluntarily consent to take part in the above study.*

Participant Signature:……………………………………… Date:……………………..

*** I have reviewed the information detailed above and confirm I have explained the purpose of the study ***

Principle Researcher Signature: ……… ……..... Date:………………………...

All participants involved in the project focus sessions were required to sign a copy of the templated consent form above. They were briefed in advance on the purpose of the study and the use of data collected herein. They were informed of their data being anonymised and were advised not to disclose any sensitive or personal data during the trials. All participants were made aware that their data will be kept in a secure and encrypted manner and may be kept for up to 3 years following the conclusion of the study, after which it will be destroyed. Participants were also informed of their right to withdraw themselves from the study, including any data of theirs collected, at any time.

The researcher holds secure copies of the signed documentation, however given the requirements for anonymity these documents will remain held in secure encrypted storage and will not be submitted within this document. If for any reason these documents need to be viewed by the participant or degree examiners then a request to do so should be made to the researcher using the contact details specified.