University of Huddersfield Repository

Eaton, Callum

Quantifying Factors of Auditory Immersion for Virtual Reality

Original Citation

Eaton, Callum (2020) Quantifying Factors of Auditory Immersion for Virtual Reality. Masters thesis, University of Huddersfield.

This version is available at http://eprints.hud.ac.uk/id/eprint/35372/

The University Repository is a digital collection of the research output of the University, available on Open Access. Copyright and Moral Rights for the items on this site are retained by the individual author and/or other copyright owners. Users may access full items free of charge; copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational or not-for-profit purposes without prior permission or charge, provided:

- The authors, title and full bibliographic details is credited in any copy;
- A hyperlink and/or URL is included for the original metadata page; and
- The content is not changed in any way.

For more information, including our policy and submission procedure, please contact the Repository Team at: E.mailbox@hud.ac.uk.

http://eprints.hud.ac.uk/

Quantifying Factors of Auditory Immersion for Virtual Reality

By

Callum Eaton

A Thesis submitted to the University of Huddersfield, carried out for the degree of

MSc by Research

January 2020

Copyright Statement

i. The author of this thesis (including any appendices and/ or schedules to this thesis) owns any copyright in it (the "Copyright") and s/he has given The University of Huddersfield the right to use such Copyright for any administrative, promotional, educational and/or teaching purposes.

ii. Copies of this thesis, either in full or in extracts, may be made only in accordance with the regulations of the University Library. Details of these regulations may be obtained from the Librarian. Details of these regulations may be obtained from the Librarian. This page must form part of any such copies made.

iii. The ownership of any patents, designs, trademarks and any and all other intellectual property rights except for the Copyright (the "Intellectual Property Rights") and any reproductions of copyright works, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property Rights and Reproductions cannot and must not be made available for use without permission of the owner(s) of the relevant Intellectual Property Rights and/or Reproductions.

ABSTRACT I

This study considers the factors that define the perception of auditory immersion, focused around the context of virtual reality. Previous literary work focuses largely on surveys and questionnaires to elicit participants perception of immersion and indeed auditory immersion, and does not consider truly what factors define this perception and how much by. This topic is discussed at length and concludes with the proposal of a Universal Immersion Paradigm; defining immersion as a consequence of three component parts, passive immersion, active immersion and the immersive system. An immersive audio survey was also conducted to ascertain professional and consumer opinions of what factors of immersion perception are most important for different content types. Results show that largely all factors questioned were perceived as important. Trends in the boxplot results showed however, that the perception of factors relating to vertical sound perception of both envelopment and localisation were rated lower on average compared with the comparable question relating to horizontal sound perception.

An initial experiment was designed to create an optimal speaker layer balance for the four '22.2' (not all recordings were presented natively in 22.2 and no subwoofers were utilised) recordings being utilised as stimuli for remaining experiments. This test showed that height layers were mixed louder than floor layers for three out of the four content stimuli, though all height and floor layer were on average mixed at a lower average loudness compared to the main layer level which was fixed throughout. The final experiment compared different speaker formats (mono, 2.0, 5.1, 9.1 and 22.2) with perceptual features highlighted throughout as potentially the most influential to the perception of immersion. Results concluded that 5.1, 9.1 and 22.2 formats were found to be significantly similar in the majority of test cases. 2.0 and mono formats found sporadic statistical similarity but were consistently rated lower than the other formats. When main effect of the format was considered with dependency on the perceptual attributes utilised, it was discovered that all stimuli results were found to be statistically significant when a Friedman test was carried out for the factors of Listener Envelopment (LEV) and Presence (Pres), but were not for Overall Tonal Quality (OTQ) and Quality of Experience (QoE). The result suggesting that the perceptual factors of OTQ and QoE are highly content specific in terms of user perception. LEV and Presence on the other hand are perhaps not as closely linked to content overall. Significant changes in perception are more clearly identified for these perceptual factors when considering different format reproductions, with no significant differences being found between 22.2, 9.1 and 5.1 across most stimuli tested, and mono being the lowest rated format for all stimuli and perceptual attributes.

CONTENTS

List of Figures and Tables

Chapter 1: Introduction

The concept of immersion has become widely used and debated over the last few years, with Head-Mounted Display (HMD) and computer technology now being able to provide the fidelity and power to render high quality virtual scenes. In regards to computer games for example, immersion is considered to be an important factor in the level of enjoyment in the experience (Nacke & Drachen, 2011). So, with the emerging prevalence of 360° video content and Virtual Reality (VR) content in general, it is now more than ever important to understand what immersion actually is, and more specifically for the audio industry, how auditory immersion can be specifically defined and quantified.

The current study has stemmed from the lack of cohesion in how to define the term immersion and other related terminology, a topic that will be explored in this work and will ultimately aim to offer a new and unified paradigm for immersion that is applicable for research and content creation, in as many fields as possible. Not only would a unified definition for the concept of immersion be an asset in the field of technical research but as Nacke (2008) attests, a quantification of what specific factors cause the perceptual feeling of immersion would be invaluable information.

Previous work of the subject of immersion such as that by Agarwal & Karahanna (2000) have focused on linking immersion to one or two of a few high level perceptual factors such as flow, cognitive absorption or presence. Not only will this work cover the literature definitions of immersion and aim to offer a new solution to the issue, experimental work has also been undertaken to attribute the importance of specific psychoacoustic factors to the perception of auditory immersion, providing a far more detailed insight into the topic, and as such will provide a strong platform for the ever continuing research in this area.

Traditionally subjective questionnaires have been used in experimentation to elicit a response from subjects as to their level of perceived immersion. This presents a number of potential flaws as it is recognised by Ermi & Mäyrä (2005) that immersion is highly influenced by not only a subjects previous experience of the virtual content itself, but may also be potentially biased in testing by their internal emotional response to the content. Jennett et al. (2008) and Mestre (2019) have suggested and tested, a link between cognitive perception and biometric data collection techniques, meaning methods such as eye tracking, pulse or Electroencephalograms (EEG) may be used in conjunction with traditional questionnaire data to help reduce potential emotional biases. In the current work, this methodology is closely considered, but ultimately is discounted in favour of more traditional scale rating, multiple comparison methods, due to the complexity of procedure and expertise required in the field to properly analyse the data and produce any meaningful conclusions.

VR is a relatively new technology in consumer markets, with the two main device manufactures; Oculus Rift released on March 28, 2016 and the HTC Vive released on April 5, 2016. This means its use in content production is still in relative infancy.

There are two main approaches to creating VR environments, these being Cave Automated Virtual Environments (CAVEs), where a scene is projected on multiple walls in a room, and Head-Mounted Displays (HMDs), where the user wears a headset which presents the scene on a LCD screen in front of the user’s eyes (Schoeffler et. al. 2015).

The differences in approach to creating virtual environments of course have their advantages and disadvantages for certain situations. A CAVE system such as those produced by, IGI (2019) are mostly suited to larger scale immersive experiences for multiple people at once. For this purpose the CAVE system allows for true interaction between participants as there is no HMD blocking natural line of sight, which dependant on the content may be hugely beneficial for a positive overall experience. Acoustically these spaces are hard to cater for in an immersive way as headphones are not optimal. Speakers are usually arranged behind or above and below the projected image and so discrepancies in acoustical performance will be highly noticeable to those participants not in the speaker sweet spot.

HMD systems on the other hand are ideally suited for a single user experience where no interaction with the outside world is required, such as in a video game. Commercially this allows audio designers a great deal more flexibility as headphone use is almost essential. Negatively, the controller systems commercially available at the current time still lack any form of tactile feedback to the user, other than basic haptics. In VR, both options for content presentation provide ample opportunity to create immersive experiences however, as with all aspects of immersion, the effectiveness of the presentation method is potentially highly content dependant.

Whilst VR is becoming an ever more popular tool for experiencing content today, the applications for VR may not be limited to commercial content like music videos or games. Virtual environments could be employed in other situations that may improve on existing technologies, or provide new and novel solutions to current problems and issues.

One potential use of VR is as an environment to conduct scientific testing. Schoeffler et. al. (2015) utilised a HMD VR environment in an experiment to test Overall Listener Experience (OLE) when comparing a virtual recreation of a physical space, also using auditory stimuli over headphones. The test concluded that VR ratings were slightly lower than that of the real room, though the stimuli that were presented to the test subject were music excerpts rather than broadband noise or single source signals; thus could explain the lower rating when OLE was the only measure. The potential validity of a virtual test environment could have a number of significant benefits for researchers. If a listening room can be accurately rendered in VR both visually and acoustically using Binaural Room Impulse Responses (BRIRs) for example, this could allow significantly greater collaboration between institutions and allow for much greater sample sizes to be gathered. Something that may allow for much more accurate information to be gathered.

Interestingly, VR environments have found use in a number of other scientific fields. The report by Minderer & Harvey (2016) comments that VR is a valuable tool in understanding neural function as it allows for precise experimental control not possible in real world approaches. Suggesting also that VR allows for a greater number of tests to be conducted than possible on a normal test subject. The comments by Minderer & Harvey (2016) bring the current study into precise focus, psychoacoustic research may benefit significantly from the advantages that virtual environments may bring in the future. The ability to have more precise control of test conditions can only mean a greater understanding of the human auditory process.

Understanding Immersion is not only important for improving content creation for the consumer, or improving the experience of a simulator. It is important to understand immersion in virtual environments as it helps us fundamentally understand our perception in the real world. It is possible in virtual environments to create situations that the laws of physics simply would not allow in the real world. By doing this, we can isolate specific elements of our sensory perception and analyse their importance in our overall quantification of the world around us.

As has already been discussed, immersion is a holistic and abstract term at best, certainly in consumer spheres. For VR purposes, immersion is what takes the experience from a gimmick to something that holds value. As Pausch, Proffitt, & Williams (1997) state, ‘Virtual Reality has generated much excitement but little formal proof it is useful’. Whilst this statement is somewhat dated given more recent studies into VR, the true usefulness of the technology is still fundamentally limited due to the lack of formal study into the quantification of immersion. Whilst the current study does not supposed to provide a flawless catagorisation of the auditory factors that contribute to the perception of immersion, it does aim to contribute to the continued research into this area.

The current work is structured into an initial literature review section with relevant discussion, followed by an overview of an immersive audio survey that was conducted as part of the project, as well as an overview of the two main experiments. Two main research questions are proposed at this juncture.

- What is the optimum layer balance for an immersive 22.2 music mix, and what perceptual motivations drive this decision?
- Can auditory perceptual factors be identified and quantified as important for the overall perception of immersion?

Chapter 2: Literature Review

The following chapter will cover the literature review and discussion around the topic of immersion and will aim to suggest a new paradigm for the study and discussion of immersion hence forth.

2.1 Defining Immersion

The Cambridge dictionary defines the term immersion as a noun, describing 'the fact of becoming completely involved in something' (Cambridge Dictionary, 2019). Perhaps for the average consumer, this definition of the term is satisfactory. Unfortunately, it falls short of describing the whole sensory experience that perceptually provides a sense of immersion, particularly in the context of VR content, and so, a new and more clearly defined definition is required.

Immersion is not a new term, and it is certainly not a new term in the world of virtual reality. Its literary origins can be traced back to the early 1960's and the work of Morton Heilig's Sensorama machine (*U.S. Patent No. 3,050,870.* 1962). The device combined stereoscopic colour display, a stereo sound system, a moving chair and odour and wind emitters, which allowed participants to experience a 'realistic' and 'immersive' motorcycle ride through New York city. Certainly for the time this setup could be considered as highly immersive. In this example the perception of immersion that is reported by Rheingold (1991) is highly linked to the sensory realism, such as the scent of cars and a pizza shop, and the feeling of wind in the face whilst experiencing the film. It may be argued then that immersion is highly linked to sensory realism, however this definition and understanding leaves the consumer or content creator with no greater understanding of the perception of immersion.

Although elements from the Sensorama machine were potentially immersive, the system was bulky, impractical and ultimately did not receive the financial backing to become commercial.

The terminology used to discuss a state of immersion vary widely depending on the field of research or even author to author within the same discipline. This makes the task of collating opinions on the subject somewhat challenging. It is convenient in the case of scholarly research and for the purpose of commercialisation to bend the meaning of the term to fit the work or product being developed. As immersion is also heavily used as a marketing term for commercial products, as such the definition has become somewhat synonymous for a product being better if it is more immersive, it ever highlights the need for clarity and understanding of what immersion means and how to quantify it. For the purpose of this work however, a more universally applicable definition of immersion was desired as aforementioned this would provide consumers, content creators and researchers a platform to continue development.

Slater (2009) provides a basic definition of immersion that states, Immersion provides the boundaries within which a sense of place can occur. Jennett et al. (2008) takes this definition somewhat further by commenting that a sense of immersion can be barriered by content type and by performance. Whilst both definitions go some way to describing the sense of immersion, they do not comment on what element of perception causes immersion, and do not quantify the bounds of this perception in any way. More widely the use of immersion is linked and quantified by easier to measure metrics, Christou (2014) for example attempted to link the perception of immersion and overall appeal in the context of video games. It found that immersion varied largely depending on the game that was played, but this level of perceived immersion did not change significantly depending on the experience of the participant. The study also suggests that appeal and immersion are highly correlated; higher levels of appeal would mean potential for higher levels of perceived immersion. Other studies have also endeavoured to link the perception of immersion to other factors such as Brown & Cairns (2004) who attempted to link immersion to a three-part semantic scale. ‘Engagement’, being the lowest level of involvement and relates to a basic investment of time and attention in the content. ‘Engrossment’, the second level where the content has a direct effect on the emotions of the user and finally, ‘Total Immersion’ which describes a feeling of presence in the virtual world and a detachment from the real world.

The term presence is not only isolated to the Brown & Cairns (2004) study. The term presence is often discussed hand in hand with immersion, and as such highlights the vastness of the problem in quantifying and fully understanding the feeling of immersion in virtual worlds, and perception of immersive content generally. In Jennett et al. (2008) presence is discussed as a phycological sense of being in a virtual environment. Slater (2009) discusses the link between the feeling of ‘being there’ in a virtual world and the concept of ‘sensorimotor contingencies’, that being the actions that we perform in the real world in order to perceive our surroundings being matched in the virtual environment. As such, in this definition presence is not necessarily linked entirely to the specific type of content the participant is engaged with, rather that the key limiting factor of the perception of presence is how satisfied sensorily the participant is and therefore how realistic their normal behaviour is matched. This principal is also discussed by Mestre (2018) who refers firstly to immersion as the sensorimotor coupling between a participant and a virtual environment and also to presence as the psychological attentional and cognitive state, in which the participant who is immersed in the virtual environment behaves in a way that is indistinguishable from reality. In this study immersion and presence are highly interrelated, a state of immersion is required to feel present in a virtual environment, but it would also be assumed that without a sense of presence, the overall perceived immersion would reduce. The problem that arises from this, is a typical definition for immersion and presence largely pertain to the same feeling, the ‘sense of being there’ and therefore in this scenario there is potentially no need to use both terms.

A differing opinion on the topic of immersion is shared by Grimshaw & Garner (2015), which refers to immersion as an inherent property of how advanced the technical system presenting the virtual environment can give the user a sensorily valid feeling of presence (the feeling of being there). This is not the only example of immersion being defined as a property of technology. Slater (2018) also defines immersion as an inherent property of a system, with higher or lower immersion being defined by the extent the system can support more natural sensory perception. In this example a higher level immersive system could be achieved by using a wider field of view heads up display, head tracking of visuals and audio, haptic feedback systems as well as improving visual fidelity to name but a few potential examples.

In this scenario there is a valid argument for the use of both presence and immersion, if presence was used to define perceptual elements and immersion was used to describe how advanced the technology was able to provide the sense of presence. It could be argued, however, that a more appropriate definition for this idea would be technical immersion or an immersive system. Whilst the immersive system is a major contributing factor in providing sensory fidelity to immersive content, immersion is not necessarily always limited by technology. For example, immersion is not necessarily limited to virtual worlds as humans can be immersed within their own imagination, such as when reading a book. It could be said that a book could not achieve a similar level of immersion as a highly advanced VR setup, which is likely true, however the book is still itself a system for providing immersion, so the immersive system is only part of the jigsaw of quantifying immersion not its sole defining factor.

Whilst immersion and presence are used interchangeably in literature to describe a sensorily valid feeling of 'being there' in a virtual environment. Both cognitively and by technological means, there is a variety of other terms that are used to describe a similar perceptual feeling. Agarwal & Karahanna (2000) defines cognitive absorption as a way to link a user's motivation and focus on a specific task. This concept of cognitive absorption is similar to the idea of immersion but refers to the potential immersiveness of a task rather than the feeling of being there within a virtual environment. This concept is highly linked to the idea of flow, 'the state in which people are so involved in an activity that nothing else seems to matter' (Csikszentmihalyi, 1990). In this sense, the feeling of 'being there' in a virtual environment may not and perhaps should not be the only goal of immersive content creation. Whilst the aforementioned study by Brown & Cairns (2004) used a semantic scale to describe the level of immersion, with total immersion describing a sense of presence, the idea of engrossment and engagement more closely align with a concept of cognitive absorption, focus and emotional investment in a task or character than simply a feeling of presence within a virtual world.

As is demonstrated by the variety of definitions and perceptual links, defining and quantifying immersion has created confusion in both a research and commercial context. The confusion seems to stem from the desire to understand an emotional state without in some cases identifying perceptual features that may cause or hinder this perception, with focus aimed more at overall experience quantification. This is not to say that immersion may not be linked to overall experience, it is simply too broad a term to be able to justify it without additional information. It is why it may be useful to look at quantifying the perception of immersion in terms of specific senses, auditory immersion, visual immersion etc. as well as immersion in a multi-sensory capacity in an attempt to more clearly outline basic perceptual features that cause this perception.

2.2 Auditory Immersion

Blauert (1997, p.2), suggests that humans are primarily visually orientated with other senses, such as auditory and tactile, being much less highly developed. This means that visual cues provide the framework for our perception of the world around us. In the context of studying immersion it is obviously important to understand the impact of visual cues on the perception of how immersive a virtual world may be, and an outline of this current understanding in literature has already been discussed in the previous subchapter. The subject of immersive audio has also become a hotly debated area in recent years, with developments in binaural, and Ambisonic rendering over headphones as well as multichannel speaker playback systems ever claiming to provide the most immersive audio experience yet. Whilst the area of multisensory immersion in the context of video games is somewhat well researched, the area of what aspects of auditory perception contribute most to immersion are as of yet largely unexplored in real depth. In auditory research, immersion itself has not necessarily been the goal, with studies such as that conducted by Schoeffler, Gernert, Neumayer, Westphal, & Herre (2015), where assessment of audio is focussed around the metric of Quality of Experience (QoE), and

Overall Listener Experience (OLE). In this experiment, when a listening test scenario was conducted in a virtual space modelled on a real space, it was discovered that there was a significant difference in perceived OLE between the virtual and real space, even with identically presented audio stimuli. Whilst this highlights the challenges of visual presentation in virtual environments, it does perhaps hint at a necessity for perceptual realism in both audio and visual playback. However, OLE or QoE, whilst being useful semantic metrics for overall experience, do not provide detailed information on any individual or directly addressable elements of human auditory perception, and how they may impact immersion from a purely auditory standpoint, as well as the impact subtle changes in audio presentation have on the perception of immersion in multisensory environments.

Consideration is also made within the audio industry towards the concept of 'spatial audio', and its relationship to immersiveness. Roginska & Geluso (2018) for example, simply defines immersion as sound coming from all around the listener. In this example it is assumed that spatial audio must be immersive but does not offer any comment on what systems may be more or less immersive. Berg & Rumsey (2001) describes listener envelopment as a way to describe a sense of immersivity in a reverberant environment, where sound seems to be arriving from all around the listener. Sazdov, Paine, & Stevens (2007) highlight however that this terms does not fully describe the spatial attributes of 3D reproduced sound, suggesting the term 'engulfment' be used more specifically to describe the sense of sound heard from above and below the listener. It is potentially important to distinguish the difference between horizontal and vertical sound envelopment when considering the components of auditory immersion, as this may allow for greater specificity in results gleaned in experimentation in this area. Traditionally in commercial content, 5.1 speaker systems have been utilised to provide a spatial sound experience, as the utilisation of surround speakers enables a greater sense of listener envelopment than a standard stereo or mono mix. When the comparative immersiveness of a virtual nature soundscape was compared with and without head tracked spatial audio by Poeschl, Wall, & Doering (2013) they found a significant difference between no sound and spatial sound conditions, also with spatial sound, presence in the virtual scene was rated higher by the participants. Whilst this example does not directly compare different types of content, it does show the potential impact spatial audio in particular can provide in boosting the perception of immersion. Further to this point, when Silzle, George, Habets, & Bachmann (2011) compared multiple speaker formats including 22.2, 9.1, 5.1 and 2.0 with various music and soundscape excerpts to asses overall sound quality (another metric aligned with OLE or QoE), they found that without a reference signal, 22.2 was not always rated as having the best overall sound quality with 9.1 often being rated as better. This experiment perhaps suggests that although the concept of an immersive system would state the 22.2 mix must be the most immersive as it utilises the highest number of speakers in this experiment, this does not necessarily hold up to scrutiny when no reference is provided and thus the perceptual differences between 9.1 and 22.2 systems are not significant in this case.

Perhaps the real key to understanding immersion from an auditory standpoint at least, is to observe immersion as a consequence of the perception of low level auditory attributes. Doing this would allow for a more easily quantifiable and understandable interpretation for researchers and content creators alike, as where before immersion was simply a holistic concept of little substance, it is now more easily accessible by more controllable low level auditory perception attributes. For example, when presenting the concept of a 22.2 speaker system, Hamasaki, Nishiguchi, Hiyama, & Okumura (2006) conducted an experiment to demonstrate the differences in low level perceptual attributes between 2 channel, 5.1 channel and 22.2 channel systems for both music and ambient sound design audio excerpts synchronised with visuals. This experiment resulted in little differences being perceived between 22.2 and 5.1 systems in attributes such as localisation accuracy, naturalness and degree of delight and interest. However, more distinct differences were seen in attributes such as presence, envelopment, depth, width, and front/back confusion. Whilst the goal of this experiment was not to outline how immersive that system or content was to the listener, the results of this experiment would suggest that the differences in the perception of low level attributes are not consistent with each other. This may then provide an avenue to explore immersion in a new and different manor, to identify and quantify auditory perceptual factors alongside the higher level perception of immersion would provide a potentially much clearer understanding of what auditory factors are most important for the perception of immersion, and also which are not.

2.3 Testing Immersion

The most common method of eliciting and recording perceptual responses whilst testing immersion is with a survey or a method of scale rating. This method of testing is prominent in psychoacoustic research, as data collected can be easily analysed statistically and it offers researchers a simple quantifiable method of testing a hypothesis. The use of surveys to test the perception of immersion is commonplace such as in the work by Tcha-Tokey, Christmann, Loup-Escande, & Richir (2016) who designed a questionnaire to test user experience in 'Immersive Virtual Environments' and measures attributes such as presence, immersion, flow and usability among others. Their results showed that the survey data was normally distributed for factors such as presence, engagement, immersion, flow, emotion and judgment but were negatively skewed for the subscales of skill, technology adoption and experience consequence. This shows both the strengths and weaknesses with surveys. They can be highly reliable for certain aspects of perception but not for others.

When testing auditory perception, multiple comparison experiments or MUSHRA (Multiple Stimuli with Hidden Reference and Anchor) tests (ITU-R BS.1534-1, 2003) are often used to elicit results from test subjects. Silzle et al. (2011) for example, used both a multiple comparison and MUSHRA tests to test overall sound quality comparing between multiple, multichannel speaker formats. They are advantageous for testing audio as it allows participants to directly compare test conditions against each other. As Silzle suggests, this methodology allows subjects with less experience to also be tested.

As the perception of immersion falls some way between a perceptual and emotional response to a piece of content, the use of simple survey or scale grading data collection methods have both advantages and disadvantages. The advantages are they are simple to analyse and collect data for in the first instance, and most participants with any experience with experiments will be used to the method meaning a reduction in potential bias. The disadvantage of these methods are, they do not necessarily allow for a full analysis of any biomechanical processes that may be contributing to overall perception; natural responses that would be impossible to elicit a response for in a questionnaire.

The use of biometric data collection alongside traditional data collection methods may serve to help garner a more complete understanding of human perception, particularly in the study of immersion and its related areas. Biometric data collection methods were used to some success by Jennett et al. (2008), who used eye tracking data combined with more traditional immersion questionnaires to test immersion linked to pupil movement around the screen. In this experiment eye movement was found to significantly increase when participants also found the content to be non-immersive and similarly eye movement decreased significantly as participants found the content they were experiencing immersive.

Other metrics have been suggested to have a link to immersion such as heightened pulse and blood pressure and higher skin conductivity. For example in an experiment conducted by Meehan (2001) the change in mean heart rate, skin conductance and skin temperature were all recorded as a way to test the physiological reaction to presence in a virtual world, between a training area and a more high stress, 'pit room' virtual environment. The results of this test showed changes in all physiological measures between the training environment and test environment with heart rate change being the most sensitive. As is acknowledged in the study whilst potential links may be drawn between physiological measures and perceptual findings on immersion and presence the reliability of these links is still up for debate as it is impossible to distinguish between the perception of immersion or presence and another factor of perception when observing physiological data.

The use of an electroencephalogram (EEG) as a method of testing presence and immersion has many of the same potential benefits and pitfalls as other biometric data collection. The underlying principal of its utilisation is that physical and neurological patterns and perception of immersion may be far less susceptible to intrasubject bias as the data is recorded directly from brain activity. This means there is less influence of emotional or social factors that may affect a participants results of a typical survey or questionnaire. Nacke (2008) for example, establishes that whilst some emotional states (such as immersion) may be able to be quantitively characterised via biometric collection methods such as an

EEG, they are not necessarily reliable in isolation. Furthermore, these measures are hard to directly link to the response of one single emotion.

Clemente, Rodríguez, Rey, & Alcañiz (2014) conducted an experiment to attempt to assess the level of presence in a virtual environment using an EEG. They measured EEG results in combination with a questionnaire to assess their level of presence during the experiment and found a link between activation of certain parts of the brain and an overall sense of presence.

In isolation this should mean all immersion and presence experiments should utilise an EEG to back up questionnaire data, however in reality there are a number of more practical limitations for most research taking place. Namely, EEG equipment is complex and requires an expert operator to use correctly and analyse the results. Furthermore, as previously mentioned, biometric data cannot necessarily be reliable and linked to the emotional or perceptual response that is being tested. This means, that whilst the utilisation of biometric data collection methods and to an advanced degree EEGs, have some specific applications where useful data can be gleaned, at this present time it is potentially not something that is practical or necessary to use in experimentation.

2.4 Immersive Playback System

When discussing the factors of auditory immersion for virtual reality, the way the audio is replicated needs to be closely considered. In the work by Larsson, Vastfjall, & Kleiner (2002) for example, it is shown that subjects rated feeling a greater sense of presence when experiencing a multimodal virtual system combining visuals and audio than just visuals alone. Also the quality of the sound replication over headphones in this case was tested between a ‘high quality’ binauralised ambisonic signal and a ‘low quality’ static stereo binauralisation. The results in this case shows, along with arguments of previous chapters that not only is sound in general important for providing presence, immersion and other important factors to the experience in virtual environments, but that the format or processing of the audio also has a potentially significant impact too.

As was previously discussed, there are two options to present audio for virtual reality content; through headphones or an array of speakers. This decision is usually dictated by the content type and to a larger extent the type of virtual environment being used (either a CAVE or HMD system) and also the technical limitations of the audio mix and technical system for reproducing it. For example a VR video game is usually enjoyed on a HMD system and the audio is designed to be reproduced over headphones. Due largely to practicality in the consumer sphere, as most people do not have access to large amounts of well-arranged speakers, to replicate what may be able to be achieved adequately through a binaural or ambisonic headphone mix. A CAVE system on the other hand may be used to display content such as an art/sound exhibition or potentially a concert type experience Lokki, Hiipakka, & Savioja (1999), IGI (2019), and as such from a practical standpoint, speakers arrays are generally preferable in this situation due to the ability to have interaction between participants in the CAVE system or simply from a design or cost standpoint..

As the current study aims to discuss the factors of auditory immersion and is not directly concerned with the practicality or feasibility of systems in the real world, it is therefore important to discuss the merits and detriments of both sound presentation methods objectively, free from the implications that practicality bring to the discussion at least initially.

Further to the discussion of the advantages and disadvantages of headphone and speaker reproduction of sound it is also important to consider the non-auditory perception of sound and how that may impact on which presentation method is most ideal. When attending a live concert or a sports event, you hear sound through your ears, but you also get a shiver down your spine as the piercing guitar solo plays, or you feel the kick drum deep in your chest. This feeling is still auditory perception and is vital for a sense of presence in the real world, it also adds a huge amount to our semantic enjoyment of a live music experience for example. It could therefore be argued that this feeling is something that is critically important in the perception of immersion in virtual environments. Certainly in experimentation conducted by Merchel & Altinsoy (2013) when concert reproductions on a 5.1 surround speaker system were paired with a vibration chair, conditions with additional physical vibration were judged more preferable than those without. If we use not only an auditory perception of sound, but a physical perception too, then this needs to be replicated somehow for VR content otherwise the consumer is simply not getting the whole acoustic experience.

The most common audio playback method for consumer VR content is via headphones. This is an issue for the idea of physical perception as whilst you may perceive some acoustic reflections from your jaw and skull, there is still large areas of the body that could not experience any physical excitation. A potential way to improve this is with audio tactile stimulation in hand controllers or other tactile transducers (Ford, Ausiello, & Barlow, 2019) (Altinsoy & Stamm, 2013). In Altinsoy & Stamm (2013) a tactile transducer was used to reproduce the vibrations of an auditory stimuli in a virtual environment and found that the tactile stimulus helps to reduce localisation blur in a corresponding auditory stimuli. Not only does this suggest the importance of tactile feedback in regards to auditory perception, but its potential importance in creating immersive content. From a strictly auditory content point of view, this tactile element of sound may be produced in the same was as in a live music environment, with a subwoofer. If a headphone setup could be supplemented with a subwoofer to provide additional physical excitation, this may be an effective way to increase the immersive capability of the playback system.

The use of speaker arrays rather than headphones for presenting audio for virtual reality means that this non-auditory sound perception can be catered for more easily, as sound vibrations can be felt through the whole body rather than just the jaw and skull as would happen with the use of headphones. Another potential advantage of speaker arrays is that when the user is statically placed in the sweet spot of a speaker array, as long as the starting head position in the virtual environment matches with the speaker array then headtracking is not required as the users head rotation in the virtual environment will not affect the auditory scene being presented over speakers. This is particularly advantageous when the auditory mix is always static such as in a concert hall recording, and thus drastically simplifies the audio processing. If there is a requirement for the auditory scene to move independently from the head rotation of the user such as in a video game where sound elements may move around the game world, then headphones are often more practical. Another advantage of the use of headphones over speakers in this case is the fact that the user is not required to stay in the sweet spot of the speaker array and can move more freely without compromising the sound quality. Again, this may not be an issue if the content requires the user to be sat down. In the same way that there are a number of methods to render audio for headphones, there is not only one speaker array arrangement that is utilised in an attempt to produce an immersive audio experience. As Hamasaki, Hiyama, Nishiguchi, & Ono (2004) attests the use of 5.1 multichannel speaker arrays have become a popular and widely used format not only for film but also for other packaged media. Systems with greater numbers of speakers have also been developed such as 9.1 and 22.2 as outlined in (Hamasaki et al., 2004). The goal of these larger speaker arrays being to increase localisation ability and envelopment for the listener, and to quote Hamasaki et al. (2004) '[the 22.2 system] can be widely applied to research on audio systems providing an exceptionally high sensation of reality'. Furthermore speaker arrays such as 9.1 and 22.2 offer a 'height layer' of speakers (five main layer and four upper layer in the case of 9.1 and ten main layer, nine upper layer and three lower layer speakers in the case of a 22.2 array) which allows for naturally elevated sound sources to be replicated at a more congruent position to a traditional 'flat' speaker array such as 5.1 or 2.0. When Hamasaki et al. (2006) conducted a semantic differential test to compare a 2.0, 5.1 and 22.2 speaker array against various low level perceptual attributes, participants were easily able to identify the 2.0 mix for all attributes and in turn these were ranked lower than both the 5.1 and 22.2. The 22.2 was rated as most preferable for all attributes with particular differences in perceived presence, depth and localisation ability. This result at least suggests that the additional height layers enhance some aspects of the auditory experience. Howie, King, & Martin (2017) conducted a study that considered if listeners could discriminate between common 3D audio reproduction methods over speakers. Those methods being NHK 22.2 Multichannel Sound, ATSC 11.1, KBS 10.2 and Auro 9.1, and the stimulus used being three different acoustic music excerpts. The results of the experiment showed that there were no significant differences between all four speaker formats in terms of ability to discriminate between them. Users were also asked to comment on any perceptual differences they noted between conditions and the most common attributes discerned were "timbre" (70% of participants), "spatial position of direct sounds" (81% of participants) and "spatial impression" (94%). This aligns with the results of the format comparison experiment conducted by Hamasaki et al. (2006) who also noticed significant differences in responses pertaining in particular to attributes such as "envelopment", "depth" and "presence". The results in both cases show consistent differences regardless of the content that was presented, and therefore suggests potentially that attributes concerning tonal quality and spatial quality are most affected by changes in reproduction method. These results also point to the fact that these factors may not be so liable to change dependant on content as potentially expected.

In the current study where auditory immersion is primarily under scrutiny, it may be as important to not only consider what semantic descriptors and low level perceptual attributes are most important for the perception of passive immersion from an auditory standpoint but it may also be possible to use difference speaker arrangements to also observe how the immersive system directly interacts with the user's ability to perceive immersion and/or presence.

2.5 Content Dependent Immersion

There is an argument to be made that as we all perceive the world around us with some individual discrepancy, this would be carried over into our perception of immersive content. In Ermi & Mäyrä (2005) the experience of playing a video game in this case is categorised as 'an ensemble comprised of the players sensations, thoughts, feelings, actions and meaning-making in a gameplay setting' and is noted therefore that the experience and by association the immersiveness of a virtual environment, game or other immersive content is not a property or cause of a particular part of the content, but is a more individual interaction between sensory elements of the content and the user themselves.

In Ermi & Mäyrä (2005) the concept of emotional and social context is discussed as a potential limiting factor for the perception of immersion. The idea being that a user's previous experience of the content or their emotional state of mind could perhaps influence their ability to be immersed in the content that they are experiencing at the time. The concept extends to the idea that 'the same activity can be interpreted as highly pleasant in some contexts but possibly unattractive in other kinds of settings' (Blythe and Hassenzahl, 2003). In the same way that emotional discrepancies may have an impact on the perceived immersion or presence perceived whilst experiencing immersive content, the requirements to provide immersion may also change substantially dependent on the type of content that is being shown, there is likely to be similarities as it is generally accepted that providing as natural a sensory experience as possible will provide the most immersive outcome.

Christou (2014) furthers this point with an experiment aimed at linking appeal and immersion for two different video games and also to observe if skill level and prior experience had an impact on both appeal and immersion. The results of this experiment observe a positive correlation effect between immersion and appeal meaning they are in some regard related. The type of content presented in this experiment did have a significant effect on both immersion and appeal however, whilst the level of experience did not have a significant effect on the overall level of immersion perceived. Christou (2014) then goes on to suggest that immersion may be a trait of the individual experiencing the content rather than a direct trait of the content itself. In the case of this experiment this observation would make sense as the test subjects experience did not have a statistically significant impact on the perception of immersion in both games. The different types of games tested did yield different immersion and appeal scores but it is unclear whether this is down to the content itself or other human factors.

2.6 The Universal Paradigm for Immersion

Whilst there are many differences in definition to the term immersion or indeed presence, there are a number of overarching similarities that mean there is a possibility of offering a new and potentially more universal definition of the term, that may be used in the future to help guide the focus of research onto the underlying elements of the perception of immersion. To this end, in the current study, a universal paradigm identifying three types of immersion is proposed; passive immersion, active immersion and the immersive system. Passive immersion describes the feeling of presence, or how much 'you feel like being there' whilst experiencing content, whether that be combined with VR or not. Passively immersive content for example, could include music or soundscape recordings, particularly if these are also paired with VR, and relates heavily to sensory perception of content.

Active immersion or cognitive absorption, relates to task-based scenarios, or other interactive immersive media such as certain types of video games or training/simulation style content. In this case the users focus is not solely on where they perceive themselves to be, but on a task or activity they are undertaking. This concept is similar to the ideas in Brown & Cairns (2004) and Agarwal & Karahanna (2000) who both highlight the significance of the ability of content to focus a user on a specific task and perhaps therefore, become absorbed to the point at which what is happening in the environment around them becomes of little importance.

The immersive system is the final part of the paradigm, it could be described as the inherent property of how advanced the system is at providing both passive, presence-based immersion and active, cognitive-absorption-based immersion. This aspect of immersion is inherited from works such as Slater (2018) and Grimshaw and Garner (2015) that have already been discussed. For the purposes of providing a more flexible and rounded definition and catagorisation of immersion, separating the concept of the immersive system from the human perception (active & passive) was felt to be particularly important. The immersive system relates to a more easily quantifiable scale for immersion as it relates more to how immersive the technology may be able to provide the other forms of immersion rather than solely providing immersion itself. This idea is backed up by work such as that by Poeschl et al. (2013) who highlight that immersion is a quantifiable aspect of technology to provide fidelity to the sensory aspects of immersive content. The only difference in the case of The Universal Immersion Paradigm proposed in this work is this immersive system aspect is defined separately to perceptual aspects of immersion.

Figure 2.1. Below, shows a basic overview and flow of the new paradigm.

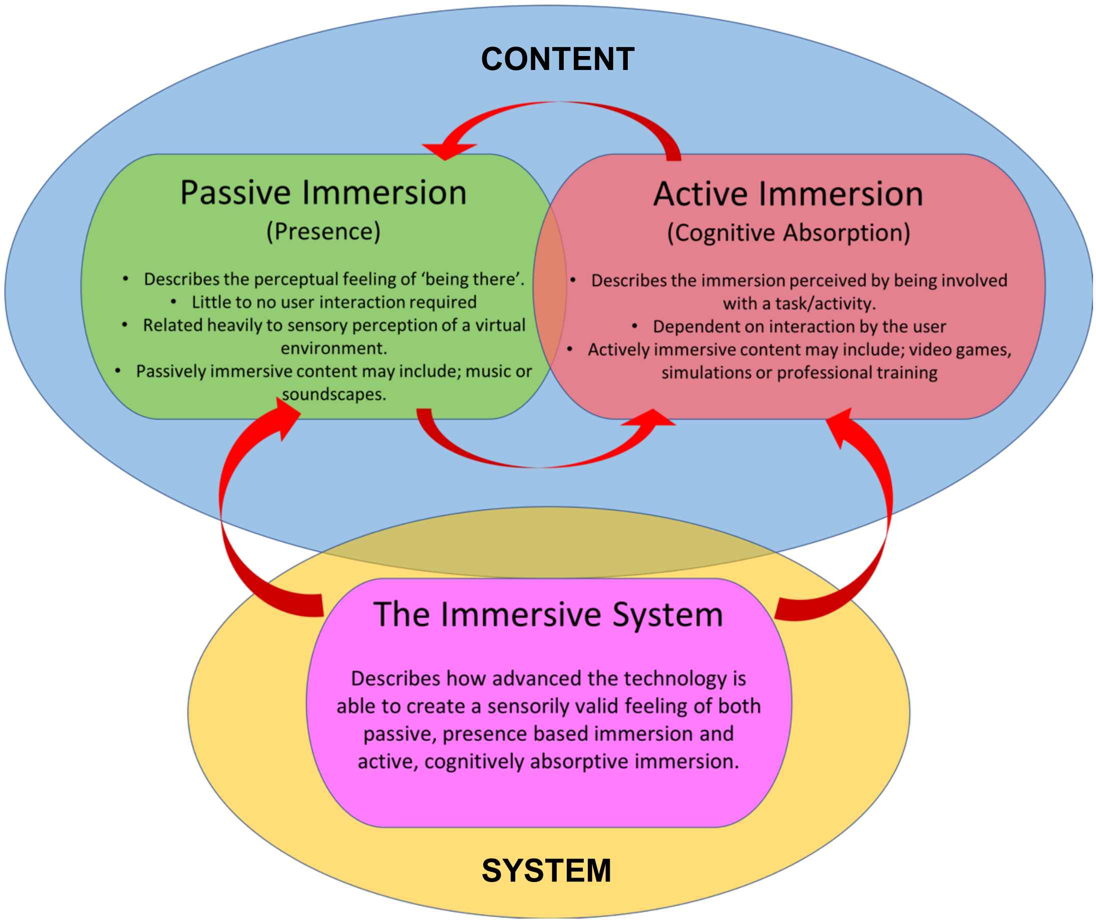


Figure 2.1. – The Universal Paradigm for Immersion

The interaction between the different elements is semantically described by the red arrows. It shows that whilst each element is definable and potentially quantifiable individually, there is a significant amount of overlap between each. In the same way as Ermi & Mäyrä (2005) define the gameplay experience model as changing certain elements of importance dependant on the content being presented, the three elements that make up the new paradigm also change in significance dependant on the type of content and how that content is being presented. Figures 2.2 and 2.3 below, show how the hypothetical relationship between active immersion, passive immersion and the immersive system change dependant on content.

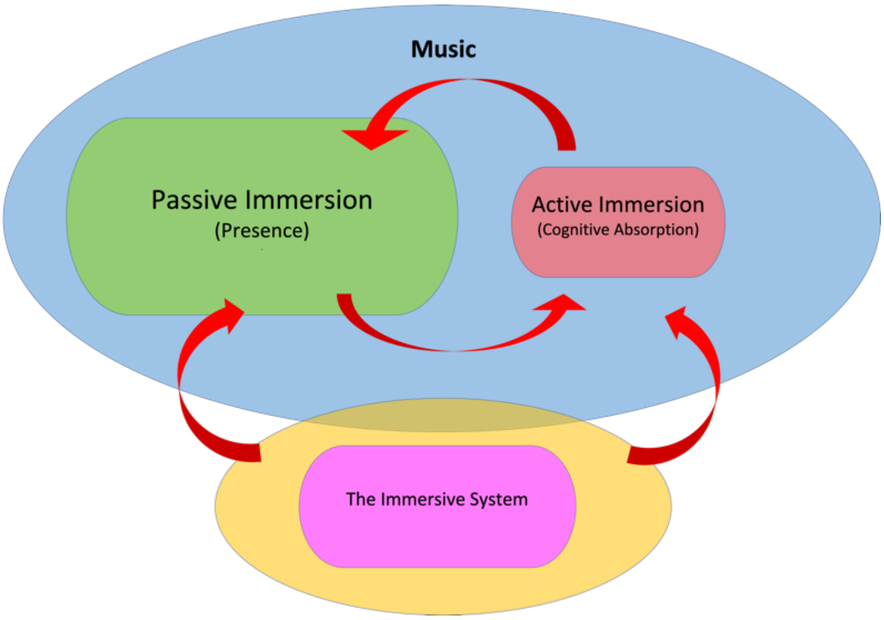


Figure 2.2. – Immersion Paradigm relationship for music content

When music content is presented to a listener, immersion in this case is mainly focused on passive immersion and making the listener feel as if they are their at a concert for example. The system presenting this is important in defining how immersive the content may be, as it could be assumed that a 22.2 rendering of a performance would be more immersive than a mono recording due to the enhanced special qualities over a mono recording. Active immersion only plays a minor role for this type of content, for example you are still required to 'pay attention' to the content, and when listening over headphones, you have to actively concentrate on the recording in a different way to experiencing music in a real world scenario.

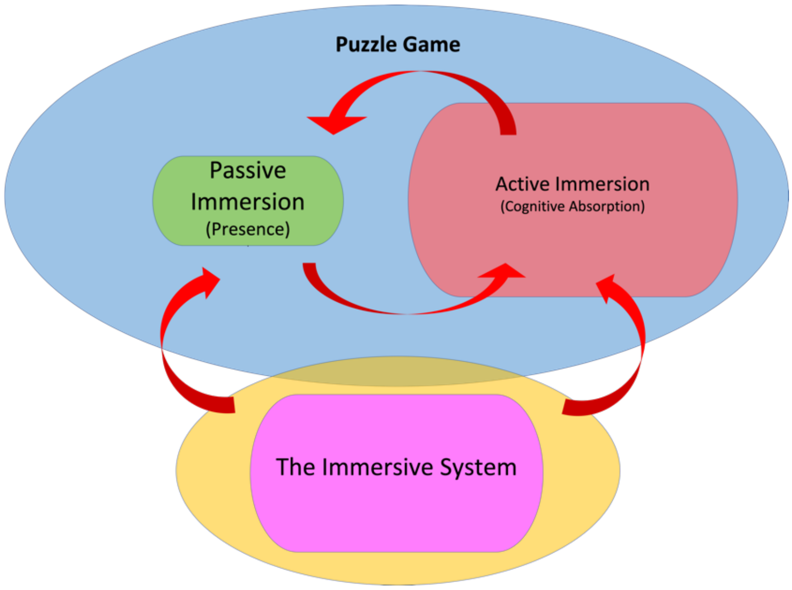


Figure 2.3. – Immersion Paradigm relationship for a Puzzle game

Figure 2.3 above shows the relationship of the paradigm in relation to a hypothetical puzzle game. In this case active immersion is the primary factor in defining the immersive experience with presence and the immersive system being of lesser importance.

This kind of catagorisation and content dependent relationship between the three elements of the immersion paradigm are of course hypothetical and not strictly to scale, and is meant more to show the changeability in categorising immersion and that this kind of multifaceted approach to defining immersion is absolutely needed. This also presents an interesting opportunity to quantify specific elements of an experience separately from others in terms of their immersiveness.

However, with all discussions of immersion and particularly auditory immersion, there is little in the way of quantification of what factors are important for providing immersion, and how they may differ for different types of content. Such information could be critically important to improving 'immersive' content in the future and this aspect of quantifying specific factors of auditory immersion in particular will be the remaining focus of this work.

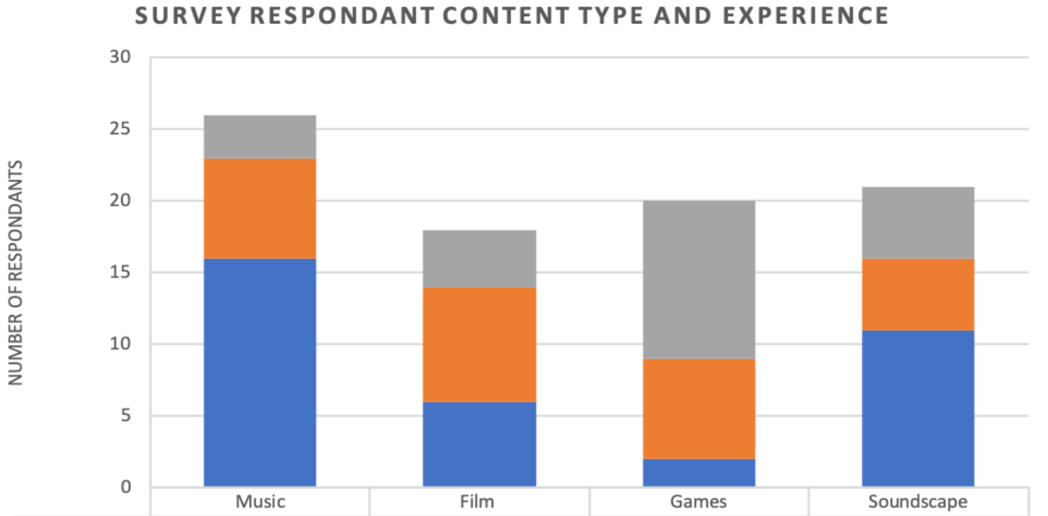
CHAPTER 3: Immersive Audio Survey

The issues with clearly defining immersion means it is imperative, at a time where the commercial popularity of VR for music in particular is on the rise, to give audio engineers and sound professionals as much practical information on how to make their work both immersive and enjoyable for the end user. As the ultimate aim of the current research is to provide some quantification to the issue of auditory immersion, it was imperative to identify parameters of auditory perception that would potentially be impactful in the perception of immersion.

It was decided that the identified auditory and technical factors of auditory perception in literature would be put in a survey where audio professionals and consumers could identify how important certain auditory factors may be for providing immersion for different types of content. The survey itself was run on google forms, and was posted on a number of Facebook groups (Ambisonics VR 360 Audio, VRContent Creators and Creations, Hey Audio Student, Virtual Reality Creative Community, Spatial Audio VR/AR/MR and Audio Engineering Society) centred around immersive audio, sound design/recording and audio engineering to attract as many professional respondents as possible.

Participants in the survey were first asked a number of catagorisation questions, age, content type for contextualization of the remainder of the survey (Games, Music, Film and Soundscapes). The participants were asked to answer the survey for the content type they were most familiar with, not necessarily on the content thy may have been working on at the time. This means that in all tables and figures as part of this chapter, when a content type is referred too, it is a catagorisation of those who answered the survey for that particular content type. Participants were also instructed they could repeat the survey again for a different content type if they also were familiar with it. Participants were then asked their experience of the content type they had previously selected (Novice, Intermediate or Expert) and their profession (Researcher, Content Creator, Consumer or Other).

At the point of writing this paper, the survey has had 85 responses. Figure 3.1 shows the number of responses for each of the four content types and shows the respective split of experience.



	Music	Film	Games	Soundscape
Novice	3	4	11	5
Intermediate	7	8	7	5
Expert	16	6	2	11

Figure 3.1. Survey respondents content type and experience

Participants were not given any kind of definition of immersion prior to the survey to base their answers around, this was done mainly to avoid biasing the answers particularly for those with novice amounts of experience. The test was also designed to be a chance to gather opinions to compare against formal literature so this separation from formal definition such as those offered by Slater (2018), Brown & Cairns (2004) or Nacke (2008) was felt to be particularly important.

Questions in the survey were split into perceptual auditory factors and technical factors that may influence auditory perception.

The identified perceptual auditory factors were as follows:

Q1: Vertical Listener Envelopment by Reverberation: (Impression of being engulfed by reverberation from above and below)

Q2: Vertical Listener Envelopment by Sound Objects: (Impression of being engulfed by sound objects from above and below)

Q3: Horizontal Listener Envelopment by Reverberation: (Impression of being surrounded by reverberation from all directions horizontally)

Q4: Horizontal Listener Envelopment by Sound Objects: (Impression of being surrounded by sound objects from all directions horizontally)

Q5: Vertical Localisation Accuracy: (How accurately, positively or negatively elevated sound objects are localised – low to high)

Q6: Horizontal Localisation Accuracy: (How accurately ear-height sound objects are localised – left to right)

Q7: Distance Localisation: (How accurately the distance of a sound object is perceived – near to far)

Q8: Apparent Source Width: (The audible impression of a spatially extended sound source)

Q9: Externalisation: (Whether a sound is perceived outside of the head)

Q10: Clarity: (How clearly sound sources can be distinguished and understood)

Participants were asked to rate each of the factors on a scale of 0-10 for how important that factor is for providing auditory immersion for the type of content they had previously selected (0 being not important at all and 10 being completely necessary for providing immersion). Box plot results for each question and content type are shown in Figure 3.2 & 3.3.

In Figure 3.2 & 3.3, there is again a minimal difference of perceived importance for certain perceptual elements dependent on content. Apparent source width, externalisation and clarity were all rated within one point of one another, suggesting that these factors in particular are less content dependent in terms of their impact on immersion. This is not the case for all aspects though, for example, Q5. Vertical localisation accuracy was scored a mean value of 7.85 for games, but only 5.35 for music suggesting it's importance for providing immersion is highly content dependent.

There is also a clear trend that vertical perception of both sound objects and reverberation were rated lower than the equivalent question regarding horizontal perception. It might generally be expected that to achieve a good level of immersion sound would need to be accurately presented both horizontally and vertically however the results of the survey would suggest that a greater importance is put on the perception of horizontal perception over vertical sound perception.

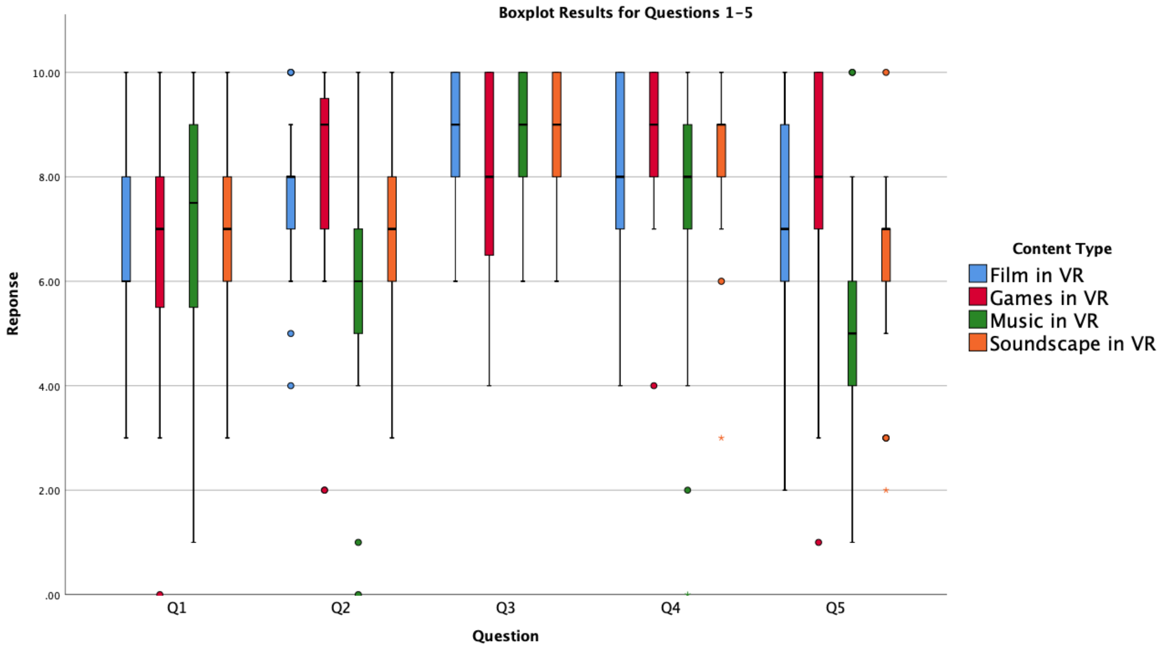


Figure 3.2. Boxplot results for survey questions 1-5

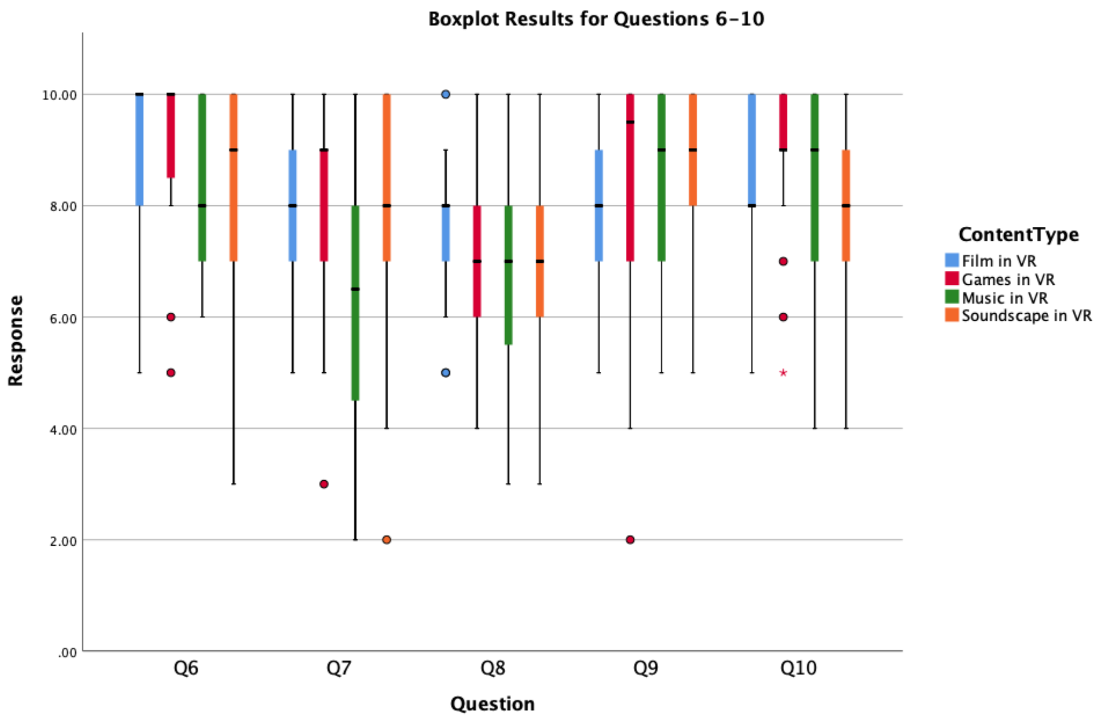


Figure 3.3. Boxplot results for survey questions 6-10

Following this participants were asked to rate the following technical factors on the same 0-10 scale, these factors were:

> *Q11: Individualised Head Related Transfer Functions (HRTFs): (HRTF's that are unique to you, not collected from a binaural dummy head)*
>
> *Q12: Head Tracking with 3 Degrees of Freedom (3DoF): (The ability of the VR system to track the position of the head in real space whilst continually keeping the positions of auditory objects in the same place with 3 degrees of freedom movement – pitch, roll and yaw)*
>
> *Q13: Head Tracking with 6 Degrees of Freedom (6DoF): The ability of the VR system to track the position of the head in real space whilst continually keeping the positions of auditory objects in the same place with 6 degrees of freedom movement – pitch, roll, yaw, forward/back, left/right and up/down)*
>
> *Q14: Latency: (The time taken for the signal to go through the system, e.g. from the computer to the VR headset)*
>
> *Q15: Headphone Equalisation: (The adjustment of a set of headphones frequency to be as close to neutral across the audible frequency spectrum in order for the HRTF spectral cues to be applied to the audio correctly)*

Boxplot results for each of the questions and content types can be seen in Figure 3.4.

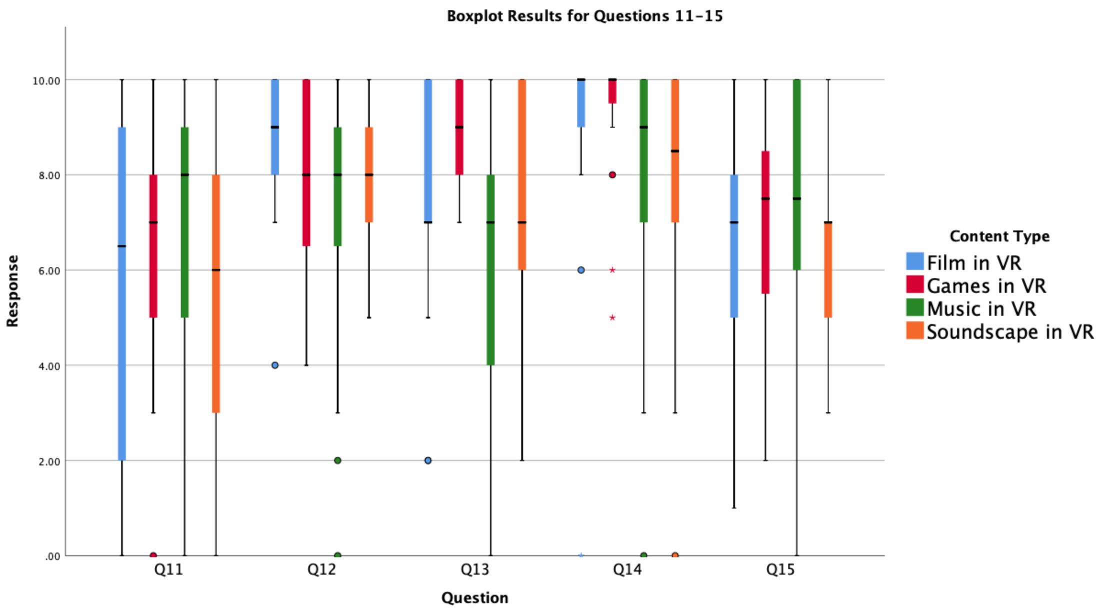


Figure 3.4. Boxplot results for survey questions 11-15

To confirm the initial observations of the data from the boxplots, a Wilcoxon signed rank test was performed on each question to test for significant differences between content types. The full results can be found in Appendix 2. Most of the results were found to have no significant differences once a Bonferroni Correction had been applied to the data. This correction was made to avoid Type-1 error where the *p* value is calculated as $p<0.05$ even when there is no significant difference between conditions. Given this correction, the results that did provide a significant difference result were as follows; Question 2, comparing Music and Games results (p=0.03). Question 5, comparing Music and Games (p=0.011). Question 7, comparing Music and Games (p=0.41). Question 13, comparing Film and Games (p=0.039) and Music and Games (p=0.007). Question 14, comparing Music and Games (p=0.043). All other comparisons returned a p>0.05 after correction, and were therefore found to be not significant results. All of the comparisons found to be significantly different bar one were comparing Music and Games responses, perhaps suggesting a particular disparity in the deemed needs for both content types to be immersive. Both Questions 2 and 7 refer to the vertical perception of sound, in both envelopment of sound objects and localisation accuracy. This perhaps makes sense as whilst vertical envelopment by reverberation may be perceived as more important for music content, the importance of envelopment of sound objects and localisation accuracy may be less important for that content type, whilst practically for game content this both factors would be considered highly important for an immersive experience, which is reflected in the boxplot results where the median response for games for both questions is significantly more than the music response. Furthermore, this results also highlights the potentially variable need for vertical sound reproduction for certain types of content. The general trend of no significance between content types for each question does highlight the general lack of concensus in both content creation and research perspectives about what factors of perception are most important for the perception of immersion and thus, provides a basis of enquiry for the remainder of the experimental portion of the report.

As has already been discussed the importance of congruency between auditory and visual perception is a high priority in literature discussion on the topic of immersion and is evidently also reflected in the opinions of survey participants.

Participants were then asked to rate on a 0-10 scale how much they thought the auditory and technical features of immersion would change depending on the content type, 0 being no change at all and 10 being a complete change. The results can be observed in Figure 3.5.

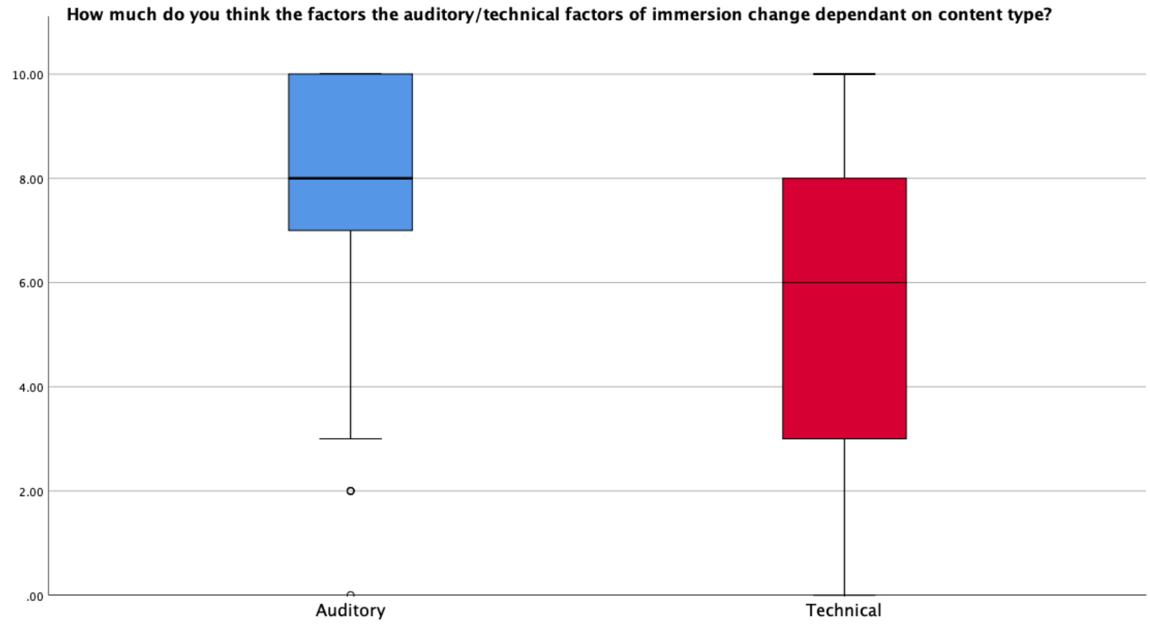


Figure 3.5. How much do you think the auditory/technical factors of immersion change dependent on content type?

Considering the original question of quantifying the factors of immersion, it can be gleaned from Figure 3.5, that the technical factors are less dependent for providing a sense of immersion compared to auditory factors, regardless of the content type. It perhaps suggests that whilst technical factors influence the perception of auditory factors, the auditory factors are perceived to more content dependent generally. Thus backing up the concept of the immersive system as a part of overall immersion perception explained previously in the literature review. Whilst this survey as a whole does provide an insight into people's expectations about what most influences the perception of immersion, it does not quantify it or address specific factors in an experimental environment and therefore it is this area that will be explored in the coming chapters.

CHAPTER 4: Experiment 1 – Optimal Speaker Layer Balance

As the both the results of the immersion survey conducted as part of the current study would attest, alongside the previous work outlined in the literature review, the potential aspects of perception that provide a sense of passive, presence-based immersion or indeed immersion as a whole may change greatly dependent on content. As the scope of the current work is simply too restricted to ascertain and quantify the factors of immersion for all types of content, with and without virtual reality, it was decided that a focus would be made on one specific content type. Whilst this potentially restricts the conclusions of the work as a whole, it is believed that this will allow the current study to come to a more finite answer to the overall research aim of the project in respect to that content type. As previously mentioned in the literature review portion of the current work, as well as the results of the immersion survey, there is potentially no significant evidence that the content type does indeed have a significant effect of the factors that define the sense of immersion, which is where this project is focused on and therefore adds to the justification of only including one type of content in the experiment phase of the current work. As alluded to previously, the results of the survey did show a disparity for vertical localisation for music content in particular which will be explored in experiment two, further rationalising the choice of this content in particular.

As the content creation expertise of the APL (Applied Psychoacoustics Lab) at the University of Huddersfield is within the field of concert hall recording, it was decided that this type of content would be the focus for the remainder of the study. The market of music recording for VR has been increasing along with VR content in general, and is also consumed by large numbers without VR too. This content type then not only provides a universal basis for quantifying auditory immersion but as the consumption of music content is not limited by the active immersion area of the 'Universal Immersion Paradigm' outlined earlier, it means the study can more easily conclude the technical and perceptual factors that lead to a more 'immersive' experience. This content type is usually experienced statically sat down and so speaker reproduction was determined to be the most optimum solution. This also allows the immersive system to be more closely considered on the impact it has on perceived immersion by comparing different speaker reproduction formats for example. The current experiment is justified further by King, Leonard & Kelly (2019) who ran an experiment to determine if there is an appropriate level for height channels as part of a 9.1 mix, motivated by the fact that determining if height channel levels set by only one mix engineer could be considered valid for use in experiments on immersive audio. In their experiment participants had to mix height channels to a determined appropriate level starting from a level ±6 dB from a reference mix by a mix engineer. Results showed a large range of reported mix levels for the three content excerpts tested, but results showed a 7-8 dB lower level mixed for the height channels than the reference mix. There are some flaws with this experiment that may have caused the result and present an opportunity in the current study. By not presenting the height channel level first at 0 dB this potentially creates a point of bias to the participant who may be influenced by the reference provided. Also, omni microphones were utilised for the height channels which potentially causes crosstalk between signals and may account in part for the range of results. This work identifies clearly the importance of creating an average mix when considering immersive audio reproduction and serves as a basis for the current experiment.

4.1 Listening Test Stimuli

As alluded to previously, concert hall music was chosen as a content focus for the experimental portion of the current work. It was decided that four music excerpts were deemed necessary, all with different numbers of musicians and instrument arrangements. It was also decided that all recordings were to be presented natively in 22.2 with potential flexibility to downmix to lower order formats for further experiments.

The first content excerpt to be captured was a brass band, identified as Brass in the rest of the report. The recording was conducted in St. Pauls Church on the University of Huddersfield campus. The full mic list and arrangement diagram can be found in Appendix 3. The technique used was PCMA-3D Lee & Gibben (2015), although due to an unavoidable technical error, the front centre floor (FCf) microphone was not operational. The subwoofer channels are also not utilised for this recording or indeed any of the remaining content excerpts. This means the Brass excerpt is actually in a 21.0 format, though as it is based on the 22.2 format, the original recording will be described as 22.2 for the purpose of the report.

The second content excerpt was also captured at St. Pauls. This recording featured a string quartet and is designated as Strings in further description. The recording technique used was again based around a modified PCMA-3D array (Appendix 3). Again, due to technical limitations, a number of 22.2 channels were not catered for. These being; FCh, BC and BCh respectively (see Figure 4.1). Again no subwoofers were utilised so the final Strings native mix is in a 19.0 format. As with the Brass excerpt, the original recording will also be designated as 22.2 in all further analysis. Whilst this may present a potential issue in data reliability, it is also felt to represent the reality of content creation limitations where technical issues, or equipment limitations may arise. This will be considered during the discussion of experimental results in future chapters of the report.

Two other recordings were utilised during the experiments, these being an orchestral excerpt of Gustav Holst's Mars suite, designated as Holst for analysis, curtesy of Howie, King, & Martin (2016) and also an orchestra and choir recording, designated as Choir, provided curtesy of Toru Kamekawa, Tokyo University of the Arts. Both recordings are natively presented in full 22.2. The recording technique used for the Holst excerpt combined a frontal sound capture array based on a 'Decca Tree' model using omni directional microphones, with height channel mics hung above the studio floor. Full lists of mics and layout can be found in (Howie et. al. 2016). The choir excerpt was also based around a Decca Tree configuration for FL, FR, FC channels and utilised A-B spaced pairs for ambience capture. A number of instrument mics were also combined into the final 22 channel output used in the current study, the layout and mic list can be found in Appendix 3.The four content excerpts whilst all being classical in composition, vary in timbre and tempo and so it is believed provide a solid content basis for experiments going forward.

4.2 Methodology

The mix process for 22.2 recordings relies on the careful balance of the height and floor speakers to provide a stable sound image and an overall enjoyable listener experience. This process is obviously highly subjective and creative, and as such, it makes experimenting with something that has only been mixed by one person potentially open to bias. The primary aim of the experiment is to ascertain an average height and floor layer balance for the four 22.2 mixes that were outlined in chapter 4.1. This can then serve as a basis for downmixing to different speaker formats in further experiments. This task is observational in its purpose and so no hypothesis or expected result is offered at this stage. The following sections of the report will outline the process undertaken and the results gathered.

4.3 Test Method

22 speakers were used in this experiment (7 Genelec 8331A speakers, 15 Genelec 8040A speakers). They were placed in the arrangement seen below in Figure 4.1. 10 loudspeakers placed around the listener at 0°, ±30°, ±60°, ±90°, ±135°, 180° azimuth angle with 0° of elevation. Seven loudspeakers were place at azimuth angles of 0°, ±30°, ±90°, ±135° with +30° of elevation. Three speakers were placed on the floor at azimuth angles of 0° and ±45°, all with -30° of elevation. These angles were chosen in accordance with the 22.2 specifications outlined in Silzle et al. (2011), where listeners were positioned in the centre of the speaker array directly facing the 0° azimuth and elevation speaker.

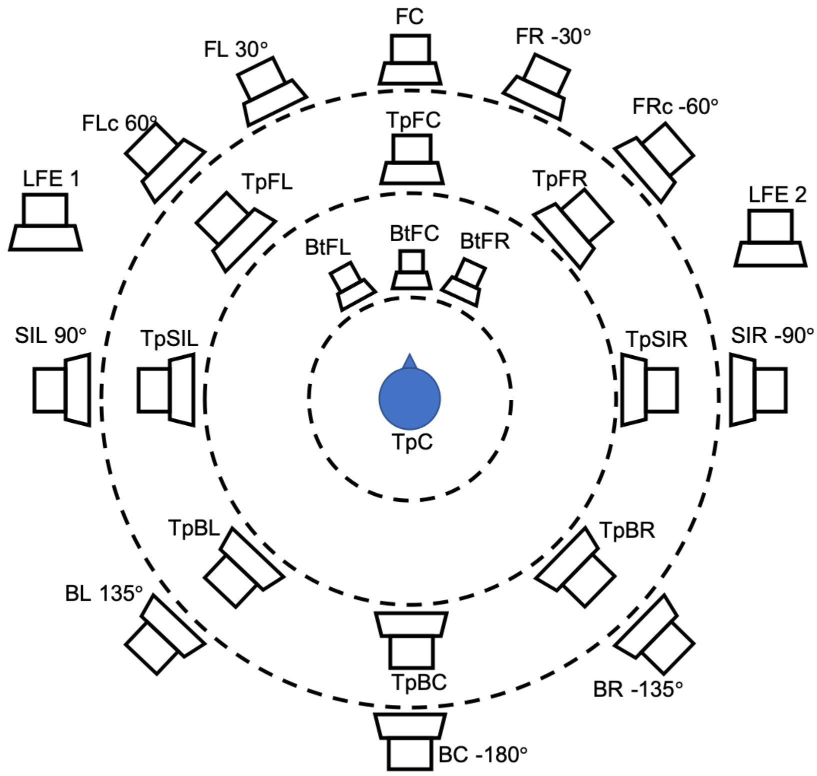


Figure 4.1. 22.2 Speaker Array Arrangement

The four excerpts used as stimuli for the experiment were previously outlined in chapter 4.1. seven participants (six male, one female) in total took part in the experiment. All participants were members of the APL and had prior experience of both listening tests and multichannel recording/mixing. The test itself was carried out in the APL's ITU-R BS.1116-compliant critical listening room at the University of Huddersfield (6.2m x 5.2m x 3.5m, RT = 0.25s, NR = 12). The Digital Audio Workstation used to manipulate the mixes was Cockos Reaper 64 (2006), where master busses for the 0° elevation layer (main layer), the -30° elevation layer (floor layer) and +30° elevation layer (height layer) were created. This method was chosen over allowing participants to control individual tracks primarily for time purposes. Participants were not expected to spend hours on each mix as would be taken in the real world and were expected to finish the mix process for all four stimuli excerpts in a one-hour test session, also allowing for breaks between stimuli. Also, as the overall aim of the test was to provide an average mix that could be used for mixing down to different speaker formats, maintaining balanced mix was felt to be of the upmost importance to not impact on the remaining experiment. A Korg nanoKONTROL 2, was connected to the Apple iMac running Reaper and was programmed to give corresponding physical interface control of both the height and floor layer master bus faders. This was done mainly to provide participants a more familiar mix experience. It also allowed the participants to mute and solo speaker layers to audition the levels they had selected, again more closely simulating a more natural mix experience under the test conditions. The audio signal chain was completed with a Merging Horus D/A converter used in coordination with Merging's ANEMAN, audio networking software to route the signals from Reaper to the correct speaker locations.

Participants were instructed to adjust a master slider for both the height and floor channels respectively to a level that they deemed most preferable. The participants began the test with only the main ear level speakers playing audio and had to mix the height and floor master faders from there. This was done to attempt to prevent participants being biased by a starting fader and loudness level for the height and floor layers and to provide a universal basis for all participants to mix around as the differences in opinion on the appropriate level for the height and floor layers was of particular interest. As previously mentioned subjects were allowed to mute and solo all layers of the 22.2 array independently to create as natural a mix process as possible. Head rotation was permitted in the experiment however listeners were instructed to remain seated where they were initially positioned in the sweet spot of the speaker array.

Subjects were required to complete this mix task for all four of the recordings as outlined in chapter 4.1. Furthermore, once participants were satisfied with their mix, they were asked to justify the reasons why they mixed both the height and floor arrays to the level they did. This was asked to ascertain if there were similarities between participants reasoning for mixing at certain levels and also to help in outlining any perceptual features that may be linked to the participants perception of their most preferable mix. The results of this questioning were collated in google forms and a discussion of the results of both portions of the experiment are to follow.

4.4 Comments

As previously stated, as part of the test, comments were collected from participants conducting the test to better understand not only the reasons behind their level selections, but to ascertain if there are any common reasons for the participants making the decisions they did. Allowing for a deeper insight into the preferable perceptual features in the mix of an auditory immersive content creation process. During the test, participants were advised to make as many comments for their decisions for the height and floor channel mix levels as they wanted but were instructed to include at least one reason for their decision to avoid participants mixing without thought to why they were doing it. To begin the review process, all comments were transferred from google forms into Microsoft Excel for more easy manipulation. The comments were then grouped depending on what overall perceptual or technical feature they were in reference to. These groups were derived from the raw data collected. The comments were not filtered by the content type they were in reference too as this was felt to potentially limit the ability to group the results, though the comments referring to the height and floor layers were separated, all comments both positive and negative were included and grouped together.

The comments were grouped into 8 categories based on the comments themselves, these were:

- *Comments referring to spaciousness or listener envelopment (which included references to a sense of envelopment or sense of space also references to spatial impression.)*
- *Comments referring to source elevation or vertical sound image spread (made reference to the level of height and floor layers causing sound image shift vertically.)*
- *Comments referring to reverberance directly (catagorised by any comment that specifically mentioned reverb or reverberance.)*
- *Comments referring to clarity, tonal quality or naturalness (which included any reference made to aspects of tonal quality.)*
- *Comments referring to the overall experience (catagorised by references made to how good or bad the mix level made to listening to the content.)*

- *Comments referring to a sense of presence or the term 'being there' (specifically catagorised by the use of the term 'presence' or phrase 'being there' as supposed to just talking about a sense of space or envelopment.)*
- *Comments referring to localisation, source width or horizontal sound image spread (included any references made regarding horizontal sound image change or localisation as these were often reported simultaneously by participants.)*
- *Miscellaneous comments, were any comment that could not be catagorised previously, examples included references to the physical perception of sound energy, environmental depth and comments referring to level balance.*

The amount of comments for each category and for both the height and floor layers were calculated and are displayed as a percentage of the total comments in the two figures below. As can be seen in Figure 4.2 the largest proportion of comments were focused around the topic of spaciousness or envelopment (36%) with references to vertical spread or source elevation (17%) and tonal quality (21%) also being most prominent in the reasoning of participants for choosing the mix level they did for the height channel mix specifically. For the floor channel mix, the results of Figure 4.3 show tonal quality being the most prominently commented on category at 40% of comments. Interestingly no references were made to either the floor layer adding to the overall experience of the mix or indeed to any comments referring to reverberance directly.

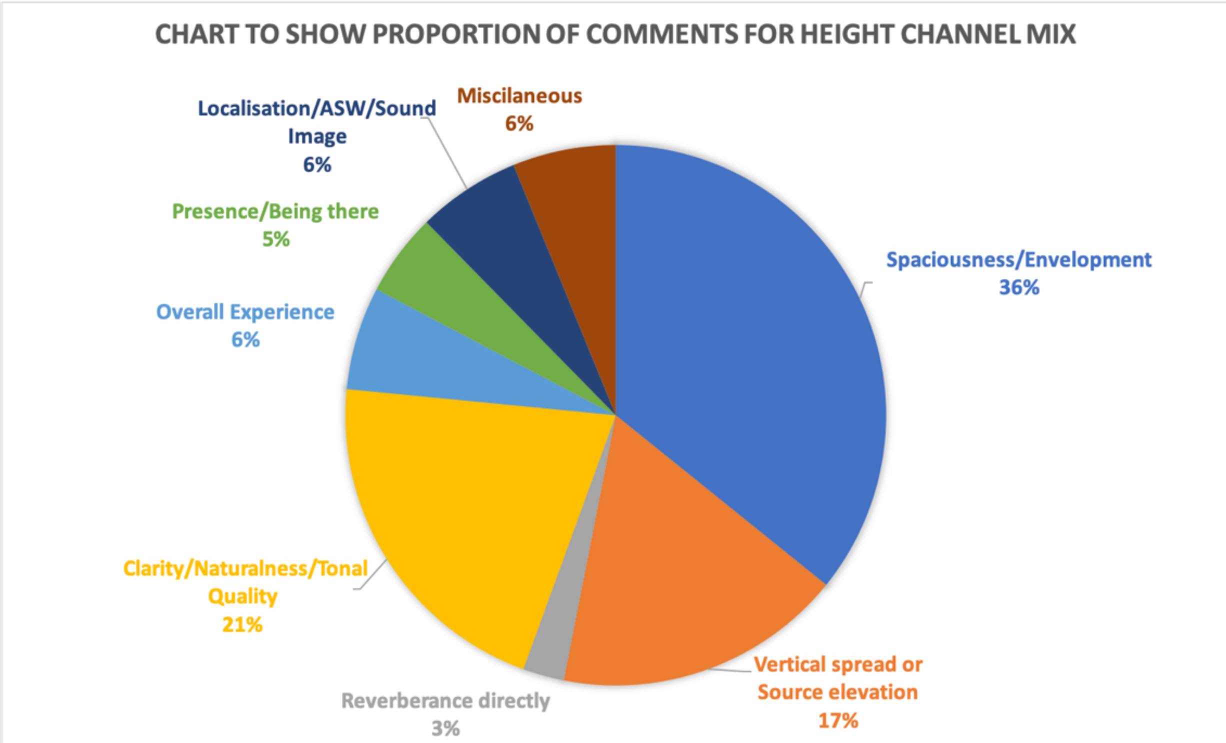


Figure 4.2. Pie chart to show proportion of comments for height channel mix

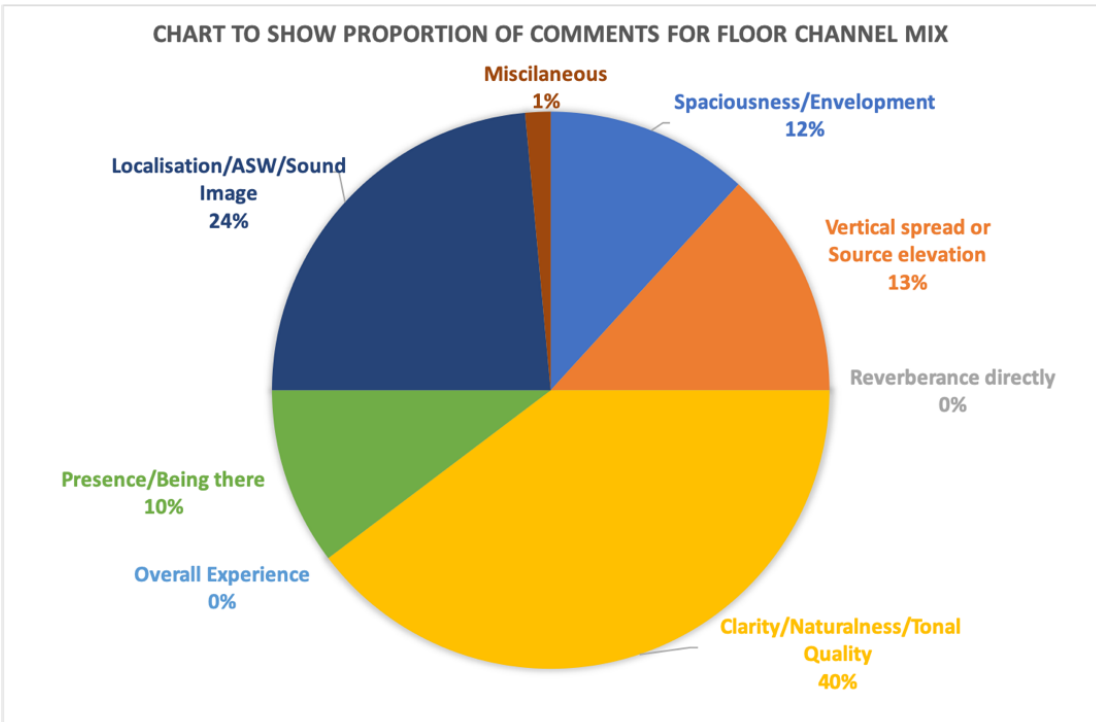


Figure 4.3. Pie chart to show proportion of comments for floor channel mix.

The differences in comments between the height and floor channels is certainly telling to what mix engineers may perceive the use of height and floor channels are. For example the high number of height channel comments regarding spaciousness and envelopment suggest the height channels may be more capable of adding or taking away the sense of envelopment than the floor channels. This may simply be down to the fact however that in the 22.2 array there are nine height channel speakers to only three floor channel speakers so this result may simply be down to that fact. The fact that the floor channels received a higher percentage of comments for both tonal quality and also presence is also of particular interest going forward. Furthermore, the floor channels seem to have a larger effect on localisation and horizontal source width than on elevation perception which may have been originally expected. This may be due to the floor channels reinforcing the frontal image of the main layer speakers and certainly within the Brass and Strings excerpts recorded the floor channel microphones were placed close at either side of the performers, which upon reproduction may have had the effect of widening the perceived sound image. Participants remarked that with the floor channels instruments often sounded more natural though with too much level may begin to sound ‘muddy’. Comments referring to presence and immersiveness were also in greater number (seven comments for floor channels (10%), four comments for height channels (5%)) perhaps suggesting that whilst the height channels may be able provide a heightened sense of envelopment in the mix, the floor channels are able to provide a more ‘natural’ and ‘present’ feeling for the participants whilst mixing.

4.5 Mix Balance Test results

As part of the mix balance experiment, the fader levels that the participants selected for the height and floor channels were recorded. Post-test these fader levels were measured and converted to long term average A-weighted SPL (LAeq), using a Casella CEL-450 sound level meter using an A weighting. The measurements were made at the listener position and measurements for each layer (Height and Floor) were taken three times to ensure measurement consistency. The results were then transferred into IBM SPSS Statistics software package for plotting and analysis. A Shapiro-Wilk test was utilised to test the normality of the data due to the small sample size (Appendix 4). Using this test the Choir Floor channel response was found to not be normally distributed ($p = 0.008$). All remaining measurements were found to be normally distributed: Brass Height (p = 0.151), Choir Height (p = 0.148), Holst Height (p = 0.966), Strings Height (p = 0.052), Brass Floor (p = 0.317), Holst Floor (p = 0.747) and Strings Floor (p = 0.521). As the purpose of the experiment was as an observation of participants mix decisions, no further statistical analysis was required. On observation of the results, a number of outliers were present in almost all responses, due to this, and the fact the median result of each condition was required for further experiments, the decision was made to utilise boxplots rather than error bars with 95% confidence intervals due to wanting to preserve the outlier data.

The boxplot below shows the results of the mix balance experiment. Note the main layer level shown by the line at 74 dB LAeq. As previously explained, this level was measured pre-test and was fixed during the experiment. The data therefore shows how the participants mixed the height and floor layer levels in relation to main layer level. It is important to also note that the figure displays level difference in SPL not the direct channel signal. This means that although the height may be lower than the main layer level for example, the interchannel level difference of each height microphone and corresponding main layer microphone may be the same or higher. For the purpose of analysis however the SPL values displayed in Figure 4.4 will be considered over the direct channel level.

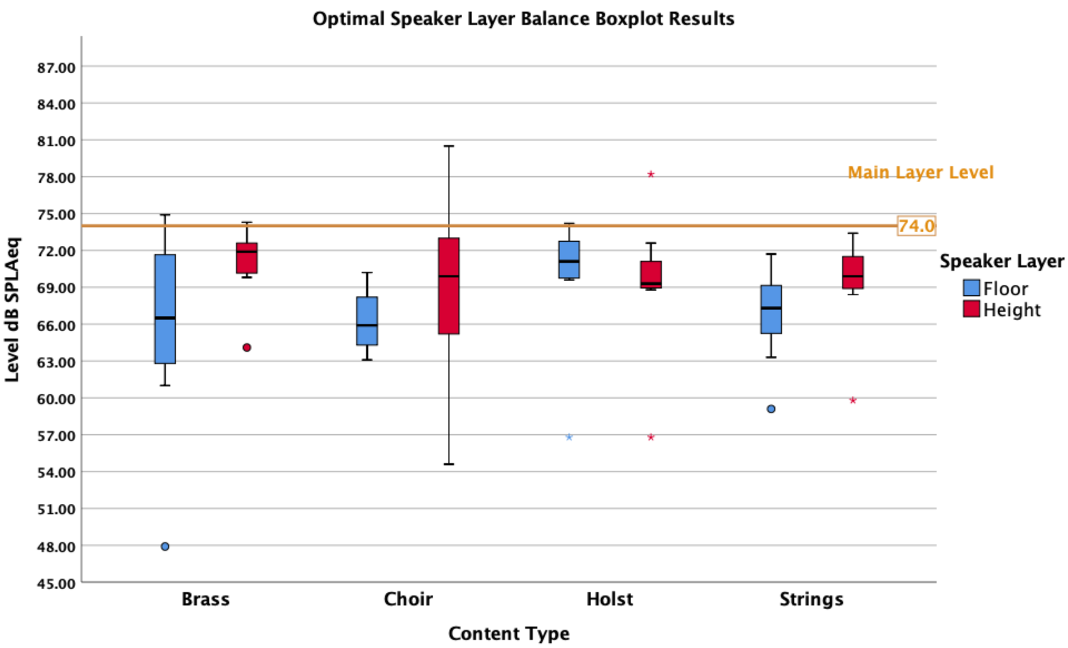


Figure 4.4. Boxplots to show the results of Optimal Speaker Layer Balance Test

For all four content excerpts, the median value of both the height and floor layers never exceeded the main layer level, and further to this point the upper quartile range of all of the layers and content types also did not exceed the main layer level. This suggests at least visually that participants utilised the height and floor channels overall to help enhance elements of the main layer mix they were initially presented with rather than building their mix from either the height or floor layers. This reason may have been due to keeping the main layer fixed; in a normal mix scenario, an engineer would have control over all elements of the mix process and so this may cause the results to differ. The consistency that the results, bar a few outlier data points maintain a level below the main layer would suggest perhaps that this would not be the case however. Although visually the ranges for each layer seem fairly compact, the median range across all layers and mixes extend from 65.9 dB LAeq for the Choir floor layer to 71.9 dB LAeq for the brass band height layer, giving a range of 6 dB LAeq, which given that figure represents sound pressure level over time, is a large difference. The range of values excluding outliers for each content excerpt also vary. The Holst orchestral recording for example has a height layer interquartile range of approximately 2 dB LAeq and a floor layer interquartile range of approximately 3 dB LAeq. Comparatively, the choir content excerpt had a much greater range of results with a height layer interquartile range of approximately 8 dB LAeq and floor layer interquartile range of approximately 3.5 dB LAeq. One difference between the two recordings is in the Holst excerpt, the height and floor layers contain mainly reverberant energy and more direct instrumentation is found more in the main layer. In the Choir recording comparatively, a number of direct instrument microphones were positioned in the height and floor layers. This perhaps explains the greater range in this excerpt as some participants liked this fact and others did not, whereas in the Holst recording the mix objective was more unanimous between participants who used the height and floor layers to 'fill out' the main layer signals. A similar observation can be made when the natural reverberance of the original recording is taken into account. The Brass, Choir and Strings excerpts are all highly naturally reverberant due to the space they were recorded in. All three of these recordings utilised a less floor layer level than height layer and comments from the previous section of the report link this to the perception of 'muddiness' and 'lack of clarity' if the level was raised to high. The

Holst recording on the other hand is comparatively dry and direct in its sound and in this case the floor layer had a higher median rating than the height channel and was linked in comments to a perception of 'natural floor reflections'. This perhaps suggests that the reverberance of the recording directly impacts the applicability of the floor layer, as with too much natural reverberance, the perception is often more negatively skewed, but with a less reverberant signal the level can be raised high and achieve a more positive overall reaction.

CHAPTER 5: Experiment 2 – Speaker Format Comparison

5.1 Methodology

In this experiment, two independent variables (IV) were used; speaker format and stimulus. The dependent variable (DV) for the experiment was the perceptual features, of which four were identified and will be explained further in the coming subchapters. To further the findings of the mix balance experiment outlined in the previous chapter, a second experiment was designed to aim to answer the overall research questions of the current work. In the literature review it was made clear that the immersive system is potentially instrumental in determining how perceptually immersive content may be. Following the Universal Immersion Paradigm theory as outlined previously, it can be suggested that with a more immersive system a greater sense of presence and reality can be perceived. Given this assumption it is important to test its validity. By comparing and analysing perceptual features alongside a variable system for presenting the same content it may be possible to ascertain how the immersive system provides the best immersive experience but also if there is a link between the system presenting the content and the perceptual experience of the listener.

5.2 Downmixing

The results of experiment 1 delivered a median mix level for the height and floor arrays for each of the four content excerpts. This median value was applied to create an averaged 22.2 master mix. At this point each of the 22.2 mixes were level matched at 77.5 dB LAeq (±1 dB) as close to each other in terms of average loudness that could realistically be achieved.

The aim of the experiment was to compare different common speaker formats and their ability to satisfy certain perceptual features associated with the perception of immersion, so downmixing of the 22.2 mixes to lower formats would be required. The formats that would need to be downmixed were; 9.1 (5+4), 5.1, 2.0 and mono. To downmix the audio the algorithm from Silzle et al. (2011) was utilised due to relative simplicity in a ‘nearest neighbour’ style downmix. The minimisation of downmix artefacts was also imperative and as this methodology had already been used to success in previous works it was felt to be trusted. The methodology was also chosen as it provided consistency for downmixing between the 9.1, 5.1 and 2.0 mixes. This was also felt to be particularly important in not creating a point of bias in participants preference of a particular downmix method or indeed the way a particular downmix methodology may rebalance the front/back sound image from the original 22.2 recordings. Tables 1 to 4 below outline the downmix algorithm used.

Table 1. Downmix Coefficients from 22.2 to 9.1.

FL = FLw + 0.7071(FL+SL) + FLf
FR = FRw + 0.7071(FR+SR) + FRf
RL = RL + 0.7071(SL+BC)
RR = RR + 0.7071(SR+BC)
FLh = FLh+ 0.7071(SLh+FCh) + 0.5(TpC)
FRh = FRh+ 0.7071(SRh+FCh) + 0.5(TpC)
RLh = RLh + 0.7071(BCh+SLh) + 0.5(TpC)
RRh = RRh + 0.7071(BCh+SRh) + 0.5(TpC)
FC = FC + 0.7071(FL+FR) + FCf
LFE1 = 0.7071(LFE1+LFE2)

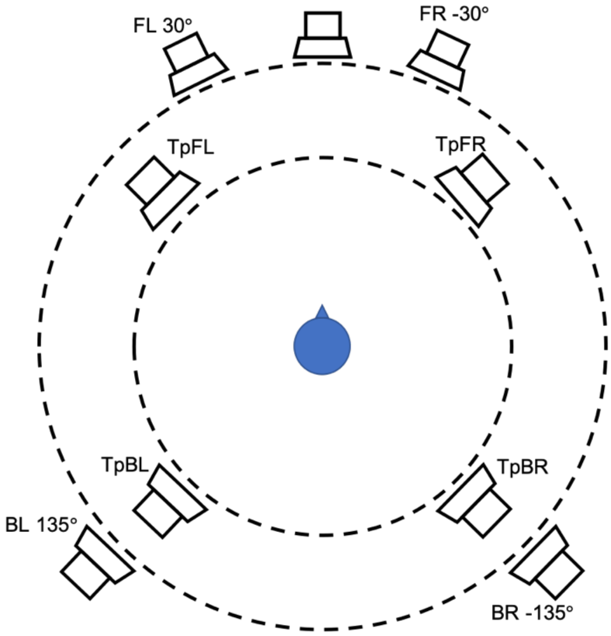


Figure 5.1. 9.1 Speaker Arrangement Used in Study

Table 2. Downmix Coefficients from 22.2 to 5.1.

FL = FLw + 0.7071(FL+SL) + FLf + FLh + 0.7071(SLh)
FR = FRw + 0.7071(FR+SR) + FRf + FRh + 0.7071(SRh)
RL = RL + 0.7071(SL+BC) + BLh + 0.7071(SLh) + 0.5(TpC) + 0.7071(BCh)
RR = RR + 0.7071(SR+BC) + BRh + 0.7071(SRh) + 0.5(TpC) + 0.7071(BCh)
FC = FC + 0.7071(FL+FR) + FCf + FCh + 0.7071(TpC)
LFE1 = 0.7071(LFE1+LFE2)

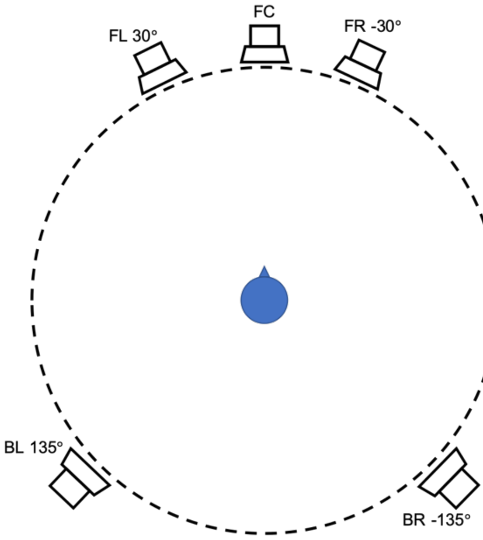


Figure 5.2. 5.1 Speaker Arrangement Used in Study

Table 3. Downmix Coefficients from 5.1 to 2.0,
mix was first downmixed using Table 2 coefficients.

FL = FL + 0.7071(FC) + 0.7071(RL) + 0.7071(LFE1)
FR = FR + 0.7071(FR) + 0.7071(RR) + 0.7071(LFE1)

Table 4. Downmix Coefficients from 2.0 to mono, first downmixed usage Table 3 coefficients.

FC = 0.7071(FL+FR)

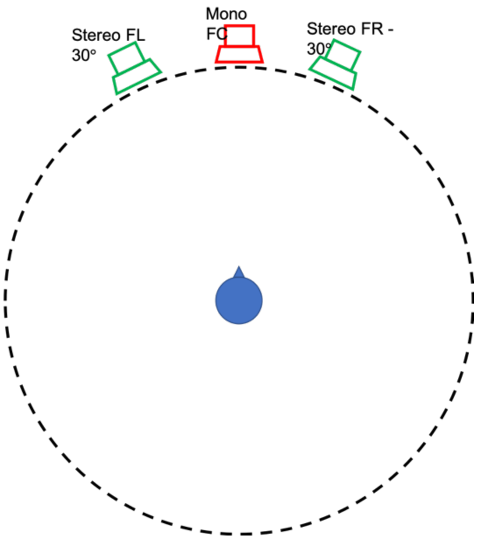


Figure 5.3. 2.0 and Mono Speaker Arrangements Used in Study

Where mixes did not have a channel present as previously explained in chapter 4.1, these channels were omitted from the algorithm. The effect was believed to be negligible, and the priority was to keep the mix balance as stable as possible. This may have an impact on the results of the experiment and so this point will be discussed later in this paper after final results have been presented. At each stage of the downmix process, the average loudness was measured and any minor adjustments to overall level were made to keep within the original 77.5 dB LAeq (±1 dB LAeq) range of the 22.2 mix.

5.3 Test Method

The aim of the experiment was to test multiple commonly used speaker formats 22.2, 9.1, 5.1, 2.0 and 1.0. against a number of perceptual features that between the survey, literature review, and the mix balance experiment have been identified as being potentially most important in determining overall auditory immersion perception. As in Experiment 1, 24 loudspeakers were arranged in the same way as Figure 4.1, with an identical signal chain as explained previously. The key difference being the interface that was used for data entry. HULTI-GEN (Huddersfield Universal Listening Test Interface Generator) Gribben & Lee (2015), in Cycling ’75 Max 7 was used to create a user interface that consisted of five sliders and corresponding play buttons (see Figure 5.4). Each of the sliders represented one of the downmixed audio files outlined in the previous subchapter and were randomised for each participant and each trial of the test, with HULTI-GEN outputting a text file post-test with all results reorganised for analysis. Each slider had a range of 0-100 with semantic labels of preference based on those outlined in the ITU-R BS.1534-1 specification. Participants were instructed to use as much of the scale as they wanted, and were not required to identify the most preferable stimuli as a score of 100 and the least preferable with a score of 0. The user interface was presented on a large TV screen in front of the listeners behind the line of the speaker array as not to cause any severe acoustic reflections. And was controlled with a mouse and keyboard placed on a small desk in front of the listener.

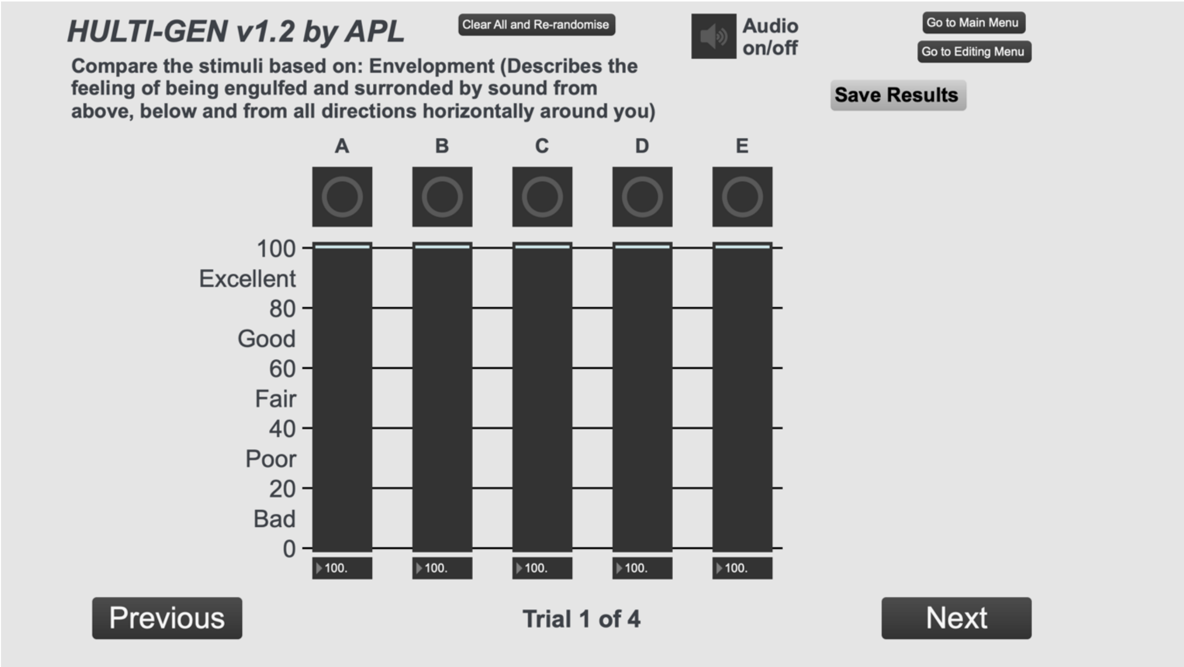


Figure 5.4. User Interface for Experiment 2

The test was designed to be multiple comparison but without a reference. As the aim of the experiment was to ascertain how much each of the downmixes were able to satisfy the definition that the participants were answering for. A reference stimulus was also not provided as this would have forced users to assume the 22.2 original mix was the most preferable as in the style of a MUSHRA (Multiple Stimulus with Hidden Reference and Anchor) test. This assumption was not desired in the current work and so no reference was included so it would be up to the discretion of each participant how to rate each of the stimuli. Each trial of the test focused on one of the four stimuli at a time, either the Brass, Holst, Strings or Choir recordings as explained previously. Each participant completed two repeats of each stimuli for each of the four perceptual features identified for the test meaning 8 trials per perceptual feature and 32 trials in total per participant. The four perceptual features that were identified as a result of the literature review and previous experiments to test against were;

- Quality of Experience (QoE): Describes the degree of satisfaction of the user towards a piece of content. (Schoeffler et. al. 2017)
- Presence (Pres): Describes the overall feeling of 'being there' whilst experiencing the content.

- Listener Envelopment (LEV): The feeling of being engulfed and surrounded by sound from above, below and all directions horizontally around the listener.
- Overall Tonal Quality (OTQ): Relates to features of sound such as timbral/tonal quality, intelligibility and clarity, not to be confused with spatial quality or envelopment.

In this experiment, 11 total participants (10 male, one female) took part. Participants all had some prior experience with spatial audio testing, and were given some time before the test began to understand the user interface and audio being presented, as well as the instructions for the test itself. Participants had to rate all of the downmixes for each recording before moving onto the next one. The order of the downmixes were randomised for each trial and the order that participants answered for each definition was also randomised between each participant. All of these measures were aimed at mitigating bias. Before each participant began the experiment they were presented with the definition they would be assessing each trial against, and any questions regarding the definitions were answered at this stage.

Participants were sat in a fixed location but were able to move their head freely. There was no visual cues utilised in the experiment. This was done as the positive impact that visual cues have on the overall perception is well regarded in literature as it stands. More in-depth study on auditory immersion as a sole focus however, has not yet been largely explored. As auditory perception is the focus of the current work, it was felt to be appropriate to omit this condition from the experiments so that any differences in perception that are noted in the results and discussion portion of the current work can be more easily attributed to auditory perceptual features rather than a positive visual bias. If the project were to continue to extend the findings of the current work, then it would certainly be recommended to repeat the experiment with a visual and no visual comparison element as this would also allow for the ascertainment of how impactful visual cues are in the perception of immersion for music content for VR.

5.4 Speaker Format Test Results

The HULTI-GEN Max 7 patch produced text files for each participants results for each condition and reordered the randomised data for easier data retrieval. The data was collated in IBM SPSS Statistics Version 26 software for analysis as was done in the first experiment of the report. From initial observation of the data, trends in responses seemed evident across different formats, however the presence of outliers was also noted. Outliers may have been caused by participants not recognising differences between different speaker formats, or perhaps certain participants utilising the scale in an incorrect way. To investigate the results of the experiment as a whole further, a number of statistical tests were performed and the results will be outlined below.

Before detailed analysis could begin, the normality of the data needed to be determined. A Shapiro-Wilk test was performed on each data set for each perceptual feature. As suspected, normality varied between test conditions with some being found normally distributed ($p > 0.05$) such as Strings_Mono_OTQ which returned $p = 0.218$, Brass_22.2_LEV ($p = 0.376$) or Choir_5.1_Pres ($p = 0.280$), and others not being found to be normally distributed ($p < 0.05$), such as, Holst_9.1_QoE ($p = 0.003$) or Strings_2.0_QoE ($p = 0.11$). Due to the large number of conditions tested further p values will not be discussed and full results can be found in Appendix 5. However, from initial observation no immediate trends could be observed in the data at this point, therefor requiring further analysis. Due to the presence of not normally distributed data, non-parametric data analysis techniques were trusted as the cause of outlier data was not yet clear so disregarding these points was not an option. The first point of analysis was to create boxplots to outline overall trends in responses. The following Figures 5.5-5.8 represent responses for each perceptual feature filtered by the stimulus type. In all cases, responses are clustered by speaker format.

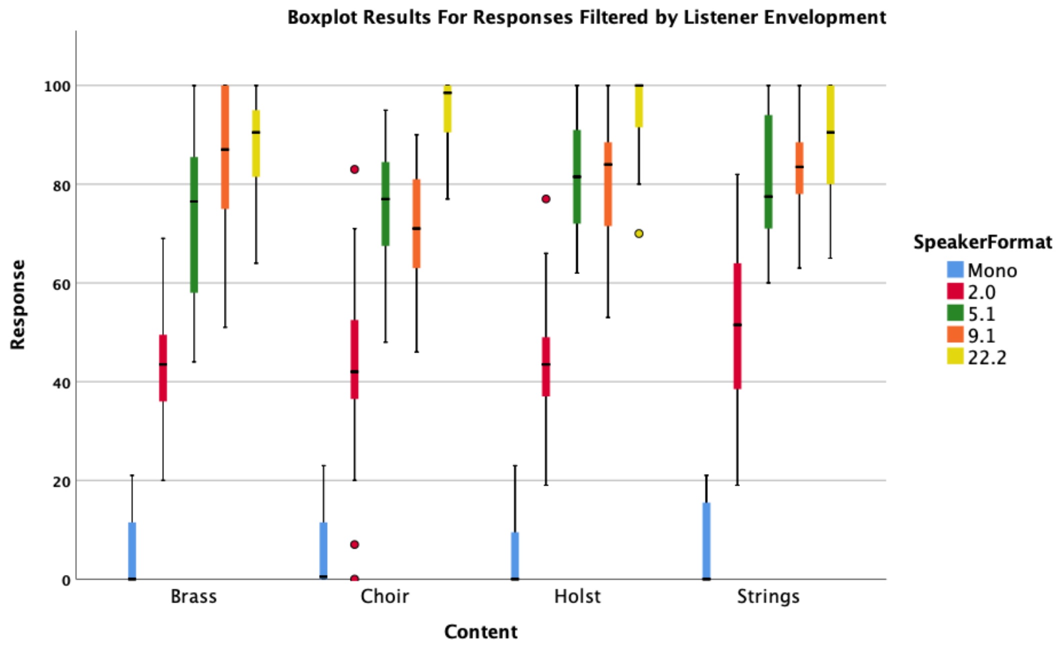


Figure 5.5. Results Filtered by Listener Envelopment

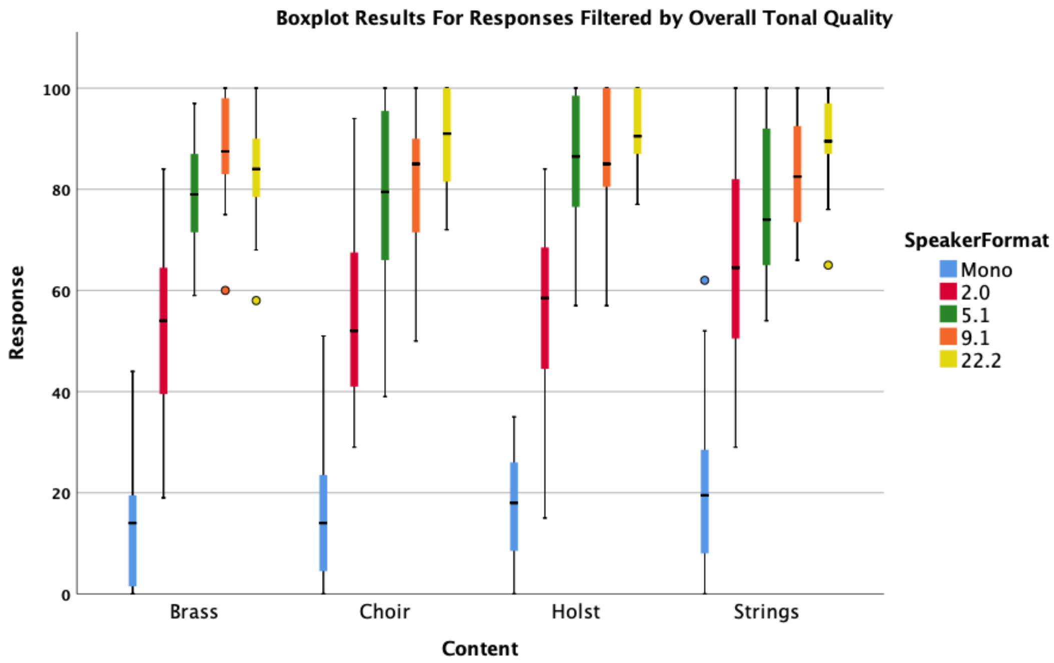


Figure 5.6. Results Filtered by Overall Tonal Quality

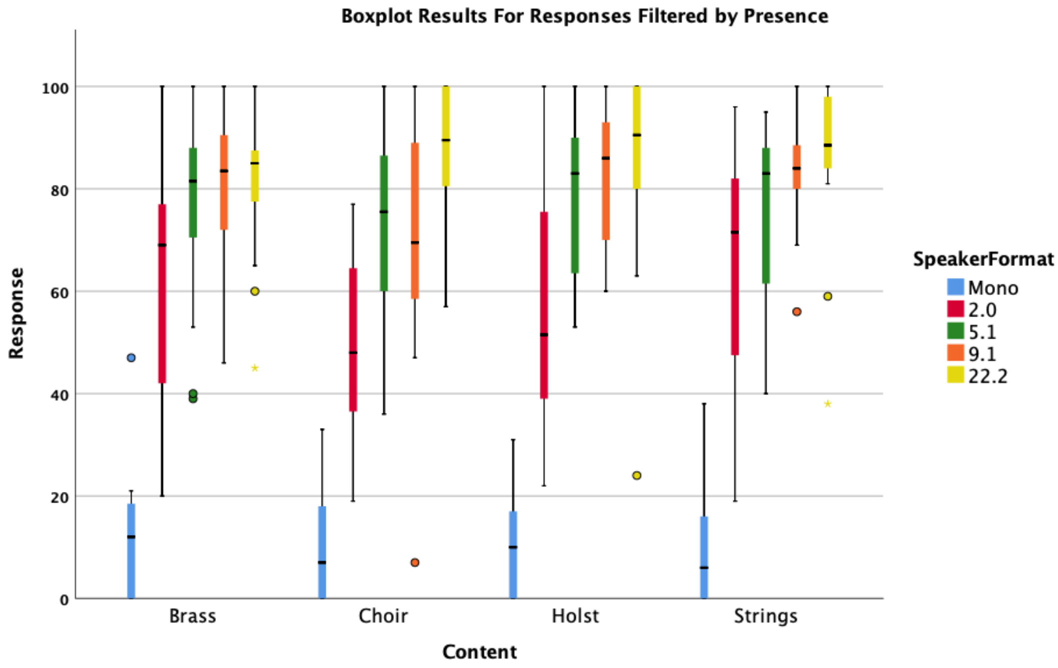


Figure 5.7. Results Filtered by Presence

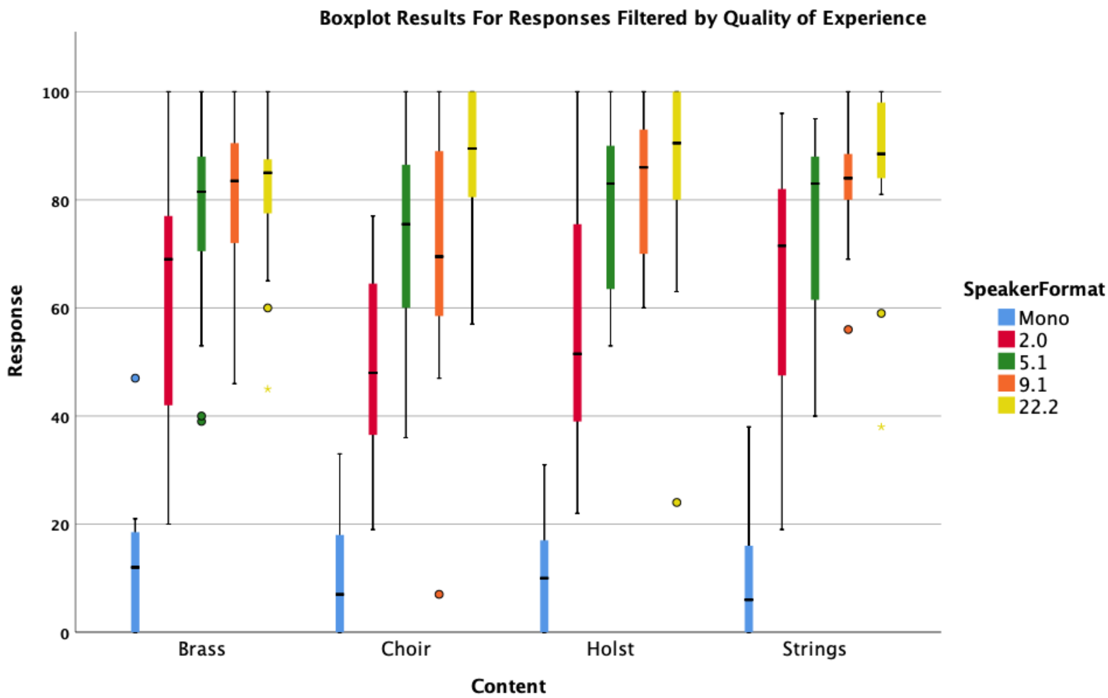


Figure 5.8. Results Filtered by Quality of Experience

There are a number of trends that are visible from the initial boxplot results. In all cases the median value of the mono speaker format is always the lowest in the 0-20 range on the scale. Similarly, the stereo 2.0 format was always being rated as second worst. What is interesting to note about the stereo format in particular is the wide range of values reported, consistently the stereo format had one of the largest ranges of responses across all content types and perceptual features. This perhaps indicates some indecision when rating the stereo channel in particular. Stereo is a format the listeners would have been extremely familiar with and so some listeners may have been biased to rate the format higher and closer to the larger formats simply due to familiarity. Further to this, the discrepancy seen may be down to the test design itself, as all formats could be compared to one another by the participants, stereo may have seemed very poor compared with 22.2 for certain attributes such as LEV whereas for other attributes such as QoE or OTQ may not have as large a discrepancy between formats. This also potentially then means the stereo format is much more changeable depending on the perceptual attribute and indeed on the participants preference in comparison to the other formats being tested. Similarly, larger ranges of responses are also seen in the 5.1, 9.1 and 22.2 channels so this is potentially not the case and only suggests some overall confusion in speaker formats above and including 2.0.

5.4.1 Friedman Test and Pairwise Comparison

Due to the previously ascertained non-parametric nature of the data set collected. The Related-Samples Friedman's Two-Way Analysis of Variance by Ranks test was utilised to test for the main effect of the format. This test allows for the comparison of multiple groups at once and compares the distribution of the data sets provided. In this case the null hypothesis (H_0) is that the distribution of scores in each data group are the same, and the alternate hypothesis (H_A) stating that at least two groups distributions differ. Data for the test was catagorised by content type with speaker format and was tested against each perceptual attribute. For example; Strings_mono, Strings_2.0, Brass_mono, Brass_2.0 and so on for all content type and speaker format combinations. The test was conducted by content type and for all perceptual attributes and in all cases the null hypothesis was rejected ($p < 0.05$). Due to this it was deemed necessary to determine if any parings from the Friedman test were statistically significantly different. The Friedman test in SPSS also generates a Pairwise comparison analysis if the initial null hypothesis is not accepted and so this data was utilised for further analysis, with particular attention to utilise the Bonferroni corrected significance value. Full pairwise comparison results can be found in Appendix 6, however in Tables 5 – 8 below, adjusted significance values for all comparisons can be observed.

Table 5. Bonferroni-corrected p values from pairwise comparison for LEV

	Brass	Strings	Holst	Choir
Mono – 2.0	0.455	.164	.357	.574
Mono – 22.2	0.000	.000	.000	.000
Mono – 9.1	0.000	.000	.000	.000
Mono – 5.1	0.000	.000	.000	.000
2.0 – 22.2	0.069	.023	.032	.164
2.0 – 9.1	0.000	.014	.007	.004
2.0 – 5.1	0.000	.000	.000	.000
22.2 – 9.1	0.801	1.000	1.000	1.000
22.2 – 5.1	0.316	1.000	.214	.004
9.1 – 5.1	1.000	1.000	.643	.164

Table 6. Bonferroni-corrected p values from pairwise comparison for OTQ

	Brass	Strings	Holst	Choir
Mono – 2.0	.574	.037	.357	.069
Mono – 22.2	.000	.000	.000	.000
Mono – 9.1	.000	.000	.000	.000
Mono – 5.1	.000	.000	.000	.000
2.0 – 22.2	.023	.643	.007	.214
2.0 – 9.1	.001	.143	.004	.069
2.0 – 5.1	.000	.000	.000	.000
22.2 – 9.1	1.000	1.000	1.000	1.000
22.2 – 5.1	.643	.244	1.000	.574
9.1 – 5.1	1.000	.989	1.000	1.000

Table 7. Bonferroni-corrected p values from pairwise comparison for Presence

	Brass	Strings	Holst	Choir
Mono – 2.0	.010	.019	.037	.124
Mono – 22.2	.000	.000	.000	.000
Mono – 9.1	.000	.000	.000	.000
Mono – 5.1	.000	.000	.000	.000
2.0 – 22.2	.574	1.000	.316	.060
2.0 – 9.1	.214	.143	.080	.060
2.0 – 5.1	.093	.003	.003	.000
22.2 – 9.1	1.000	1.000	1.000	1.000
22.2 – 5.1	1.000	.404	1.000	.801
9.1 – 5.1	1.000	1.000	1.000	.801

Table 8. Bonferroni-corrected p values from pairwise comparison for QoE

	Brass	Strings	Holst	Choir
Mono – 2.0	.164	.124	.357	1.000
Mono – 22.2	.000	.000	.000	.000
Mono – 9.1	.000	.000	.000	.000
Mono – 5.1	.000	.000	.000	.000
2.0 – 22.2	.012	.512	.023	.027
2.0 – 9.1	.006	.003	.016	.019
2.0 – 5.1	.002	.000	.000	.000
22.2 – 9.1	1.000	.891	1.000	1.000
22.2 – 5.1	1.000	.143	.188	.244
9.1 – 5.1	1.000	1.000	.244	.316

The results in the tables above show some interesting trends in the data which back up the initial observations of the boxplots. Across the data set as a whole, statistically significant differences were always found when comparing the 22.2 format with 5.1 and 9.1 formats across all perceptual features. 5.1 and 9.1 formats were also found to consistently statistically similar apart from the Choir stimulus for Listener Envelopment where the adjusted $p = 0.004$. The mono format was never found to be statistically similar to either the 5.1, 9.1 and 22.2 formats across all content types and perceptual features ($p < 0.05$). Initial analysis would suggest with some certainty that the mono format was rated significantly worse than those formats mentioned. Mono was also found to be statistically similar to 2.0 ($p > 0.05$), for QoE and LEV across all content types. For OTQ, similarity was also found for all content types apart from Strings ($p = 0.037$). For Presence, Brass ($p = 0.010$), Strings (0.019) and

Holst (p = 0.037) were also found to be statistically different. Further discussion of these points and others raised throughout the results phase of the report can be found in Chapter 6.

5.4.2 Further Analysis

Whilst initial pairwise comparisons alluded to some trends in the data, it was decided a further Friedman test would be conducted to investigate the effect of content type for each format and for each perceptual feature. This means for example, all the 2.0 format results for each content type would be directly compared alongside each perceptual feature, to test how responses differed depending on the content. Whilst each of the stimuli were not directly compared during the experiment itself, the same semantic scale was used so some indirect comparison can be made to potentially identify how content type influences the perception of individual speaker formats and perceptual features. Initially a Friedman test was trusted to test overall similarity between results, as in the previous sub-chapter, where data was not found to be significant a pairwise comparison was conducted to identify key significant differences.

Results of the Friedman test identified a number of conditions that were found to have no significant differences between content types. These were; mono format for LEV (p = 0.377), Presence (p = 0.267) and QoE (p = 0.241). 2.0 format for LEV (p = 0.116). 5.1 for OTQ (p = 0.071) and Presence (p = 0.123). 9.1 format for OTQ (p = 0.060) and Presence (p = 0.106). 22.2 for Presence (p = 0.901) and QoE (p = 0.078). All other conditions were found to be significantly different, in this case a further pairwise comparison was conducted. Results showed even when results did not satisfy the Friedman test, pairwise comparison still showed that the majority of conditions were significantly similar ($p > 0.05$).

Further to this, a Spearman's Correlation Test was performed on the data set to test the strength of correlation between content types for each perceptual feature, regardless of format results of which can be found in the table below. The Spearman test is the non-parametric equivalent of the Pearson Test, and so was chosen due to the non-parametric nature of the data found previously. The purpose of running this test is to observe if the trend of format difference found previously was consistent between content types.

Table 9. Spearman's Rho Correlation Analysis

	LEV	OTQ	Presence	QoE
Strings – Brass	0.457**	0.748**	0.696**	0.735**
Strings – Choir	0.579**	0.850**	0.757**	0.808**
Strings – Holst	0.583**	0.801**	0.839**	0.845**
Brass – Choir	0.842**	0.781**	0.661**	0.683**
Brass – Holst	0.856**	0.823**	0.675**	0.772**
Choir – Holst	0.897**	0.862**	0.805**	0.791**

**. Correlation is significant at the 0.01 level (2-Tailed)

The general trend of the correlation data shows largely strong positive correlation between all content types and for all perceptual features, an exception being the comparison of Strings & Brass for LEV with a correlation value of 0.457. This result may be due to both stimuli not having a full 22 speaker arrangement for the '22.2' format, however as this test is regardless of format this cannot be guaranteed as the result as when the Brass and Strings stimuli are compared to the other two stimuli much stronger correlation values are observed. The Choir & Holst stimuli for all perceptual features are highly correlated in terms of overall distribution and show the least deviation in their correlation coefficients between perceptual features. Overall the strongest correlation values are on average found for the OTQ and QoE perceptual features with LEV and Presence finding lower correlation values on average. This suggests that OTQ and QoE show a more consistent distribution of results meaning that whilst the speaker format still has an impact on the perception of the perceptual feature it is less variant depending on the stimuli. When compared to LEV and Presence where correlation results were lower on average and more deviant, it suggests that these perceptual features whilst still being influenced by speaker format as seen in the earlier analysis, are more content dependent overall.

Whilst the results of the initial Friedman and pairwise comparison tests and the correlation analysis provide a good insight into the results of the experiment, a further main effect test was desired to ascertain what influence perceptual feature had on content type regardless of speaker format and also speaker format regardless of content type.

For the content main effect analysis, the Freidman test was once again trusted. All responses were grouped by content type, regardless of speaker format. Analysis was filtered by perceptual feature and the results were; for LEV, $p = 0.84$, for OTQ, a significant effect was found $p = 0.029$, for Presence $p = 0.082$ and QoE, $p = 0.001$. Pairwise comparison showed however that all conditions were found to be significantly similar when the adjusted p for Bonferroni correction was considered. When the main effect of speaker format was considered regardless of content, for all perceptual features no significance was found in the Friedman test. Pairwise comparisons showed consistently that the 22.2 and 5.1 conditions were significantly similar for all perceptual features, and also that the 5.1 and 9.1 conditions were also significantly similar when the adjusted p value was considered. Full results of all main effect tests can be found in Appendix 8.

CHAPTER 6: Discussion

In this chapter, the results of both experiments will be summarised and analysis of the findings will be presented. Furthermore, comparisons to existing literature will be drawn and discussed to affirm the legitimacy of the results found in the current work.

6.1 Experiment 1 Discussion

In Experiment 1 – Optimal Speaker Layer Balance, the research objective 'What is the optimum layer balance for an immersive 22.2 music mix, and what perceptual motivations drive this decision?' was addressed. The results show that for all layers and all stimuli, the median level never exceeded the main layer level of 74 dB LAeq. The range of results for both height and floor layers as seen in Figure 4.4 are also notably similar with around a 5-10 dB LAeq range in most cases. Notably the results that do not conform to this trend are the Brass Floor layer with a range of ~14 dB LAeq and Choir Height Layer with a range ~26 dB LAeq. When considering why these results may have occurred, it is important to consider the content itself. The Brass stimuli is highly naturally reverberant, with the floor microphones positioned only approximately 1m from the performers due to space limitations. This means the impact of the floor layer in this case is not necessarily beneficial to the mix as a whole and was highly subjective in terms of what sounded most preferable. Also 40% of comments regarding the floor channels mentioned tonal quality as a reason for the mix decision and so this impact on tonal quality specifically could be attributed to this variance in result. The other stimuli floor layers were not as reverberant as the Brass stimuli so it could be assumed that high levels of reverberance in the floor channel is not universally preferable.

In the other stimuli tested, the height channels focus primarily on capturing ambient information about the room the recording is made in. In the case of the Choir stimulus, the height channels also feature direct instrument signals. This similarly makes the preference on how much level is appropriate potentially highly subjective. When Howie et al. (2017) looked at perceptual differences between formats, it was found the most common attributes reported were timbre (70%), spatial position of direct sounds (81%) and Spatial impression (94%). These factors were also found to be important to participants when mixing the stimuli, and therefore could potentially be most important to consider in the content creation process to guarantee a generally preferable result. Most interestingly is although the range of values were sizable for some speaker layers, the median values for the height and floor layers between content were relatively consistent. For the height layer the range of median values was 69.3-71.9 dB LAeq (2.6 dB LAeq) with an average of 70.25 dB LAeq. The floor layer median range was also similarly narrow at 66.5-71.1 dB LAeq (4.6 dB LAeq range) with an average of 67.7 dB LAeq.

This result presents an opportunity to offer an equation to define an optimum speaker layer balance for 22.2 recordings. This equation can be seen below. T refers to the total mix, *M* is the main layer level, *H* is the height layer level and *F* is the floor layer level, all measured in dB LAeq.

$$\boldsymbol{T = M + H + F}$$

$$\boldsymbol{H = M - 3.75dB\ LAeq}$$

$$\boldsymbol{F = M - 6.3dB\ LAeq}$$

Figure 6.1. Proposed Equation for optimum 22.2 layer balance

This equation means that regardless of the main layer level, a balance of height and floor layer can be achieved and according to the results of the current experiment, will achieve the most generally preferable balance. This result obviously needs to be caveated by the fact that in the current study, only seven participants were able to mix only four pieces of content, and those content pieces were all of similar concert hall recordings. However, whilst the scope of the current study is too limited to conclude the validity of this equation in a wider context than it is presented in.

6.2 Experiment 2 Discussion

Experiment 2 – Speaker Format Comparison, the research question 'Can auditory perceptual factors be identified and quantified as important for the overall perception of immersion?', was explored. The experiment utilised the mixes created as a part of experiment one to downmix the stimuli to different formats. Meaning they could be tested alongside perceptual features identified throughout the study in the literature review, immersive audio survey and experiment one as potentially the most influential factors on the perception of auditory immersion, which is the primary focus of the study as a whole. When the boxplots produced are compared to those in Hamasaki et al. (2006), Schoeffler, Silzle, & Herre (2017) and Silzle et al. (2011) who all compared speaker formats for different perceptual features, consistencies can be clearly seen.

In Silzle et al. (2011) overall sound quality; a perceptual factor most associated with Quality of Experience in the current study, was assessed with a two channel, five channel, nine channel and 22 channel setup without a reference. In the results of that study the two channel result was rated as least preferable, with five, nine and 22 channel excerpts being rated consistently better with the 22 channel being rated as most preferable, with some exceptions where the nine channel setup is rated highest. Comparably the result from the current study when comparing the results to Figure 5.8. for Quality of Experience, mono is rated as least preferable. This format wasn't included in the Silzle paper, however, the 2.0, 5.1,9.1 and 22.2 speaker formats follow a near identical pattern, although the overall rating is slightly inflated. It may also be due to the balance of the height and floor layers as part of the overall mix. In the current study this mix was somewhat conservative, with all height and floor layers being mixed lower than the main layer, which may be more practical as mentioned previously. The balance information is not disclosed in the Silzle work and so the observed differences may also be due to their samples having a more exaggerated height and floor layer, which in turn would potentially account for the differences in observed result. In the Silzle et. al. (2011) work, participants were also conducted a MUSHRA test with the 22 channel stimuli as a reference so it is possible when conducting the test without a reference similarly to the experiment carried out in the current study, the scale was utilised differently. In this case however even without a reference there was still a significant difference found between the 9 and 22 channel stimuli, a result also found in (Schoeffler et al., 2017). In the current study significant differences were not found, so it is also possible that this difference in result is due to the height and floor layers being mixed for optimal balance.

Interestingly in the boxplots for the current work, is a trend in the data for the Choir content for Listener Envelopment, Presence and Quality of Experience perceptual features where the 9.1 median is rated lower than both the 22.2 and 5.1 speaker formats. This is not found to be statistically significant in the pairwise comparison results, however this visual trend may be, as previously alluded to, that the choir stimulus has direct rather than reverberant signals present in the height layer. When these signals are mixed down, it may have shifted the source image that would be present in the 9.1 mix, where height channels are cut from 8 channels in the 22.2 mix to 4 in the 9.1 mix. This issue may not be as easily perceived in 5.1 mix as all height channels are mixed down into the main layer, possibly explaining why this stimulus suffered this visual anomaly in the data.

The initial Friedman and multiple comparison results support the initial observations made. Consistent significant similarities are found between the 5.1, 9.1 and 22.2 formats across all stimuli and perceptual features tested. The exception to this being the Choir Listener Envelopment case where whilst the 22.2 format was found to have no statistically significant difference to both the 9.1 ($p = 0.164$) and 5.1 ($p = 1.000$) respectively, when the 5.1 and 9.1 formats were compared with each other there was found to be a statistically significant difference ($p = 0.004$). As previously alluded too, this result may be due to the impact of the mixdown for this perceptual feature as previously discussed, this recording utilised a large amount of direct sound in the height channels and when downmixed to 5.1, a lot of information may be missing from the height layer, thus lowering the perceived LEV. In fact, when results of other content type and perceptual feature combinations are also analysed, 5.1, 9.1 and 22.2 comparisons are always found to have no significant differences. This at least suggests that no perceptual difference is perceived when comparing responses between these three formats. As discussed previously in the chapter, this result may be due to height and floor layer level determined from experiment one. The conservative levels chosen by participants may have meant that less differences could be perceived between these formats. Furthermore, as discussed, previous studies where this mix method was potentially not utilised produced a different result. It may be potentially argued that the levels for all layers should have no SPL difference, and that by doing this, more obvious differences would be observed. However, in practise, all mixes are subject to some adjustment based on the preference of the sound engineer. Therefore, the method used in the current study of finding an average mix by a number of engineers is considered to be more practically valid. Furthermore, as alluded to in Chapter 4, King et al. (2019) also conducted an experiment to determine optimum layer balance with a result that in general found that participants mixed height layer level 7-8 dB lower than the reference mix utilised. This would further back up the results shown in the current work that on average participants mixed the height and floor layers lower than the main layer level and as such creates an important practical distinction between the current study and previous work. Conducting the experiment again with a mix that has no level differences between layers may yield a different result but is another test condition in itself.

When addressing the implications this has on the research questions in the study as a whole, this result would suggest that 5.1 is as able as 9.1 and 22.2 to provide as good a sense of all 4 perceptual features tested, at least for the types of content tested in this study. This contrasts to the results of the study by Schoeffler et al. (2017) who when comparing a stereo, 'surround' and '3D Audio' stimuli for the perceptual features of Basic Audio Quality (BAQ) found there to be significant differences between the different audio presentation formats for all stimuli tested. This may suggest some error in the current work, however these differences may also be attributed to slight differences between experimental conditions rather than an error. The test conducted in the Schoeffler et al.'s study was a MUSHRA style test with '3D Audio' as a reference, so it is expected to be rated as 100 on the MUSHRA scale. As part of the same study a test was also conducted to investigate Overall Listener Experience (OLE) without using a reference, similarly to the current study and that conducted by Silzle et al. (2011) and in this case the variance of OLE ratings was much higher than compared to the BAQ results, as well as the increments from surround to 3D audio being notably lower for the OLE test compared to the BAQ test. The pattern of responses were shown to be highly similar to the data collected in the current study, with the 5.1 mix being found to have no significant difference to the 3D formats of 9.1 and 22.2, even though different perceptual features have been assessed. This perhaps then suggests that all perceptual attributes follow a similar trend as the ones tested. Of course, the only way to expand this statement is to conduct further experiments on alternate perceptual features, something that will be discussed in the further work section of the coming chapter.

When analysis was performed on formats individually the similarities in the data continued to be evident. It is likely that the anomalies discussed previously are simply due to low sample size, with further participants this error would likely reduce or if not would confirm the anomalies as something more significant. In the scope of the current study, as previously stated, the Choir stimulus presents its height channels differently than the other mixes so may explain why it is found to be different so often. It is also interesting to note that the mono signal was always found to have significantly similar responses according to the Freidman and pairwise comparison tests conducted. This was always rated lowest however this suggests strongly that this is always the case.

The final point of analysis is of the main effect tests (Appendix 8). When stimulus main effect was considered, Friedman tests on LEV and Presence retained the null hypothesis and were found to be statistically similar. Whilst OTQ and QoE were not found to be similar. From this along with the previous data, this potentially points to quality of experience not necessarily being highly linked to the perception of presence and to a wider extent immersion. It also suggests that although participants appreciation of envelopment was consistent between content, the tonal quality of the stimulus was more variant. From a practical sense this potentially means, the aspects of perception relating to qualities such as presence and envelopment are more linked to the technical format the audio is presented in, but are relatively consistent between different content types. However perceptual features such as tonal quality of overall quality of experience is linked to participants personal preference of the content they are experiencing.

CHAPTER 7: Conclusion

7.1 Conclusions

As immersion and more specifically, auditory immersion has become an ever more discussed topic in consumer and research circles it is important to understand what perceptual features underline this perception. It is for this reason in the current work literature was reviewed and summarised by the proposal of a new Universal Immersion Paradigm, consisting of components, passive immersion relating to the feeling of presence or ‘being there’, active immersion or cognitive absorption; the feeling of becoming lost in a task or activity and finally the immersive system, or how the technology itself can facilitate the perception of active and passive immersion. This paradigm gives a clearer focus to further research as aspects of immersion can be considered individually this way rather than as a whole topic. To further the literature review, an immersive audio survey was conducted to ascertain real opinions of consumers and professionals on what perceptual and technical factors are most important to the perception of immersion for different content types. Results showed an overall similarity of results between content types, which as discussed may be down to a relatively low sample size of 85, or perhaps hints at the similarity in perceived effect of different perceptual features on the overall perception of immersion. All perceptual and technical factors were also overall all rated as important to this perception which further underlines the confusion in the area of research and as such, presents a clear gap for the current study.

Concert hall music recordings were selected as a content focus of the study due to the experience of content creation of this type within the research group. An initial experiment was designed to ascertain an average optimal 22.2 speaker layer balance for each of the four stimuli gathered. Upon analysis of the results of this experiment, a formula for an average optimal speaker layer balance was proposed. The results are limited however, due to the limited scope of the current project, in terms of style of the stimuli, and number of stimuli being tested. Furthermore as the number of participants who conducted the experiment was only seven, to confirm the legitimacy of this proposed formula a significant amount of further work would have to be conducted, though it is hoped that this initial finding can serve as a basis to do so. Comments from this experiment also highlighted the differences in motivation for mixing height and floor layers of the 22.2 stimuli. With the height layer being more associated with spaciousness and envelopment perception and the floor layer being most associated with aspects of tonal quality and naturalness.

The individual average 22.2 mixes were used as a basis for downmixing the stimuli to 9.1, 5.1, 2.0 and mono for the second main experiment. The goal of which was to ascertain what impact speaker format, or more broadly the immersive system had on the participants perception of features that were identified as potentially most influential in the perception of passive, presence based auditory immersion. Results showed in the majority of analysis no significant differences were found between the 5.1, 9.1 and 22.2 conditions for all stimulus and perceptual feature combinations apart from a few outlier cases mainly concerning the Choir stimulus, though this potential anomaly here is due to the recording style itself and then how the downmix process may have impacted on the quality of the stimulus. This suggests in terms of the results collected as part of the current experiment that 22.2 and 9.1 are not statistically significantly superior to 5.1 for the perceptual features tested. When the main effect of the stimulus was analysed, results were found to be statistically similar for LEV and

Presence perceptual features but not for OTQ and QoE, suggesting perhaps that some perceptual features are equally important to different stimuli, however some perceptual features such as quality of experience are more influenced by the individual participants personal perception and are not so unanimously aggreged upon.

In terms of the impact of the current work on the overall debate around auditory immersion, the fact that stimulus is only found to have a significant effect on listener perception on some, but not all perceptual features tested, means this provides a clear opportunity for further research. Certainly it is hoped that the Universal Immersion Paradigm, outlined in this work provides a logical framework to better understand the complexity of immersion in a more targeted and manageable way, by breaking down immersion perceptions component parts.

As the title of the work outlines one of the main areas of interest in the current work was an attempt to quantify factors of auditory immersion for VR. Whilst the experimental work in this paper did not focus directly on the visual aspects of VR, important auditory factors were found and quantified for four musical excerpts. The results of experiment 1 outlined a proposed optimum layer balance for the mixing of auditory work for 22.2 speaker arrays. Whilst commercially, the use of such speaker arrays is limited, the results also glean what is important to the mix engineer when creating a 3D audio mix. The fact that reverberance was seemingly one of the most important factors to engineers whilst mixing height channels and tonal clarity was important whilst mixing floor channels, can provide a good starting point for mix engineers when creating work for VR too. The results of Experiment 2 also answer to some extent the research question proposed in the title of the current work. Four auditory factors were identified and showed clear differences in results between content and speaker format types. This information can be utilised by VR content creators who are deciding what audio format to present their work in, and therefore understand what perceptual features are most important to cater for, whether that be tonal quality or a sense of presence. Given that the computing power given to audio presentation in the VR format is always going to be less than the visual fidelity, knowing what perceptual features have the biggest impact on an overall sense of immersion is vastly important, and it is hoped the current work may be a catalyst for further quantification in this area. Furthermore, the Universal Immersion Paradigm proposed allows work in multiple specific areas of immersion as a whole, whether that be passive, active or the immersive system. This means that specific audio experiments can have as direct impact on the understanding of immersion as a whole.

7.2 Further Work

The study in its current form, whilst providing an initial insight and basis for further research into defining and quantifying factors of auditory immersion, is far from complete and certainly further study in this area is required. Further stimulus types and genres of content such as film or games must also be explored in order to provide a more universal appraisal of the research question at hand. Further participants in the current study would also have helped to improve reliability in both Experiment 1 and 2. In the immersive audio survey, only 85 participants took part, which is a relatively small sample. Increasing the number of participants here may have shown more clear differences in responses for certain questions. Experiment 1 only utilised seven participants and Experiment 2 used 11. Increasing the sample size of both experiments would be preferable to improve the credibility of results and would be a first priority in any further study.

In regards to Experiment 1, more stimuli comprising of different genres of music would be able to make the proposed formula more universally reliable. Increasing participant numbers as well as repeating the test may also yield more reliable data. Furthermore, allowing participants full control over the mix rather than just on a layer by layer basis is not ideal and typical in a professional environment, whilst intentionally done to ensure balance was ensured within each layer and also that the mixes could be completed in a reasonable time frame, if the experiment was to be redone, allowing participants more free control may have led to a different result.

In Experiment 2, a key area that requires further work and somewhat limits the results and conclusions of the current study is the lack of visual cues. The inclusion of visual cues would have allowed for comparison between visual and no visual conditions to analyse the impact this key factor in immersive content for VR is concerned. It must be said however, that as the focus of the content chosen and the study as a whole was on factors of auditory immersion specifically, the methodology used is still believed to be valid due to the current studies focus on auditory perceptual factors. To increase the validity of the results, further participants could have been utilised, potentially mitigating some further outliers from the data set and leading to more significant differences between different speaker formats. Speakers were utilised in the current study in part to do with the additional physical perception of sound they allow over headphones. A large proportion of content for VR is designed with the use of headphones in mind however. To this end, repeating the test with a binauralised version of each speaker format over headphones would be useful to compare with the current results, as potentially results may differ using this method. This does add complexity to the methodology as there are a significant number of methods of auralisation methods to choose from and each may present a different result. This depth of choice and variance underlines the scale of the task of quantifying the factors of immersion for virtual reality and it is why the results of the current experiment can only be considered truly within the scope they are set out in.

References

Agarwal, R., & Karahanna, E. (2000). Time Flies When You're Having Fun: Cognitive Absorption and Beliefs about Information Technology Usage. *MIS Quarterly*, *24*(4), 665. https://doi.org/10.2307/3250951

Altinsoy, M. E., & Stamm, M. (2013). *Touch the sound: The role of audio-tactile and audio-proprioceptive interaction on the spatial orientation in virtual scenes*. *19*, 5.

Berg, J., & Rumsey, F., (2001). "Verification and Correlation of Attributes Used for Describing the Spatial Quality of Reproduced Sound" *Proceedings of the 19th Audio Engineering Society Conference. Surround Sound: Techniques, Technology and Perception*, pp234–251.

Brown, E., & Cairns, P. (2004). A grounded investigation of game immersion. *Extended Abstracts of the 2004 Conference on Human Factors and Computing Systems - CHI '04*, 1297. https://doi.org/10.1145/985921.986048

Blauert, J. (1997). *Spatial Hearing: The Psychophysics of Human Sound Localization* (Rev. ed.). Cambridge, Massachusetts, London, England : The MIT Press.

Blythe, M., and Hassenzahl, M. "The Semantics of Fun: Differentiating Enjoyable Experiences," in Blythe, M.A., Overbeeke, K., Monk, A.F., and Wright, P.C. (eds.). Funology: From Usability to Enjoyment. Kluwer Academic Publishers, Dordrecht, (2003), pp. 91–100.

Cambridge Dictionary. (2019). *Immersion* . Retrieved from https://dictionary.cambridge.org/dictionary/english/immersion.

Christou, G. (2014). The interplay between immersion and appeal in video games. *Computers in Human Behavior*, *32*, 92–100. https://doi.org/10.1016/j.chb.2013.11.018

Clemente, M., Rodríguez, A., Rey, B., & Alcañiz, M. (2014). Assessment of the influence of navigation control and screen size on the sense of presence in virtual reality using EEG. *Expert Systems with Applications*, *41*(4), 1584–1592. https://doi.org/10.1016/j.eswa.2013.08.055

Csikszentmihalyi, M. Flow: The Psychology of Optimal Experience, Harper and Row, New York, 1990.

Ermi, L., & Mäyrä, F. (2005). *Fundamental Components of the Gameplay Experience: Analysing Immersion*. 14.

Ford, A., Ausiello, L., & Barlow, C. (2019). *Characteristics of Whole-body Vibration Induced Through Low-frequency Sound Exposure*. 8.

Grimshaw, M. & Garner, T., *Sonic Virtuality* 198 Madison Avenue, New York: Oxford University Press. (2015)

Hamasaki, K., Hiyama, K., Nishiguchi, T., & Ono, K. (2004). *Advanced Multichannel Audio Systems with Superior Impression of Presence and Reality*. 12.

Hamasaki, K., Nishiguchi, T., Hiyama, K., & Okumura, R. (2006). *Effectiveness of height information for reproducing presence and reality in multichannel audio system*. 15.

Heilig. M. L.. (1962). *U.S. Patent No. 3,050,870*. New York : U.S. Patent and Trademark Office.

Howie, W., King, R., & Martin, D. (2016). A Three-Dimensional Orchestral Music Recording Technique, Optimized for 22.2 Multichannel Sound. *Los Angeles*, 10.

Howie, W., King, R., & Martin, D. (2017). Listener Discrimination Between Common Speaker-Based 3D Audio Reproduction Formats. *Journal of the Audio Engineering Society*, *65*(10), 796–805. https://doi.org/10.17743/jaes.2017.0030

HTC Vive. (n.d.). In *Wikipedia.* Retrieved November, 19, 2017, from https://en.wikipedia.org/wiki/HTC_Vive

IGI. (2019). *CAVE Systems*. Retrieved from https://www.werigi.com/cave.

ITU-R BS.1534-1 (2003). Method for subjective assessment of intermediate quality level of coding systems. Intern. Telecom Union: Geneva, Switzerland

Jennett, C., Cox, A. L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., & Walton, A. (2008). Measuring and defining the experience of immersion in games. *International Journal of Human-Computer Studies*, *66*(9), 641–661. https://doi.org/10.1016/j.ijhcs.2008.04.004

King, R., Leonard, B., & Kelly, J. (2019). Height channel signal level in immersive audio—How much is enough? 4.

Larsson, P., Vastfjall, D., & Kleiner, M. (2002). *Better Presence and Performance in Virtual Environments by Improved Binaural Sound Rendering*. 8.

Lee, H. K., & Gibben, C. (2015). Effect of Vertical Microphone Layer Spacing for a 3D Microphone Array. *Journal of the Audio Engineering Society*, *62*(12), 870–884. https://doi.org/10.17743/jaes.2014.0045

Lokki, T., Hiipakka, J., & Savioja, L. (1999). *Immersive 3-D Sound Reproduction in a Virtual Room*. 6.

Meehan, M. (2001). *Physiological Reaction as an Objective Measure of Presence in Virtual Environments*. 142.

Merchel, S., & Altinsoy, M. E. (2013). *Auditory-Tactile Music Perception*. *19*, 8.

Mestre, D.R. (2019) Immersion and presence

Mestre, D. R. (2018). Presence in Virtual Reality: Insights from fundamental and applied research. *Electronic Imaging*, *2018*(3), 433-1-433–435. https://doi.org/10.2352/ISSN.2470-1173.2018.03.ERVR-433

Minderer, M., Harvey, C.D. (2016). Neuroscience: Virtual reality explored . *nature* , *533*(1), 324-325. doi: 10.1038/nature17899.

Nacke, L. (2008). *Flow and Immersion in First-Person Shooters: Measuring the player's gameplay experience*. 8.

Nacke, L., & Drachen, A. (2011). *Towards a Framework of Player Experience Research*. 6.

Oculus Rift. (n.d.). In *Wikipedia.* Retrieved November, 19, 2017, from https://en.wikipedia.org/wiki/Oculus_Rift

Pausch, R., Proffitt, D., & Williams, G. (1997). Quantifying immersion in virtual reality. *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques - SIGGRAPH '97*, 13–18. https://doi.org/10.1145/258734.258744

Poeschl, S., Wall, K., & Doering, N. (2013). Integration of spatial sound in immersive virtual environments an experimental study on effects of spatial sound on presence. *2013 IEEE Virtual Reality (VR)*, 129–130. https://doi.org/10.1109/VR.2013.6549396

Reaper 64. (2006). Retrieved from https://www.reaper.fm/

Rheingold, H. (1991). *Virtual Reality*, Simon & Schuster, New York, N.Y.

Roginska, A., & Geluso, P. (Eds.). (2018). *Immersive Sound, The Art and Science of Binaural and Multi-Channel Audio*. New York, NY: Routledge.

Sazdov, R., Paine, G., & Stevens, K. (2007). *Perceptual Investigation into Envelopment, Spatial Clarity, and Engulfment in Reproduced Multi-Channel Audio*. 11.

Schoeffler, M., Gernert, J. L., Neumayer, M., Westphal, S., & Herre, J. (2015). On the validity of virtual reality-based auditory experiments: A case study about ratings of the overall listening experience. *Virtual Reality*, *19*(3–4), 181–200. https://doi.org/10.1007/s10055-015-0270-8

Schoeffler, M., Silzle, A., & Herre, J. (2017). Evaluation of Spatial/3D Audio: Basic Audio Quality Versus Quality of Experience. *IEEE Journal of Selected Topics in Signal Processing*, *11*(1), 75–88. https://doi.org/10.1109/JSTSP.2016.2639325

Sensorama. (n.d.). In *Wikipedia.* Retrieved April 17. 2019, from https://en.wikipedia.org/wiki/Sensorama

Silzle, A., George, S., Habets, E. A. P., & Bachmann, T. (2011). *Investigation on the Quality of 3D Sound Reproduction*. 8.

Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1535), 3549–3557. https://doi.org/10.1098/rstb.2009.0138

Slater, M. (2018). Immersion and the illusion of presence in virtual reality. *British Journal of Psychology*, *109*(3), 431–433. https://doi.org/10.1111/bjop.12305

Sutherland, I. (1968). A head-mounted three-dimensional display. In Proceeding of the Fall Joint Computer Conference. AFIPS Conference Proceedings, vol. 33. AFIPS, Arlington, VA., 757- 764.

Tcha-Tokey, K., Christmann, O., Loup-Escande, E., & Richir, S. (2016). Proposition and Validation of a Questionnaire to Measure the User Experience in Immersive Virtual Environments. *International Journal of Virtual Reality*, *16 (1)*, 33–48. https://doi.org/⟨hal-01404497⟩

Appendix

Appendix 1 – Factors of Auditory Immersion Survey

Factors of Auditory Immersion Survey

My name is Callum Eaton, I am currently studying for a MSc by Research at the University of Huddersfield.

Working within the Applied Psychoacoustics Lab (APL), supervised by Dr. Hyunkook Lee, my project aims to quantify factors of auditory immersion and test the link between perception of immersion and overall user experience.

The purpose of this survey is to gather consumer and industry professional opinions on the factors that pertain to auditory immersion and immersion as a whole, within the context of spatial audio and Virtual Reality (VR) as a delivery format.

Thank you for your time in completing this survey.

Please direct any enquiries or questions to: callum.eaton@hud.ac.uk

To conform to the GDPR all data from this survey will be stored anonymously, if for any reason you no longer want your responses to be included in the survey results then please contact the email above and your data will be erased. All data will be destroyed when no longer required for the current project.

*Required

1) Pick one of the following content types to contextualise the remainder of the questions. *

Pick the content type you are most familiar with. If you are familiar with multiple content types, then feel free to repeat the test!

- ○ Film in VR
- ○ Games in VR
- ○ Music in VR
- ○ Soundscape in VR

2) Age *

- 0-24
- 25-44
- 45-64
- 65-74
- 75+

3) Experience of content chosen in question 1. *

- Novice (Limited experience)
- Intermediate (Good overall understanding and experience)
- Expert (Professional in creating/delivering this content type)

4) Profession *

- Researcher
- Content Creator/Producer
- Consumer
- Other

NEXT

Factors of Auditory Immersion Survey

Perceptual Factors

5) For each of the following perceptual auditory factors, rank how important you believe that factor is for providing auditory immersion for the content type selected in Question 1

0 being not important at all for providing immersion and 10 being completely necessary for providing immersion

If you do not understand a factor then please leave it blank, do not guess a response.

Vertical Listener Envelopment by Reverberation
(Impression of being engulfed by reverberation from above and below)

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Vertical Listener Envelopment by Sound Objects
(Impression of being engulfed by sound objects from above and below)

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Horizontal Listener Envelopment by Reverberation
(Impression of being surrounded by reverberation from all directions horizontally)

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Horizontal Listener Envelopment by Sound Objects

(Impression of being surrounded by sound objects from all directions horizontally)

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Vertical Localisation Accuracy

(How accurately positively or negatively elevated sound objects are localised - low to high)

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Horizontal Localisation Accuracy

(How accurately ear-height sound objects are localised - left to right)

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Distance Localisation Accuracy

(How accurately the distance of a sound objects is perceived - near to far)

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Apparent Source Width

(The audible impression of a spatially extended sound source)

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Externalisation

(whether a sound is perceived outside the head)

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Clarity

(How clearly sound sources can be distinguished and understood)

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Other perceptual features - (if no then leave blank)

Please list any other perceptual feature you would like and give a score from 0-10 in brackets. e.g. localisation (5). (Please enter responses in all lower case)

Your answer

BACK NEXT

Never submit passwords through Google Forms.

Factors of Auditory Immersion Survey

Technical Factors

6) For each of the following technical factors, rank how important you believe that factor is providing auditory immersion for the content type selected in Question 1

0 being not important at all for providing immersion and 10 being completely necessary for providing immersion

If you do not understand a factor then please leave it blank, do not guess a response.

Individualised Head Related Transfer Function

(HRTF's that are unique to you, not collected from a binaural dummy head)

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Head Tracking with 3 Degrees of Freedom

(The ability of the VR system to track the position of the head in real space whilst continually keeping the positions of auditory objects in the same place with 3 Degrees of freedom movement - pitch, roll and yaw)

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Head Tracking with 6 Degrees of Freedom

(The ability of the VR system to track the position of the head in real space whilst continually keeping the positions of auditory objects in the same place with 6 Degrees of freedom movement - pitch, roll, yaw, forward/back, left/right and up/down)

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Latency

(The time taken for a signal to go through the system, e.g. from the computer to the VR headset)

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Headphone Equalisation

(The adjustment of a set of headphones frequency response to be as close to neutral across the audible frequency spectrum in order for HRTF spectral cues to be applied to the audio correctly)

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Other technical features - (if no then leave blank)

Please list any other perceptual feature you would like and give a score from 0-10 in brackets. e.g. audio presentation format (5). (Please enter responses in all lower case)

Your answer

BACK NEXT

Never submit passwords through Google Forms

Head Tracking with 6 Degrees of Freedom

(The ability of the VR system to track the position of the head in real space whilst continually keeping the positions of auditory objects in the same place with 6 Degrees of freedom movement - pitch, roll, yaw, forward/back, left/right and up/down)

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Latency

(The time taken for a signal to go through the system, e.g. from the computer to the VR headset)

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Headphone Equalisation

(The adjustment of a set of headphones frequency response to be as close to neutral across the audible frequency spectrum in order for HRTF spectral cues to be applied to the audio correctly)

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Other technical features - (if no then leave blank)

Please list any other perceptual feature you would like and give a score from 0-10 in brackets. e.g. audio presentation format (5). (Please enter responses in all lower case)

Your answer

BACK NEXT

Never submit passwords through Google Forms.

Factors of Auditory Immersion Survey

*Required

Content Dependent Immersion

7) On the scale below, how much do you think the auditory factors that provide immersion change dependent on content *

	0	1	2	3	4	5	6	7	8	9	10	
Not at all	○	○	○	○	○	○	○	○	○	○	○	Completely different

8) On the scale below, how much do you think the technical factors that provide immersion change dependent on content *

	0	1	2	3	4	5	6	7	8	9	10	
Not at all	○	○	○	○	○	○	○	○	○	○	○	Completely different

BACK NEXT

Never submit passwords through Google Forms.

Factors of Auditory Immersion Survey

*Required

Physical Perception of Immersion

9) How much do you agree with the following statement: "Immersion can be lost when audio is reproduced over headphones, due to a loss of the physical perception of sound." e.g. Feeling the bass frequencies in your chest at a live concert. *

- Completely Disagree
- Somewhat Disagree
- Neither Agree nor Disagree
- Somewhat Agree
- Completely Agree
- Dont Know

BACK NEXT

Never submit passwords through Google Forms.

Appendix 2 – Immersion Survey Wilcoxon Test Results

Q1 Wilcoxon Test Results

	Q1Games - Q1Film	Q1Music - Q1Film	Q1Soundscape - Q1Film	Q1Music - Q1Games	Q1Soundscape - Q1Games	Q1Soundscape - Q1Music
Z	-.281[b]	-.694[c]	-.259[b]	-1.030[c]	-.573[c]	-.415[b]
Asymp. Sig. (2-tailed)	.779	.488	.796	.303	.567	.678

Q2 Wilcoxon Test Results

	Q2Games - Q2Film	Q2Music - Q2Film	Q2Soundscape - Q2Film	Q2Music - Q2Games	Q2Soundscape - Q2Games	Q2Soundscape - Q2Music
Z	-.571[b]	-2.107[c]	-.775[c]	-2.809[c]	-1.429[c]	-1.341[b]
Asymp. Sig. (2-tailed)	.568	.211	.439	.030	.153	.180

Q3 Wilcoxon Test Results

	Q3Games - Q3Film	Q3Music - Q3Film	Q3Soundscape - Q3Film	Q3Music - Q3Games	Q3Soundscape - Q3Games	Q3Soundscape - Q3Music
Z	-1.237[b]	-.212[c]	-.320[c]	-1.535[c]	-1.644[c]	-.484[b]
Asymp. Sig. (2-tailed)	.216	.832	.749	.125	.100	.628

Q4 Wilcoxon Test Results

	Q4Games - Q4Film	Q4Music - Q4Film	Q4Soundscape - Q4Film	Q4Music - Q4Games	Q4Soundscape - Q4Games	Q4Soundscape - Q4Music
Z	-.223[b]	-1.627[c]	-.159[b]	-2.050[c]	-.212[c]	-1.561[b]
Asymp. Sig. (2-tailed)	.823	.104	.874	.242	.832	.118

Q5 Wilcoxon Test Results

	Q5Games - Q5Film	Q5Music - Q5Film	Q5Soundscape - Q5Film	Q5Music - Q5Games	Q5Soundscape - Q5Games	Q5Soundscape - Q5Music
Z	-1.630[b]	-1.880[c]	-1.070[c]	-3.110[c]	-2.423[c]	-1.013[b]
Asymp. Sig. (2-tailed)	.103	.060	.285	.011	.092	.311

Q6 Wilcoxon Test Results

	Q6Games - Q6Film	Q6Music - Q6Film	Q6Soundscape - Q6Film	Q6Music - Q6Games	Q6Soundscape - Q6Games	Q6Soundscape - Q6Music
Z	-.475[b]	-.699[c]	-.592[c]	-1.586[c]	-1.775[c]	-.685[b]
Asymp. Sig. (2-tailed)	.635	.485	.554	.113	.076	.494

Q7 Wilcoxon Test Result

	Q7Games - Q7Film	Q7Music - Q7Film	Q7Soundscape - Q7Film	Q7Music - Q7Games	Q7Soundscape - Q7Games	Q7Soundscape - Q7Music
Z	-.052[b]	-2.529[c]	-.086[b]	-2.703[c]	-.429[c]	-2.046[b]
Asymp. Sig. (2-tailed)	.958	.069	.932	.041	.668	.245

Q8 Wilcoxon Test Results

	Q8Games - Q8Film	Q8Music - Q8Film	Q8Soundscape - Q8Film	Q8Music - Q8Games	Q8Soundscape - Q8Games	Q8Soundscape - Q8Music
Z	-.806[b]	-1.226[b]	-1.290[b]	-.493[b]	-.239[b]	-.418[c]
Asymp. Sig. (2-tailed)	.420	.220	.197	.622	.811	.676

Q9 Wilcoxon Test Results

	Q9Games - Q9Film	Q9Music - Q9Film	Q9Soundscape - Q9Film	Q9Music - Q9Games	Q9Soundscape - Q9Games	Q9Soundscape - Q9Music
Z	-.641[b]	-1.444[b]	-1.382[b]	-.623[b]	-.829[b]	.000[c]
Asymp. Sig. (2-tailed)	.522	.149	.167	.533	.407	1.000

Q10 Wilcoxon Test Results

	Q10Games - Q10Film	Q10Music - Q10Film	Q10Soundscape - Q10Film	Q10Music - Q10Games	Q10Soundscape - Q10Games	Q10Soundscape - Q10Music
Z	-.709[b]	-1.348[c]	-1.098[c]	-1.930[c]	-2.006[c]	-.172[c]
Asymp. Sig. (2-tailed)	.479	.178	.272	.322	.269	.864

Q11 Wilcoxon Test Results

	Q11Games - Q11FIlm	Q11Music - Q11FIlm	Q11Soundscape - Q11FIlm	Q11Music - Q11Games	Q11Soundscape - Q11Games	Q11Soundscape - Q11Music
Z	-1.105[b]	-.882[b]	-.312[b]	-.607[b]	-.525[c]	-1.253[c]
Asymp. Sig. (2-tailed)	.269	.378	.755	.544	.599	.210

Q12 Wilcoxon Test Results

	Q12Games - Q12Film	Q12Music - Q12Film	Q12Soundscape - Q12Film	Q12Music - Q12Games	Q12Soundscape - Q12Games	Q12Soundscape - Q12Music
Z	-1.303[b]	-2.540[b]	-1.539[b]	-1.091[b]	-.078[c]	-1.420[c]
Asymp. Sig. (2-tailed)	.193	.066	.124	.275	.938	.156

Q13 Wilcoxon Test Results

	Q13Games - Q13Film	Q13Music - Q13Film	Q13Soundscape - Q13Film	Q13Music - Q13Games	Q13Soundscape - Q13Games	Q13Soundscape - Q13Music
Z	-2.721[b]	-1.261[c]	-.694[b]	-3.254[c]	-2.086[c]	-1.801[b]
Asymp. Sig. (2-tailed)	.039	.207	.487	.007	.222	.072

Q14 Wilcoxon Test Results

	Q14Games - Q14Film	Q14Music - Q14Film	Q14Soundscape - Q14Film	Q14Music - Q14Games	Q14Soundscape - Q14Games	Q14Soundscape - Q14Music
Z	-1.166[b]	-1.419[c]	-1.035[c]	-2.027[c]	-2.105[c]	-.153[b]
Asymp. Sig. (2-tailed)	.244	.156	.301	.043	.212	.879

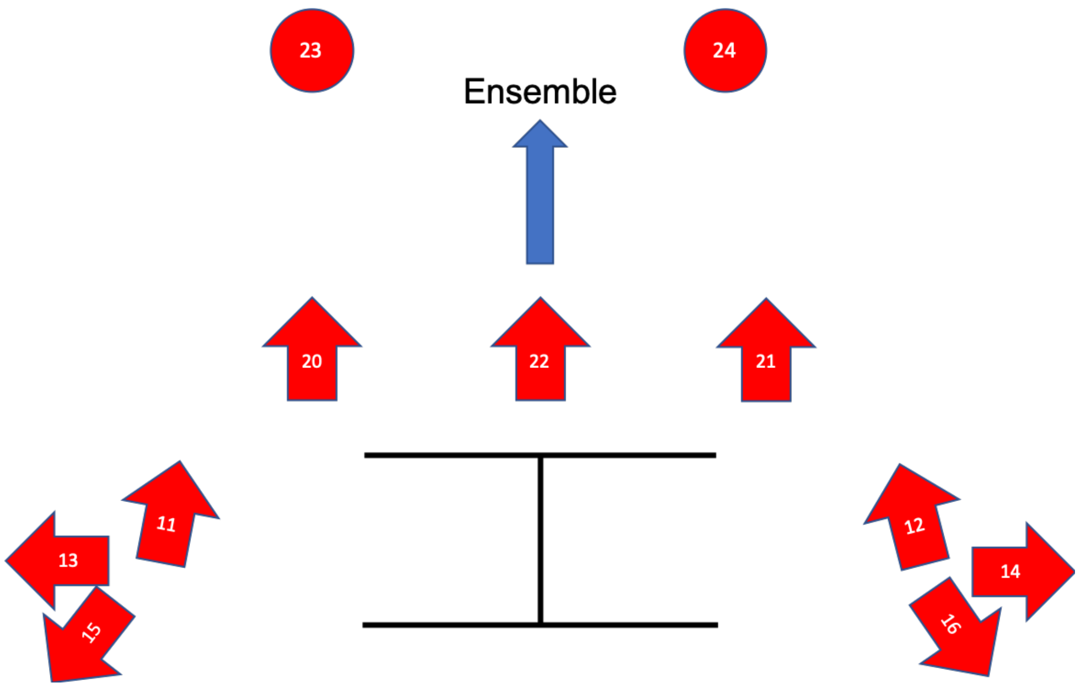
Q15 Wilcoxon Test Result

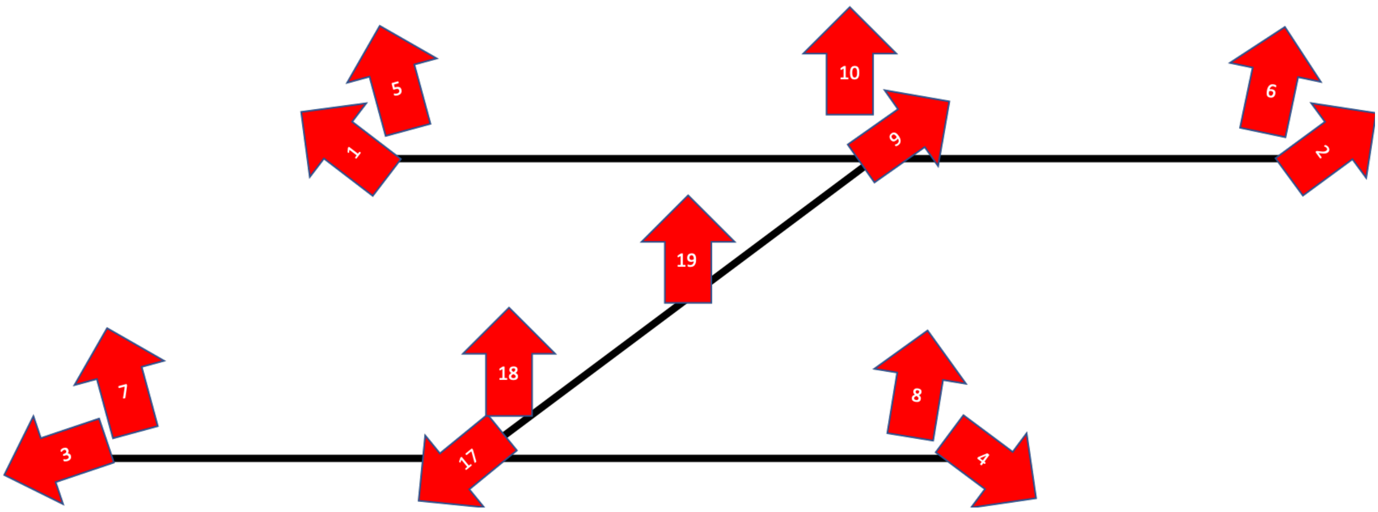
	Q15Games - Q15Film	Q15Music - Q15Film	Q15Soundscape - Q15Film	Q15Music - Q15Games	Q15Soundscape - Q15Games	Q15Soundscape - Q15Music
Z	-.352[b]	-.563[b]	-.431[c]	-.364[b]	-.959[c]	-1.294[c]
Asymp. Sig. (2-tailed)	.725	.574	.666	.716	.338	.196

Appendix 3 – Microphone Lists and Diagrams

Brass Microphone List/Diagram

Channel	Mic	Speaker Channel	Polar pattern
1	Sennheiser mkh8040	FL	Cardioid
2	Sennheiser mkh8040	FR	Cardioid
3	Sennheiser mkh8040	RL	Cardioid
4	Sennheiser mkh8040	RR	Cardioid
5	Sennheiser mkh8050	FLh	Hypercardioid
6	Sennheiser mkh8050	FRh	Hypercardioid
7	Sennheiser mkh8050	RLh	Hypercardioid
8	Sennheiser mkh8050	RRh	Hypercardioid
9	Neumann km184	FC	Cardioid
10	Neumann km184	FCh	Cardioid
11	DPA 4011	FLw	Cardioid
12	DPA 4011	FRw	Cardioid
13	Schoeps ccm8	SL	Figure-of-8
14	Schoeps ccm8	SR	Figure-of-8
15	Schoeps mk4	SLh	Cardioid
16	Schoeps mk4	SRh	Cardioid
17	Neumann km184	BC	Cardioid
18	Neumann km184	BCh	Cardioid
19	Rode nt5	VOG	Cardioid
20	AKG 414	FLf	Cardioid
21	AKG 414	FRf	Cardioid
22	N/A	FCf	
23	AKG 414	Spot 1	Hypercardioid





Strings Microphone List/ Diagram

Channel	Mic	Speaker Channel	Polar Pattern
1	DPA 4011A	FL	Cardioid
2	DPA 4011A	FR	Cardioid
3	DPA 4011A	RL	Cardioid
4	DPA 4011A	RR	Cardioid
5	DPA 4081C	FLh	SuperCardioid
6	DPA 4081C	FRh	SuperCardioid
7	DPA 4081C	RLh	SuperCardioid
8	DPA 4081C	RRh	SuperCardioid
9	DPA 4011A	FC	Cardioid
10	N/A	FCh	
11	DPA 4060A	FLw	Omni
12	DPA 4060A	FRw	Omni
13	DPA 4011A TL	SL	Cardioid
14	DPA 4011A TL	SR	Cardioid
15	DPA 4018C	SLh	SuperCardioid
16	DPA 4018C	SRh	SuperCardioid
17	N/A	BC	
18	N/A	BCh	
19	DPA 4018C	VOG	SuperCardioid
20	AKG 414	FLf	Cardioid
21	AKG 414	FRf	Cardioid
22	AKG 414	FCf	Cardioid

*Note – Microphone arrangement is the same as the Brass diagram above minus missing microphones

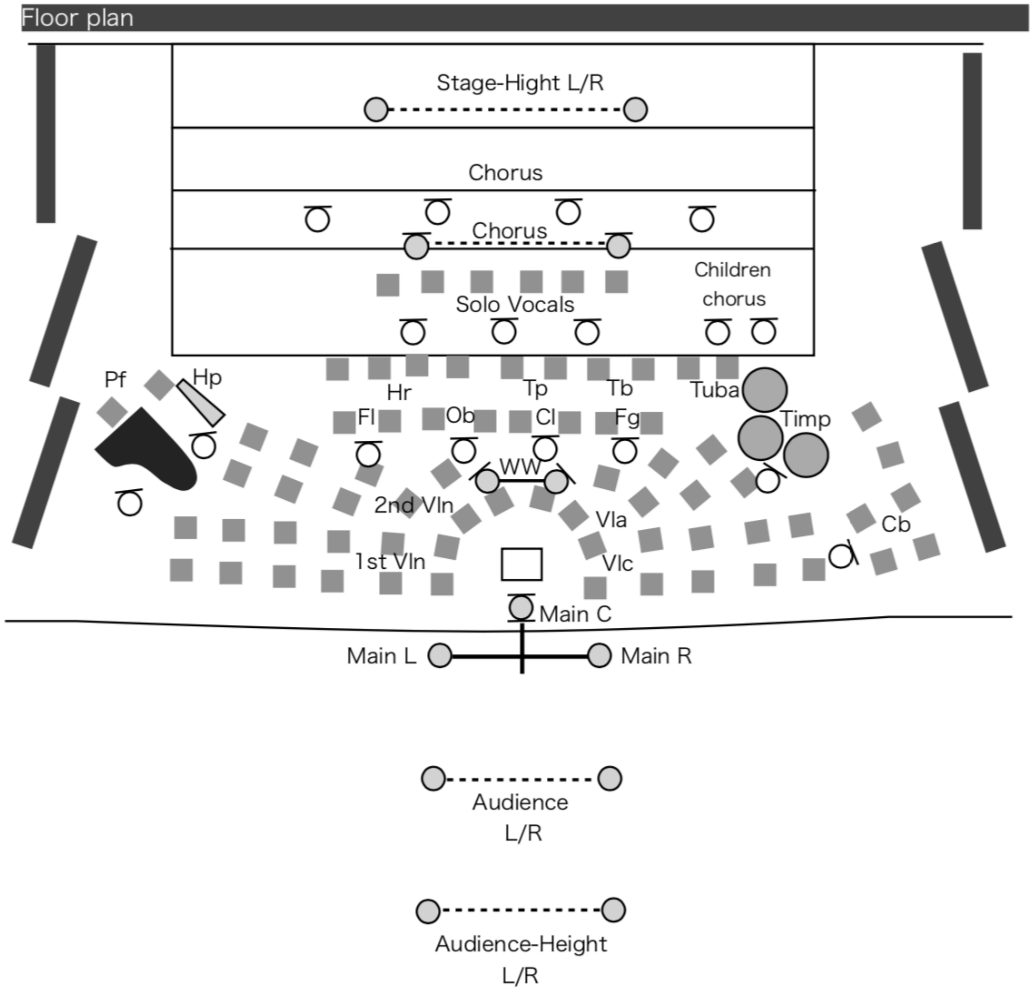
Choir Microphone List/ Diagram – Diagram Provided with Permission of Professor Toru Kamekawa, Tokyo University of the Arts

FL	Main L, Wood Winds L, Pf, Hp
FR	Main R, Wood Winds R
FC	Main C, SoloVocal 2, (SoloVocal 1*, SoloVocal 3*)
LFE1	n/a
BL	Audience L
BR	Audience R
FLc	SoloVocal 1*, Chorus spot 2
FRc	SoloVocal 3*, Chorus spot 3
BC	Audience L+R
LFE2	n/a
SiL	Chorus spot 1
SiR	Chorus spot 4
TpFL	Chorus (hanging) L
TpFR	Chorus (hanging) R
TpFC	Chorus (hanging) L+R
TpC	Stage-Height L+R
TpBL	Audience-Height L
TpBR	Audience-Height R
TpSiL	Stage-Height L
TpSiR	Stage-Height R
TpBC	Audience-Height L+R
BtFC	Clarinet, (Oboe*), ContraBass, (Timpani*)
BtFL	Flute, Oboe*
BtFR	Bassoon, Timpani*

(*using Pan pot. – This was done prior to receiving the individual project stems)

Kaido Tosei

Sogakudo 2015.11/28



Hanging microphones

	inst	mic	line
1	Main L	DPA4006	Main3pL-1
2	Main R	DPA4006	Main3pL-2
3	Main C	CCM68	Main3pC-1
4			
5	Audience L	DPA4006	AudFrontL
6	Audience R	DPA4006	AudFrontR
7	Chorus L	CMC64	Stage 4
8	Chorus R	CMC64	Stage 6
9	WW L	CMC64	Stage 2
10	WW R	CMC64	Stage 5
11	Stage-Height L	MKH8020	Stage 1
12	Stage-Height R	MKH8020	Stage 3
13	Audience-Height L	CMC62	AudRearL
14	Audience-Height R	CMC62	AudRearR

Stage microphones

	inst	mic	stand
1	Pf	CMC64	ST210
2	Hp	CMC64V	MF209
3	Fl	CMC64	ST210
4	Ob	CMC64	ST210
5	Cla	CMC64	ST210
6	Fg	CMC64	ST210
7	CB	CU41	ST210
8	Timp	CU41	ST210
9	Solo1	CMC64	Schoeps
10	Solo2	CMC64	Schoeps
11	Solo3	CMC64	Schoeps
12	Chorus 1	CMC64V	MF221
13	Chorus 2	CMC64V	MF221
14	Chorus 3	CMC64V	MF221
15	Chorus 4	CMC64V	MF221
16	ChildrenChorus L	CMC64V	MF209
17	ChildrenChorus R	CMC64V	MF209

Appendix 4 – Normality Test for Experiment 1

Tests of Normality Speaker Layer Balance Test

	ContentType	Shapiro-Wilk Statistic	df	Sig.
HeightValue	Brass	.860	7	.151
	Choir	.859	7	.148
	Holst	.981	7	.966
	Strings	.811	7	.052
FloorValue	Brass	.898	7	.317
	Choir	.729	7	.008
	Holst	.952	7	.747
	Strings	.926	7	.521

Appendix 5 – Experiment 2 Shapiro-Wilk Test Results

Experiment 2 – Shapiro-Wilk Test Results

	Attribute	Shapiro-Wilk Statistic	df	Sig.
Strings_Mono	LEV	.752	20	.000
	OTQ	.938	20	.218
	Pres	.834	20	.003
	QoE	.832	20	.003
Strings_2.0	LEV	.970	20	.762
	OTQ	.970	20	.753
	Pres	.944	20	.285
	QoE	.867	20	.011
Strings_5.1	LEV	.949	20	.346
	OTQ	.939	20	.226
	Pres	.940	20	.243
	QoE	.945	20	.292
Strings_9.1	LEV	.888	20	.025
	OTQ	.889	20	.026
	Pres	.752	20	.000
	QoE	.892	20	.030

Strings_22.2	LEV	.910	20	.063
	OTQ	.928	20	.142
	Pres	.897	20	.036
	QoE	.923	20	.113
Brass_Mono	LEV	.761	20	.000
	OTQ	.890	20	.027
	Pres	.820	20	.002
	QoE	.884	20	.021
Brass_2.0	LEV	.968	20	.709
	OTQ	.982	20	.953
	Pres	.944	20	.291
	QoE	.949	20	.351
Brass_5.1	LEV	.880	20	.017
	OTQ	.912	20	.069
	Pres	.923	20	.112
	QoE	.916	20	.082
Brass_9.1	LEV	.899	20	.039
	OTQ	.963	20	.601
	Pres	.904	20	.049
	QoE	.765	20	.000
Brass_22.2	LEV	.951	20	.376
	OTQ	.972	20	.793
	Pres	.893	20	.031
	QoE	.876	20	.015
Choir_Mono	LEV	.755	20	.000
	OTQ	.914	20	.076
	Pres	.831	20	.003
	QoE	.804	20	.001
Choir_2.0	LEV	.964	20	.628
	OTQ	.964	20	.632
	Pres	.938	20	.216
	QoE	.924	20	.118
Choir_5.1	LEV	.962	20	.587
	OTQ	.928	20	.142
	Pres	.913	20	.072
	QoE	.944	20	.280
Choir_9.1	LEV	.790	20	.001
	OTQ	.891	20	.028
	Pres	.869	20	.011
	QoE	.864	20	.009

Choir_22.2	LEV	.953	20	.412
	OTQ	.930	20	.153
	Pres	.945	20	.296
	QoE	.910	20	.064
Holst_Mono	LEV	.665	20	.000
	OTQ	.950	20	.369
	Pres	.852	20	.006
	QoE	.777	20	.000
Holst_2.0	LEV	.959	20	.518
	OTQ	.966	20	.668
	Pres	.921	20	.102
	QoE	.944	20	.288
Holst_5.1	LEV	.954	20	.439
	OTQ	.908	20	.057
	Pres	.925	20	.123
	QoE	.942	20	.256
Holst_9.1	LEV	.705	20	.000
	OTQ	.900	20	.041
	Pres	.766	20	.000
	QoE	.833	20	.003
Holst_22.2	LEV	.951	20	.390
	OTQ	.909	20	.060
	Pres	.913	20	.074
	QoE	.956	20	.459

Appendix 6 – Experiment 2 – Friedman/Pairwise Comparison Results

Strings for Listener Envelopment

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary

Total N	20
Test Statistic	60.203
Degree Of Freedom	4
Asymptotic Sig.(2-sided test)	.000

Pairwise Comparisons

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Strings_Mono-Strings_2.0	-1.200	.500	-2.400	.016	.164
Strings_Mono-Strings_22.2	-2.725	.500	-5.450	.000	.000
Strings_Mono-Strings_5.1	-2.800	.500	-5.600	.000	.000
Strings_Mono-Strings_9.1	-3.275	.500	-6.550	.000	.000
Strings_2.0-Strings_22.2	-1.525	.500	-3.050	.002	.023
Strings_2.0-Strings_5.1	-1.600	.500	-3.200	.001	.014
Strings_2.0-Strings_9.1	-2.075	.500	-4.150	.000	.000
Strings_22.2-Strings_5.1	.075	.500	.150	.881	1.000
Strings_22.2-Strings_9.1	.550	.500	1.100	.271	1.000
Strings_5.1-Strings_9.1	-.475	.500	-.950	.342	1.000

Strings For Overall Tonal Quality

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary

Total N	20
Test Statistic	57.914
Degree Of Freedom	4
Asymptotic Sig.(2-sided test)	.000

Pairwise Comparisons

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Strings_Mono-Strings_2.0	-1.450	.500	-2.900	.004	.037
Strings_Mono-Strings_22.2	-2.375	.500	-4.750	.000	.000
Strings_Mono-Strings_5.1	-2.675	.500	-5.350	.000	.000
Strings_Mono-Strings_9.1	-3.500	.500	-7.000	.000	.000
Strings_2.0-Strings_22.2	-.925	.500	-1.850	.064	.643
Strings_2.0-Strings_5.1	-1.225	.500	-2.450	.014	.143
Strings_2.0-Strings_9.1	-2.050	.500	-4.100	.000	.000
Strings_22.2-Strings_5.1	.300	.500	.600	.549	1.000
Strings_22.2-Strings_9.1	1.125	.500	2.250	.024	.244
Strings_5.1-Strings_9.1	-.825	.500	-1.650	.099	.989

Strings For Presence

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary

Total N	20
Test Statistic	53.985
Degree Of Freedom	4
Asymptotic Sig.(2-sided test)	.000

Pairwise Comparisons

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Strings_Mono-Strings_2.0	-1.550	.500	-3.100	.002	.019
Strings_Mono-Strings_22.2	-2.325	.500	-4.650	.000	.000
Strings_Mono-Strings_5.1	-2.775	.500	-5.550	.000	.000
Strings_Mono-Strings_9.1	-3.350	.500	-6.700	.000	.000
Strings_2.0-Strings_22.2	-.775	.500	-1.550	.121	1.000
Strings_2.0-Strings_5.1	-1.225	.500	-2.450	.014	.143
Strings_2.0-Strings_9.1	-1.800	.500	-3.600	.000	.003
Strings_22.2-Strings_5.1	.450	.500	.900	.368	1.000
Strings_22.2-Strings_9.1	1.025	.500	2.050	.040	.404
Strings_5.1-Strings_9.1	-.575	.500	-1.150	.250	1.000

Strings For Quality of Experience

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary

Total N	20
Test Statistic	63.446
Degree Of Freedom	4
Asymptotic Sig.(2-sided test)	.000

Pairwise Comparisons

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Strings_Mono-Strings_2.0	-1.250	.500	-2.500	.012	.124
Strings_Mono-Strings_22.2	-2.225	.500	-4.450	.000	.000
Strings_Mono-Strings_5.1	-3.075	.500	-6.150	.000	.000
Strings_Mono-Strings_9.1	-3.450	.500	-6.900	.000	.000
Strings_2.0-Strings_22.2	-.975	.500	-1.950	.051	.512
Strings_2.0-Strings_5.1	-1.825	.500	-3.650	.000	.003
Strings_2.0-Strings_9.1	-2.200	.500	-4.400	.000	.000
Strings_22.2-Strings_5.1	.850	.500	1.700	.089	.891
Strings_22.2-Strings_9.1	1.225	.500	2.450	.014	.143
Strings_5.1-Strings_9.1	-.375	.500	-.750	.453	1.000

Brass For Listener Envelopment

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary

Total N	20
Test Statistic	69.929
Degree Of Freedom	4
Asymptotic Sig.(2-sided test)	.000

Pairwise Comparisons

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Brass_Mono-Brass_2.0	-1.000	.500	-2.000	.046	.455
Brass_Mono-Brass_22.2	-2.350	.500	-4.700	.000	.000
Brass_Mono-Brass_9.1	-3.225	.500	-6.450	.000	.000
Brass_Mono-Brass_5.1	-3.425	.500	-6.850	.000	.000
Brass_2.0-Brass_22.2	-1.350	.500	-2.700	.007	.069
Brass_2.0-Brass_9.1	-2.225	.500	-4.450	.000	.000
Brass_2.0-Brass_5.1	-2.425	.500	-4.850	.000	.000
Brass_22.2-Brass_9.1	.875	.500	1.750	.080	.801
Brass_22.2-Brass_5.1	1.075	.500	2.150	.032	.316
Brass_9.1-Brass_5.1	.200	.500	.400	.689	1.000

Brass For Overall Tonal Quality

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary

Total N	20
Test Statistic	65.542
Degree Of Freedom	4
Asymptotic Sig.(2-sided test)	.000

Pairwise Comparisons

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Brass_Mono-Brass_2.0	-.950	.500	-1.900	.057	.574
Brass_Mono-Brass_22.2	-2.475	.500	-4.950	.000	.000
Brass_Mono-Brass_9.1	-2.925	.500	-5.850	.000	.000
Brass_Mono-Brass_5.1	-3.400	.500	-6.800	.000	.000
Brass_2.0-Brass_22.2	-1.525	.500	-3.050	.002	.023
Brass_2.0-Brass_9.1	-1.975	.500	-3.950	.000	.001
Brass_2.0-Brass_5.1	-2.450	.500	-4.900	.000	.000
Brass_22.2-Brass_9.1	.450	.500	.900	.368	1.000
Brass_22.2-Brass_5.1	.925	.500	1.850	.064	.643
Brass_9.1-Brass_5.1	.475	.500	.950	.342	1.000

Brass For Presence

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary

Total N	20
Test Statistic	48.687
Degree Of Freedom	4
Asymptotic Sig.(2-sided test)	.000

Pairwise Comparisons

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Brass_Mono-Brass_2.0	-1.650	.500	-3.300	.001	.010
Brass_Mono-Brass_22.2	-2.600	.500	-5.200	.000	.000
Brass_Mono-Brass_9.1	-2.800	.500	-5.600	.000	.000
Brass_Mono-Brass_5.1	-2.950	.500	-5.900	.000	.000
Brass_2.0-Brass_22.2	-.950	.500	-1.900	.057	.574
Brass_2.0-Brass_9.1	-1.150	.500	-2.300	.021	.214
Brass_2.0-Brass_5.1	-1.300	.500	-2.600	.009	.093
Brass_22.2-Brass_9.1	.200	.500	.400	.689	1.000
Brass_22.2-Brass_5.1	.350	.500	.700	.484	1.000
Brass_9.1-Brass_5.1	.150	.500	.300	.764	1.000

Brass For Quality of Experience

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary

Total N	20
Test Statistic	58.670
Degree Of Freedom	4
Asymptotic Sig.(2-sided test)	.000

Pairwise Comparisons

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Brass_Mono-Brass_2.0	-1.200	.500	-2.400	.016	.164
Brass_Mono-Brass_22.2	-2.825	.500	-5.650	.000	.000
Brass_Mono-Brass_9.1	-2.925	.500	-5.850	.000	.000
Brass_Mono-Brass_5.1	-3.050	.500	-6.100	.000	.000
Brass_2.0-Brass_22.2	-1.625	.500	-3.250	.001	.012
Brass_2.0-Brass_9.1	-1.725	.500	-3.450	.001	.006
Brass_2.0-Brass_5.1	-1.850	.500	-3.700	.000	.002
Brass_22.2-Brass_9.1	.100	.500	.200	.841	1.000
Brass_22.2-Brass_5.1	.225	.500	.450	.653	1.000
Brass_9.1-Brass_5.1	.125	.500	.250	.803	1.000

Choir For Listener Envelopment

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary

Total N	20
Test Statistic	74.937
Degree Of Freedom	4
Asymptotic Sig.(2-sided test)	.000

Pairwise Comparisons

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Choir_Mono-Choir_2.0	-.950	.500	-1.900	.057	.574
Choir_Mono-Choir_5.1	-2.150	.500	-4.300	.000	.000
Choir_Mono-Choir_22.2	-2.725	.500	-5.450	.000	.000
Choir_Mono-Choir_9.1	-3.925	.500	-7.850	.000	.000
Choir_2.0-Choir_5.1	-1.200	.500	-2.400	.016	.164
Choir_2.0-Choir_22.2	-1.775	.500	-3.550	.000	.004
Choir_2.0-Choir_9.1	-2.975	.500	-5.950	.000	.000
Choir_5.1-Choir_22.2	-.575	.500	-1.150	.250	1.000
Choir_5.1-Choir_9.1	-1.775	.500	-3.550	.000	.004
Choir_22.2-Choir_9.1	1.200	.500	2.400	.016	.164

Choir For Overall Tonal Quality

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary

Total N	20
Test Statistic	58.412
Degree Of Freedom	4
Asymptotic Sig.(2-sided test)	.000

Pairwise Comparisons

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Choir_Mono-Choir_2.0	-1.350	.500	-2.700	.007	.069
Choir_Mono-Choir_5.1	-2.500	.500	-5.000	.000	.000
Choir_Mono-Choir_22.2	-2.700	.500	-5.400	.000	.000
Choir_Mono-Choir_9.1	-3.450	.500	-6.900	.000	.000
Choir_2.0-Choir_5.1	-1.150	.500	-2.300	.021	.214
Choir_2.0-Choir_22.2	-1.350	.500	-2.700	.007	.069
Choir_2.0-Choir_9.1	-2.100	.500	-4.200	.000	.000
Choir_5.1-Choir_22.2	-.200	.500	-.400	.689	1.000
Choir_5.1-Choir_9.1	-.950	.500	-1.900	.057	.574
Choir_22.2-Choir_9.1	.750	.500	1.500	.134	1.000

Choir For Presence

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary

Total N	20
Test Statistic	60.902
Degree Of Freedom	4
Asymptotic Sig.(2-sided test)	.000

Pairwise Comparisons

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Choir_Mono-Choir_2.0	-1.250	.500	-2.500	.012	.124
Choir_Mono-Choir_5.1	-2.625	.500	-5.250	.000	.000
Choir_Mono-Choir_22.2	-2.625	.500	-5.250	.000	.000
Choir_Mono-Choir_9.1	-3.500	.500	-7.000	.000	.000
Choir_2.0-Choir_5.1	-1.375	.500	-2.750	.006	.060
Choir_2.0-Choir_22.2	-1.375	.500	-2.750	.006	.060
Choir_2.0-Choir_9.1	-2.250	.500	-4.500	.000	.000
Choir_5.1-Choir_22.2	.000	.500	.000	1.000	1.000
Choir_5.1-Choir_9.1	-.875	.500	-1.750	.080	.801
Choir_22.2-Choir_9.1	.875	.500	1.750	.080	.801

Choir For Quality of Experience

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary

Total N	20
Test Statistic	53.005
Degree Of Freedom	4
Asymptotic Sig.(2-sided test)	.000

Pairwise Comparisons

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Choir_Mono-Choir_2.0	-.550	.500	-1.100	.271	1.000
Choir_Mono-Choir_5.1	-2.050	.500	-4.100	.000	.000
Choir_Mono-Choir_22.2	-2.100	.500	-4.200	.000	.000
Choir_Mono-Choir_9.1	-3.175	.500	-6.350	.000	.000
Choir_2.0-Choir_5.1	-1.500	.500	-3.000	.003	.027
Choir_2.0-Choir_22.2	-1.550	.500	-3.100	.002	.019
Choir_2.0-Choir_9.1	-2.625	.500	-5.250	.000	.000
Choir_5.1-Choir_22.2	-.050	.500	-.100	.920	1.000
Choir_5.1-Choir_9.1	-1.125	.500	-2.250	.024	.244
Choir_22.2-Choir_9.1	1.075	.500	2.150	.032	.316

Holst For Listener Envelopment

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary

Total N	20
Test Statistic	70.123
Degree Of Freedom	4
Asymptotic Sig.(2-sided test)	.000

Pairwise Comparisons

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Holst_Mono-Holst_2.0	-1.050	.500	-2.100	.036	.357
Holst_Mono-Holst_5.1	-2.525	.500	-5.050	.000	.000
Holst_Mono-Holst_22.2	-2.750	.500	-5.500	.000	.000
Holst_Mono-Holst_9.1	-3.675	.500	-7.350	.000	.000
Holst_2.0-Holst_5.1	-1.475	.500	-2.950	.003	.032
Holst_2.0-Holst_22.2	-1.700	.500	-3.400	.001	.007
Holst_2.0-Holst_9.1	-2.625	.500	-5.250	.000	.000
Holst_5.1-Holst_22.2	-.225	.500	-.450	.653	1.000
Holst_5.1-Holst_9.1	-1.150	.500	-2.300	.021	.214
Holst_22.2-Holst_9.1	.925	.500	1.850	.064	.643

Holst for Overall Tonal Quality

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary

Total N	20
Test Statistic	66.450
Degree Of Freedom	4
Asymptotic Sig.(2-sided test)	.000

Pairwise Comparisons

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Holst_Mono-Holst_2.0	-1.050	.500	-2.100	.036	.357
Holst_Mono-Holst_22.2	-2.750	.500	-5.500	.000	.000
Holst_Mono-Holst_5.1	-2.825	.500	-5.650	.000	.000
Holst_Mono-Holst_9.1	-3.375	.500	-6.750	.000	.000
Holst_2.0-Holst_22.2	-1.700	.500	-3.400	.001	.007
Holst_2.0-Holst_5.1	-1.775	.500	-3.550	.000	.004
Holst_2.0-Holst_9.1	-2.325	.500	-4.650	.000	.000
Holst_22.2-Holst_5.1	.075	.500	.150	.881	1.000
Holst_22.2-Holst_9.1	.625	.500	1.250	.211	1.000
Holst_5.1-Holst_9.1	-.550	.500	-1.100	.271	1.000

Holst For Presence

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary

Total N	20
Test Statistic	54.338
Degree Of Freedom	4
Asymptotic Sig.(2-sided test)	.000

Pairwise Comparisons

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Holst_Mono-Holst_2.0	-1.450	.500	-2.900	.004	.037
Holst_Mono-Holst_22.2	-2.525	.500	-5.050	.000	.000
Holst_Mono-Holst_5.1	-2.775	.500	-5.550	.000	.000
Holst_Mono-Holst_9.1	-3.250	.500	-6.500	.000	.000
Holst_2.0-Holst_22.2	-1.075	.500	-2.150	.032	.316
Holst_2.0-Holst_5.1	-1.325	.500	-2.650	.008	.080
Holst_2.0-Holst_9.1	-1.800	.500	-3.600	.000	.003
Holst_22.2-Holst_5.1	.250	.500	.500	.617	1.000
Holst_22.2-Holst_9.1	.725	.500	1.450	.147	1.000
Holst_5.1-Holst_9.1	-.475	.500	-.950	.342	1.000

Holst For Quality of Experience

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary

Total N	20
Test Statistic	69.664
Degree Of Freedom	4
Asymptotic Sig.(2-sided test)	.000

Pairwise Comparisons

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Holst_Mono-Holst_2.0	-1.050	.500	-2.100	.036	.357
Holst_Mono-Holst_22.2	-2.575	.500	-5.150	.000	.000
Holst_Mono-Holst_5.1	-2.625	.500	-5.250	.000	.000
Holst_Mono-Holst_9.1	-3.750	.500	-7.500	.000	.000
Holst_2.0-Holst_22.2	-1.525	.500	-3.050	.002	.023
Holst_2.0-Holst_5.1	-1.575	.500	-3.150	.002	.016
Holst_2.0-Holst_9.1	-2.700	.500	-5.400	.000	.000
Holst_22.2-Holst_5.1	.050	.500	.100	.920	1.000
Holst_22.2-Holst_9.1	1.175	.500	2.350	.019	.188
Holst_5.1-Holst_9.1	-1.125	.500	-2.250	.024	.244

Appendix 7 – Experiment 2 Further Analysis

Mono For Listener Envelopment – Friedman/Pairwise Comparison

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary

Total N	20
Test Statistic	3.096[a]
Degree Of Freedom	3
Asymptotic Sig.(2-sided test)	.377

Mono For Overall Tonal Quality – Friedman/Pairwise Comparison

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary

Total N	20
Test Statistic	9.204
Degree Of Freedom	3
Asymptotic Sig.(2-sided test)	.027

Pairwise Comparisons

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Choir_Mono-Brass_Mono	.025	.408	.061	.951	1.000
Choir_Mono-Holst_Mono	-.400	.408	-.980	.327	1.000
Choir_Mono-Strings_Mono	.975	.408	2.388	.017	.102
Brass_Mono-Holst_Mono	-.375	.408	-.919	.358	1.000
Brass_Mono-Strings_Mono	.950	.408	2.327	.020	.120
Holst_Mono-Strings_Mono	.575	.408	1.408	.159	.954

Mono For Presence – Friedman/Pairwise Comparison

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary

Total N	20
Test Statistic	3.947[a]
Degree Of Freedom	3
Asymptotic Sig.(2-sided test)	.267

Mono For Quality of Experience – Friedman/Pairwise Comparison

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary

Total N	20
Test Statistic	4.196[a]
Degree Of Freedom	3
Asymptotic Sig.(2-sided test)	.241

2.0 For Listener Envelopment – Friedman/Pairwise Comparison

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary

Total N	20
Test Statistic	5.921[a]
Degree Of Freedom	3
Asymptotic Sig.(2-sided test)	.116

2.0 For Overall Tonal Quality – Friedman/Pairwise Comparison

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary

Total N	20
Test Statistic	10.846
Degree Of Freedom	3
Asymptotic Sig.(2-sided test)	.013

Pairwise Comparisons

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Brass_2.0-Choir_2.0	-.325	.408	-.796	.426	1.000
Brass_2.0-Holst_2.0	-.600	.408	-1.470	.142	.850
Brass_2.0-Strings_2.0	1.275	.408	3.123	.002	.011
Choir_2.0-Holst_2.0	-.275	.408	-.674	.501	1.000
Choir_2.0-Strings_2.0	.950	.408	2.327	.020	.120
Holst_2.0-Strings_2.0	.675	.408	1.653	.098	.589

2.0 For Presence – Friedman/Pairwise Comparison

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary

Total N	20
Test Statistic	12.437
Degree Of Freedom	3
Asymptotic Sig.(2-sided test)	.006

Pairwise Comparisons

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Choir_2.0-Holst_2.0	-.825	.408	-2.021	.043	.260
Choir_2.0-Strings_2.0	.975	.408	2.388	.017	.102
Choir_2.0-Brass_2.0	1.400	.408	3.429	.001	.004
Holst_2.0-Strings_2.0	.150	.408	.367	.713	1.000
Holst_2.0-Brass_2.0	.575	.408	1.408	.159	.954
Strings_2.0-Brass_2.0	-.425	.408	-1.041	.298	1.000

2.0 For Quality of Experience – Friedman/Pairwise Comparison

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary

Total N	20
Test Statistic	16.091
Degree Of Freedom	3
Asymptotic Sig.(2-sided test)	.001

Pairwise Comparisons

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Holst_2.0-Choir_2.0	.250	.408	.612	.540	1.000
Holst_2.0-Brass_2.0	.675	.408	1.653	.098	.589
Holst_2.0-Strings_2.0	1.475	.408	3.613	.000	.002
Choir_2.0-Brass_2.0	.425	.408	1.041	.298	1.000
Choir_2.0-Strings_2.0	1.225	.408	3.001	.003	.016
Brass_2.0-Strings_2.0	.800	.408	1.960	.050	.300

5.1 For Listener Envelopment – Friedman/Pairwise Comparison

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary

Total N	20
Test Statistic	17.705
Degree Of Freedom	3
Asymptotic Sig.(2-sided test)	.001

Pairwise Comparisons

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Choir_5.1-Strings_5.1	1.175	.408	2.878	.004	.024
Choir_5.1-Holst_5.1	-1.250	.408	-3.062	.002	.013
Choir_5.1-Brass_5.1	1.575	.408	3.858	.000	.001
Strings_5.1-Holst_5.1	-.075	.408	-.184	.854	1.000
Strings_5.1-Brass_5.1	-.400	.408	-.980	.327	1.000
Holst_5.1-Brass_5.1	.325	.408	.796	.426	1.000

5.1 For Overall Tonal Quality – Friedman/Pairwise Comparison

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary

Total N	20
Test Statistic	7.023[a]
Degree Of Freedom	3
Asymptotic Sig.(2-sided test)	.071

5.1 For Presence – Friedman/Pairwise Comparison

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary

Total N	20
Test Statistic	5.777[a]
Degree Of Freedom	3
Asymptotic Sig.(2-sided test)	.123

5.1 For Quality of Experience – Friedman/Pairwise Comparison

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary

Total N	20
Test Statistic	8.423
Degree Of Freedom	3
Asymptotic Sig.(2-sided test)	.038

Pairwise Comparisons

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Choir_5.1-Holst_5.1	-.275	.408	-.674	.501	1.000
Choir_5.1-Brass_5.1	.900	.408	2.205	.027	.165
Choir_5.1-Strings_5.1	.925	.408	2.266	.023	.141
Holst_5.1-Brass_5.1	.625	.408	1.531	.126	.755
Holst_5.1-Strings_5.1	.650	.408	1.592	.111	.668
Brass_5.1-Strings_5.1	.025	.408	.061	.951	1.000

9.1 For Listener Envelopment – Friedman/Pairwise Comparison

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary

Total N	20
Test Statistic	16.445
Degree Of Freedom	3
Asymptotic Sig.(2-sided test)	.001

Pairwise Comparisons

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Brass_9.1-Strings_9.1	.375	.408	.919	.358	1.000
Brass_9.1-Choir_9.1	-1.175	.408	-2.878	.004	.024
Brass_9.1-Holst_9.1	-1.250	.408	-3.062	.002	.013
Strings_9.1-Choir_9.1	-.800	.408	-1.960	.050	.300
Strings_9.1-Holst_9.1	-.875	.408	-2.143	.032	.193
Choir_9.1-Holst_9.1	-.075	.408	-.184	.854	1.000

9.1 For Overall Tonal Quality – Friedman/Pairwise Comparison

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary

Total N	20
Test Statistic	7.423[a]
Degree Of Freedom	3
Asymptotic Sig.(2-sided test)	.060

9.1 For Presence – Friedman/Pairwise Comparison

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary

Total N	20
Test Statistic	6.120[a]
Degree Of Freedom	3
Asymptotic Sig.(2-sided test)	.106

9.1 For Quality of Experience – Friedman/Pairwise Comparison

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary

Total N	20
Test Statistic	9.057
Degree Of Freedom	3
Asymptotic Sig.(2-sided test)	.029

Pairwise Comparisons

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Brass_9.1-Choir_9.1	-.300	.408	-.735	.462	1.000
Brass_9.1-Strings_9.1	.675	.408	1.653	.098	.589
Brass_9.1-Holst_9.1	-1.025	.408	-2.511	.012	.072
Choir_9.1-Strings_9.1	.375	.408	.919	.358	1.000
Choir_9.1-Holst_9.1	-.725	.408	-1.776	.076	.455
Strings_9.1-Holst_9.1	-.350	.408	-.857	.391	1.000

22.2 For Listener Envelopment – Friedman/Pairwise Comparison

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary

Total N	20
Test Statistic	9.808
Degree Of Freedom	3
Asymptotic Sig.(2-sided test)	.020

Pairwise Comparisons

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Brass_22.2-Choir_22.2	-.175	.408	-.429	.668	1.000
Brass_22.2-Holst_22.2	-.900	.408	-2.205	.027	.165
Brass_22.2-Strings_22.2	1.025	.408	2.511	.012	.072
Choir_22.2-Holst_22.2	-.725	.408	-1.776	.076	.455
Choir_22.2-Strings_22.2	.850	.408	2.082	.037	.224
Holst_22.2-Strings_22.2	.125	.408	.306	.759	1.000

22.2 For Overall Tonal Quality – Friedman/Pairwise Comparison

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary

Total N	20
Test Statistic	9.628
Degree Of Freedom	3
Asymptotic Sig.(2-sided test)	.022

Pairwise Comparisons

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Brass_22.2-Strings_22.2	.250	.408	.612	.540	1.000
Brass_22.2-Choir_22.2	-.475	.408	-1.164	.245	1.000
Brass_22.2-Holst_22.2	-1.175	.408	-2.878	.004	.024
Strings_22.2-Choir_22.2	-.225	.408	-.551	.582	1.000
Strings_22.2-Holst_22.2	-.925	.408	-2.266	.023	.141
Choir_22.2-Holst_22.2	-.700	.408	-1.715	.086	.518

22.2 For Presence – Friedman/Pairwise Comparison

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary

Total N	20
Test Statistic	.582[a]
Degree Of Freedom	3
Asymptotic Sig.(2-sided test)	.901

22.2 For Quality of Experience – Friedman/Pairwise Comparison

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary

Total N	20
Test Statistic	6.820[a]
Degree Of Freedom	3
Asymptotic Sig.(2-sided test)	.078

Appendix 8 – Experiment 2 Main Effect Results

Main Effect Content For Listener Envelopment

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary

Total N	100
Test Statistic	6.647[a]
Degree Of Freedom	3
Asymptotic Sig.(2-sided test)	.084

Main Effect Content For Overall Tonal Quality

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary

Total N	100
Test Statistic	9.006
Degree Of Freedom	3
Asymptotic Sig.(2-sided test)	.029

Pairwise Comparisons

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Brass-Choir	-.080	.183	-.438	.661	1.000
Brass-Strings	.330	.183	1.807	.071	.424
Brass-Holst	-.470	.183	-2.574	.010	.060
Choir-Strings	.250	.183	1.369	.171	1.000
Choir-Holst	-.390	.183	-2.136	.033	.196
Strings-Holst	-.140	.183	-.767	.443	1.000

Main Effect Content For Presence

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary

Total N	100
Test Statistic	6.715[a]
Degree Of Freedom	3
Asymptotic Sig.(2-sided test)	.082

Main Effect Content For Quality of Experience

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary

Total N	100
Test Statistic	16.952
Degree Of Freedom	3
Asymptotic Sig.(2-sided test)	.001

Pairwise Comparisons

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Choir-Holst	-.180	.183	-.986	.324	1.000
Choir-Brass	.485	.183	2.656	.008	.047
Choir-Strings	.635	.183	3.478	.001	.003
Holst-Brass	.305	.183	1.671	.095	.569
Holst-Strings	.455	.183	2.492	.013	.076
Brass-Strings	.150	.183	.822	.411	1.000

Main Effect Format For Listener Envelopment

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary

Total N	80
Test Statistic	264.405
Degree Of Freedom	4
Asymptotic Sig.(2-sided test)	.000

Pairwise Comparisons

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Mono-M_2.0	-1.050	.250	-4.200	.000	.000
Mono-M_22.2	-2.637	.250	-10.550	.000	.000
Mono-M_5.1	-2.725	.250	-10.900	.000	.000
Mono-M_9.1	-3.525	.250	-14.100	.000	.000
M_2.0-M_22.2	-1.587	.250	-6.350	.000	.000
M_2.0-M_5.1	-1.675	.250	-6.700	.000	.000
M_2.0-M_9.1	-2.475	.250	-9.900	.000	.000
M_22.2-M_5.1	.088	.250	.350	.726	1.000
M_22.2-M_9.1	.888	.250	3.550	.000	.004
M_5.1-M_9.1	-.800	.250	-3.200	.001	.014

Main Effect Format For Overall Tonal Quality

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary

Total N	80
Test Statistic	240.741
Degree Of Freedom	4
Asymptotic Sig.(2-sided test)	.000

Pairwise Comparisons

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Mono-M_2.0	-1.200	.250	-4.800	.000	.000
Mono-M_22.2	-2.575	.250	-10.300	.000	.000
Mono-M_5.1	-2.850	.250	-11.400	.000	.000
Mono-M_9.1	-3.312	.250	-13.250	.000	.000
M_2.0-M_22.2	-1.375	.250	-5.500	.000	.000
M_2.0-M_5.1	-1.650	.250	-6.600	.000	.000
M_2.0-M_9.1	-2.112	.250	-8.450	.000	.000
M_22.2-M_5.1	.275	.250	1.100	.271	1.000
M_22.2-M_9.1	.738	.250	2.950	.003	.032
M_5.1-M_9.1	-.462	.250	-1.850	.064	.643

Main Effect Format For Presence

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary

Total N	80
Test Statistic	214.187
Degree Of Freedom	4
Asymptotic Sig.(2-sided test)	.000

Pairwise Comparisons

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Mono-M_2.0	-1.475	.250	-5.900	.000	.000
Mono-M_22.2	-2.519	.250	-10.075	.000	.000
Mono-M_5.1	-2.781	.250	-11.125	.000	.000
Mono-M_9.1	-3.225	.250	-12.900	.000	.000
M_2.0-M_22.2	-1.044	.250	-4.175	.000	.000
M_2.0-M_5.1	-1.306	.250	-5.225	.000	.000
M_2.0-M_9.1	-1.750	.250	-7.000	.000	.000
M_22.2-M_5.1	.263	.250	1.050	.294	1.000
M_22.2-M_9.1	.706	.250	2.825	.005	.047
M_5.1-M_9.1	-.444	.250	-1.775	.076	.759

Main Effect Format For Quality of Experience

Related-Samples Friedman's Two-Way Analysis of Variance by Ranks Summary

Total N	80
Test Statistic	236.541
Degree Of Freedom	4
Asymptotic Sig.(2-sided test)	.000

Pairwise Comparisons

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Mono-M_2.0	-1.012	.250	-4.050	.000	.001
Mono-M_22.2	-2.431	.250	-9.725	.000	.000
Mono-M_5.1	-2.700	.250	-10.800	.000	.000
Mono-M_9.1	-3.325	.250	-13.300	.000	.000
M_2.0-M_22.2	-1.419	.250	-5.675	.000	.000
M_2.0-M_5.1	-1.687	.250	-6.750	.000	.000
M_2.0-M_9.1	-2.312	.250	-9.250	.000	.000
M_22.2-M_5.1	.269	.250	1.075	.282	1.000
M_22.2-M_9.1	.894	.250	3.575	.000	.004
M_5.1-M_9.1	-.625	.250	-2.500	.012	.124