Personality and cognitive factors in the assessment of multimodal stimuli in immersive virtual environments

Ph.D. Thesis

Acoustics Research Centre, University of Salford
Greater Manchester

Author

John William Bailey

Supervisor

Dr. Bruno M. Fazenda

Co-supervisor

Prof. Trevor J. Cox

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

September 2019

Declaration of Authorship

I, John William Bailey, declare that this thesis titled, 'Personality and cognitive factors in the assessment of multimodal stimuli in immersive virtual environments' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: _____

Date: _____

"Le bon sens est la chose du monde la mieux partagée; car chacun pense en être si bien pourvu, que ceux même qui sont les plus difficiles à contenter en toute autre chose n'ont point coutume d'en désirer plus qu'ils en ont."

"Good sense is, of all things among men, the most equally distributed; for every one thinks himself so abundantly provided with it, that those even who are the most difficult to satisfy in everything else, do not usually desire a larger measure of this quality than they already possess."

René Descartes

Abstract

Literature in the study of human response to immersive virtual reality systems often deals with the phenomenon of *presence*. It can be shown that audio and imagery with spatial information can interact to affect presence in users of immersive virtual reality. It has also been shown that there is variation between individuals in the experience of presence in VR. The relationship between these effects has hitherto not been fully explored. This thesis aims to identify and evaluate the relationships between spatial audio rendering and spatial relationships between audio and visual objects and cognitive and personality differences which account for variation in the experience of presence in VR with spatial audio. This thesis compares measures of audiovisual quality of experience with an existing model of presence in a factor-analytical paradigm. Scores on these dimensions were compared between environments which are similar or dissimilar to pre-exposure conditions and compared between when participants believed they were listening to real-world or headphone rendered audio events. Differences between audiovisual treatments, including audio rendering methods and audiovisual spatial relationships, were compared with differences attributed to cognitive and personality factors identified as significant predictors using hierarchical modelling. It was found that audiovisual quality of experience relates to subscales of presence by being independent of reported visual realism and involvement, but combines linearly with these factors to contribute to 'spatial presence', a dimension of overall presence which is identified as the largest component in the construct. It was also found that, although manipulation of the spatial information content of audiovisual stimuli was a predictor of audiovisual quality of experience, this effect is overshadowed by inter-participant variation. Interactive effects between extraversion, empathy, ease of resolving visual detail, and systematisation and are better predictors of quality of experience and spatial presence than the changes to spatial information content investigated in this work. Anchoring biases are also identified which suggest that novel environments are rated higher on audiovisual quality than those geometrically similar to the pre-exposure environment. These findings constitute support for a novel framework for assessing propensity for presence in terms of an information-processing model.

Acknowledgements

I would like to thank my supervisors, Dr. Bruno Fazenda and Professor Trevor Cox for their unwavering and continued support throughout the production of this thesis. Particularly, I would like to thank Dr. Fazenda for giving me the opportunity to study here at the University of Salford and for believing in me from the beginning. I would also like to thank my parents and wider family who are a never ending font of support and goodwill, and seem to be O.K. with me repeatedly going back to school and not getting a proper job. I would also like to thank all the people who passed through G10, G11 and the 'fish bowl'. Your company, conversation, problems and solutions made my time here a fun and rewarding place to learn and I really hope, although some of us are widely flung across the world, that we will continue to keep in touch long into the future. If you've ever asked me a difficult question, listened to me lay out an intractable problem or endured me agonizing over a graph, you will have helped me understand something better and for that I am in your debt. You know who you are, and the debt is redeemable at your convenience. I am also indebted to Professor Andrew Jackson for his valuable insight into statistics and experimental design, and the time he was willing to spend sharing that insight. Finally, I would like to express my sheer inability to adequately acknowledge the immense contribution of Dr. Claire Gwinnett. She is the centre of my world and without her I would not have done anything. Every step towards this doctorate was taken with her carrying, pushing, pulling and shoving me to be the person she thought I could be. Thank you for everything. I love you.

Contents

List of Figures

List of Tables

Author Publications

First Author

- Bailey, J. W., Fazenda, B. M., The effect of reverberation and audio spatialization on egocentric distance estimation of objects in stereoscopic virtual reality. The Journal of the Acoustical Society of America 141, 3510, 2017
- Bailey, J. W., Fazenda, B. M. , The Effect of Visual Cues and Binaural Rendering Method on Plausibility in Virtual Environments, Audio Engineering Society Convention 144, May 2018, Milan, Italy
- Bailey, J. W., Fazenda, B. M., Personality and Listening Sensitivity Correlates of the Subjective Response to Real and Simulated Sound Sources in Virtual Reality, 2018 AES International Conference on Spatial Reproduction - Aesthetics and Science, July 2018, Tokyo, Japan

PART I. INTRODUCTION, LITERATURE REVIEW AND MATERIALS

1

Introduction to Thesis

Studies of multimodal immersive virtual environments (VEs) are commonly framed as a study into the phenomenon of *presence*, the experience of being spatially located and present within the virtual world. Studies of presence usually pertain to qualities of stimuli, where others might attempt to account for intrinsic factors which might predispose individuals to the experience of presence.

The work presented in this thesis which constitutes a contribution to knowledge covers three main areas. The first is the support of the assertion that responses to multimodal stimuli in VR are different when pre-exposure geometry is dissimilar from the VE and where pre-exposure geometry is similar. This difference manifests in lower ratings of audiovisual quality metrics in similar geometry conditions. This is likely due to a perceptual anchoring to the expectation of acoustic response introduced by matching pre- and during exposure imagery.

Secondly, this thesis reports on the relationship between the subjective perception of audiovisual stimuli and reported presence. Perceived quality of spatial audio rendering and spatial relationships between audio and visual stimuli are shown to contribute to the experience of spatial presence.

Thirdly, this thesis attempts to account for the variation observed in presence and responses of audiovisual spatial quality observed in the data that are not explained by the manipulation of independent variables related to the stimuli presented, and to allow comparisons between effects observed within treatment manipulations and those observed when characterising variation due to intrinsic properties of the subject as a receiver of information. Using information processing theory interpretations of the constructs identified and used in this thesis, results that account for differences in the data on an inter-participant level can be understood as a set of interactions

between organisation, sensitivity (detection threshold) and specificity (error rejection) to information within multimodal stimuli that elicit the internal representation of a somatic (bodily) state that is recognised as spatial presence. Presence is selected as a response parameter of interest due to the body of literature that exists around this construct, which is reviewed in section 2.3.1. Presence, as defined below in section 1.1, refers to the sense that one is physically present in an environment and its elicitation is assumed to be the goal of a successful immersive virtual reality system.

It can be argued that the study of aspects of the head related transfer function (HRTF) and accuracy of room simulation and reproduction are mature fields and suffer from diminishing returns. The use of increased realism is, in part, driven by increased capability in reproduction hardware. Early studies which had limited availability to high performance real time processing focussed primarily on efficient methods [1] and reduced bandwidth stimuli [2]. The current state of the literature benefits from increases in processing power and techniques such as partitioned convolution to enable low latency auralisation of large FIR filters [3] at high sample rates. This has naturally produced a body of literature investigating high realism stimuli using measured responses and acoustic simulations [4] [5] [6]. It has been demonstrated that similarity to the response to simulated sources that one experiences from real sources converges the experience of the stimulus [7][8]. However, this finding in reality should not be surprising. As synthetic stimuli are processed to have more of the perceptually relevant features which are associated with realism, there is no surprise that they are perceived as more realistic. This work, therefore, aims to investigate phenomena associated with lower levels of physical realism in spatial auralisation. Where there is less divergence between reality and simulation, it can be expected that there would be more agreement between subjects that high quality stimuli are of high quality. However, when auralisation is imperfect, it provides an opportunity to identify where there is variation in the assessment of stimuli.

If we extend this case to the multimodal immersive environment, we see similar stories emerging. The theoretical case of Gibson's classification of vision types by degrees of freedom [9], and work demonstrating that increases in modality are measurable on a psychophysiological level provide weight to the argument that simply increasing features or modes is associated with greater perceptual performance. However, there have been shown to be greater nuance that can be found within the understanding of these phenomena. The inter-participant variance observed in

studies often suggests noisy data. In terms of response to stimuli, that noise can be thought of as the transfer function of the human participant as the stimulus is encoded and perceived, a response formulated, and a motor process is executed. This is essentially the basis of signal detection theory, a branch of statistics used in cognitive science and perception which assumes a portion of stochastic noise in the response of a subject to a stimulus [10]. In terms of unimodal audio stimuli, basic audio quality and spatial audio quality are often used to assess audio systems in terms of degradation. Procedures such as multiple source hidden reference (MUSHRA) [11], and the triple stimulus method described in ITU BS.116-3 [12] are a based on a unidimensional model of 'basic audio quality'. This is described in ITU documentation in terms of degradation causing annoyance. Conversely, spatial audio quality taxonomies [13] [14] tend to be multidimensional. However, studies have suggested that spatial audio quality may be unidimensional and dominated semantically by envelopment [15]. The work in this thesis focusses on multimodal stimuli, and existing taxonomies of spatial audio may not have been fit for purpose. The descriptors of audiovisual quality of experience that are introduced in section 4.2.1 and used throughout are taken from a range of existing literature concerned with differing aspects of multimodal spatial stimuli which were initially hypothesised to be independent: Audiovisual fusion [16], externalisation [17], plausibility [18], sense of localisation [19][20], and awareness of headphones [21].

The psychology literature reviewed in this work can be thought of as modelling the stimulus/response transfer function of humans in relation to phenomelogical traits which subjects might display. It is in this context that the work which is presented in this thesis is intended to be read; that much of the study of spatial audio and spatial audio in VR has focussed on 'signal domain features', those which can be controlled and improved on by an engineer. Conversely, there is a body of work which seeks to classify 'receiver domain features', which measure and classify the response of subjects to given, often simplified inputs. In addition, of all the works concerning immersive VEs referenced in this literature review, few papers compare responses to an actual real world reference. As such, the aims of this thesis are to investigate both effects and interactions between the following:

1. The effect of similarity between pre-exposure real world geometry and virtual environment geometry on audiovisual quality of experience and presence
2. The effect of modification of low-realism spatial rendering on audiovisual quality of experience and presence

3. The effect of differences in sensory response and personality on audiovisual quality of experience

The use of low quality rendering was selected for two reasons. The first is the reason described above that it can be argued that increasing salient signal features is known to produce better results with diminishing returns. Secondly, is the prevalence of object based simplified rendering models which are available to the development community, a discussion of which can be found in section 3.2. It was therefore decided that an investigation into such low realism, real time processors would have greater value than simply aiming to quantify effects at the maximum possible level of rendering realism. To achieve the overall aims of the study, the following objectives were identified.

Objectives to Aim 1: The effect of similarity between pre-exposure real world geometry and virtual environment geometry on audiovisual quality of experience and presence

The outcomes from this aim form a novel contribution to knowledge in determining the influence of anchoring bias in the perception of spatial audiovisual stimuli due to pre-exposure similarity to virtual environments.

a . To identify the effect of modification of acoustic response accuracy on quality of experience responses in immersive virtual environments with similar and dissimilar pre-exposure environment geometry

b . To identify the effect of audio-visual ambiguity (where audio and visual sources may not be co-located) on the quality of experience responses in both similar and dissimilar pre-exposure conditions

c . To identify the effect of changes to the explicitness of the spatial relationship between audio and visual components of a stimulus on quality of experience responses in immersive virtual environments with similar and dissimilar pre-exposure and environment geometry

Objective to Aim 2: The effect of modification of low-realism spatial rendering on audiovisual quality of experience and presence

The outcomes from this aim form two novel contributions to knowledge. Firstly, determining the relationship between audiovisual quality of experience and an existing model of presence. Secondly, determining the relationships between spatial relationships between auditory and visual stimuli and auditory spatial cue content using low-realism room models in HMD VR in terms of AV-QoE and dimensions of presence.

a . To identify differences in quality of experience responses relating to stimuli when subjects believe stimuli to be emitted within the real room or simulated over headphones

b . To identify the effect of the following manipulations on the reporting of presence in VR environments:

- Manipulation of acoustic response accuracy
- Manipulation of audio-visual ambiguity.

c . To identify the relationship between quality of the audiovisual experience and reported presence

Objectives to Aim 3: The effect of differences in sensory response and personality on audiovisual quality of experience

The outcomes from this aim constitutes a novel contribution to knowledge in refining relationships between individualising characteristics and subjective response of presence and audiovisual quality of experience. Additionally, this contribution attempts to posit a causal hypothesis for the relationships observed.

a . To identify potential relationships between personality and cognitive dimensions which may improve any relational models between signal domain features identified above

1.1 Definitions of terms

Throughout this thesis, there will be multiple references to terms which have specific meanings in the context of the document which may be used interchangeably in casual discourse. For the purposes of clarity these terms are defined below.

- *Presence* - Presence is the feeling or sensation of being 'in' a place or location and physically present in this space [22]. This sensation is ordinarily transparent and unnoticed, except in cases such as depersonalisation disorders [23]. The experience of presence in VR is an active area of research and proposed factor structures and modes of assessment are discussed in section 2.3.1. It should be noted that presence is distinct from other evaluative measures of a VR experience such as realism and plausibility, due to being a physical/psychological response to the experience.

- *Immersive* - The term 'immersive' is used in this work as defined by Witmer [24]. This definition places it as related to, but distinct from *presence*. Where presence is the sense of 'being there', immersion is the effect of having the signals to the senses replaced by those created by the virtual reality system. This distinction is important to recognise as it is possible to be immersed without the experience of presence. Furthermore, the feeling of immersion is related more to the *immersive* properties of the reproduction technology.

- *Virtual Reality* - In the context of this thesis, virtual reality is taken to refer to any audiovisual system which attempts to replace real-world sensory inputs with those generated within a computer. The methods for this might be achieved using projections, structures of electronic screens or head mounted displays. For the purposes of this thesis, this definition does not include ordinary displays which do not aim to situate a user in a simulated environment, or are not intended to elicit presence. Further discussion can be found in section 2.3

- *Virtual Environment* - A virtual environment is defined as any environment which is simulated within a computer. For the purposes of this thesis, they are considered separate from virtual reality in that VR refers to the modality of consumption whereas the virtual environment is the content which is being consumed.

2

Literature Review

Contents

2.1 Introduction

In this chapter work will be reviewed in the general areas of spatial audio rendering and the special case of spatial audio in VR. This chapter will also cover concepts surrounding immersive multimodal virtual reality systems and technologies and the perception of stimuli within immersive multimodal systems and the experience of presence that is elicited by these systems. Finally, research on the differences between individuals which account for differences in response to stimuli

will be discussed and the physiological evidence supporting these categorisations will be presented. The aim of this chapter is to give the reader an understanding of the context and issues surrounding the work described within this thesis and to identify literature which has inspired, contributed to and informed the contributions which are detailed in this document.

2.2 Spatial audio and virtual acoustics

The simulation of the acoustic cues which humans use to localise sound in three dimensions is an active and mature area of research. Although strictly comprising of two separate corpi of knowledge, the study of three dimensional localisation, and the implementation of spatial audio processing often rests on an intersection of virtual acoustic simulation and psychoacoustically informed reproduction.

2.2.1 Reproduction systems

2.2.1.1 Speaker based systems

The spatial reproduction of sound can be divided into two approaches: Speaker based and headphone based systems. Speaker based systems rely on multi-channel arrays of loudspeakers, arranged in such fashion as to surround the listener to some extent and approximate a soundfield by controlling the amplitude, and in some cases the timing, of sound objects within a scene. The most simple and ubiquitous spatial format is stereo, which relies on a combination of time and amplitude differentials between reproduction channels depending on soundfield capture or synthesis [25]. Higher order systems for home reproduction exist, such as surround sound [26]. However, these systems are not typically intended for use in accurate soundfield reproduction, but the playback of film, television and, to a lesser extent, musical content. More accurate spatial reproduction is achieved through systems such as vector base amplitude panning, ambisonic multichannel arrays or via wavefield synthesis.

Vector base amplitude panning extends ordinary two channel stereo and allows for the generation of a phantom source within a pair or triangular segment of an arbitrarily sized two or three dimensional speaker array. In its simplest form, VBAP takes the form of a simple panning law in which the channel gains for loudspeakers

oriented at 90 deg are the Cartesian components of a unit length vector with a given azimuthand elevation. Compensation for non-orthogonal positioning of speakers is achieved by scaling gains appropriately. In the three dimensional case, gains are taken to be the linear combination of orthogonal base gains and loudspeaker location vectors, with scaling performed in three dimensions [27]. As VBAP extends channel panning based on sine laws to arbitrarily sized arrays in three dimensions, ambisonics [28] extends the sum and difference stereophony proposed by Blumlein [29] to accommodate multichannel arrays and height information. As with mid-side encoding, ambisonic soundfields are reproduced as a summation of a unidirectional signal with signals in which the polarity and gains of signal content are directionally dependent. The lowest commonly used order of ambisonics, often referred to as B-Format encodes one channel for each cardinal axis (x,y,z) and the omnidirectional (w) component. Synthetic ambisonic soundfields are rendered by generating encoding coefficients based on spherical harmonic decomposition and applying weighting and polarity inversion to the signal encoded to produce a given directional component. As order increases, spatial resolution is increased, as is the radius of the effective listening area [30, 31]. Once encoded, the whole soundfield may be rotated arbitrarily. [32].

Wavefield synthesis (WFS) is based on the Huygens principle, that any wave-front can be approximated by a combination of secondary sources. An arbitrary wave-front is reproduced by controlling gains and time delays of individual components of large speaker arrays [31, 33]. This allows for arbitrary curvature of the emitted wave-front, enabling the positioning of virtual sound sources from within the boundaries of the listening area [34]. This technique can produce convincing spatialisation, however spatial aliasing is observed at frequencies above equation 2.1 where Δx represents speaker placement interval, α_{max} is the maximum angle between any loudspeaker and the receiver [35], and c being the speed of sound.

$$f_{al} = \frac{c}{2\Delta x \sin \alpha_{max}} \tag{2.1}$$

Other than acknowledging the concepts behind audio spatialisation, and for providing familiarity with techniques that are referenced later in this work, a complete description of these approaches to speaker based spatial audio rendering lies outside

the scope of this project. Further, and more detailed, descriptions can be found within the literature on the subjects of VBAP [27], Ambisonics [30, 36], and WFS [35].

2.2.1.2 Headphone based systems

The reproduction of spatial sound over headphones is based on the ability to either reproduce or simulate the effect of human physiology on an acoustic pressure wave as it enters the auditory canal. The physical separation of the ears and the baffling produced by the human head, reflections from shoulders and absorption from hair, in addition to filtering due to the complex morphology of the pinnae produce a direction dependent transfer function. This transfer function is referred to as the head related transfer function (HRTF). The relative changes in frequency and time response between the received signals, when processed within the brain, elicit a sensation of localisation and externalization in the listener. Binaural signals may be captured using dummy heads or in-ear microphones to record the sound pressure variations at an anatomically analogous or actual human ear canal . However, although these captured signals can be realistic when played back over headphones, the rotation of the soundfield is not possible. This precludes the use of head tracking, the absence of which introduces breakdowns in externalization and plausibility [37]. Binaural synthesis is achieved through the convolution of a signal with the impulse response or transfer function observed at the entrance to the ear canal at a given azimuth and elevation. Ideal circumstances would see the use of individualised HRTFs recovered from the listener. In the majority of cases, this is not feasible. As such, it is common to see the use of non-individualised HRTF filters. Early work in this field appeared to support the general use of non-individualised filter sets [38]. However, more recent work has linked performance of these filters to perceptual similarity [8]. It has been shown that although individualisation increases quality of experience, the ability for users to reliably identify individualised HRTFs is low [7] and that common spectral cues are the dominant features in eliciting externalisation and localisation [8]. There has been shown to be errors in localisation which are associated with HRTF mismatching [39] but these shortcomings have been shown to be surmountable due to learning effects when subjects are given spatial feedback [40][41]. In cases where the performance of signal processing hardware is insufficient to allow convolution with an actual measured response, parametric approximations derived through various methods have been implemented [1][42][43]. Similarly, many processing approaches

assume a point source emission model, however, propagation from volumetric acoustic sources has been modeled and demonstrated in real-time VEs [44]. Although anechoic HRTFs are commonly used, and may be suitable for some outdoor scenes, it is also common to find binaural room impulse responses (BRIRs), binaurally captured room measurements, which provide the binaural representation of the HRTF in an echoic environment in a given position and rotation within that space. Such BRIRs may provide greater realism than anechoic HRTFs. However, they suffer from the same shortcoming as binaurally captured sound recordings in that they provide a representation of a static position and orientation within a given space. The use of virtual speaker rendering allows the production of a binaural soundfield which allows the listener to be presented with binaural renderings appropriate for both rotation and position within a space with arbitrary room response [45]. It is this class of signal processors which have given rise to a conflation of spatial audio processing and virtual acoustics in the marketing of spatial audio renderers for use in the development of computer games and virtual reality. Such processing may be achieved through the use of multiple BRIRs which are appropriately applied for a given listener position or orientation. Similarly, room responses may be simulated using methods described in section 2.2.2

2.2.2 Virtual Acoustics

2.2.2.1 Artificial Reverberation

Statistical methods for artificial reverberation aim to leverage the stochastic nature of a diffuse impulse response [46]. The parameterisation of these class of reverberators tend to focus on *a posteriori* qualities of a recreated space such as reverberation time (RT_{60}), perceptual mixing time (pre-delay) [47] and diffusion [48], as opposed to the *a priori* parameters of physical modelling techniques described below. Implementations of this class of reverberation include Schroeder and Moorer reverberators [49, 50], feedback delay networks [51], and convolution with temporally shaped noise [46]. The aim of these topologies is, with the exception of convolution with noise, to generate a diffuse decay by way of recursive delays, all-pass filters or comb filters, constructed as to minimise time coherence of individual feedback nodes. Interaural decorrelation may be achieved by alternately inverting the polarity of some summation points in the network or through the use of low frequency modulation of delay parameters for one or more channels [52]. This class of reverberator is useful

due to low computational cost and simple parameterisation allow for implementations to be used in low cost consumer electronics. They also facilitate design of arbitrary responses on aesthetic grounds, rather than physical accuracy in addition to providing some level of approximation of real spaces. This notwithstanding, it a has been shown that emulations of real spaces constructed using these methods are discernable from measured responses [53].

2.2.2.2 Wave Based Methods

Wave based simulation methods fall into two classifications, finite element methods or finite difference time domain methods (FDTD). Where finite element methods operate in the frequency domain, FDTD is a time domain process. An orthoganal mesh of nodes is constructed which represent the propagation of velocity and pressure components throughout the space. This topology can be realised using the digital waveguide approach in which bidirectional delay lines are interconnected in a grid formation, with the receiver signal to be taken as the summation of signals at a given node of the mesh at each time-step [54]. Finite element methods rely on descretisation of an enclosed volume or boundary surface in order to numerically solve the Helmholz-Huygens integral in the frequency domain to derive the transfer function for an arbitrary surface or space [55]. The boundary element method (BEM) [56] and finite element method (FEM) [57] can produce simulations of high accuracy, particularly at low frequency. However, computation time grows rapidly with bandwidth, as an increasingly fine descretisation of the volume or boundary is required, rendering these methods unsuitable for real time simulation of dynamically changing systems.

2.2.2.3 Ray Based Methods

A ray based method for simulation of acoustic response was first proposed by Schroeder [49]. This is achieved by the computation of delay times for a large number of emitted ray paths which are reflected either specularly or modified by some random diffraction or diffusion parameter. The output signal is comprised of the summation of signals delayed and attenuated appropriately for ray paths which interact with a receiver volume. Rays may be terminated after a given number of reflections or when the energy discontinuity percentage reaches a determined threshold [58]. In the image method, sources are approximated as a point-source emitting spherical pressure waves. Vectors describing the positions of a source and receiver

pair are subtracted producing a vector which describes the path of the acoustic pressure wave from source to receiver. With image expansion, assuming rigid boundaries of an enclosed space, a reflection image is placed beyond the boundary, unfolding the reflection path producing a single vector describing the apparent point of origin of the reflection image. Reflection images are then themselves reflected, producing an impulse response whose length is determined by the attenuation of images produced by spherical spreading of the wave-front due to the geometry of the room and absorption effects accounted for at room boundaries [59]. Although the production of reflection images within arbitrary polyhedra has also been described [60] with further improvements to image calculation methods for complex enclosures having been developed [61], the level of sophistication in calculating wavefront paths within complex enclosures offered by these techniques is matched by an increase in the complexity of the algorithm required. Due to the complexity of these algorithms, the use of this method is restricted to simple geometries.

2.2.2.4 Hybrid Methods

In order to gain both the benefits of the realism offered by physical models and computational efficiency of statistical methods, it is possible to combine both by computing the specular and diffuse components of the impulse response separately and combining them later in the signal path [62]. The late part of the signal can be further parameterised as to respond to data which is precomputed from a wave-based simulation to achieve position dependent acoustic response composed of parametric units [63]. Methods such as these are employed in proprietary spatialisers designed for first person perspective and virtual reality, with the statistical component provided by a reverberator topology such as a feedback delay network. To reduce the number of convolutions required, as the number of sources multiplies the number of convolutions required for individual image paths, spatialisation is efficiently achieved by encoding the soundfield in ambisonic representation and the resultant soundfield is decoded to fixed rendering points around the listener. These rendering points are in turn convolved with appropriate HRTFs to produce a hybrid synthetic BRIR, limiting the number of convolutions for HRTF convolution to twice the rendering point count [64].

2.3 Virtual reality

Immersive virtual reality was first proposed and implemented by Sutherland [65, 66]. In the intervening years since the first head mounted display, computer generated imagery and audio signal processing has improved dramatically. The miniaturisation of electronic devices and the increase in computing power that is readily and commercially available, driven by the adherence to Moores Law by semiconductor manufacturers [67, 68], has allowed what was once an uncommon and esoteric medium to be all but ubiquitous in the current age of mobile computing. The 'Ultimate Display' is now a common feature of the multi purpose device that is familiar and in every day use. In addition to this, recent years have witnessed a surge of dedicated hardware devices aimed at bringing immersive virtual reality into every day use. Arguably the most well known of these devices is the Oculus Rift, itself becoming a byword for the HMD. Oculus were purchased by Facebook in 2013 and afforded the device a high profile throughout its development, first releasing Developer Kit models designed for content creation and prototyping, before releasing the commercial model in 2016. In addition to the release of Oculus [69], Sony [70], HTC [71] and Samsung have all brought to market HMD solutions for their own proprietary platforms. In addition, low cost HMDs were developed by RAZER and Sensics under the name OSVR to provide an accessible and open source solution for experimentation, development and research [72]. Currently, most low cost VR solutions offer three degrees of freedom within a VE, updating rendering to account for head rotation However, low cost untethered 6-dof VR became commercially available in early 2019. Dedicated commercial HMD solutions often include fused gyroscopic/accelerometer and optical tracking to provide limited translation within the virtual world. At the time of writing, room scale VR is becoming more widespread through the use of larger scale tracking systems which are implemented in to systems such as the HTC Vive and Oculus Quest [73]. Although VR is now synonymous with the head mounted display, there have been other approaches to VR which have relied on surrounding the user with an array of static displays in order to provide immersion[74]. These systems had the advantage of being able to be constructed from existing technologies, when HMDs were not as available as today. It has also been shown that Cave systems outperform HMDs in the induction of presence [75]. However, with modern 'six degrees of freedom' (6-dof) systems and wider fields of view, this may need to be revisited.

2.3.1 Presence and VR systems

The premise of a virtual reality system is to replace the sensory inputs to an individual provided by the real world with ones generated within a computer on to transducers [65]. The sensation of being immersed within a virtual environment to the degree to which one accepts the virtual environment as the one in which the participant is located has been described as that of presence [76]. Presence is distinguished from the quality of immersion, although the two are often conflated, in that immersion refers to properties of a VE presentation technology or modality; one that allows the participant to be surrounded by the computer generated environment. Presence is the successful qualitative experience of a participant using an immersive technology [77, 78]. The experience of presence within VE contexts can be further deconstructed into two complementary metrics. Those of *place illusion* (PI) and *plausibility illusion* (Psi) [21]. Place illusion is best described as the effect produced by tracking and reproducing the orientation and motion of a subject within the virtual space; a refinement in previous hypotheses of presence effects which linked the capacity for agency within the VE as the precursor to the sensation of presence [79]. This concept was built on the theory of perception and sensory integration posited by Gibson [9, 80]. Gibson distinguishes between modalities of vision which are action dependant: Snapshot and aperturevision occurs when the head and eyes are static; ambient vision occurs where the head and eyes are given rotational freedom; finally ambulatory vision is the process where the subject is given translational freedom within the space visualised. All provide different, but complimentary, information streams to the subject about the scene to be understood. This idea of cognition via action and locomotion within a space has been argued as the theoretical basis of VE design and provides the mechanism for the acceptance of the virtual environment by the subject [81]. This classification by degrees of freedom has been supported by work shown to produce distinct responses from participants in comparisons between photographs, 360 photography and ambulatory HMD VR, with the 6-dof HMD producing responses closer to those produced by exposure to real environments [82]. Parallel to PI, Psi refers to the extent that the subject accepts the events and stimuli within the VE as real. This is a subtle distinction and may exist simultaneously with the knowledge that the virtual experience is virtual. Typically, presence is described as an emergent state of consciousness arising from the experience of an immersive experience. Within the PI/Psi framework, plausibility illusion can be understood as the emergent cognitive experience component of

the experience of presence. Of the two components, it relies on a greater number of variables and is thought to be the least stable of the two [21]. Psi as a property of the experience of presence is characterised by the automatic reaction by participants to stimuli in a realistic manner. The assessment of the degree of presence experienced by an individual in a VE can be performed using a variety of techniques, both direct and indirect. Direct methods may employ psychophysiological measurements, however, indirect data collection via self-reporting questionnaire is a very common practice in psychological research and this is no less true in the study of presence. Spagnolli and Bracken [83] cite six distinct questionnaire designs which have been adopted by at least two presence in VE studies at time of publication: Presence Questionnaire [PQ] [22]; ITC Sense of Presence Inventory [ITC-SOPI] [84]; Immersive Tendency Questionnaire [ITQ] [22]; Slater-Usoh and Steed Questionnaire [SUS] [85]; Igroup Presence Questionnaire [IPQ] [79]; and the Presence and Reality Judgement Questionnaire [PRJ][86]. The authors argue that these tests can be differentiated by the assumptions made about which variables are either dependent or independent on the perceptual model that describes the measure of presence experience. In the first category, it is argued that presence is considered an internal state of the subject who has a predetermined propensity for this state being elicited by a VE, which is considered to be understood as a condition opposite to the natural sensory condition. The second category treats presence as a emergent consequence of the immersive quality of the presentation medium, necessarily experienced but at an intensity correlated with the efficacy of the equipment used [83]. The former approach, which considers the prime factor in the experience of presence to be individual susceptibility, seems to downplay the contribution of sensory replacement and shifts the responsibility for the extent of immersion onto the individual using the VE. However, although suspension of disbelief may contribute to presence [21], the theories of vision and spatial perception which inform VR suggest that the prime factor in illiciting presence should be the efficacy of the technology. Additionally, the former assumption diminishes the measure of engagement and presence as a tool for improving the application of VR technology as it excuses poor performance on low receptivity of the subject. There also exist schisms in the methods of describing presence. These testing schemes present presence as both unidimensional and multivariate quantities. If a unidimensional test is used, it has been argued that it may be useful to measure other cognitive effects, which might infer or represent precursors to the emergence of presence. Early work in the study of presence identified factors such as attentiveness to the current place and time and the degree to which

time seemed to pass. The degree of presence felt is also known to be a predictor for the likelihood of motion sickness symptoms experienced by an individual using VR [87]. Fontaine [87] also argues that it is the distribution of attentive focus that produces varying degrees of presence experience. It is suggested that novel environments demand a wider attentive distribution, generating a greater sense of presence by virtue of attending to the current environment. Witmer & Singer [22], however, suggest that presence is related to selective attention, where attention can be more readily directed by stimulus signals which are meaningful, a model also suggested by Treisman [88]. In this application of the attentional resources model, presence is the result of stimuli forming meaningful representation through coherence of sensory signals from the VE [22]. As such, the emergence of presence is dependent on the ability of the VE designer to construct a world where sensory cues which inform the user of their environment are meaningfully and plausibly implemented and encoded into the audiovisual content displayed within the VE.

The scales suggested in the above cited works, however, vary in the ability to account for variation within the sample used to construct them. The unidimensional Witmer & Singer Presence Inventory [22] uses only *a-priori* descriptors as subscales, which are assumed to linearly sum to a measure of presence. No further analysis is performed to determine if there are latent variables which can inform more targeted assessment of presence. The ITC-SOPI, IPQ and PRJ were subjected to factor analysis on construction. However, they demonstrate differing efficacy as assessment tools. The authors of the ITC-SOPI identify a four-factor model, listing sense of physical space, engagement, ecological validity and negative effects as independent contributing factors to the experience of presence. However, factor analysis suggests this model accounts for only around 38% of the observed variation in the sample, with percentage of variation explained ranging from 14% - %5 between factors I and IV [84]. The low explanatory power of the model may be an artefact of a large number of initial variable classifications but this does not mitigate the fact that this scale may be susceptible to noise within respondents. The PRJ identifies three constituent factors in the experience of presence, accounting for 53% of the variance in response observed: Reality judgement (24%), internal/external correspondence (17%), and attention/absorption (11.5%) [86]. Although this model has greater explanatory power, it was found in *post-hoc* analysis that these subscales were not independent and that factors II and III were significantly correlated with factor I at $R^2 = 0.33$ and $R^2 = 0.25$ respectively. In the construction of the IPQ, second order factor analysis was used to further reduce the dimensionality of the response space,

resulting in a three factor model which accounts for 64% of observed variation. The latent variables identified by this scale are: Spatial presence (40%), involvement (13%) and experience of realism (11%) [79]. This model appears to account for a greater amount of variation as it separates out spatial representation, task and attention and realism in the analysis of presence, factors predicted to be salient by *a-priori* discourse. It is notable that spatial representation is a greater contributor than the sum of the other two factors, indicating that stimuli which contribute to a plausible representation of space within a VE are important considerations in the design of such environments.

2.3.2 3D audio in virtual reality

The use of sound in virtual environments, particularly those in virtual reality contexts, can be understood within the framework of analysis used for video game audio. However, the consumption modality of VR necessitates certain limitations of user perspective that are not required in traditional game environments. Game audio is often thought of as extending cinematic practices providing indications of diegetic activity within the game world and non-diegetic emotional exposition and scene-setting providing a narrative context for the scenario and activity in the game [89]. Whalen [90] argues that the use of audio within games more closely mirrors that of animation, rather than operating on a general filmic theory of composition and sound design. The use of sound to provide life and realism to a necessarily unrealistic representation of a fictive object, despite recent advances in real time rendering technologies, is a representational device used since the earliest examples of animated film [91]. Although, this was borrowed from the practices employed in theatres before the advent of synchronous sound recording in film, the use of overtly gestural audio cues became more associated with animation practises than those employed within live action film production. The argument being, that the unreality of the game environment necessarily requires corresponding audio cues to realise the environment and its constituents in a way which live action film is not so constrained [90]. This analysis of the function of game audio, and by extension virtual reality audio serves as a basic framework for an understanding of the function of sound within a virtual environment. Sound in the film, animation, and now the game environment, informs us though isomorphic representation of the corresponding visual object of the kinematic, material and mechanical properties of what it is that we are seeing in the virtual scene. Sounds need not even be truly representational of

the object in question. What a particular object sounds like may not be the main concern of a sound designer. It is argued that what is more relevant is that an audio event should be semantically encoded with the properties and behaviours belonging to the object as presented in the scene; that it tells us whether a thing is heavy or light, hard or soft, fast or slow [89].

The association of congruent sounds to actions or events within a VE is, however, only part of the process of constructing a convincing virtual auditory world. The use of signal processing to synthesise the spatial cues that humans use to perceive the auditory scene in three dimensions is critical to the design of truly convincing virtual environments. Spatial audio can be implemented in various ways within VR, although two reproduction modalities are commonly used and are often associated with particular presentation systems. VR systems which rely on image projection often use speaker based 3D audio systems such as VBAP [92, 93] and wavefield synthesis [94–96] using the projection screens to hide loudspeakers. Head mounted display (HMD) virtual reality systems often utilise headphone reproduction. Speaker based systems benefit from being extensible and flexible to implement and can allow for good spatial performance, hybrid systems may even be employed to leverage the benefits of each spatialisation scheme while minimising artefacts that are associate with a particular technique [97]. However, when using HMD VR, the convenience of being untethered by headphones is made redundant by way of being tethered by the HMD itself. Headphone reproduction also benefits from isolating the listener from the acoustic environment in which they might be situated and does not require the listener to stay within a fixed listening position. The use of headphones to deliver sound in VR lends itself almost exclusively to binaural techniques. The convolution of monaural sound sources with a head related transfer function (HRTF) filter is a well described and often employed process to implement spatial audio in VR [98–100]. As discussed above in section 2.3, there has been an increase in the availability of low cost commercial HMD systems in recent years. The effects of this increased access to HMD equipment can be seen reflected in the literature and illustrates this association between HMD VR and binaural rendering over headphones. Publications in peer-reviewed journals discussing behavioural response to audiovisual stimuli in virtual environments were reviewed. Criteria for inclusion were that they were behavioural studies on some aspect of presence, spatial audio quality or multimodal perception; used some form of multimodal stimulus which obfuscated outside stimuli; and had audio implemented in the experimental design. Technical

papers detailing novel presentation media or methods were omitted. The full list of these publications can be found in appendix A.

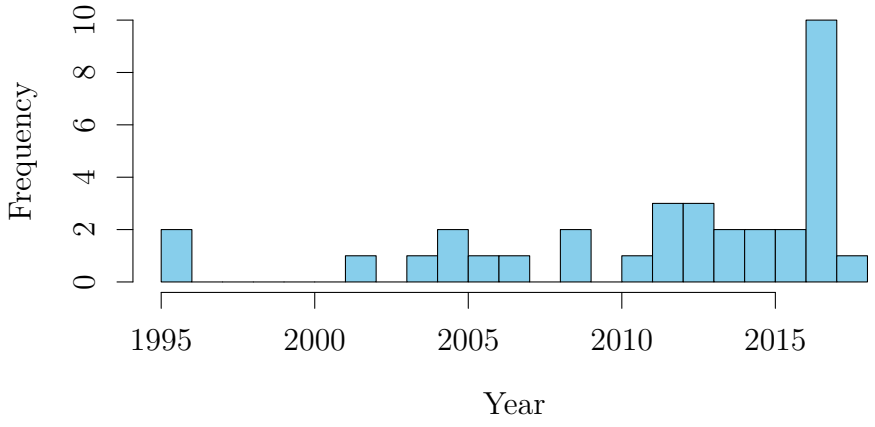


Figure 2.1: Frequency of behavioural studies using immersive VEs and spatial audio by year

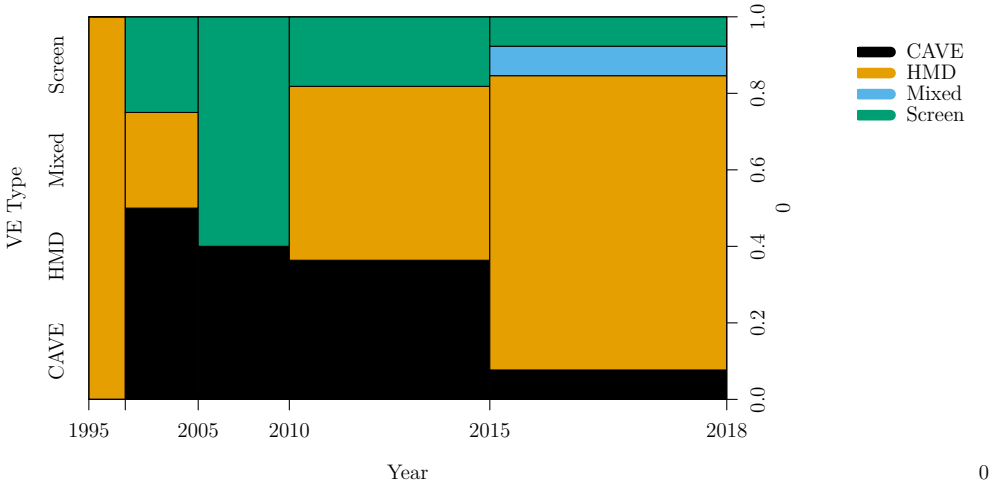


Figure 2.2: Frequency of behavioural studies using immersive VEs and spatial audio by VE type and year. Right side y-axis shows proportion of representation within a 'year' column

Figure 2.1 shows the frequency of publication dates of the studies included. There is a clear peak of publications in this area since 2015. Forty seven studies published between 1995 and 2018 in the area of perception of virtual environments were reviewed. Figure 2.2 shows the proportion of publications in the area of behavioural

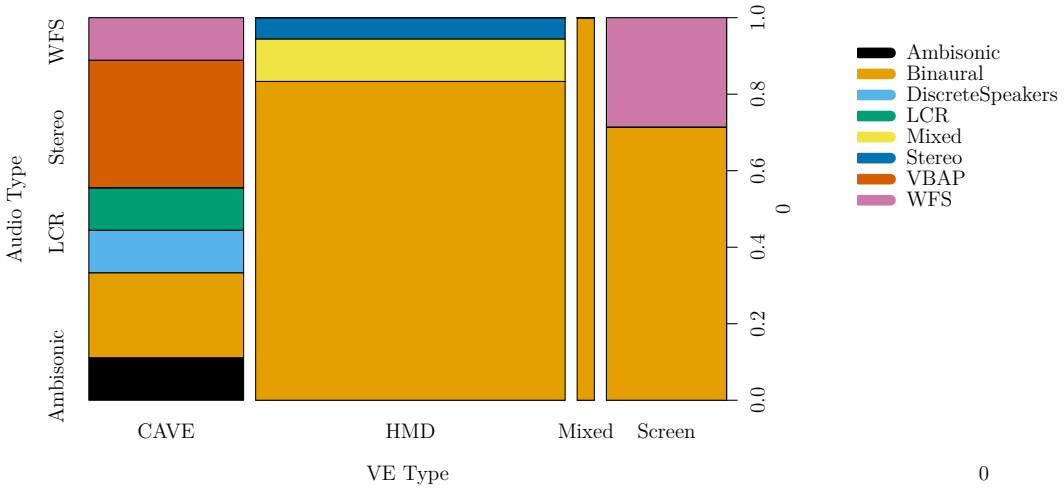


Figure 2.3: Frequency of behavioural studies using immersive VEs and spatial audio by audio reproduction type and VE type. Right side y-axis shows proportion of representation within a 'year' column

response in immersive virtual reality by year from 1995 to 2018 as a function of display type. It can be seen that the proportion of HMD based has increased with a commensurate decrease in the publication of work investigating CAVE or screen based reproduction systems. Similarly, this effect can be observed in the field of audio in VR. Thirty five studies reviewed were on behavioural response to auditory stimuli in VR and tabulated by year and audio rendering method. Figure 2.3 shows that for these studies, HMD VR is most associated with binaural reproduction, with CAVE systems employing the greatest variety of techniques. Figure 2.4 shows audio rendering representation as a function of year. It can be seen that the studies published between 2015 and 2018 all employed binaural rendering. This analysis cannot be considered fully exhaustive, however, and was limited in scope. However, despite the significant limitations in sample size, it is illustrative of the practices that are currently used in the field. The increase of access to affordable HMD equipment is likely spurring greater activity in this area. The nature of the medium which is driving this activity lends itself to certain rendering methods over others, namely binaural synthesis via headphone reproduction.

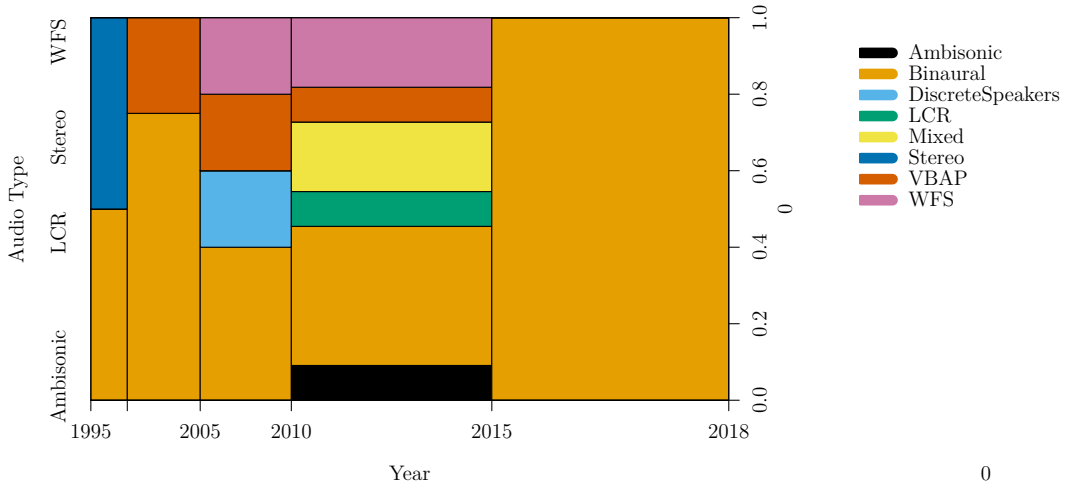


Figure 2.4: Frequency of behavioural studies using immersive VEs and spatial audio by audio reproduction type and year. Right side y-axis shows proportion of representation within a 'year' column

2.3.3 Three dimensional vision in VR

Human vision is binocular and this feature allows for the perception of depth by stereopsis; the integration of two images. The integration of the disparate individual monocular images produces a phantom image that appears to originate from the centroid of the interpupillary distance. This point is referred to as the cyclopean eye. The degree of binocular disparity provides one of the major cues for the judgment of distance in stereoscopic vision [101]. However, retinal disparity does not, in and of itself, produce the sole distance cue which enables the estimation of egocentric distance [102]. Fixation upon an object within the horopter, the static field of view, requires both accommodation of the crystalline lens and independent orientation of each eye by the extraocular muscles to converge the lines of sight on the object of interest [103]. Thus, lens accommodation and ocular convergence provide a non-retinal distance cue, which is integrated into the retinal disparity cue to produce a sense of distance. Altering either the apparent interpupillary distance or the apparent focal length of a fixation point by way of prismatic lenses causes error in distance estimation, with independent accommodation/convergence alteration producing the highest order of error [104], demonstrating the extraretinal nature of the distance estimation cue.

In head mounted display (HMD) based VR systems, however, accomodation and

convergence cues cannot be reproduced; only binocular disparity cues for a fixed convergence form the basis of the stereopsis employed. It has been shown that the quality of the visual perception of space within virtual environments presented using HMDs is lower than that experienced in the real world [22], with estimations of distance to previously seen locations being shorter in VEs than those made in the real world . This spatial distortion effect has been accounted for using fMRI imaging [105], where it is argued that distant objects in VEs are processed as objects within action space or personal space [106]; an area in which the subject would be physically capable of manually manipulating the object. This is evidenced by activity in the motor cortex which is normally absent in the cognition of objects within vista space, the area outside of action space. It is hypothesized that this phenomenon is accounted for by the disparity between motor-sensory ocular cues and binocular visual cues, which are usually integrated to judge distance of objects in real world scenarios [ibid.]. It has long been known that disparity between focal accommodation and binocular convergence reduces accuracy of distance estimation [104] and, due to the fixed focal plane presented by HMD VR and the proximity of the screen to the users eyes, such disparity takes place in this presentation modality.

2.3.4 Cross modal effects

Humans make use of the integration of multiple sensory inputs to inform perception of their surroundings. It has been shown that visual information is used to calibrate auditory localisation [92] and create contextualisation for the plausibility of synthetic auralisation [18]. Conversely, adding audio content to a visual stimulus can affect the perception of the integrated experience. Auditory accompaniment of stereoscopic visuals has been demonstrated to increase the perceived quality of a computer generated image [107], while bimodal textual presentation has been shown to increase engagement in interaction in the context of educational role playing games [108]. In the auditory domain, assessment of spatial audio processing algorithms in multimodal environments has been shown to be affected by the presence of visual stimuli, with authors concluding that in cases where visual and auditory stimuli are combined, cross modal assessment criteria should be used [109]. In addition, binaurally rendered audio has also been shown to affect the proprioceptive senses when viewing video suggesting that subjects are in motion [110, 111].

It is clear that auditory and visual modalities are interactive with respect to each other. It has long been known that the auditory system can be tricked into hearing

incorrect vocalisation by altering the motion of the mouth in an accompanying image [112]. Similar anomalies can be demonstrated with respect to localisation, with the visual localisation capturing the auditory localisation. However, the direction of this appropriation of location has been shown to be dependent on the strength of the visual localisation. For low contrast, blurry objects, the auditory localisation is integrated into the localisation estimation with greater weight [113]. In terms of temporal localisation, although variably dependent on attentional capacity [114] and light intensity, the time response for visual processing has been determined to have a minimum critical time of 34ms over a 1 minute arc area of the visual field [115]. However, it has been demonstrated that audition holds primacy over some aspects of visual perception, with auditory stimuli able to distort the perceived timing of visual stimuli by up to 100ms. It can be argued that sensory integration is performed dynamically, with the individual determining the salience and corresponding sensory weighting of each modality contextually. Dynamic sensory integration is also implicated in the continuous recalibration of visual and auditory localisation cue parsing. It has been shown that early blind subjects display lower levels of auditory localisation accuracy than those with normal vision [116]. Although, It was originally assumed that the cognition of cross-modal sensory inputs was performed solely on a spatial basis, with proximity of the stimulus determining the integration effect perceived by the subject, it has been argued and demonstrated that the integration of auditory and visual sensory events is also processed at an object-based level, wherein a sensory object with perceptual primacy captures the corresponding sensory cue in the different modality [117]. Differences have been found in event related potentials (ERPs) recovered from participants exposed to spatially co-located or dislocated stimuli, suggesting mutiple pathways for processing but some which depend on spatial co-location of sources [118]. Further integration has been demonstrated with sensory cues related to action. Studies on the activation of mirror neurons of monkeys, neurons that fire when actions are performed or observed, have shown that there exists both modality independent and modality specific pathways for the cognition of sensory inputs when relating those inputs to action events [119]. This modality independent pathway provides some physiological basis for a shared attentional processing model, at least with respect to sounds that are perceived as being gestural in nature. Despite this, measurements of ERP activity in humans have shown differing activity in response to multi-modal stimuli based on expectation and congruence, indicating that the amount of sensory integration experienced between auditory and visual pathways is dependent on the prior experiences of the

subject, with combinations of cues which were unlikely to be related showing less than additive response compared to linearly additive increases in measured responses with what was considered a congruent combination of stimuli, a combination of the plausibility effect mentioned above [120].

2.3.5 The effect of multimodal stimuli on engagement and presence

The effect of multimodal stimuli in virtual environments has been studied in a wide variety of presentation formats; from two dimensional screen projection to immersive presentation. Studies have focussed mainly on distinct strands for the assessment of sensory integration in the VE and game experience. One thread focuses on the experiential quality provided by the game or environment to the subject, the other focuses on proxies to quantify perceptual effects; either through physiological measurement or task performance. Although methods may contain elements of both, there is distinction between research in this field between quantified qualitative and purely quantitative approaches. Byun and Loh [108] describe the effect of redundant audio content on user engagement in game-based learning environments. The findings were that the presence of audio voice-overs, which reiterated textual information presented to the users increased engagement with the game-based task, as measured using a self-reported questionnaire focusing on their experience of completing the task; itself a modified form of the Game Engagement Questionnaire [121]. The findings suggest that bi-modal presentation of textual content increases engagement with such content. In terms of incongruent auditory and visual stimuli, it has been shown that there are differing responses to distraction stimuli depending on the modality employed. Studies on the efficacy of brand placement within games has shown that auditory distractions have a greater disruptive effect on recall of visual information than visual distraction, especially when the recall target objects are familiar in nature to the subject [122], echoing the work of Turatto et al. [117] who demonstrated the ability of contrary auditory events to detract from visual processing. In fact, this tendency for auditory stimuli to diminish visual attention supports the variant of the irrelevant probe technique used by Kober and Neuper [123] to measure the extent of perceptual weighting during presence illusion experienced by participants using a projection based VR system. However, it was demonstrated using this technique, that ERPs associated with auditory attention during distraction were lower in participants reporting a higher level of presence. The suggestion from

this work is that the degree to which distracting or contrary auditory stimulus is attended to is dependent on the degree of immersion in the visual. As such, it might be inferred that the propensity of auditory distraction to affect visual processing is more dependent on how engaging a visual stimulus or task is to the participant.

In terms of the technological contributors to presence, Hendrix [124] assesses presence as a function of head tracking, geometric field of view, stereopsis and spatial sound processing. It was concluded that these factors all contribute to increase in presence, however the relationships between them was not further explored. It was also suggested that audio spatialisation increases presence but not perceived realism. This finding is supported by Larsson et al. [125] who claim that alteration to the spatial rendering of immersive scenes induces further spatial effects which may be related to the experience of presence. Auditory spatial cues have been shown to induce circular vection [99], decreasing the exposure time required and increases the convincingness of this effect, in still, seated participants [110]. There is no consensus as to whether spatial audio contributes to multimodal search task times [126][92], However, this may be an artefact of reproduction methods as differences in localisation and distance perception error rates can be seen when comparing studies which have used differing rendering techniques [127][95], suggesting that technological factors may be a significant contributor along with receptiveness to immersion.

2.4 Individual differences and perception

Differences in the performance of immersive technologies when technological factors contributing to presence are held constant may be explained by differences in the way perceptual information is processed by an individual. Such differences may be in the form of perceptual differences or cognitive differences both being affected by personality differences. Understanding these factors may give insight into the experience of users of these technologies. Theories of perception which preceded the early theories of differentiation in perceptual processing preference were largely attentional models. Titchener [128] and James [129] provide good overviews of the state of the art in perceptual studies at the birth of the field. Focus was drawn on categorising stimuli and treating the experience of perception as generally homogenous within the population, with the exception of physical pathology. Differentiation of perception was treated as an attentional construct with the experience of

perceived stimuli defined by previous experiential context. Later, gestalt theorists such as Perls [130] conceived of perception as an integrated phenomenon wherein discrete sensory modalities contributed to a unitary experience of the outside world which existed at the contact-boundary of the organism and the external world; this boundary being in and of itself a psychological construct. Although gestalt theories of consciousness supposed contribution from all available senses, selective attention was accounted for by the active suppression of modalities by an organism as stimulus response informed by the motivations of that organism [131], in contrast to the active direction of attention suggested by James [129]. It has been noted, however, that such unifying models of perception that were developed in the absence of strict empiricism assumed isomorphic analogy between the external world and the internal perceptual representation within the consciousness of the organism [132]. Conversely, the psychophysical approach attempted to view perception in an empirical and positivist manner, measuring variance in perceptual response as a psychophysical function dependent on stimulus, conditions and time with no assumption of universality across the population [133]. The differences in perceptual faculties measured using such techniques have in some cases been referred to as cognitive styles [134]. Psychophysical methods informed factor analytical research on a multitude of cognitve style dimensions such as tolerance of perceptual instability [135], visual field independence [136], and category width [137]. However, it has been noted that this era of cognitive style research produced mainly descriptive, dichotic cognitive models with dimensionality that was heavily informed by the methods of measurement used to determine effects [138].

Although such scales often relied on self reported data in the form of questionnaires, it was identified that any difference in information processing should translate into differentiation in performance of tasks which favour or hinder those with differing cognitive style. Broverman [139] utilised a two stage methodolgy, first screening for styles which demonstrated difference in cognition based on conceptual versus perceptual bases and preference for automatisation of response, then subjecting participants to various tests to measure cognitive ability and response time, with the intention of determining intra-variability in task performance as an interactive function of cognitive style and stress. Variation in the ability to resolve visual stimuli has been decribed using a multipass model of visual cognition and embedded

figures tests [140]. In terms of conceptual processing, differences within the analytical/relational and category width dimensions have been identified using sorting tasks. Messick [141] uses a modified Kagan association task [142], whereby visual patterns are associated with nonsense syllables, with participants instructed to classify novel variations of these patterns in relation to these nonsense labels, with the aim in determining the main level of focus for classification of visual information; gross structure or granular detail. The effects of these dimensions on a conceptual level have been investigated using a similar method aimed at word similarity and classification sorting [143]. Tolerance to ambiguity has been attempted to be measured using a variety of strategies. Apparent motion [135], gradually changing picture series and the Stroop interference test, used in conjunction with ambiguous narrative recall, suggested an independence of visual ambiguity and conceptual ambiguity [144], although most modern definitions of this scale refer to an intolerance to conceptual and perceptual instabilities of multiple dimensions: complexity, unfamiliarity and insolubility [145].

Cognitive styles have been used recently to frame investigations into variation in areas such as creativity [146], management [147], cognitive ability [148] and in understanding internal spatial cognition strategies [149]. Current discourse also focusses on arguments for and against a unitary or multidimensional model of cognitive style and the relationship to other descriptive psychological constructs. There has been some criticism of cognitive styles, with claims that some dimensions correlate significantly to personality type dimensions. Specifically, extraversion was shown to correlate significantly with the Activist/Reflector dichotomy proposed by Honey and Mumford [150][151]. However, it was shown that the holist/analytical and verbal/imagery dichotomies do not significantly correlate with any Eysenck dimensions of personality or the Big Five personality type components [152][153], suggesting that these may be information processing dimensions which are independent of personality. There are studies which have suggested that field dependence, itself considered a correlate of the holist/analyst dichotomy, is significantly related to cognitive ability. However, this does not disqualify it as a distinct dimension of information processing style. These have been performed through meta analysis [154] and by correlating task performance using instruments designed to measure ability and style [155]. Although more recent studies have failed to replicate these results [153], it has been shown that these dimensions may be linked to academic performance [156][157].

Neural imaging studies have allowed psychologists to attempt to find correlations in neural activity and variation in dimensions of cognitive style. Mitchell [158] reviews functional magnetic resonance imaging (fMRI), structural magnetic resonance imaging (MRI) and diffusion tensor imaging (DTI) studies pertaining to cortical stimulation sensitivity, limbic stimulation sensitivity and connectivity between the prefrontal cortex and amygdala, and generalised white matter connectivity and offers the results as a neural basis for the descriptive model of personality proposed by Eysenck. Kennis, et al. [159], similarly use a review of neuroimaging studies to support a modified form of Grays reinforcement sensitivity theory (RST), in which observed neurological systems of interacting regions correlate to the behavioural approach system (BAS), behavioural inhibition system (BIS), fight flight freeze system (FFFS) and a system of cognitive restraint. The RST and Eysenck/Big five frameworks of personality treat neuroticism and extraversion as orthogonal personality dimensions. However, the RST, while taking a more functional or procedural point of view, differentiates between low level and high level threat sensitivity and adds the independent dimension of self control, although self control may be identified with the dimension of conscienciousness in the big five model. That distinct neural correlates for descriptive personality typing have been identified lends credence to the use of certain bipolar continua, particularly the extraversion-introversion (E-I) dichotomy that is used in the aforementioned typing frameworks, but also typing schemes including, but not exclusively, the MBTI, Five-Factor model and Analytical Psychology [160]. In addition to neuroimaging studies on the neural correlates of personality dimensions, information processing stlyes have also been investigated. Tolerance for uncertainty or ambiguity has been suggested to have a developmental basis, with adolescents displaying greater ambiguity tolerance than adults as well as lower risk aversion [161]. Risk aversion and uncertainty tolerance has been shown to have a different neural basis in adolescents and adults [162], although observation of insular activation during exposure to uncertainty has suggested that there may be some common neural basis in uncertainty intolerance [163]. It has been shown that habitual preference for verbal and visual processing was correlated with differences in activation detected using fMRI and that performing dual modality tasks increases load on mid-frontal areas associate with working memory [148]. Similar results targeting areas associated with processing of images and words found that taking in information outside of an individuals preferred modality increases activity on the associated area [164]. Structural MRI has suggested correlates in variance of structural composition of grey matter mass with further personality dimensions of novelty

seeking, reward dependence, persistence and harm avoidance [165] and some association with responses on the empathising-systematizing model [166]. That there has been shown to be differences in cortical organisation between individuals who display divergent behaviours is not, in and of itself, surprising. However, studies such as these lend weight to the categorisations that have been developed, in that they reflect real physiological differences and are derived from physical variation in neural structure.

Ford [167] reviews the use of holist/analyst and field dependence to differentiate task performance in data search and information-space tasks [168]. It was at that time, however, noted that although these information processing schemas showed promise in the understanding of the cognition of virtual environments, there was a conspicuous paucity of studies investigating these factors as predictors. In the intervening period, cognitive style has been investigated as a factor affecting the perception of multimedia quality [169], the effect on learning within VEs [170] and, via cognitive ability, navigation tasks [171]. As yet, these cognitive styles have not been investigated as a predictor of presence, despite the understanding of presence as being an outcome of sensory information processing in users. Other psychological attributes such as empathy and locus of control have been correlated with scores on the ITQ [172], which provides insight into the variation found within presence response in the use of immersive media. In understanding virtual reality under a Gibsonian paradigm (Section 2.3.1), and viewing the experience of presence as a stimulus integration process, it may provide benefit to the understanding of human response to VR to investigate the effect that cognitive processing differences have on the experience of presence and its relationship to the externalisation of audio sources presented in such immersive virtual environments.

The evaluation of spatial audio reproduction often focusses on the physical or signal feature aspects of three dimensional sound. The use of head-related transfer functions and virtual acoustic techniques for eliciting externalisation of auditory sources is well described using a variety of techniques [98][62][173][36]. It is understood that the use of non-individualised HRTF sets produces some degree of externalisation and localisation [174] which allows for the widespread use of spatial audio processing, particularly in the area of immersive virtual reality (VR). Within this field, the study of *presence* is considered one of the fundamental aspects of understanding human response to immersive virtual environments [79][111][175]. This has some parallel with the phenomenon of externalisation in the study of spatial audio as it has been

shown that the experience of externalisation can be influenced by plausibility and expectation in addition to signal domain features [176]. The experience of presence, much like externalisation, has proponents which argue for a 'signal-domain' origin, one that arises from the quality of the immersive world which is created [83]. However, there have been shown to be personality correlates which predispose individuals to experience presence more than others. It has been identified that individuals who score highly on empathy scales and those who display immersive tendencies show greater predisposition to the experience of presence [22][172]. The relationship between personality and cognitive traits has been investigated, albeit overwhelmingly in= the visual domain. As reveiwed by Kober [177], it can be seen that results in this field are often not conclusive, often contradictory, and assessed using disparate configurations of presentation technologies. There have also been attempts made to unify individual differences and *a priori* prediction of presence [178]. However, once again, this is limited largely to the effect of stimuli presented in the visual domain, and assumes a additive relationship between orthogonal factors as the unified predictor of presence. As theories of presence in VR tend to rest on gestalt ideas of perception, [9][80], greater focus on the contribution of auditory stimuli in a multimodal context within the paradigm of participant classification would form a contribution to the existing body of knowledge. Additionally, in the context of gestalt theories of perception, it may be instructive to identify relationships between holist/systematising tendencies and the experience of externalisation of audio sources in immersive virtual environments.

2.5 Summary and Conclusions

In this section literature has been reviewed in the following areas of study: Spatial audio rendering and its use in virtual reality; the perception of multimodal stimuli in virtual reality; presence and its assessment of using self report questionnaire; and differences in perception of stimuli in virtual environments as a function of psychological and cognitive attributes. It has been shown that although spatial processing of audio stimuli in immersive virtual environments has been asserted to contribute to presence, the nature of this contribution and its relationship to other aspects of presence is not clear. Particularly, the contribution of perception of spatial audio has not been integrated into existing models of presence, except for superficial recognition of its importance [124]. Existing models of presence focus on visual

quality and spatial information, or high level personality factors such as enjoying narrative [22] [84] [85] [79] [86]. Integrating the perception of auditory stimuli into an existing model of presence would contribute a contribution to a field which has neglected an important sensory modality. This is a line of inquiry which will be followed in later chapters in this document.

Although there is a body of work which attempts to quantify the effects of individual differences in the reported experience of presence, the solution presented in the literature are somewhat inconsistent. Differing results reported in studies where personality and cognitive factors have been analysed [179] [180] [181] [182] [183] [177] have as yet been unable to consistently implicate factors, despite there being some evidence that this area of study has potential to account for noise evident in perceptual response data. This may be due to problems in analytical methodology, or issues with the selection of independent variables for study. Additionally, in terms of audio, no work investigating individual differences and the subjective grading of audio was found. As such, this thesis attempts to address these gaps in knowledge. Finally, it was shown that there are indications that the nature of VR research is converging to a state which increasingly studies immersive virtual environments using commercially available head mounted displays and, as such, the use of HRTF based auralisation is common commensurate with this display medium.

3

Materials and methods

Contents

3.1 Statistical methods

3.1.1 Mixed and multilevel linear models

Many of the analyses described in this thesis make use of a variant of ANOVA designed to compensate for the lack of independence that arises in repeated measures samples. In situations where data are sampled from both repeated and independent measures, the assumptions of both approaches to ANOVA are violated. The use of repeated measures from participants can be approached in both a multivariate or univariate fashion. However, if two sets of treatments utilise independent samples of a population, then random, within participant, effects cannot be universally compensated for between determining fixed effects. Similarly, the use of an independent

measures approach would fail to take into account correlated errors due to individual task performance. In a mixed effects approach, covariance within groups, between groups and within the population is accounted for, meaning these random effects do not contaminate the estimates of coefficients [184]. A mixed effect model may take the form

$$y = \alpha + \beta_{im} x_{im} + \epsilon_m \quad | \quad \epsilon = \mathcal{N}(\mu_m, \sigma_m^2) \tag{3.1}$$

With α the intercept of the model and β_{nm} the n^{th} coefficient for the m^{th} nested condition. This model may be referred to as a varying intercept model, where the error term ϵ is parameterised with the mean and variance of the m^{th} group in which fixed effects are nested. Linear mixed effects models can be analysed using a multi-level approach, where a baseline model is compared to models of increasing order in which add predictors for main effects and interactions successively [185]. Assessment of model fit is based on the Akaike information criterion ($AIC = 2k - 2ln(\hat{L})$) [186], a quantity which applies a penalty to the log-likelihood for k parameters used in the model, thus preventing overfitting. The relative contribution to goodness of fit of a fixed or random effects predictor can be assessed with a likelihood test [184]. However p-values can be estimated under the assumption that L is distributed as χ^2 with k degrees of freedom [187]. *Post hoc* contrasts may be performed using generalised linear hypothesis tests for simultaneous comparisons which are appropriate for models with parameter estimates derived from correlated data, such as are found within mixed linear models [188].

3.1.2 Wilcoxon signed-rank test

The Wilcoxon test can be used for simple one-way, two sample comparisons in non-normal data. Data are ranked ordinally and the test statistic W is computed using:

$$W = \sum_{i=1}^{N_r} [sign(x_{2,i} - x_{1,i}) R_i] \quad [189] \tag{3.2}$$

The expected value for W is 0 under H_0 and is distributed with a variance of:

$$\sigma^2 = \frac{N_r(N_r + 1)(2N_r + 1)}{6} \quad [189] \tag{3.3}$$

Where N_r is the highest rank value and R_i is the i^{th} ranked value. As N increases, W converges to a normal distribution [189].

3.1.3 Kruskal-Wallis ANOVA by Ranks

The Kruskal-Wallis test is a non-parametric alternative to analysis of variance (ANOVA) which can be used where one way comparisons are performed across k groups. As with the Wilcoxon test, N samples are ranked and the test statistic H is evaluated by:

$$H = \frac{12}{N(N+1)} \sum_{j=1}^{k} \frac{{R_j}^2}{n_j} - 3(N+1) \quad [190] \tag{3.4}$$

where n_j is the number of observations in the j^{th} group of k and R_j is the sum of the ranks in the j^{th} group. If the groups are samples from the same population, H will be distributed as χ^2 with $df = k - 1$. H_0 is rejected if H exceeds a critical value defined by α and df.

3.1.4 Pearson's χ^2 test of independence

Pearson's χ^2 test of independence is a statistic used to identify frequencies of measurements which differ from the expectations set by a model of the data. The test is limited to two dimensional contingency tables, with the model (E) taking the form:

$$E_{ij} = \frac{\sum O_i \times \sum O_j}{n} \quad [189] \tag{3.5}$$

Where O is the contingency table of n observations with rows and columns indexed with i and j. $\sum O_i$ and $\sum O_j$ represent row and column sums respectively. The χ^2 statistic is calculated as:

$$\chi^2 = \sum \frac{(O_{ij} - E_{ij})}{E_{ij}} \quad [189] \tag{3.6}$$

The significance of the χ^2 statistic is dependent in the number of degrees of freedom present in the comparison; the p value being given as the integral of the probability density function of the χ^2 distribution for a given number of degrees of freedom whose lower limit is the Pearson χ^2 statistic.

Pearson's χ^2 test assumes that the sample size is large and that at least 80% of the cells of the contingency table exceed a value of 5 [189].

3.1.5 Measures of effect size

Cramér's V - Cramér's V is a measure of association between categorical variables that is derived from the χ^2 statistic [191][192].

$$V = \sqrt{\frac{\chi^2/N}{min(p-1, k-1)}} \tag{3.7}$$

Where N is the sample size and p an k are the row and column sizes. Nominally, V expresses a value between 0 (no association) and 1 (equality). However it has been shown that the value of V is dependent on the degrees of freedom of the contingency table from which it is derived [193]. Examples of thresholds for interpretation are found in table 3.1.

Table 3.1: Interpretation thresholds for Cramér's V at k degrees of freedom

d.f.	Small	Medium	Large
1	0.1	0.3	0.5
2	0.07	0.21	0.35
3	0.06	0.17	0.29
4	0.05	0.15	0.25
5	0.05	0.13	0.22

Generalised eta squared (η^2_G) - Generalised eta squared (η^2_G) is a modification of the effect size η^2 which takes in to account variance that may result from within subject and between subject random effects. It is recommended for repeated measures designs and those with mixed or nested designs. η^2 is ordinarily defined as

$$\eta^2 = \frac{\mathrm{SS}_{\text{effect}}}{\mathrm{SS}_{\text{total}}} \tag{3.8}$$

However, factorial designs are recommended to use partial eta squared (η^2_P), which includes sums of squares from interaction terms in the denominator. As mixed effects models are used in the body of this work, η^2_G is used. This is defined as

$$\eta^2_G = \frac{\mathrm{SS}_{\text{effect}}}{\delta \times \mathrm{SS}_{\text{effect}} + \mathrm{SS}_{\text{measured}}} \tag{3.9}$$

Where δ is a dummy variable with value of 0 or 1 indicating whether a factor was manipulated within a nested group (1) or between a nesting factor (0) meaning that between nested groups individual differences can be taken in to account [194].

Freeman's theta (θ) - Freeman's θ is a measure of effect size that is used in conjunction with non-parametric tests for ranked data. It is defined as a measure of association between categorical factors and a ranked numeric variable. It identifies the presence of systematic regularity in order values between classes [195]. The suggested interpretation of this statistic is that if $\theta = 0$, there is stochastic equality between classes. If $\theta = 1$, this is an indication that there exists classes which contain ranked values which are consistently higher or lower than other classes in the comparison. Freeman's theta can be calculated using the equation

$$\theta = \frac{\sum_{i<N} \sum_{j<k} \Delta_{ij}}{1/2(k(k-1))} \tag{3.10}$$

where $\Delta_{ij} = |L_{ij} - L_{ji}|$, the absolute value of differences between ranked observations L paired across i data points in the j^{th} group of k factors.

3.1.6 Principal component analysis and dimensionality reduction

Factor analysis is a group of techniques which are used to determine if a large set of variables are related or independent, the results of which can be used to build a model of a process from a group of speculatively selected variables (exploratory factor analysis) or to validate an existing model (confirmatory factor analysis) [189]. Exploratory techniques include principal component analysis (PCA), multidimensional scaling (MDS), partial least-squares regression (PLS) and stepwise model selection. Principal component analysis is a technique that is used to reduce the dimensionality of a dataset and to produce orthogonal bases which describe the majority of the variation in a sample within a reduced number of principal components. Each k^{th} component is an eigenvector corresponding to the k^{th} largest eigenvalue λ of the covariance matrix of the input variables. This eigenvector is equivalent to a linear combination of m input vectors and weights such that $\boldsymbol{\alpha_k X} = \sum_{j=1}^{m} \alpha_{jk} X_j$ where $\boldsymbol{\alpha}_k$ represents a vector of loadings of input variables on to the derived orthogonal components. [196]. Where PCA uses a singular value decomposition of a covariance matrix, MDS is a technique which uses the piece-wise distances between the co-ordinates of the input variables on a hyper-plane of M dimensions of input variables as its initial stage. As with PCA, the matrix of distances are subjected to singular value decomposition. In the case of Euclidian distance being used as the initial method for transforming the data, the output of MDS is equivalent to

PCA. However, it may be appropriate to use other methods such as Manhattan, great circle or mean squared error. The selection of distance metric will affect the outcome of the analysis and may be beneficial or detrimental to the interpretation of the results, constituting an important *a priori* assumption about the data [197]. Partial least squares (PLS) regression is a technique which has been used in the construction of models of spatial quality perception [198][199] and is a useful tool for reducing a large number of independent variables to predict one or more dependent variables. PLS uses the covariance matrices of both independent and dependent variables to provide vectors of coefficients to give linear functions which are predictive solutions which can be assessed with the usual tools of regression analysis. It is possible to achieve a similar goal using principal component regression, using an extracted feature space as independent variables for regression analysis, however this is limited to a univariate output [196]. For the purposes of this work, PCA and principal component regression was used in the analysis of responses. PCA was selected over PLS as the methodology used in the analysis was heavily focussed on the relationships between responses. As noted, PCA and MDS can give equivalent outputs given the assumption of Euclidian distance in MDS. Given this fact, the selection of the technique for the identification of relationships between variables and individuals is somewhat academic. The use of principal components in regression analysis, however, is advantageous in that where collinearity between independent variables is observed, significant vectors extracted by PCA can be used as predictors in regressions. As such, PCA was selected as the dimensionality reduction technique for the analysis of responses in this work. This is used in conjunction with stepwise approaches for model selection where it was not deemed appropriate to enter variables into PCA. Stepwise model selection is used to remove redundant terms and obtain an optimal model from a large number of predictors by iteratively adding or removing independent variables [200] and assessing change in goodness of fit using a penalised information criterion such as AIC [186]. Information criteria assess model fit with a penalty for number of terms in order to prevent over-fitting. This technique is widely used [177][201][108][202], and is a convenient way to eliminate non-significant or collinear terms. However, it lacks the ability to discern latent structure within a dataset and is used only where elimination of terms is required.

3.1.7 Signal detection theory

Signal detection theory (SDT) measures were used to evaluate whether participants experienced the rendered audio as truly realistic [18]. The sensitivity metric d' is defined as the difference between the inverse z score transforms (Φ^{-1}) of the probability of a true positive and the probability of a false positive, as in equation 3.11 where $H = \frac{S_{correct}}{S}$ and $F = \frac{N_{incorrect}}{N}$ with S denoting the number of signal trials and N denoting the number of noise trials. For the purpose of this study, signal trials are taken to be instances where participants are exposed to simulated audio through headphones and noise trials are instances where audio is emitted from loudspeakers. This quantity essentially expresses the difference in probabilities in units of standard deviations.

$$d' = \Phi^{-1}(H) - \Phi^{-1}(F) \tag{3.11}$$

d' is sensitive to high levels of bias in respondents. Bias (c) (Equation 3.12) is an independent quantity at low values, but high values for this statistic are associated with anomalously high d' [203]. Bias may refer to a perceptual bias, such as the erroneous perception of a signal, or a response bias. However, this statistic cannot differentiate the two [204]. Using these two statistics, it is possible to estimate the ability of an individual to distinguish two stimuli and identify their behavioural bias in the case of perceptual ambiguity.

$$c = -\frac{\Phi^{-1}(H) - \Phi^{-1}(F)}{2} \tag{3.12}$$

A more complete discussion on SDT can be found in Green & Swets [10] and Stanislaw and Tordorov [203].

3.2 Signal processing and audio rendering

Digital signal processing was designed to provide a synthetic binaural room impulse response (BRIR) for the real time spatialisation of audio in a VR test environment. Signal processing was based on the naive image source method as described by Allen [59]. HRTFs were sourced from the MIT KEMAR data set [205]. The plug-in was

developed and prototyped in managed C# in the scripting layer of Unity Editor, with the signal processing modules subsequently ported to native C++ with the image generation and parameter calculation remaining in the scripting layer (Fig 3.1).

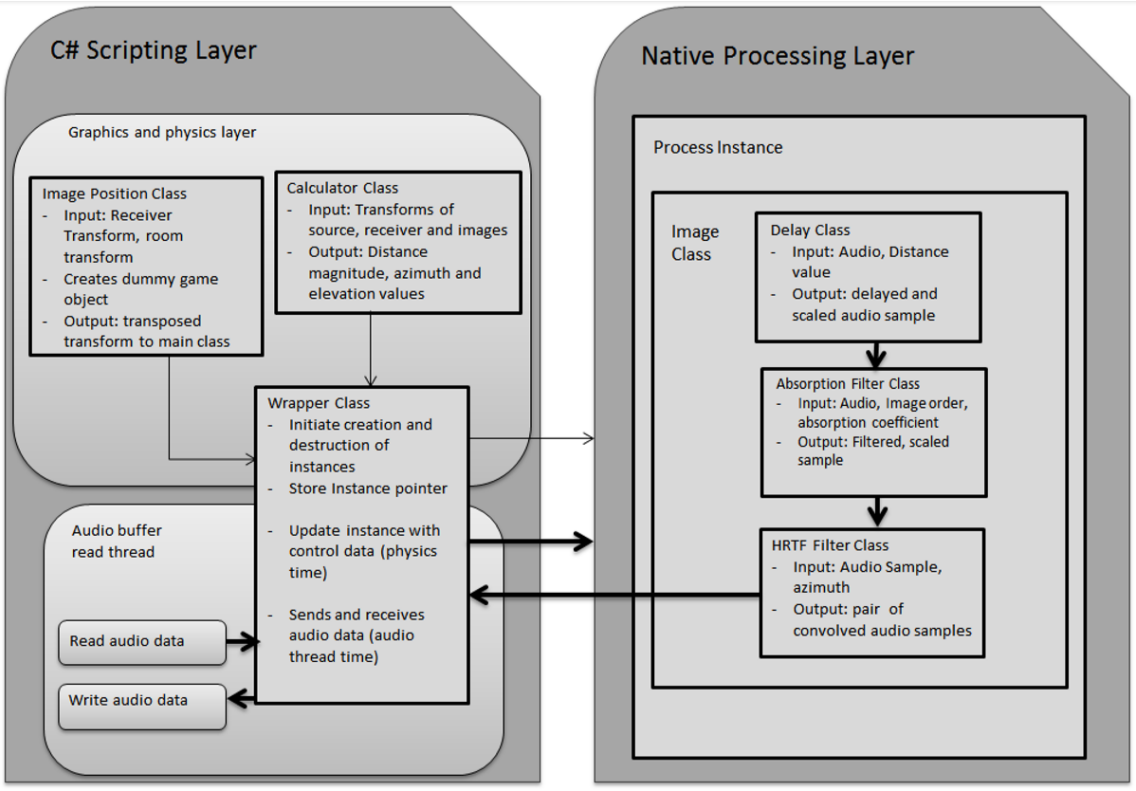


Figure 3.1: Structure of Signal Processing Used in the Native Processing Version of the Spatialiser

Figure 3.2: Large reverberant room (5s RT_{60})

Signal processing was designed to apply discrete spatialisation on each reflection image. Image prioritisation and maximum order length can be configured at build

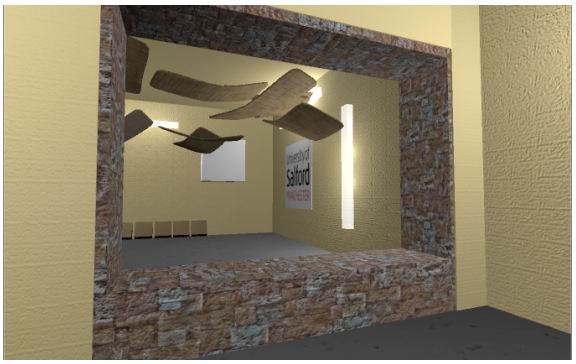


Figure 3.3: VE recreation of large reverberant room (5s RT_{60})

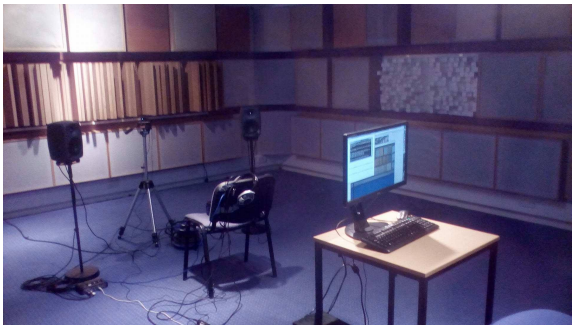


Figure 3.4: Medium RT_{60} room (270ms RT_{60})

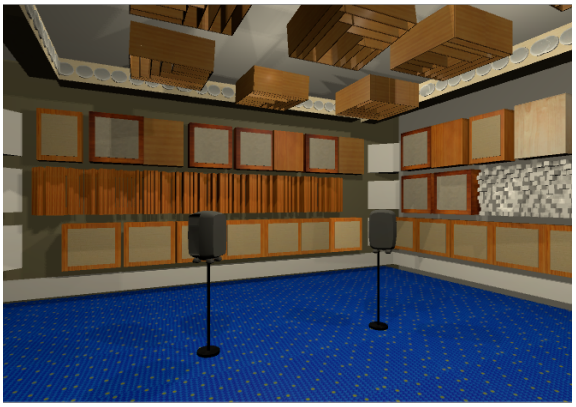


Figure 3.5: VE recreation of medium RT_{60} room (270ms RT_{60})

time. Inclusion of height rendering, azimuth rendering, order truncation and mean surface absorption can be configured at runtime. The design is such that implementation of further granular parameter control is possible and limited only to the requirements of the project. The synthesized room response was generated in such a way as to allow for the observation of the effect of early reflection components and late reverberation independently. Spatial audio processing used in experiments in this thesis was performed using a hybrid method, after Heinz [62], with HRTF processing and early reflection components generated using an image source model controlled by the transform of the participant avatar and the geometry of the virtual

Figure 3.6: Low RT_{60} room ($90ms$ RT_{60})

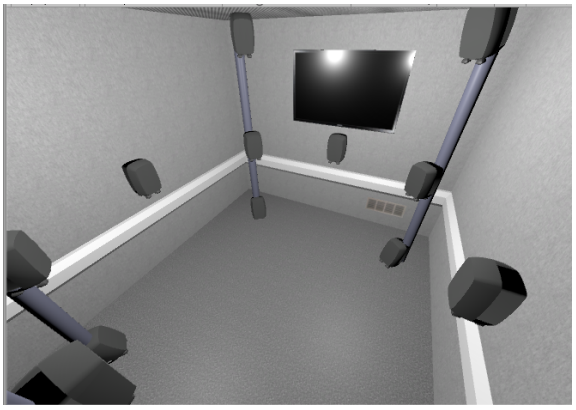


Figure 3.7: VE recreation of low RT_{60} room ($90ms$ RT_{60})

space, and late reverberation produced using the included reverb processor in Unity [206], configured using the geometry and apparent surface properties of the virtual space. A script was also written to configure the Unity Editor reverb plug-in based on the parameters of the image source based DSP to produce appropriate decay [207] and mixing times [47, 97] when late reverberation is required. Reverberation time was calculated using Sabine's equation with mean surface absorption, again, calculated from the apparent surface material textures applied to the boundaries of the room or to match known reverberation times where known. The mixing time of the discrete and diffuse components was determined by the mixing time prediction equation described by Lindau [208] [Equation 3.13].

$$t_{mp50} = 20 \cdot V/S + 12 \tag{3.13}$$

These parameters were applied to the decay time and reverb delay parameters of the Unity reverberation effect. In conditions where pre-exposure and virtual environment geometry were similar, absorption parameters were computed from the known decay time value of that real environment. HRTFs were sourced from the KEMAR compact dataset with both azimuth and elevation and processing was implemented

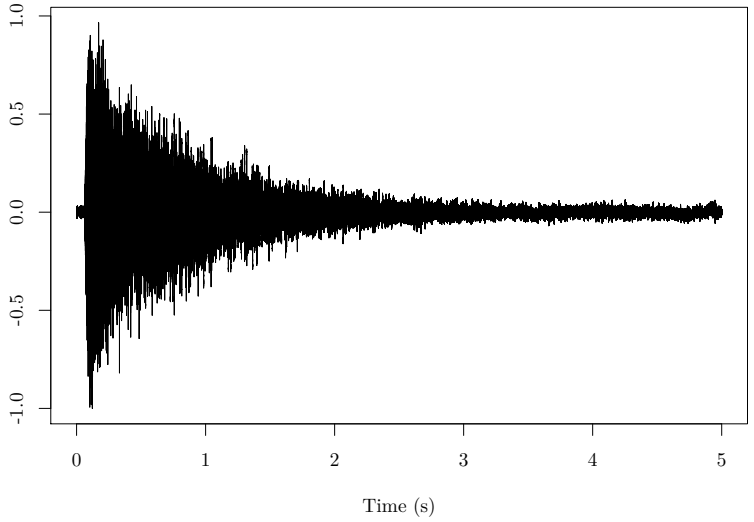


Figure 3.8: Impulse response of large, long ($5s$) RT_{60} space shown in figure 3.2

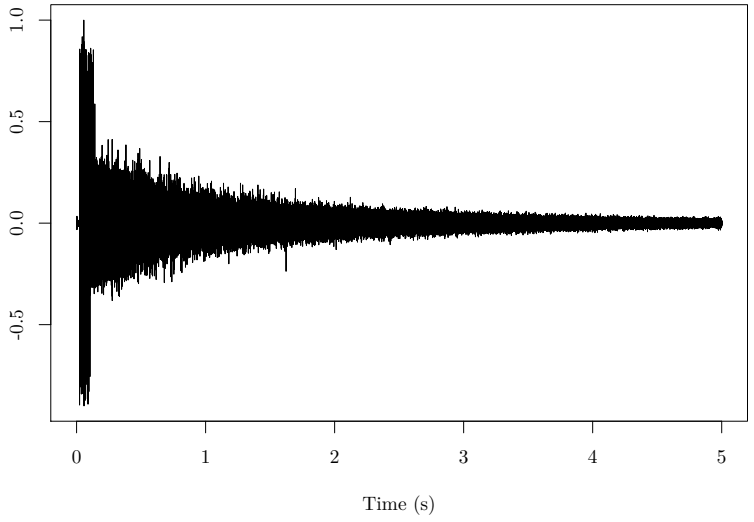


Figure 3.9: Modelled impulse response using parameters of large, long ($5s$) RT_{60} space shown in figure 3.2 used with virtual environment shown in figure 3.3

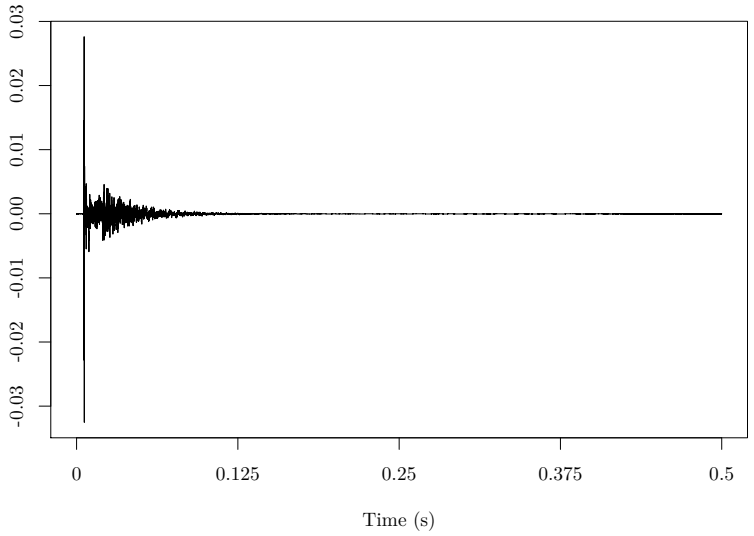


Figure 3.10: Impulse response of medium ($270ms$) RT_{60} space shown in figure 3.4

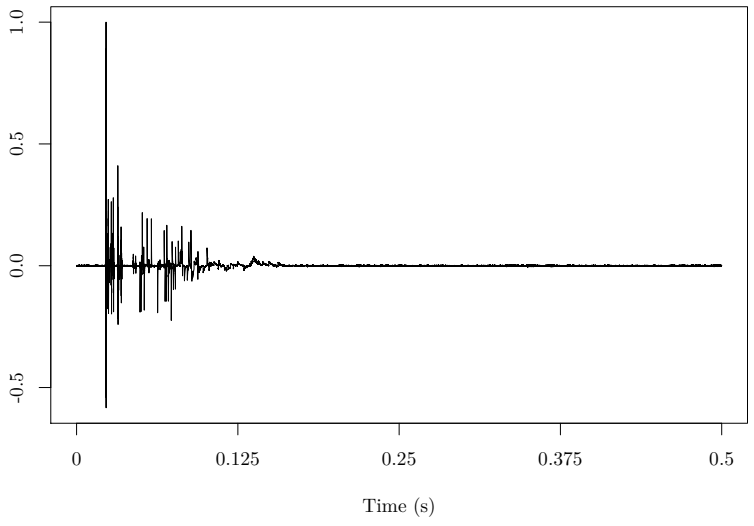


Figure 3.11: Modelled impulse response using parameters of medium ($270ms$) RT_{60} space shown in figure 3.4 used with virtual environment shown in figure 3.5

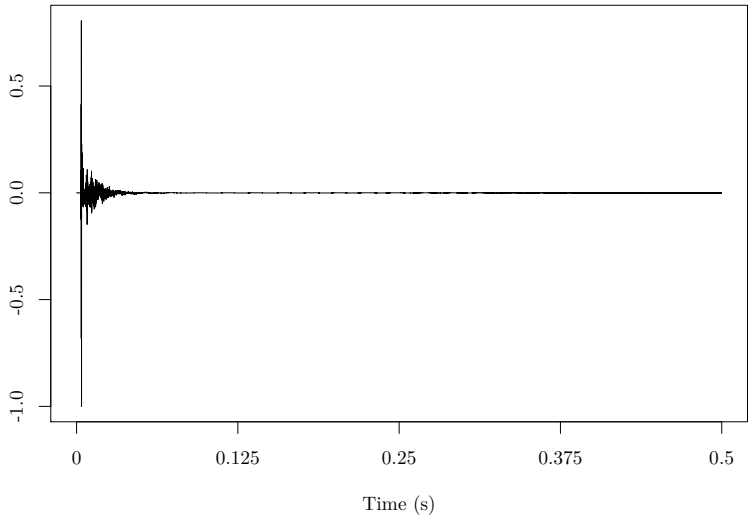


Figure 3.12: Impulse response of small, short ($90ms$) RT_{60} space shown in figure 3.6

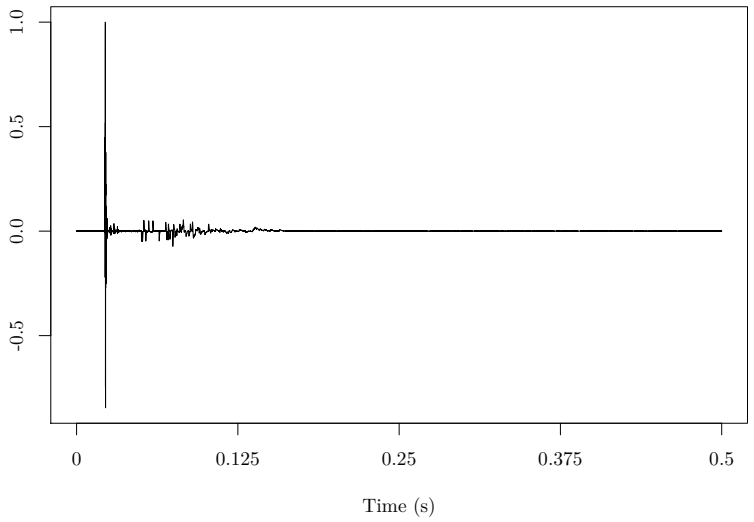


Figure 3.13: Modelled impulse response using parameters of small, short ($90ms$) RT_{60} space shown in figure 3.6 used with virtual environment shown in figure 3.7

in C++ using the Unity native audio processing API. Participants completed the tasks under the following conditions either with or without late reverberation applied to the audio:

- No spatialisation
- HRTF processing applied to direct sound
- HRTF processed direct sound plus 1st order discrete reflections
- HRTF processed direct sound plus 1st and 2nd order discrete reflections
- HRTF processed direct sound plus 1st, 2nd and 3rd order discrete reflections

Figures 3.2, 3.3, 3.4, 3.5, 3.6 and 3.7 show comparisons of real and virtual environments used in this work for geometrically similar conditions. Comparisons between the impulse responses obtained for real rooms and rooms modelled using the method described above, without spatial processing applied, are shown in figures 3.8, 3.9, 3.10, 3.11, 3.12, and 3.13. It can be seen that this method of simulation does not produce impulse responses which match those which were obtained from deconvolution of chirp signals in equivalent real spaces, particularly for shorter reverberation times. Differences in the relative gains of the early and late components can be observed between real and equivalent modelled responses. In addition, the implementation of the image source method described above does not take in to account diffusion. The large and medium sized rooms used in this study were both treated with diffusive structures to maximise scattering of early reflections. This effect is not reproduced in the model used in this work and is particularly evident in the impulse response shown in figures 3.10 and 3.11. In the measured response, early reflections are almost completely suppressed and a dense, diffuse component follows after a visible gap in energy over time. In the case of the modelled response, decay time appears comparable. However, the density of the response is lower, and individual reflections can be seen until $100ms$. Similar discrepancies in figures 3.12 and 3.13 in terms of energy distribution can be seen. In the medium RT_{60} case, the modelled reflections are too energetic, despite mean absorption for being calculated from the target decay time and room volume. The low RT_{60} environment suffers from the reverse issue. The mean absorption calculated by solving the Sabine equation using the given volume, surface area and target decay time has produced early

reflection components with far too little energy. Additionally, the mean free path equation for perceptual mixing time has greatly overestimated the pre-delay required for the FDN used for late reverberation, resulting in a gap between the early and late parts. The responses shown in figures 3.2 and 3.3 showing impulse responses for the $5s$ RT_{60} environments demonstrate a dramatic difference in energy of the late component. This was attributed to the default output of the FDN reverb used in unity. Although it can be seen that the IR modelling used failed to produce similar responses to those of the room, signal processing was not adjusted to more closely match real world conditions. This decision was taken for two reasons, the first was ecological validity. A review of the tools available for implementing spatial audio in VR applications was performed. The tools reviewed were commercially available products maintained in June 2016. Processors were all measured using sine sweep deconvolution [209] and measurements were taken to determine change in impulse and frequency response with respect to azimuth and distance. Measurement was performed on the following spatialisers:

1. Oculus Native Spatializer
2. Impulsonic Phonon 3D
3. Google Cardboard Audio SDK
4. TwoBigEars 3Dception
5. VisiSonic RealSpace3D

Processors 1 and 2 employ HRTF processing only and appear to utilise different HRTF sets. Inspection of the dependencies of spatialiser 2 reveal that HRTFs from the CIPIC database are used [210]. Processors that employ simulated acoustic responses (3-5), all use cuboid bounding boxes to define the space to be simulated and generate discrete, identifiable reflections (Fig 3.15). Late reverberation is controllable independently of room geometry, via API or GUI. These features suggest the use of the image source method for the generation of the early reflection component of the synthetic BRIR [59]. In the case of Spatialiser 3, reducing the quality parameter changes the azimuth dependent frequency response of the input signal. The 'medium' setting appears to filter with lower spectral resolution than the 'high' setting. Measurement of the 'Eco' setting suggested a simple low pass filter, applied to direct and reflected signals, whose cut off frequency is dependent on emitter azimuth [Fig 3.16].

From these analyses, it can be determined that spatialisation, where it is implemented beyond simple filtering, is likely based on the image source method. Spatializers 3 and 4 render reflections to a very low order [Fig. 3.14]. However, spatializer 5 can produce reflections up to 10th order, dependent on surface reflectivity settings [Fig. 3.15].

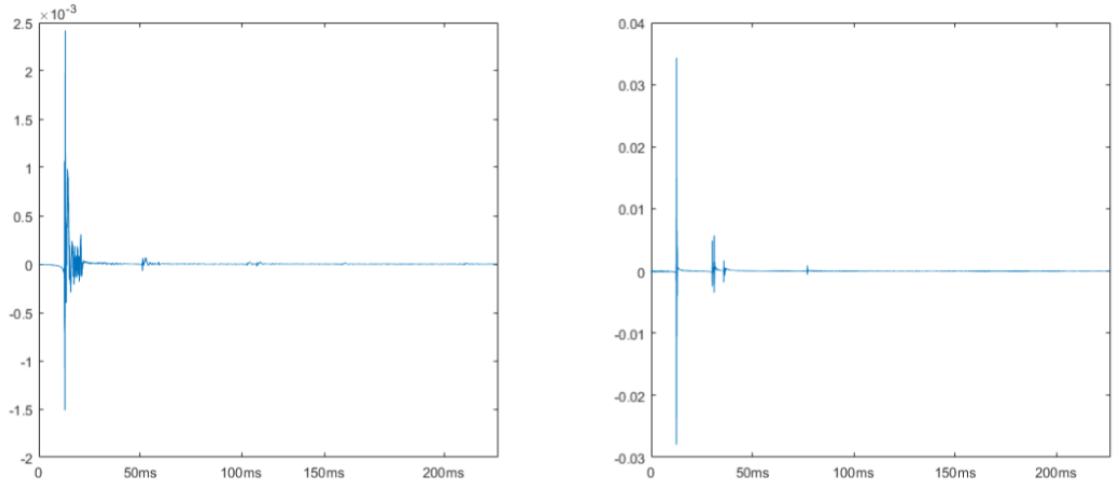


Figure 3.14: Impulse responses for Google Cardboard "High Quality" (left) and TwoBigEars 3Dception (right). Source at 1m.

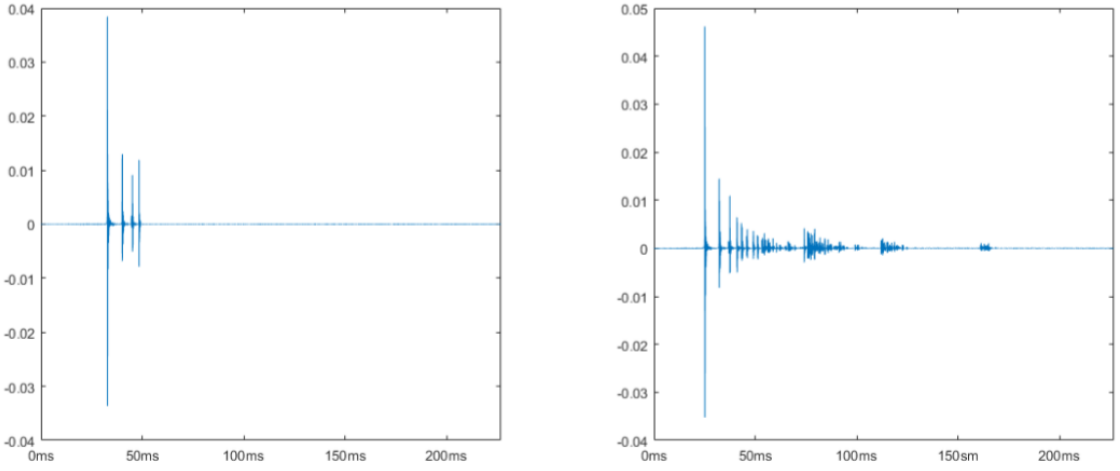


Figure 3.15: Impulse responses for RealSpace3D, 1st order maximum (left), 7th order maximum (right). Source at 1m.

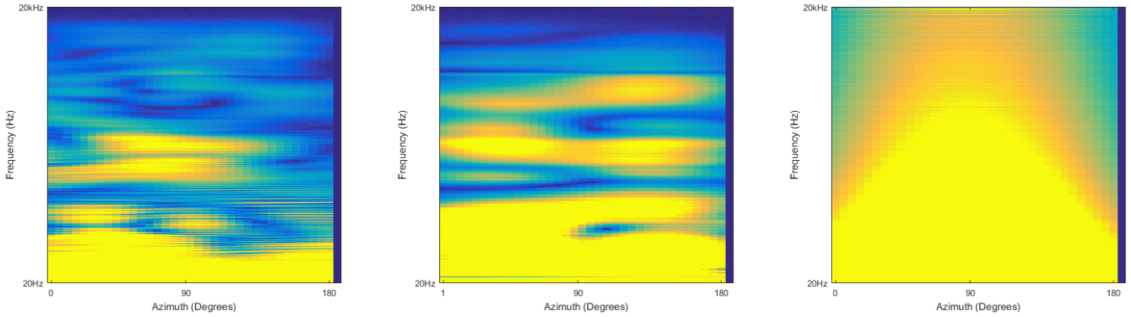


Figure 3.16: Ipsilateral hemisphere frequency response by angle of Google Cardboard Audio SDK using 'High Quality' (left), 'Medium Quality' (centre), and 'Eco Quality' (right)

As discussed in section 2.1, realistic simulations are often used in research. However, it is clear that limitations in hardware and the availability of low realism audio simulations in active use supports the use of this class of processor.

Secondly, while in using imperfect simulations it could be hypothesised that all participants would rate stimuli low on scales relating to realism, such unrealistic stimuli offer range within response scales to identify individuals or groups who rate low realism stimuli highly. The use of physically perfect simulations might obfuscate patterns in the collected data by hiding the differences in the way that individuals respond. It is not within the scope of this study to evaluate or validate the signal processing methods used, but to evaluate the human participants who are exposed to the simulated environments. If all participants respond similarly to stimuli with a high degree of physical accuracy, the aims and objectives set out in section 1 would not be served and conclusions might be drawn on type II errors.

PART II. EXPERIMENTAL WORK

4

Quality of Experience in HMD VR Environments an its Dependence on Spatial Audio Modelling

Contents

4.1 Introduction

This chapter will cover experimental work conducted to assess the relationship between five factors that were identified as salient in the assessment of immersive audiovisual scenes (section 2.1). These factors will be investigated in the context of virtual environments with divergence between pre- and post- exposure geometry and similarity between pre- and post- exposure geometry. Work will also be described in assessing the differences between responses given when subjects believe stimuli to be real or simulated. Finally, these factors will be contrasted with extant measures of presence to identify any relationships between these scales and the experience of presence. The work in this chapter focusses on the modulation of responses to these questionnaire items by changes to impulse response content using a simplified, computationally efficient, spatial renderer designed to replicate indoor space and the effect of changing the level of explicitness in the spatial relationship between visual and auditory stimuli.

4.2 Impulse response simulation and visual co-location and subjective response in dissimilar virtual environments

4.2.1 Materials and methods

Virtual environment - To facilitate the collection of data, a virtual environment was constructed in Unity 3D [206] and presented using an OSVR HDK 1.6 headset. A test area of 40m x 16m x 7m was built and textured with materials that were deemed to have familiar acoustic properties and visually set an expectation of acoustic response for participants. Wall surfaces were textured with bare brick, flooring was textured as metal tread plate and the ceiling was textured as untreated wooden planks. Absorption coefficients for these materials were obtained from Vorlander [55] and used to calculate RT_{60} and discrete reflection gains in the

image source model described below. Participants were situated on an elevated pedestal 1.5m from the ground and the HMD viewer was positioned at a sitting height of 1.2m. Emitted audio was either male speech, female speech, repeated noise bursts or a synthesized sound effect. Audio was randomly selected at runtime. During training and test phases, instructions were given to participants in-game by a diegetic source in a fixed position within the game. This was designed to allow for direction of participants with minimal external interaction and also to act as an auditory anchor, giving participants a reference for loudness and DRR of sources at a given visual distance for each treatment. Audio levels were calibrated so that human speech samples had a nominal loudness of 68dBA from an emitter at 1m from the audio listener object. This level was obtained through informal experiment. A male volunteer was measured speaking at a comfortable volume at 1m within an anechoic chamber. The room response was synthesised as described in section 3.2. For this study, spatial response data, in the form of estimated locations of rendered audio ere discarded and not analysed. Tasks were performed under three levels of ambiguity of the spatial relationship between visual an audio objects. Low ambiguity conditions used spatially co-located audio and visual objects. Medium ambiguity stimuli had multiple visual objects with only one object being associated with an audio source. The ambiguity introduced being that multiple plausible sources could be the emitter. High ambiguity had audio and visual objects dislocated in azimuth and depth with no spatial association between events.

Low ambiguity (LA) task: Single co-located source - Participants were drawn from the staff and students of the School of Computing, Science and Engineering at the University of Salford. 30 participants (25 male, 5 female) took part. The task required focusing on and estimating the distance to target objects. Participants were instructed to use head movements to centre a reticule on targets and distance was then estimated by holding a button on a controller for a length of time to determine a corresponding estimation of distance. Training was given in the form of two fixed position objects, at 5m and 10m, which participants were allowed to aim and range at. Visual and auditory feedback were given when a correct aim and distance estimation was made. This training was open ended and the decision to continue to the trial phase was left to the participant. There was no minimum time enforced, but the participant was briefed not to skip this phase. Figures 4.1 and 4.2 show an image of the virtual environment used and a schematic of the virtual environment. During the trial phase, spheres appeared in random locations within the

space and participants were instructed to attempt to destroy the objects by aiming and ranging. Target objects emitted audio continuously while instantiated. Objects would remain in place for ten attempts or until a correct estimation was made. Ten targets were instantiated for each participant within each treatment condition. Training was repeated for each treatment.

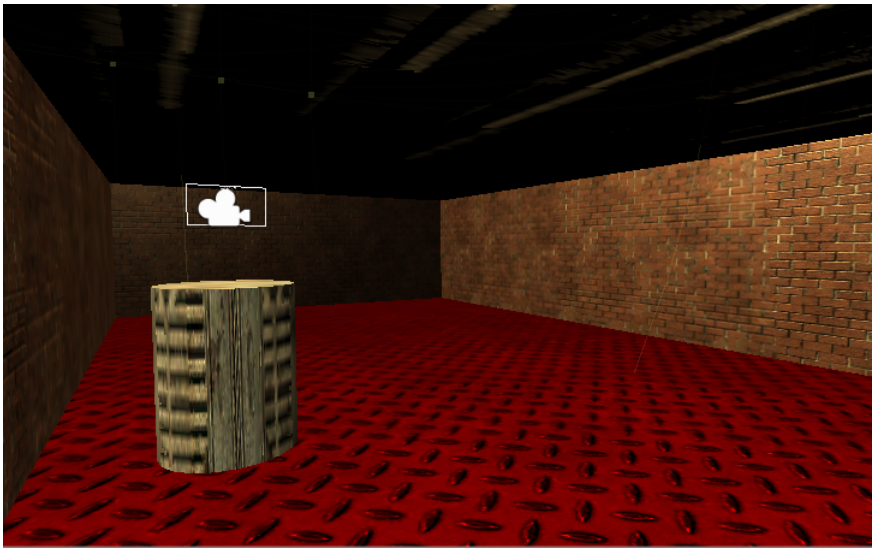


Figure 4.1: Image of the unreal (dissimilar) test environment

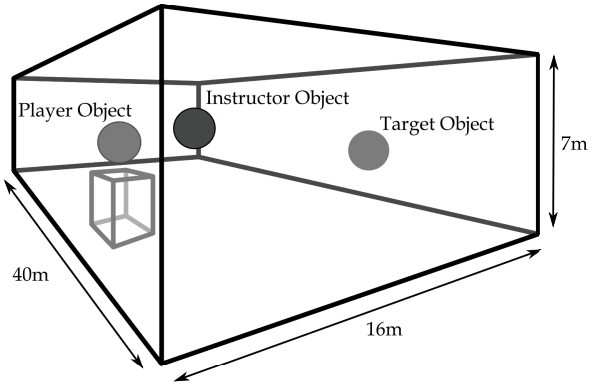


Figure 4.2: Schematic view of test environment

Target positions, shot positions and player position and orientation was recorded for analysis.

Medium ambiguity (MA) task: multiple visual source, single co-located audio source - Participants were drawn from the staff and students of the School of Computing, Science and Engineering at the University of Salford. 20

participants (15 male, 5 female) took part. No training for the task was given, only verbal explanation of the task from the researcher and in-game from a diegetic source. Participants who felt that they had not understood the task at the outset were allowed to restart their first set and data for that set was not recorded. The method was a VE adaptation of a method to test the proximity image effect in anechoic and reverberant environments [211] [212], modified so that participants could see all potential audio sources, by lowering the sources to 1m below participant eye level. Participants were presented with five co-linearly arranged cubes radially oriented from the participant. Audio was rendered from a position that corresponded to one of the locations of three possible objects (Figure 4.3) The actual rendering position was selected in real time at random from a 7×15 grid of possible rendering points between 4 - 18m in distance. The limitation in distance was designed to allow for any number up to the maximum of incorrect options to be positioned between the participant and the correct stimulus position. Participants were instructed that any of the five cubes would be the object emitting sound. However, only the middle three objects would be used as valid rendering positions. This was intended to allow for either under estimation or over estimation of all possible trial positions. Participants were instructed to select the sound emitting object by highlighting the object using a game-pad controller. Each participant completed twenty two repetitions of the task for each of the five audio conditions in which they were exposed.

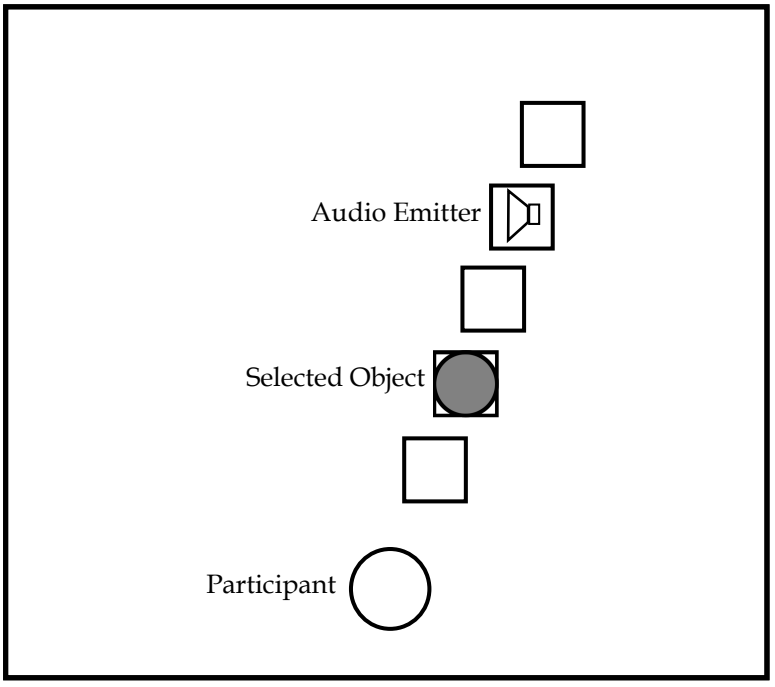


Figure 4.3: Schematic view of multiple possible source, single co-located audio emitter environment

Audio stimuli emitted in the VE were the same audio samples used in the experiment described in section 4.2.1. Objects were oriented radially at random angles between ±60 at random minimum distances between 1m and 10m and with random spacing between 0.5m and 2m. Participants completed the task impulse response conditions as described in section 4.2.1, with two groups of 10 completing the task with or without reverberation and all participants completing the task in all early impulse response conditions.

High ambiguity (HA) task: Single spatially separated sources - Participants were drawn from the staff and students of the School of Computing, Science and Engineering at the University of Salford. 20 participants (14 male, 6 female) took part. As in the MA task, no training portion was given. However, participants who felt that they had not understood the task at the outset were allowed to restart the task and data for that set was not recorded. Participants were presented with a single visual object and an unseen audio source emitting sound within the space. Audio rendering position was selected in real time at random from the grid of possible positions used in the MA task. The visual object could be moved radially from the participant using the thumb controls on a game pad, with the object following the orientation of the participant's head in azimuth and elevation. Participants were instructed to identify the location of an unseen audio source and position the object so that it was co-located with the audio source. Once the object was positioned as intended, a button was pressed on a game pad and another object/audio source pair was generated with random positions in the space. Each participant completed twenty two repetitions of the task for each of the five audio conditions in which they were exposed.

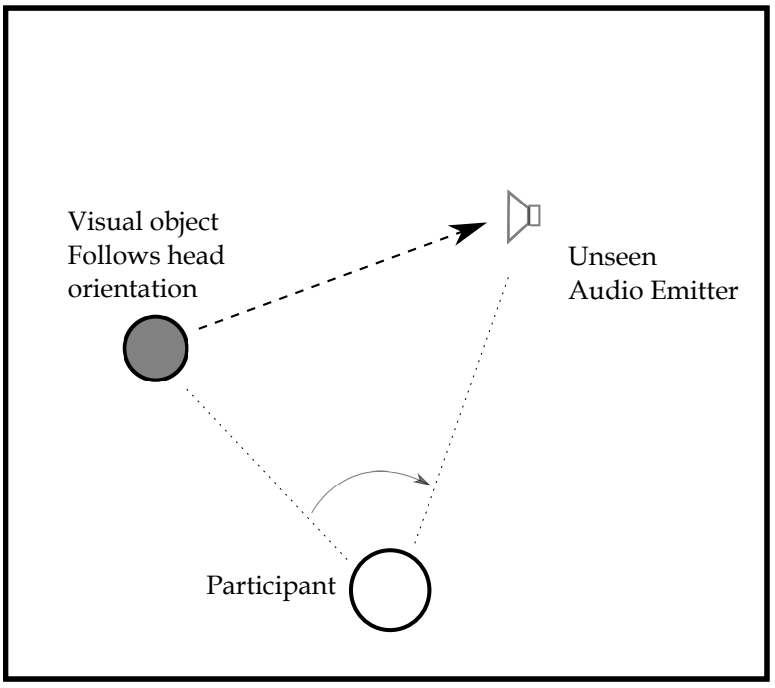


Figure 4.4: Schematic view of Single spatially separated source emitter environment

Self reported audiovisual quality ratings - After each condition, participants were asked to complete a questionnaire about their audiovisual experience while performing the task. To limit participation time, the questionnaire was limited to five items asking participants to report the extent to which:

- Audio appeared to 'belong' to the objects in the room (audiovisual fusion)[16]
- Audio was plausibly emitted in the room (environmental plausibility)[18]
- Audio was easy to localise in the VE (localisation)[19, 20]
- Audio was experienced as outside of the head (externalisation)[17]
- The participants were conscious of sound being emitted from headphones as they performed the task (awareness of headphones)[21]

Responses were collected using a 7-point Likert scale ranging from "not at all" to "fully". In the case of the HA task, participants were asked to respond to questions referencing the point at which they had made the decision that the visual object and audio source were co-located.

4.2.2 Results

PCA of questionnaire responses -

Questionnaire responses were subjected to exploratory factor analysis by way of principal component analysis [PCA] to reduce the dimensionality of the data to facilitate simplification and further analysis. PCAs were performed using the FactoMineR package in R. The analysis suggests that the most of the variance within the response sample can be described using two components. Components one [PC1] and two [PC2] are the only components with Eigenvalues over 1 and explain 73.4% of the variation in responses [Figure 4.5]. The third component in this solution was inspected but did not appear to contain meaningful information. AoH and Ext are both positively loaded on to this component and all other factors have negligible loadings onto this dimension. As the third component is after the point of inflection on the scree plot and has an eigenvalue of <1 and provides meaningless data, the two component model was selected.s Variable loadings for dimensions 1 and 2 are presented in figure 4.6 and table 4.1. The first component has high loadings for environmental plausibility, audiovisual fusion and perceived localisation. Externalisation is loaded with large effect on to the positive sign of component one and with medium effect on the negative sign of component two. Awareness of headphones has almost the inverse loading.

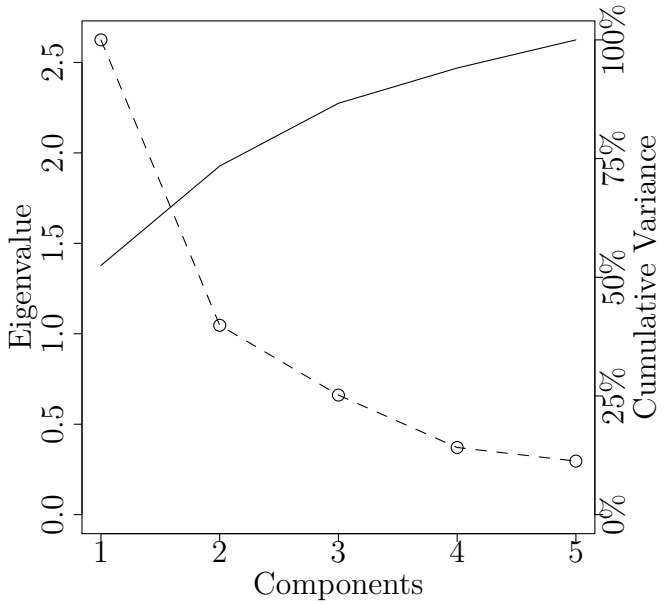
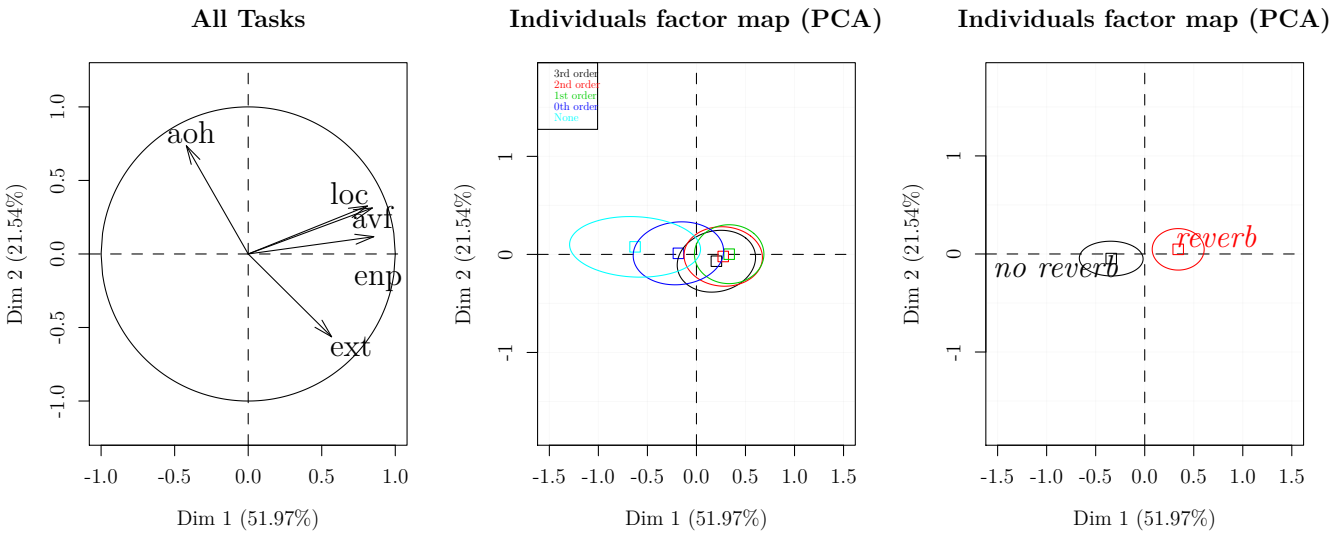


Figure 4.5: Scree plot and cumultive variance of PCA of all questionnaire responses



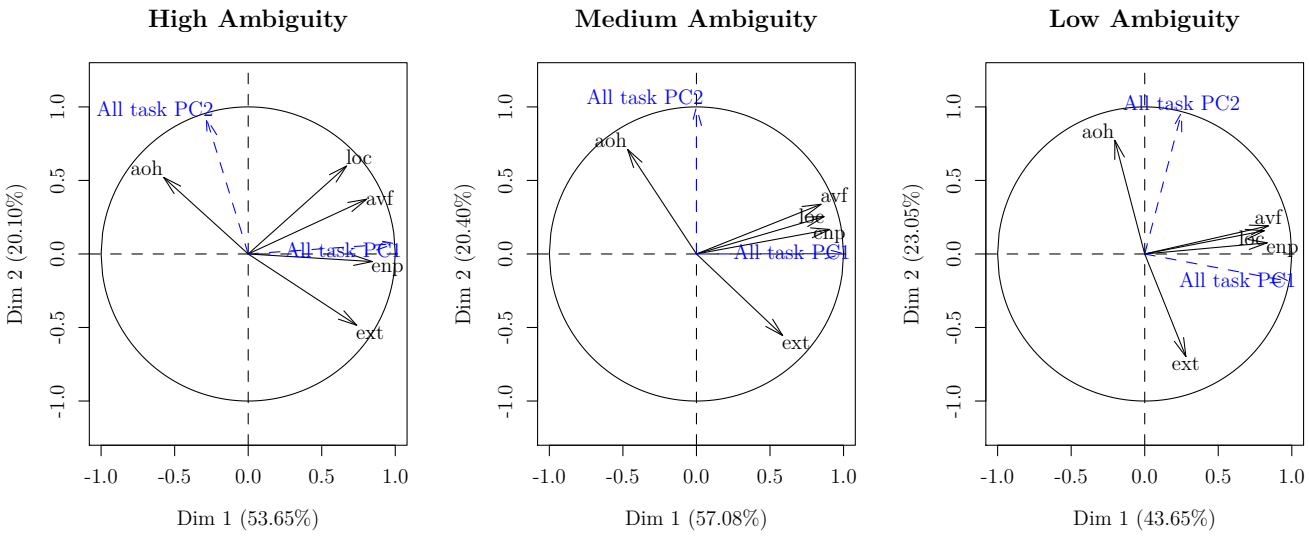
Quality of experience:

aoh - awareness of headphones avf - audiovisual fusion

enp - environmental plausibility ext - externalisation

loc - localisation

Figure 4.6: PCA factor map of all questionnaire responses - Barycentres for individuals for early IR and reverberation conditions.



Quality of experience:

aoh - awareness of headphones avf - audiovisual fusion

enp - environmental plausibility ext - externalisation

loc - localisation

Figure 4.7: PCA factor map of questionnaire responses performed by audio-visual co-location task. Factor loadings for all combined tasks are plotted as a supplementary variable (not entered into PCA)

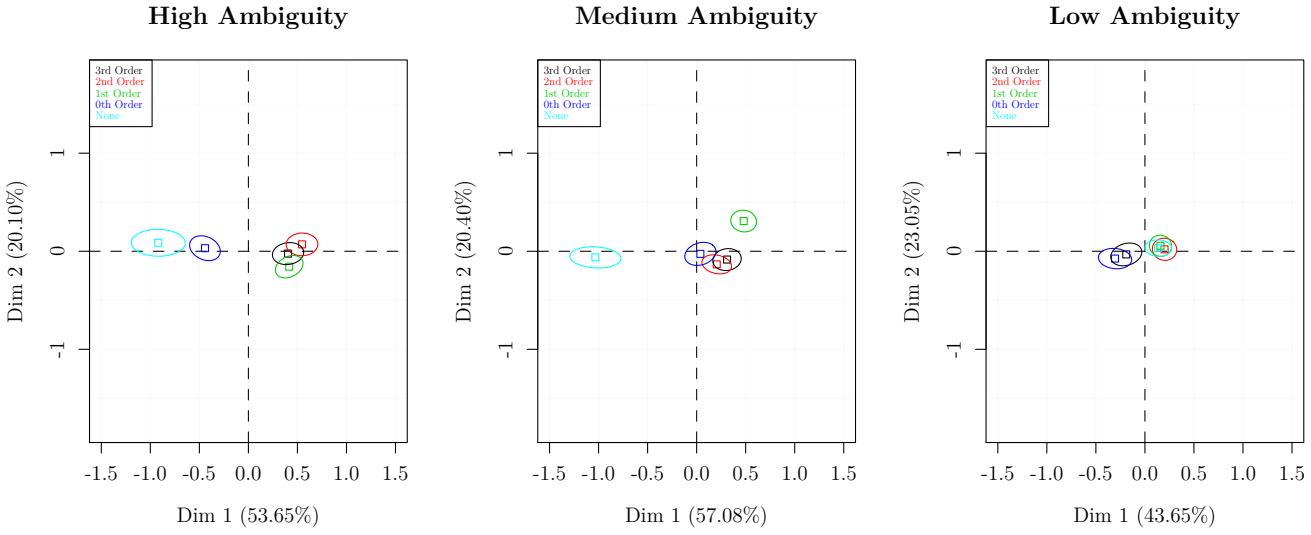


Figure 4.8: Barycentres of individuals on extracted feature space, separated by co-location ambiguity and early impulse response content

Table 4.1: Loading table for PCA of questionnaire responses

	Dim.1	Dim.2
avf	0.86	0.24
enp	0.79	0.15
loc	0.78	0.17
ext	0.34	-0.70
aoh	-0.29	0.73

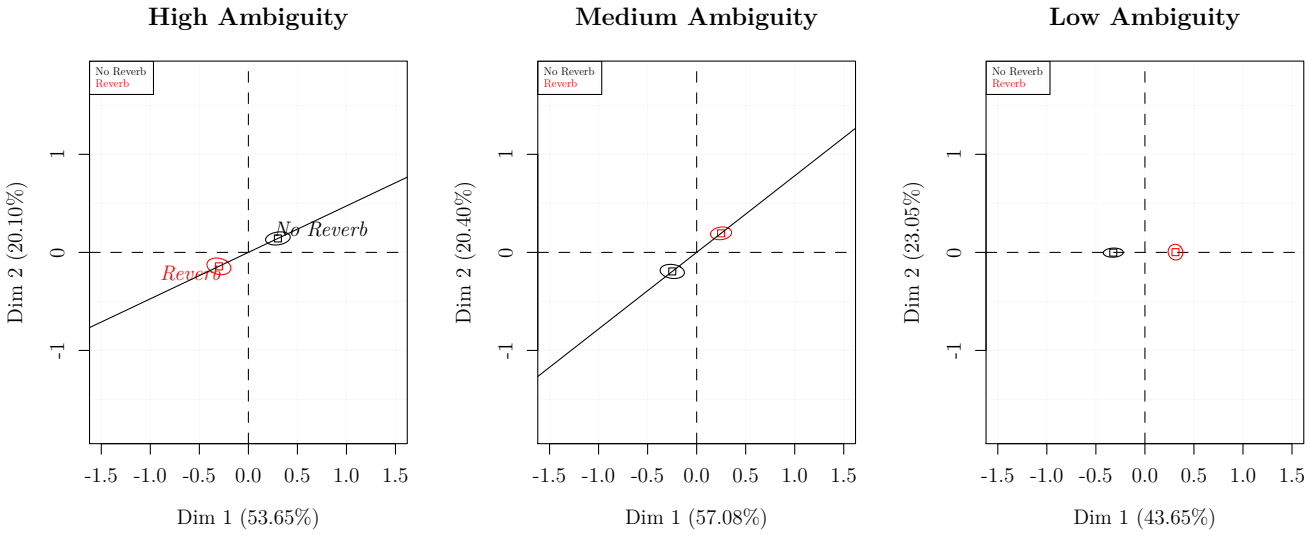


Figure 4.9: Barycentres of individuals on extracted feature space, separated by co-location ambiguity and late reverb presence

PCAs were performed on each task group response set separately. Factor maps are presented in figure 4.7. factor maps demonstrate increased correlation with environmental plausibility and perceived localisation and audiovisual fusion in the medium and low ambiguity tasks in comparison with the high ambiguity task. Similarly, as ambiguity decreases, awareness of headphones and externalisation become more independent of the other variables, but remain roughly inversely correlated as loading on to PC2 for each task set changes.

Difference in barycentre of individuals in the PCA by early impulse response conditions can be observed between tasks (Figure 4.8). When the relationship between audio and visual stimulus is ambiguous, the presence of reflections results in a positive response on the first dimension for that task. In the medium ambiguity case, it is the presence of a HRTF which results in a positive response on the first axis. In the third, low ambiguity task, there is no relationship between the position of

the barycentre of the response on the first dimension and the early impulse response treatment. With the inclusion of reverb (Figure 4.9), for medium and low ambiguity tasks, reverb produces a positive loading on dimension 1. This is reversed in the high ambiguity task. There is also a reversal on resultant loading on to dimension 2 between high and medium ambiguity between reverb conditions.

Wilcoxon signed-rank tests suggest that there is significant difference along PC1 in reverb conditions (All data: $p = 1.556 \times 10^{-12}$; LA, MA, HA: $p < 2.2 \times 10^{-16}$). ANOVA of PC1 scores on all data suggest significant difference between early impulse response treatments, but with small effect size ($F= 452$, $p = 0.0014$, $\eta_G^2 = 0.051$). Table 4.2 shows pairwise contrasts of PC1 for all data, p-values suggest difference between conditions with and without HRTF filtering.

Table 4.2: Pairwise comparisons (Wilcoxon signed-rank p-values) of PC1 by early impulse response condition with Bonferroni-Holm p-value correction

	3rd Order	2nd Order	1st Order	0th Order
2nd Order	0.63			
1st Order	0.41	1		
0th Order	9.7×10^{-8}	1.7×10^{-9}	1.1×10^{-13}	
None	5.1×10^{-6}	1.7×10^{-9}	1.4×10^{-8}	1

It can be shown that the effect of the presence or absence of HRTF filtering dominates the variance between audio factors and that the granularity between early reflection conditions is not significant to warrant this fine level of factorisation. As such, data were refactored into the following conditions:

- Anechoic audio: Audio is processed with HRTF filtering only
- Early component only: Audio is processed with HRTF filtering and spatialised early reflections up to 3rd order are included
- Late component only: Audio is processed with HRTF filtering and the stochastic portion of the impulse response is included after a delay corresponding to the perceptual mixing time
- Full IR simulation: Audio is processed with HRTF filtering and both early and late components are included

This factorisation was subjected to hypothesis testing using the Kruskal-Wallis test and was found to have no significant differences between factors Kruskal-Wallis chi-squared = 5.0944, df = 3, p-value = 0.165)

4.2.3 Discussion

Factor analysis of self reported metrics show two significant dimensions in the responses. The first dimension appears to correspond to the representation of the audiovisual objects within the space. The sense that the audio and visual components are a unified event, that the emission of audio is within the presented environment and that the auditory events have definite localisable position are all heavily loaded on to this dimension. This could be argued to be an extrinsic factor which depends on the composition of the simulated impulse response, as evidenced by the difference measured on this dimension when both reverberation and reflection are used as predictors and from the between task comparisons of the barycentric plots of the PCAs. The second dimension seems to represent the intrinsic response of the participant to the auditory component of the virtual environment. No significant difference was observed in this dimension for either reverberation or reflection order. These dimensions, an extrinsic representational component and an intrinsic reactive component, appear to correspond to dimensions of presence identified in the study of visual representation of VEs by Slater [21], place Illusion (PI) and plausibility (Psi); the former constituting the effect of stereopsis, orientation and positional tracking, and the latter an involuntary response related to both the overall effect of immersion and the individual predisposition to the experience of presence. The level of independence of these dimensions may be related to the level of ambiguity that is provided to the coherence of the audiovisual event. The change in loadings of the questions on to the extracted dimensions, and to each other, between tasks appears to show a transitional relationship as ambiguity in audiovisual coherence is changed. In the HA task, there is less inter-variable correlation, with most responses loading weakly on to orthogonal dimensions, with the exception of environmental plausibility. However, as ambiguity decreases, audiovisual fusion, environmental plausibility and localisation begin to correlate. In the LA task, we then see the percepts in the analysis above take form. This suggests that the assessment of spatial audio quality in VR has a strong visual component, and the independence of externalisation and perceived representational realism is determined by the strength of the visual cue. It is curious that, in the high ambiguity task, there is common loading

on to dimension 2 for awareness of headphones, localisation and audiovisual fusion. It must be noted that this cannot necessarily be interpreted as a loading on to the percept described above, but indicates a degree of correlation which is independent of other inter-variable correlations which load onto the first principal component. This notwithstanding, it can be demonstrated that awareness of headphones and externalisation maintain a very similar relationship between tasks and, to a great extent, remain consistent in the degree of correlation with the all task dimensions, which show progressive clockwise rotation as ambiguity of audiovisual congruence is decreased. In addition, in the high ambiguity condition, the orthogonality of the all task principal components, when plotted as supplementary variables, is slightly undermined.

The relationship between questionnaire responses and audio treatment can be seen to be a function of ambiguity of audiovisual coherence. In the HA task, positive bias is observed in responses loaded on to PC1 for treatments which include early reflection components. Negative bias is observed in treatments which have no reflections, with no reflections or HRTFs producing the greatest negative shift. In the MA task, a similar relationship is observed, however the HRTF only condition (0th order) is neutral on PC1. In the LA condition, there is no clear association between early impulse treatment and response to questions associated with PC1. This suggests that there may be an effect of ambiguity on the importance of early impulse response cues. In the case of HA, reflections produce a stronger representational response, however as ambiguity decreases, the association of the auditory stimulus to the apparent visual source dominates. The interaction between this parameter and reverberation is curious. In the HA task, reverb was associated with weaker localisation, belonging to the visual object and plausibility in the environment. However, in MA and LA tasks, this relationship is reversed. This may be due to the strength of the visual cue changing the weighting of participants' reliance on auditory localisation cues. Reverberation may be, to some extent, masking ILD and spectral cues, degrading localisation, which is loaded on to PC1. In the presence of some indication of audio source emission, this degradation of auditory localisation could be less perceptually relevant. That, as ambiguity decreases, localisation becomes more correlated with plausibility and audiovisual fusion, quality of experience metrics which are by their nature cross-modal, it may be considered that the observed effect on variable loadings is due to a shift from audio only localisation to fused audiovisual localisation. It is important to note that the differences observed in the barycentric plots are small

(<± 1 unit) but statistically significant.

4.3 Quality of experience ratings in similar virtual environments

4.3.1 Introduction

In the previous section, audiovisual quality assessments were made in the context of an unreal and dissimilar virtual environment. This section describes an investigation in to quality of experience responses in the context of 'similar' virtual environments, defined as virtual environments in which the geometry of the reproduced scene is similar to that of the real environment in which the user is located.

4.3.2 Materials and methods

Participants - 28 participants were recruited from the students and staff at the University of Salford with informed consent in accordance with University of Salford ethical guidelines. For the purposes of the analysis of the data collected in the following, the data from the 16 participants collected in work described in section 4.2 was used in the comparison between dissimilar and similar environments.

Virtual Environments - Virtual environments were built in Unity Editor [206] and presented using an HTC Vive head mounted display. Environments were constructed as to be visually analogous to the space in which the experiments were taking place. Two environments were used. Firstly, a large, reverberant room was used for the purposes of the experiment to provide an environment with a similar RT_{60} to the unreal virtual environment used in section 4.2. The room used had a mid-frequency RT_{60} of 2.7 seconds and was treated with diffusive elements to minimise coherent reflections. A plan of the room is provided in figure 4.10 and an image of the space used is shown in figure 4.11. Secondly, a smaller, low reverberation environment was used. This space had a broadband RT_{60} of 270ms. The virtual environments were constructed using a combination of geometric primitives in Unity

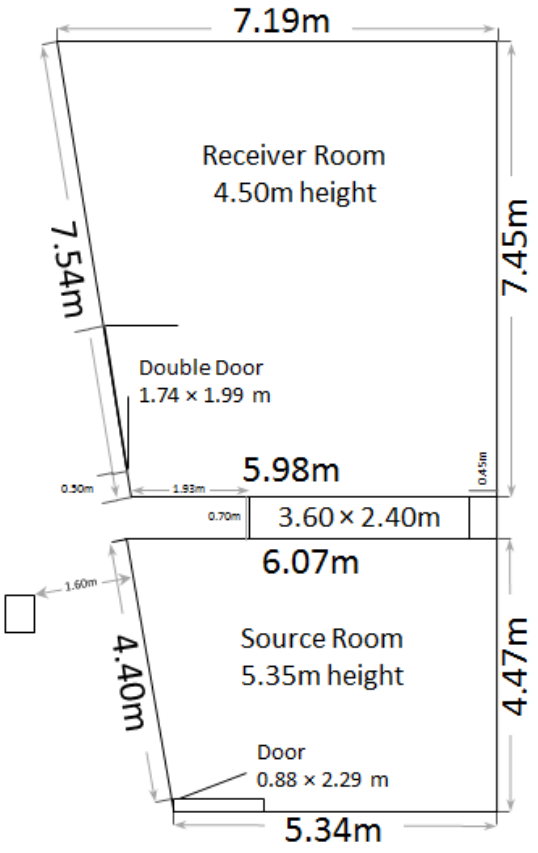


Figure 4.10: Plan of large reverberant room

Figure 4.11: Large reverberant room (5s RT_{60})

and simple 3D models created in Blender [213]. An image of the virtual space is shown in figure 4.12

Auralisation of auditory stimuli was achieved using the processing described in section 3.2. The room volume was set to match that of the space used and the absorption coefficients were selected such that, in conjunction with room volume, total

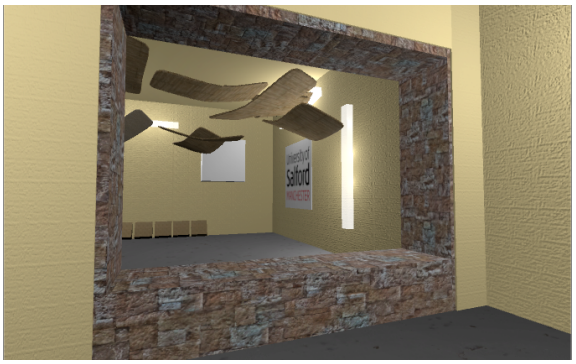


Figure 4.12: VE large reverberant room ($5s$ RT_{60})

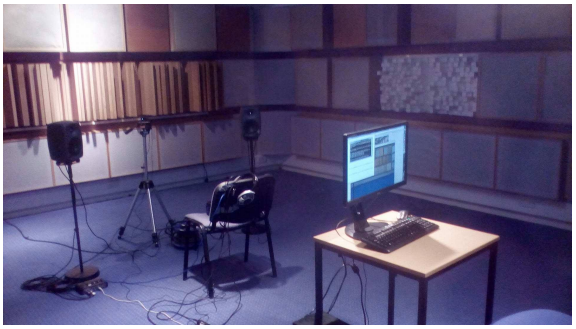


Figure 4.13: Medium RT_{60} room ($270ms$ RT_{60})

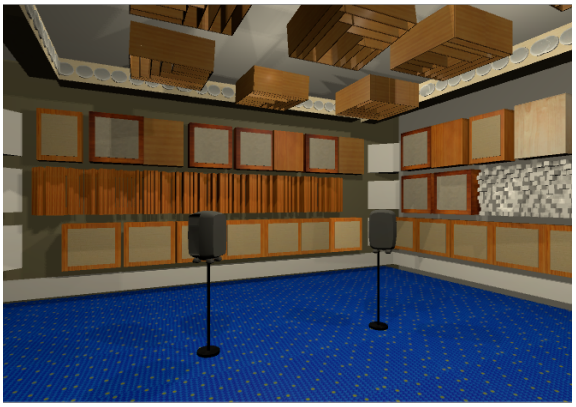


Figure 4.14: VE recreation of medium RT_{60} room ($270ms$ RT_{60})

Figure 4.15: Impulse responses of large modelled space and large reverberant room

decay time matched that of the known value for the space. Impulse responses of the modelled room and real room are shown in figure 4.16

Informed by the analysis presented in section 4.2.2, audio was factored into four conditions: Anechoic with HRTF, early only, late only and full IR simulation.

Figure 4.16: Impulse responses of smaller modelled space and low reverberation room

Tasks and questionnaires - Participants were placed in the virtual space in a position corresponding to their position in the real world environment. Participants were seated in the mid-line of the room, 2.25m from the back wall of the environment. Once the experiment had started, participants were presented with audiovisual stimuli consisting of floating spheres and audio rendered at a point within the virtual room. Audio and visual stimuli were factored as in section 4.2, with low, medium and high levels of co-location ambiguity. During stimulus exposure, participants were asked, via an in-VE questionnaire to respond to QoE scales as described in section 4.2.1 (Figure 4.17). Participants were given a wireless controller which acted as a laser pointer within the VE. Responses were given by directing the pointer to the number which corresponded to the Likert scale level which most fit the response of the subject.

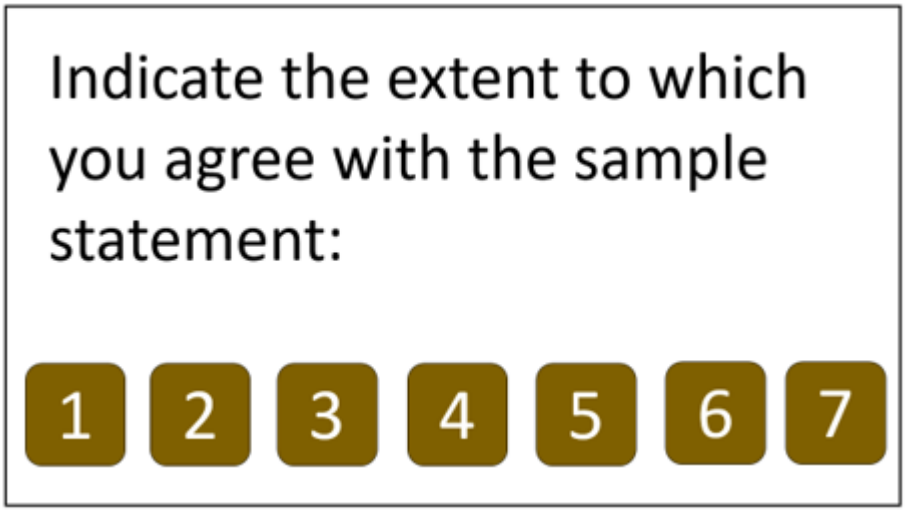
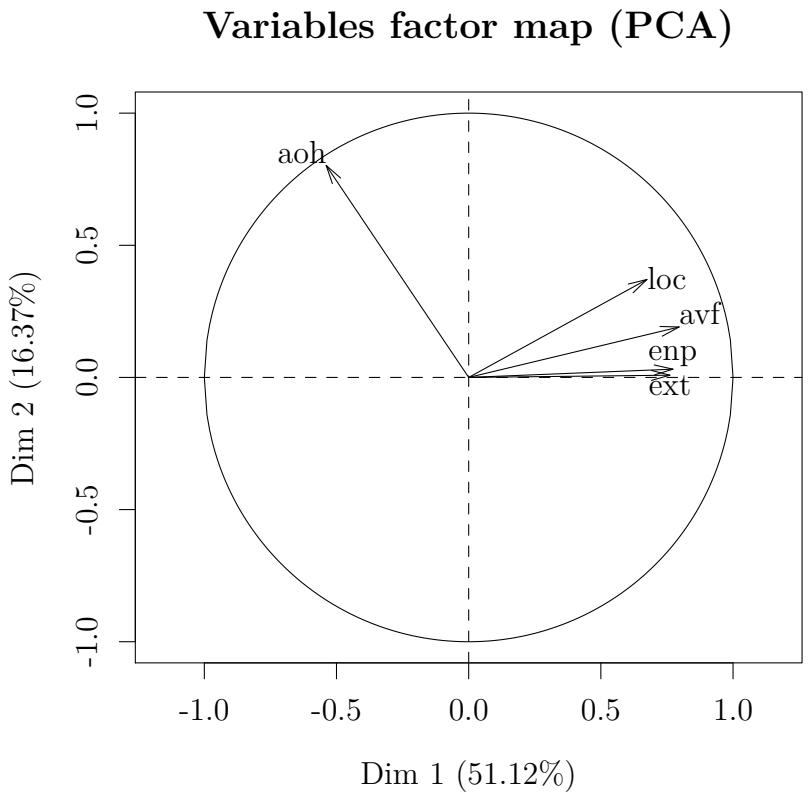


Figure 4.17: Questionnaire response object for stimulus wise response

4.3.3 Results

Responses in both dissimilar and similar environments were subjected to principal component analysis for the purposes of dimensionality reduction. Factor loadings are shown in figure 4.18. Factor loadings, eigenvalues and parallel analysis suggest a one factor model in which audiovisual fusion, environmental plausibility, localisation and externalisation constitute a unitary percept of quality which represents 59.27% of the variance in the sample. Figure 4.19 shows the scree plot for the PCA and

parallel analysis. Component one is the only factor in the model in which eigenvalues exceed that of noise. Figure 4.20 shows loadings for individual PCA solutions for each environment. It can be seen that in the case of the dissimilar environment, the solution approaches simple structure, with externalisation and awareness of headphones being independent of the remaining factors. In the case of the similar environments, simple structure is not achieved and externalisation is more dependent on the responses for other variables.



	Dim.1	Dim.2
avf	0.8	0.19
enp	0.77	0.031
loc	0.67	0.37
ext	0.76	0.0084
aoh	-0.54	0.8

Quality of experience:
aoh - awareness of headphones
avf - audiovisual fusion
enp - environmental plausibility
ext - externalisation
loc - localisation

Figure 4.18: Factor loadings for PCA of questionnaire responses in similar and dissimilar environments

The extracted first component was subjected to testing for normality using the Shapiro-Wilk test and was found to violate this assumption ($p = 0.0002$). In addition, Bartlett's test for sphericity showed that data were heteroskedastic ($p = 0.018$). As such, it was deemed that non-parametric tests were appropriate for further analysis. The Kruskal Wallis test was used to test for differences between audio

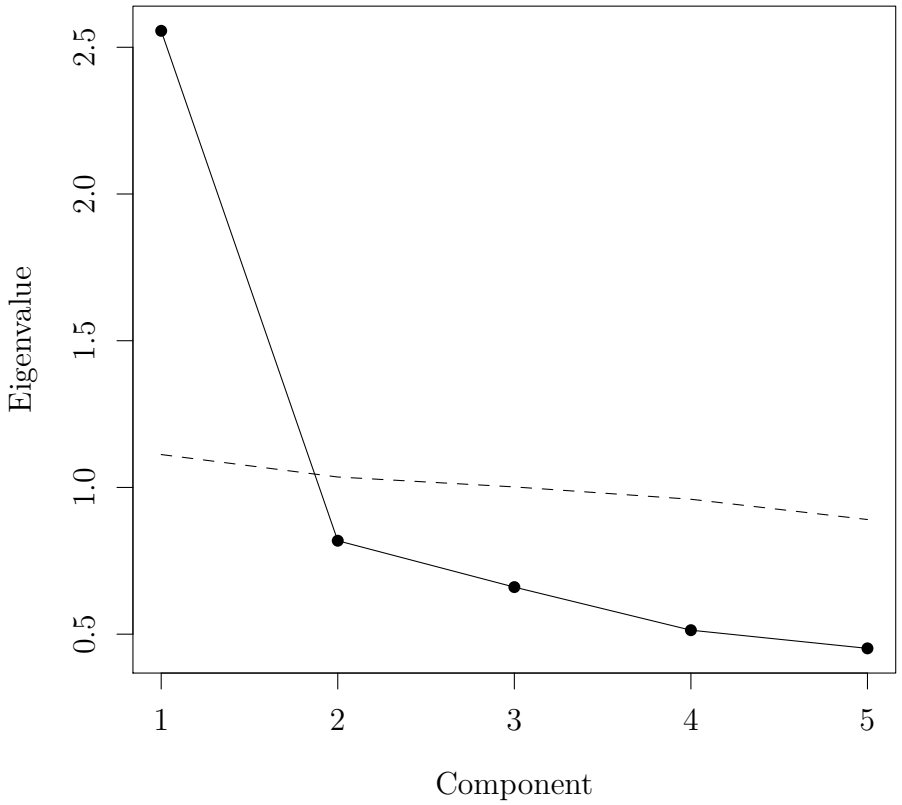


Figure 4.19: Scree plot for PCA of questionnaire responses in similar and dissimilar environments

conditions, visual conditions, environments (Table 4.4). All factors were found to have significant effects in both simple and two way interactions.

Interactive effects - Kruskal-Wallis tests statistics and p-values were computed for two way interactions between design variables for PC1 scores extracted from Likert responses (Table 4.4). Significant effects were found for all two way interactions. Visualisation of the audio × environment interaction shows the change in response profile between environment and audio condition. Responses made after stimulus exposure in the unreal environment show little difference between auditory conditions. Median anechoic and early IR responses are almost identical with late and full IR simulation responses lying within the upper interquartile range. Pairwise contrasts using Wilcoxon signed rank and Bonferroni-Holm p-correction show that this level of interaction is not significant between audio factors ($p = 1$). In the

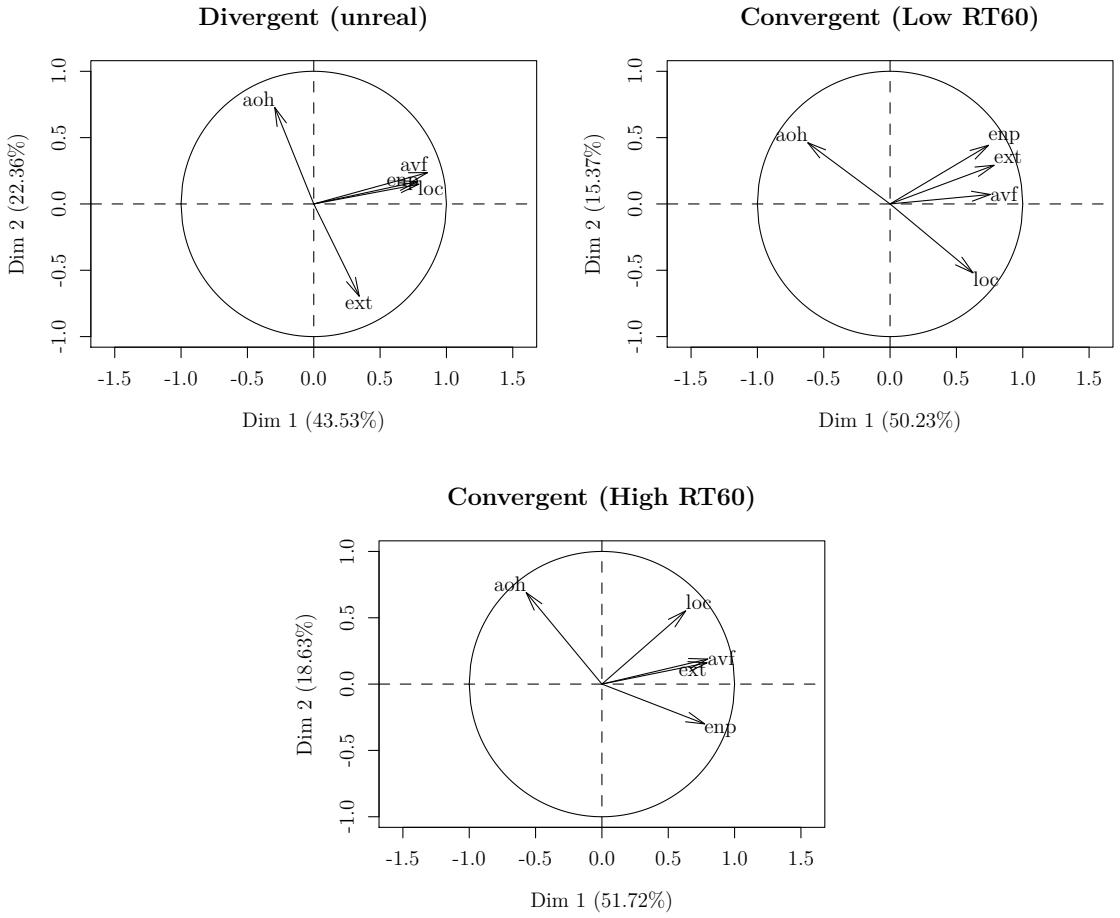


Figure 4.20: Factor loadings for PCA of questionnaire responses in similar and dissimilar environments, with responses separated by environment

real/virtual similar environments, anechoic stimuli have responses that are significantly lower than full IR simulation in both low RT_{60} ($p = 0.008$) and high RT_{60} ($p = 8.4 \times 10^{-8}$) conditions. In the low RT_{60} conditions, difference is only observed between anechoic and full IR stimuli, in high RT_{60} conditions, anechoic stimuli are significantly different to all other conditions, however all levels of simulation have no significant difference after p-value correction ($p > 0.05$). Figure 4.21 shows boxplots for the interaction between acoustic response simulation conditions and environmental similarity conditions. In the case of the similar environments, the distribution of energy appears to determine which part of the impulse response is more important for higher responses of quality, In the long RT room, the late part appears to be more important. However, in the short RT environment both parts of the impulse

response are required for significantly higher assessments of quality. Interaction

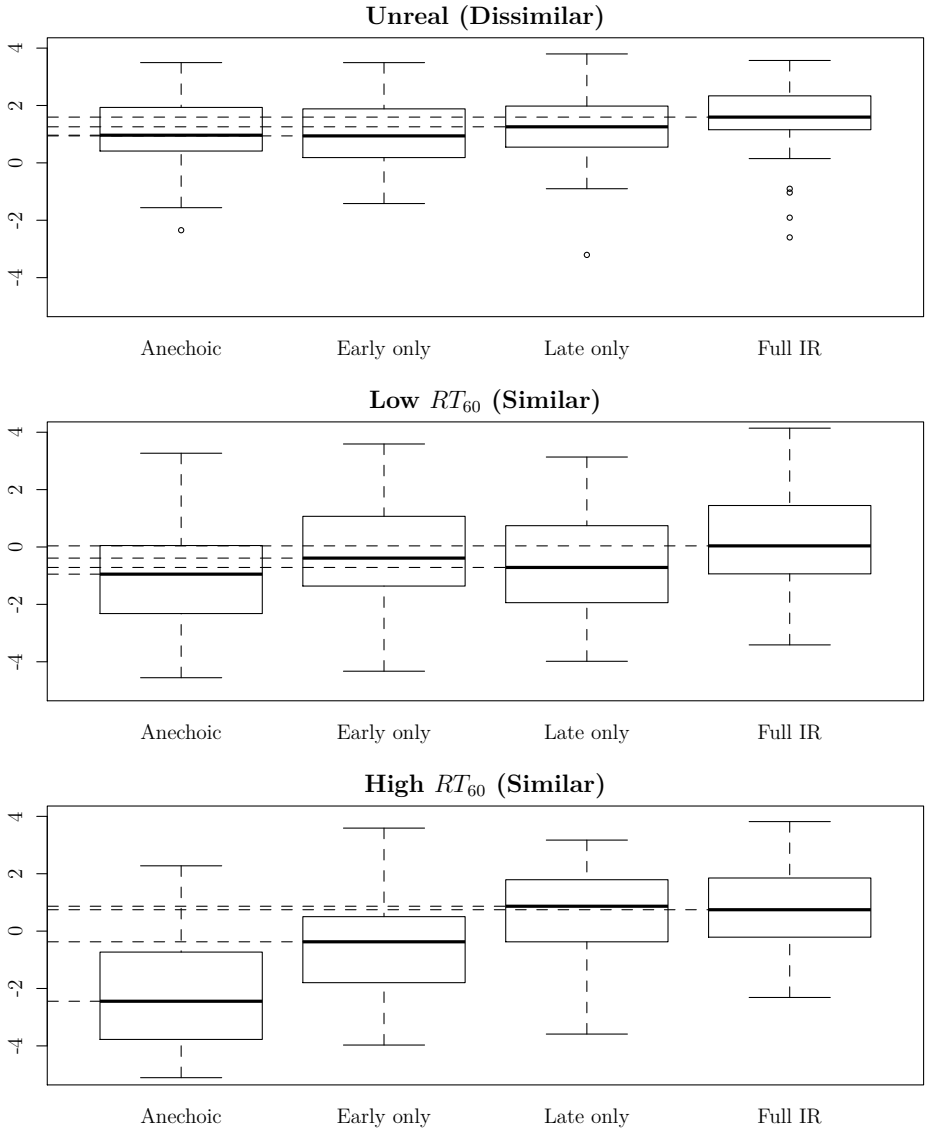


Figure 4.21: Boxplot of interactive effect on quality responses between environment and audio factors

between co-location ambiguity and environment is illustrated in figure 4.22. As before, the unreal environment stimuli were rated higher overall and have no significant difference between co-location ambiguity conditions when subjected to pairwise contrasts ($p > 0.05$). In addition, there is no significant difference between co-location ambiguity conditions in the high RT_{60} real/virtual similar environment. However, in the low RT_{60} environment, stimuli which were presented with a high level of co-location ambiguity resulted in significantly lower overall quality responses (Table 4.3). Two way interaction between audio and audiovisual co-location conditions is visualised in figure 4.23. Pairwise contrasts suggests that there is only difference

Table 4.3: Subset of pairwise contrasts for low RT_{60} stimuli across audiovisual co-location conditions. Wilcoxon signed rank with Bonferroni-Holm p-value correction

	Low RT High Ambiguity	Low RT Mid Ambiguity
Low RT Low Ambiguity	9.1×10^{-9}	1
Low RT Mid Ambiguity	5.5×10^{-9}	-

in terms of interaction in the case of mid ambiguity anechoic stimuli in comparison with mid ambiguity full IR stimuli (p = 0.009). Differences in this interaction are better explained by the simple effect of audiovisual co-location on responses.

Simple effects -

Figures 4.24, 4.25, and 4.26 show differences between audio conditions, co-location conditions and environmental similarity conditions, respectively. In the case of audio factors, overall quality of experience responses are lower for anechoic (HRTF only) stimului and increase progressively through 'early only', 'late only' and 'full IR' conditions. Pairwise Wilcoxon signed rank tests demonstrate that anechoic stimuli are significantly different to all other conditions. There is no significant difference between either partial simulation condition and that full simulation is not significantly different to late reverberation only. This pattern of association is comparable with results observed in figure 4.8.

Table 4.4: Kruskal-Wallis tests results for design factors

Factor	K.W. χ^2	d.f.	p-value	Freeman's θ
Audio	37.5	3	3.56×10^{-8}	0.223
Co-location	54.58	2	1.4×10^{-12}	0.277
Environment	67.74	2	1.95×10^{-15}	0.474
Audio×Environment	128	11	$< 2.2 \times 10^{-16}$	0.353
Co-location×Environment	128.87	8	$< 2.2 \times 10^{-16}$	0.374
Co-location×Audio	93.74	11	3×10^{-15}	0.295

Table 4.5: Wilcoxon signed rank p-values for pairwise contrasts of overall quality of experience by audio condition

	anechoic	early only	late only
early only	0.00464	-	-
late only	0.00039	0.27987	-
full	5.6×10^{-8}	0.00450	0.06273

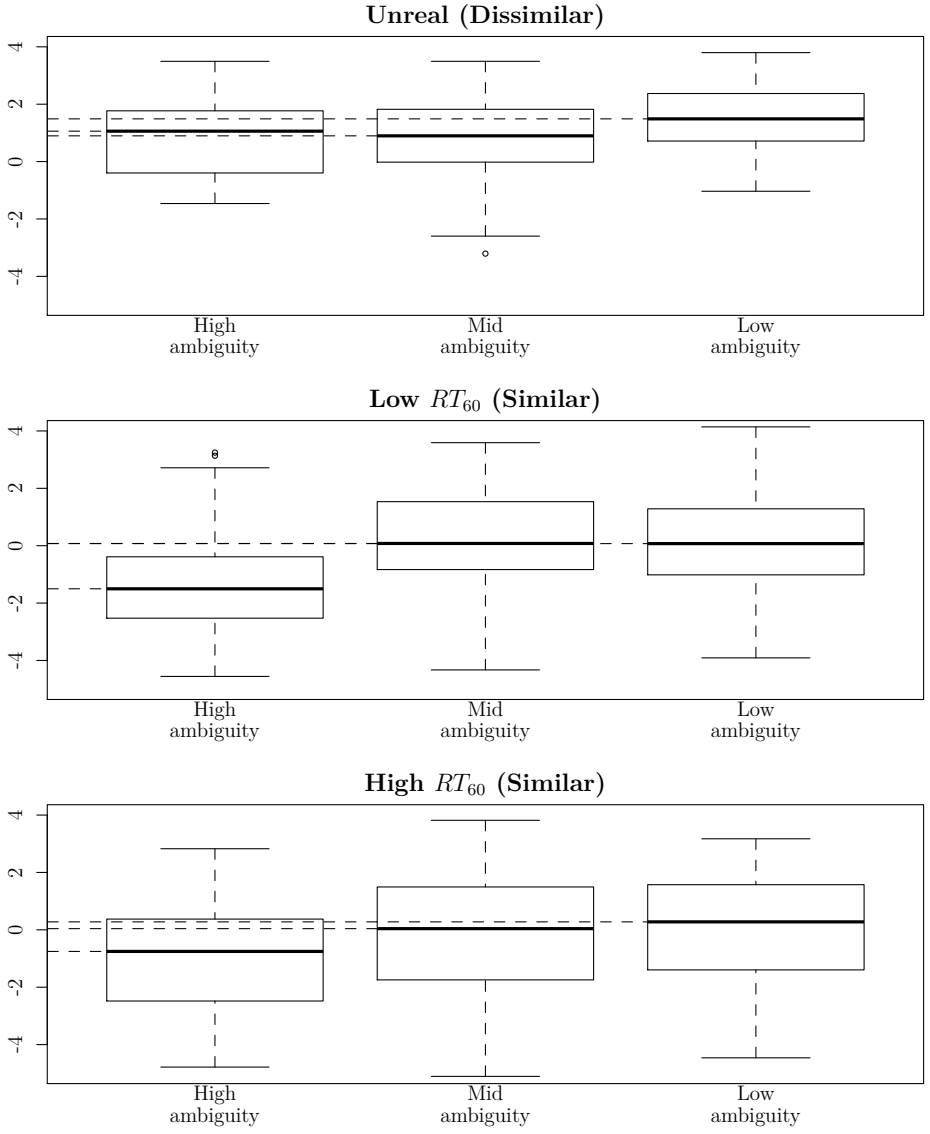


Figure 4.22: Boxplot of interactive effect on quality responses between environment and audiovisual co-location ambiguity factors

Simple effects for co-location factors, again, show comparable response profiles as seen in section 4.2.2, where responses for high audiovisual relationship ambiguity are significantly lower than where audio-visual co-location is more explicit (table 4.6).

Table 4.6: Wilcoxon signed rank p-values for pairwise contrasts of overall quality of experience by audiovisual co-location ambiguity condition

	High ambiguity	Mid ambiguity
Mid ambiguity	6.4e-08	-
Low ambiguity	5.9e-12	0.15

Simple effects for virtual/real environment similarity show higher responses for the unreal environment, where there is no relationship between virtual and real world

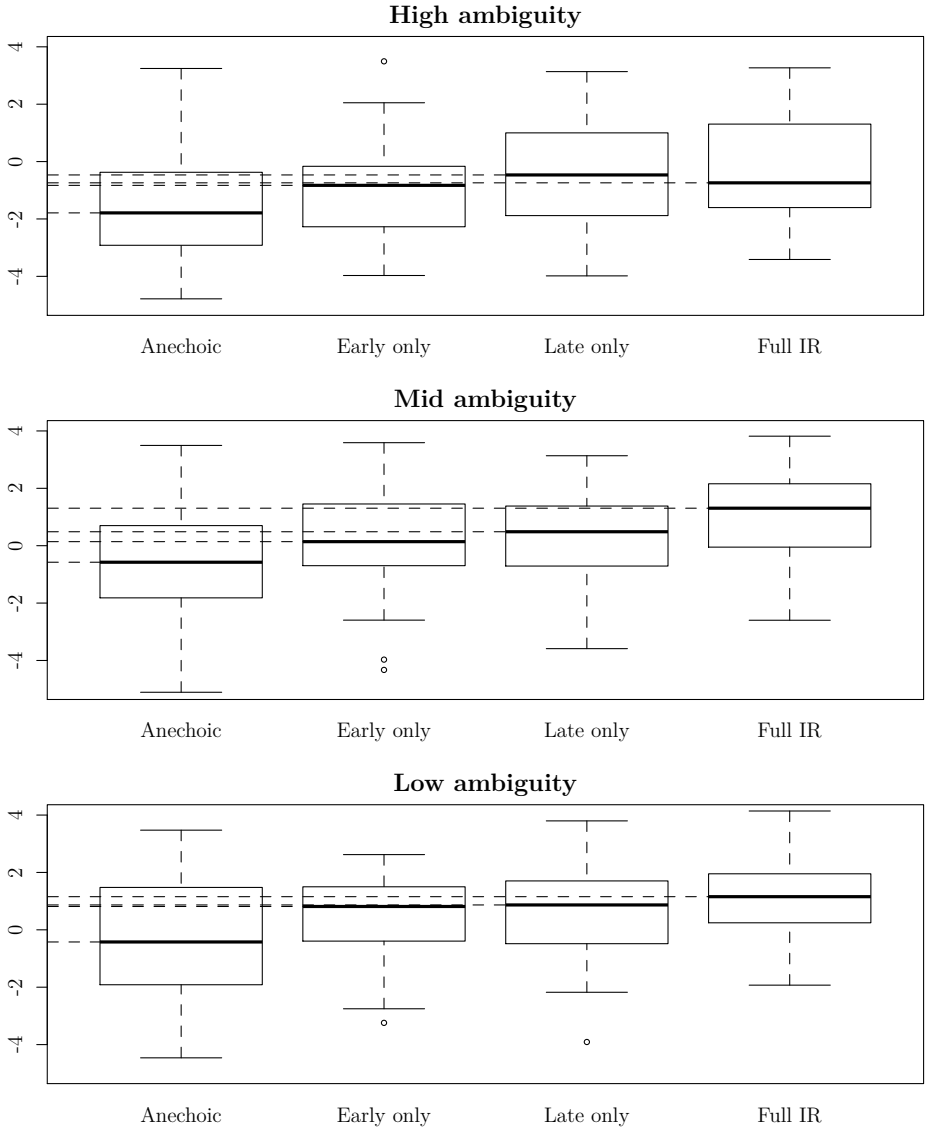


Figure 4.23: Boxplot of interactive effect on quality responses between audiovisual co-location ambiguity and audio factors

geometry. Stimuli presented in the unreal condition show a lesser degree of variance in addition to rating higher (Figure 4.26). Pairwise contrasts suggest that both similar environments are not significantly different and that responses for the unreal environment are significantly higher (Table 4.7).

Table 4.7: Wilcoxon signed rank p-values for pairwise contrasts of overall quality of experience by real/virtual similarity condition

	Unreal	Low RT_{60}
Low RT_{60}	3.6×10^{-15}	-
High RT_{60}	3.3×10^{-10}	0.77

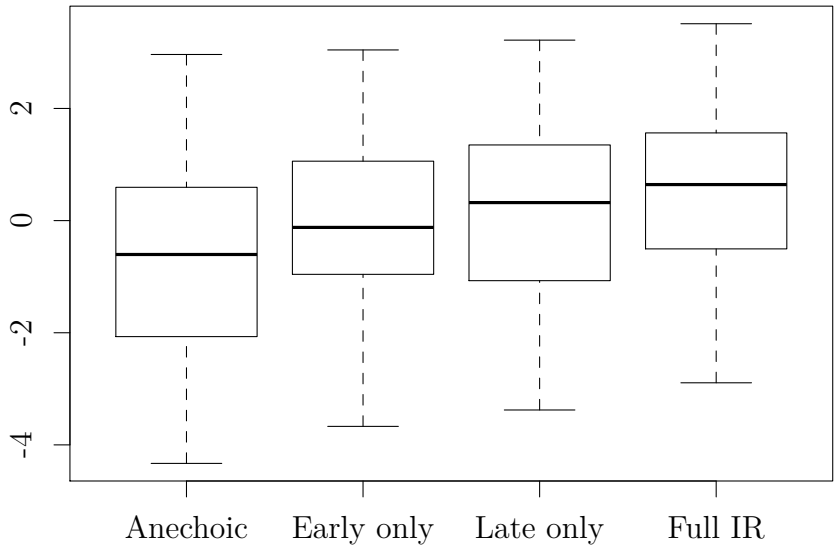


Figure 4.24: Overall audiovisual quality responses from PCA of questionnaire data by audio conditions

4.3.4 Discussion

Comparison of quality of experience responses for both real/virtual dissimilar and similar environments suggest that there may be differences in the profile of responses given as a function of visual anchoring for the perception of plausibility, externalisation, audiovisual fusion and localisation in immersive VR environments. In addition, comparison of the PCA factor maps suggests that, as sample size increases and greater factorisation of the data is introduced, variation in the patterns of loading for individual items converge in to a simple structure one factor model and that although there is difference in the greatest contributor to variance in the measured responses for individual factors, overall, the items used to determine quality of experience constitute a unitary model of overall quality in the variety of stimulus exposures used in these observations. As such, although as in section 4.2.2, it can be shown that some fatcors may result in greater or less independence of responses, overall this combination of factors is able to assess overall quality of experience despite individual item loading.

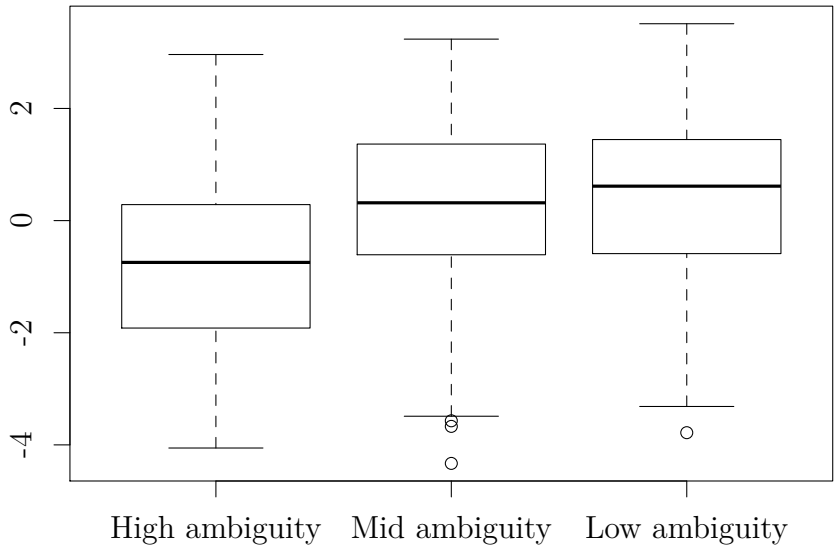


Figure 4.25: Overall audiovisual quality responses from PCA of questionnaire data by audiovisual co-location ambiguity

In the case of dissimilar environments, where there is no association with the geometry of the pre-stimulus environment and the reproduced environment, responses rated more highly overall with θ values suggesting that ranked responses in this comparison are highly likely to be higher from this sample than others in the comparison. A hypothetical explanation for this is that in unreal and dissimilar environments, there is no *a-priori* mental model on to which to base decisions of whether an auralisation is real or unreal, and that the context of the experience of the virtual experience creates a bias for the acceptance of the rendered source as real. This hypothesis would explain the smaller variance observed in the dissimilar environment condition, especially in light of the lack of effect as a function of impulse response content observed in comparison with the clear effect demonstrated in the similar environment with comparable RT_{60}. In the case of similar rooms, it can be shown that with simplified room modelling and partitioning of the impulse response that the late component of the IR is the most important component for externalisation, plausibility, audiovisual fusion and subjective localisability in cases where the reverberation time is long. In the case of shorter reverberation, partitioning the IR is ineffective and both components equally contribute to improvement of audiovisual quality of experience over anechoic renderings.

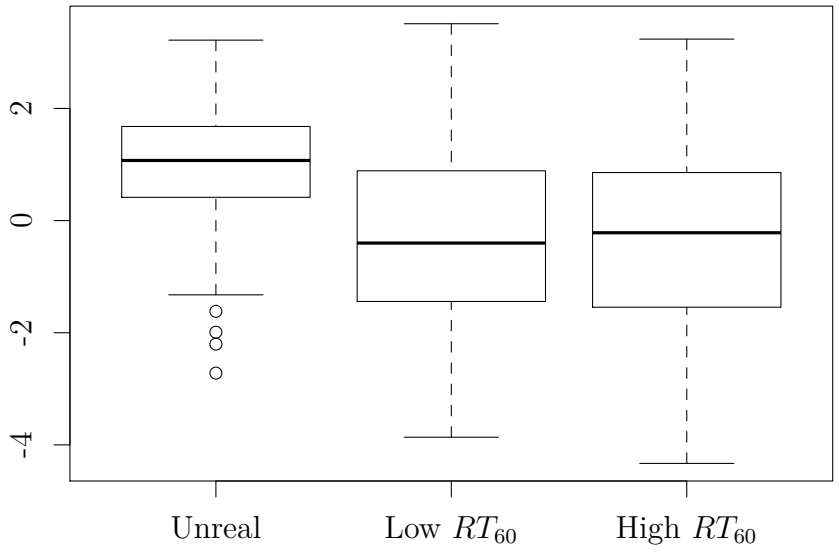


Figure 4.26: Overall audiovisual quality responses from PCA of questionnaire data by environment condition

The effect of co-location ambiguity between auditory and visual stimuli such that overall audiovisual quality is reduced in conditions where auditory sources are not rendered to be co-located. This also constitutes a damping effect on the increase in responses associated with addition of impulse response simulation component. In the cases where there is some spatial association between visual and multimodal stimuli, full IR simulation results in a significant increase in comparison to anechoic stimuli ($p = 0.00961$). In cases where audiovisual co-location is explicit, there is no difference between factors ($p > 0.05$). However, this is due to the overall increase attributed to the low ambiguity condition and the increased variance observed in the low ambiguity anechoic coondition. The net effect can be described as ambiguity of audiovisual spatial association modulating the effect of spatial room response rendering on overall ratings of quality.

4.4 Quality of experience ratings and judgement of real and unreal sources

4.4.1 Introduction

In section 4.3 work was conducted in which quality of experience metrics were compared between similar and dissimilar environments. In addition, loading factor maps were compared between dissimilar and similar virtual environments. It was hypothesised that in the case of the dissimilar environments, auditory stimuli were treated as real objects despite the intrinsic knowledge that the rendered objects were in the virtual environment only, constituting a suspension of disbelief which appeared to be statistically independent of audio rendering treatment in terms of the content of the simulated impulse response. There follows in this section, work investigating quality of experience responses to both real and rendered sources in similar virtual environments in order to identify differences in quality of experience response which may be associated with the perception of audio as a real source while within a VE and to determine if this might be modulated by the accuracy of the room simulation.

4.4.2 Materials and methods

4.4.2.1 Participants

Twenty-four participants aged between 19 and 37 volunteered to take part in this experiment (16 male and 8 female). Participants were not compensated for their time. After taking part in one set of conditions, the subjects were given the opportunity to volunteer to complete more trials at a later date. As such, not all participants completed the same number of trials under the same number of conditions. All data collection was undertaken in compliance with University of Salford ethical guidelines.

4.4.2.2 Virtual Environments (VEs)

Virtual environments were built in Unity Editor [206] and presented using an HTC Vive head mounted display. Environments were constructed as to be visually analogous to the space in which the experiments were taking place. Two spaces were used for the purposes of data collection. The first was medium sized ($6m \times 7m \times 3.4m$) room which is acoustically treated for the purposes of subjective audio evaluation

testing and has a broadband RT_{60} of approximately 230ms. The second space was a small ($3m \times 3m \times 2.5m$) acoustically treated booth designed for speaker based spatial audio reproduction with a broadband RT_{60} of approximately 90ms. Comparisons of real and virtual spaces are given in figures 4.13, 4.14, 4.27 and 4.28. Audio from loudspeakers was emitted from either one of a pair of Geneec 8030a studio monitors positioned at ± 45 deg with respect to the listener. In the larger of the rooms, speakers were positioned at 2 metres from the central listening position, to correspond to source-receiver positions in the soundfield measurements used for ambisonic rendering. In the smaller space, loudspeakers were positioned 1.4m from the listening position.

Figure 4.27: Low ($90ms$) RT_{60} room

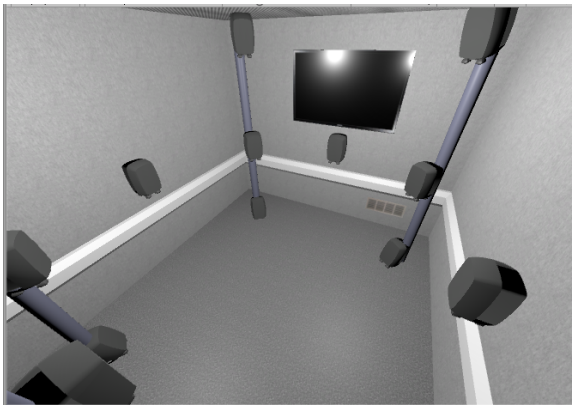


Figure 4.28: VE recreation of ($90ms$) low RT_{60} room

4.4.2.3 Auralisation

Spatial rendering of auditory stimuli was achieved in two ways intended to contrast between a comparatively high and low level of physical accuracy. In the first case, auralization was achieved by convolving stimuli with ambisonic impulse responses

measured at the listening position from source positions described in [214]. Decoding and rendering for playback was performed using the GoogleVR AudioSoundfield object from the GVR Unity SDK [215]. A discussion of ambisonic decomposition and rendering can be found in section 2.1. Although is has been demonstrated that B-Format directional encoding results in low levels of accuracy of response in the spatial domain [216], the time-energy content of such measurements can be considered to be a complete representation of the impulse response of the room. As such, the physical accuracy of audio rendered by this method was considered high for the purposes of this study. Low physical accuracy stimuli were processed using an image source algorithm which assumed a 'shoebox approximation' of the space for the early component of the impulse response and used the inbuilt reverb processor in Unity Editor for the diffuse component of the IR. The direct path and early reflections were convolved with HRTFs from the KEMAR compact dataset [205]. RT_{60}s for each space were calculated from the W component of the measured responses and used to determine average absorption coefficients for the image source model. The spectral shape of the W channel was also used to determine a rough approximation of the overall frequency response of the modelled output. Headphone playback was achieved using Sennheiser HD800 headphones. To reduce the opportunity for consistent level differences to be used as a cue, playback level was randomised between 54dBA and 68dBA at the listening position for both headphone and speaker playback. Audio that was rendered over headphones was also processed to take into account the impulse response of the loudspeakers used and the transmission of the sound through the headphone ear-cup. Speaker impulse responses were those used by Hughes et al [214]. Full details of the generation of the B-format room responses can be found in [217]. Filters to account for the transmission of sound through the headphones were obtained by first recording the measurement signal at 0.5m at 90 deg azimuth through the ipsilateral inner ear microphone of a Brüel & Kjær head and torso simulator (HATS). A second signal was recorded with the headphones in place and these two signals were deconvolved to produce an impulse response which approximated the occlusion effect of the headphones. Audio which was to be rendered over headphones was first filtered with these two impulse responses in MATLAB before further processing.

4.4.2.4 Stimulus exposure and response task

After entering the real room on which the virtual environments was based, participants were presented with the virtual environments described above, with loudspeakers positioned in the real space corresponding to source locations in the VE. Audio samples included in the experiment were a combination of male speech, female speech and short (0.5 sec) white noise bursts. Participants were asked to indicate if an auditory stimulus originated from the loudspeakers or the headphones by way of a gamepad controller. In instances where participants perceived the sound to originate from the speakers, participants were instructed to push a thumbstick away from them and press a button. In cases where sound was perceived as originating from the headphones, the instruction was to pull the thumbstick towards them and push the button. Participants completed the task with audio presented over headphones spatially rendered using B-format impulse responses or using the object based image source model described in section 3.2. For a discussion of B-format, see section 2.1. In total there were 108 singular decisions to be made by participants under each audio condition with a 50% chance of auditory stimuli being presented over headphones or loudspeakers which was randomised at runtime. Each speaker/headphone decision was made under one of three visual conditions:

- Both speakers and virtual room visible
- Speakers invisible and virtual room visible
- No visual cues, head mounted display showing black screen

4.4.3 Results and discussion

Five point QoE questionnaire results across all conditions including both headphone and loudspeaker decisions are presented in table 4.8. Questionnaire responses were subjected to principal component analysis (PCA) for the purposes of dimensionality reduction. PCA vector maps for aggregated data, responses for 'loudspeaker' decisions and responses for 'headphone' decisions were generated (Fig. 6.9).

In all three cases there are only two significant dimensions (Eigenvalues greater than 1), accounting for between 72% - 78% of the variation in the data. Attention to playback media and reported externalisation are inversely correlated, with audiovisual fusion and environmental plausibility independent of these factors. Loadings for the

Table 4.8: Descriptive statistics for ITQ responses

	1st Q	Median	Mean	3rd Q
A/V fusion	3	5	4.62	6
Atten. to Playback	2	3	3.65	5
Env. Plausibilty	4	5	4.8	6
Externalisation	4	6	5.16	7
Localisation	4.25	6	5.2	6

aggregated data (Figure 4.29) share a similar structure to those observed in section 4.2.2. Independent PCA of responses associated with loudspeaker (Figure 4.30) and headphone (Figure 4.31) judgement, again show varying patterns of loading but with similar gross structure. When loudspeaker and headphone judgements are aggregated, there appears to be a smaller degree of independence between the representational features and externalisation. Although it was argued in section 4.3.3 that the structures observed in earlier work with similar division of variance and loadings (Section 4.2.2), it may be argued that this analysis suggests two percepts. It can be shown that the variance across both dimensions can be can be explained by difference in responses for judgements of whether stimuli are emitted in headphones or loudspeakers. Figure 6.10 shows individuals and confidence ellipses on the extracted feature space. Data was tested for normality using the Shapiro-Wilk test and for sphericity using Bartlett's test with factorisations for all independent variables. Data were shown to be normal for dimension 1 ($p = 0.07$) and exhibit homogeneity of variance between all design factors ($p > 0.05$). Along dimension 2, data were shown to be non-normal and to violate the assumption of sphericity. However, using a mixed effects model, it is possible to demonstrate that the violations of normality are due to intra-participant variability. Comparison was made using ANOVA of a intercept only model of dimension 2 and an intercept only model with random effects by participant. The random effects model was found to fit the data significantly better than the intercept only model (Table 4.9). Removing participant level effects results in normality of dimension 2 and sphericity between groups for all independent variables ($p > 0.05$).

Scores for dimension 1 and dimension 2 were subjected to hierarchical regression analysis to identify fixed effects while controlling for participant level random effects (Table 4.10 & 4.11). Significant difference was found between responses associated with judgements for whether audio was emitted from either speakers or headphones.

Table 4.9: Multilevel ANOVA of inetrcept only model fit and random effect predictor of participant on dimension 2 scores

	Model	df	AIC	logLik	Test	L.Ratio	p-value
Intercept only	1	2	356.3	-173.3			
Random effects within participant	2	3	337.8	-165.9	1 vs 2	14.8	1×10^{-4}

However, no significant difference was found for simple fixed effects for the other independent variables.

Table 4.10: Multilevel mixed effects ANOVA between fixed effects on dimension 1 between independent variables

	Model	df	AIC	logLik	Test	L.Ratio	p-value
Intercept only	1	3	491.4				
HP/SPK judgement	2	4	430.9	-211.5	1 vs 2	62.5	<.0001
Room	3	5	432.7	-211.4	2 vs 3	0.153	0.69
Rendering	4	6	434.3	-211.2	3 vs 4	0.397	0.52

Table 4.11: Multilevel mixed effects ANOVA between fixed effects on dimension 2 between independent variables

	Model	df	AIC	logLik	Test	L.Ratio	p-value
Intercept only	1	3	337.8	-165.9			
HP/SPK judgement	2	4	326.5	-159.2	1 vs 2	13.3	0.0003
Room	3	5	326.2	-158.1	2 vs 3	2.33	0.13
Rendering	4	6	328.2	-158.1	3 vs 4	0.02	0.87

Differences between judgements correspond to higher scores for audiovisual fusion (AVF), environmental plausibility (ENP), localisation (LOC) and externalisation (EXT). Conversely, audio judged as originating from headphones shows lower scores on these measures and higher for awareness of headphones (AoH). This analysis suggests that when presented with possible real or unreal sources in a virtual environment, the perception that the audio is a real source is associated with higher rating on the scales used. It is significant that environmental plausibility is loaded heavily on to dimension 1 in this analysis, as participants were instructed to respond with reference to the virtual environment, not the real environment as an intrusive stimulus. Within the context of similar/dissimilar real to virtual environment geometry, this may be interpretable as a visual anchoring effect carrying over from the pre exposure to the virtual room and causing subjects to attribute perceived

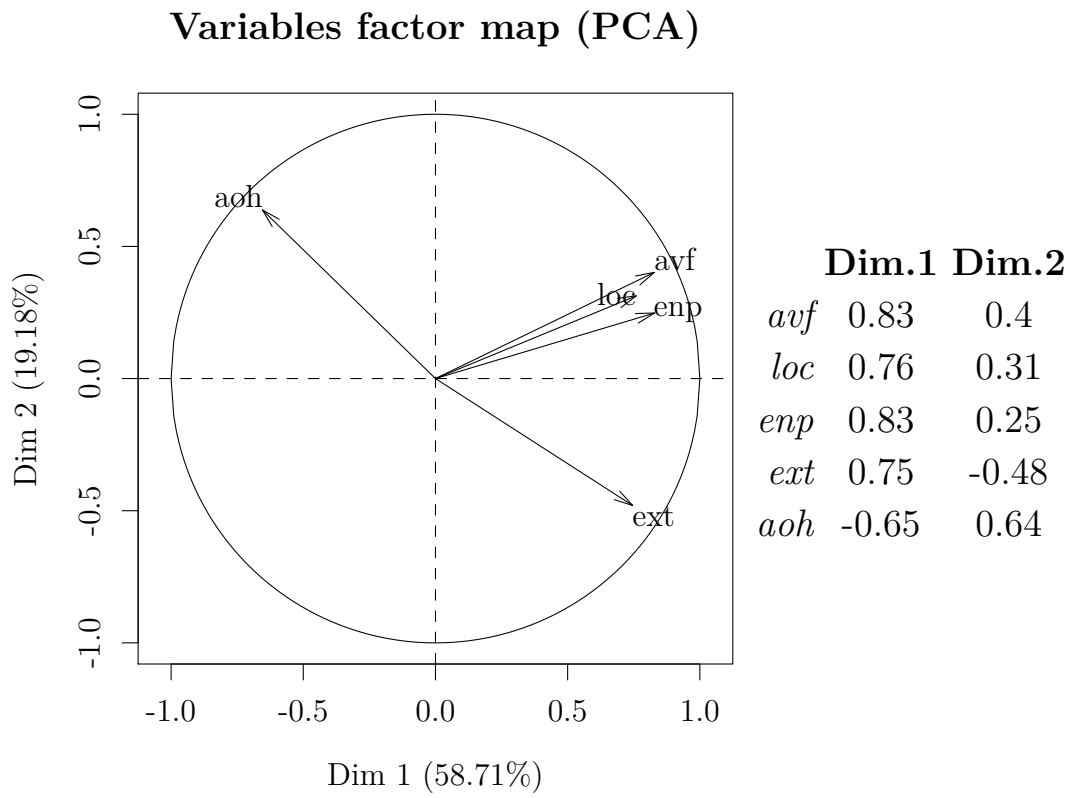
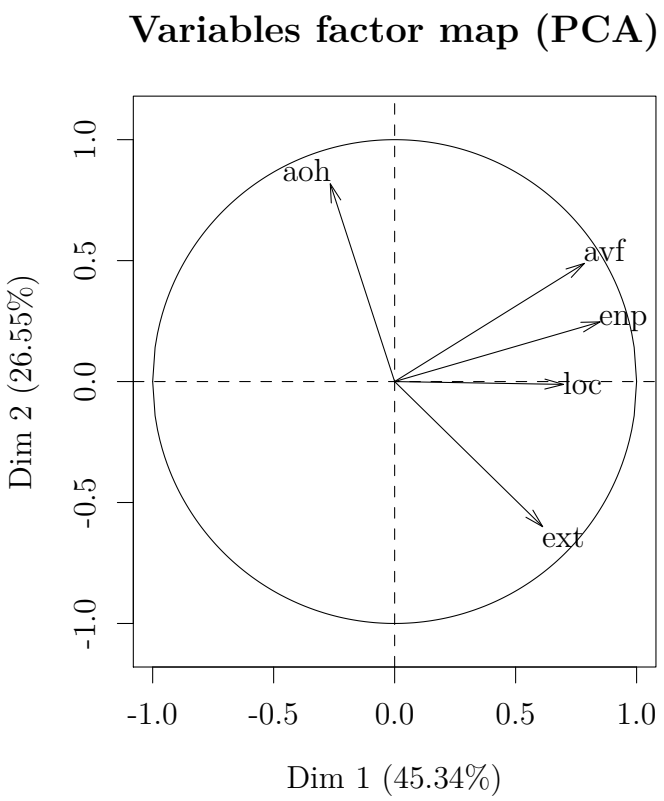


Figure 4.29: PCA factor loadings for both loudspeaker and headphone judgements

loudspeaker sources to virtual diegetic events, despite consciously attributing these qualities, possibly representing a category bias related to the semantics associated with absolute realism. Differences in dimension 2 are smaller and appear to be more related to externalisation. This may be due to the anchoring effect of close comparison between real and rendered sources, highlighting deficits in reproduction due to the non-individualised HRTF filters used in rendering.

Analysis of clustering of individuals on the extracted dimensions from both subsetted PCAs suggested no significant difference between groups on dimensions 1 or 2 (Table 6.6) suggesting no significant difference in response between reverb time and rendering type conditions.

As the differences between responses loaded on to PC1 of the principal components



	Dim.1	Dim.2
avf	0.78	0.49
loc	0.7	-0.012
enp	0.85	0.25
ext	0.61	-0.6
aoh	-0.27	0.82

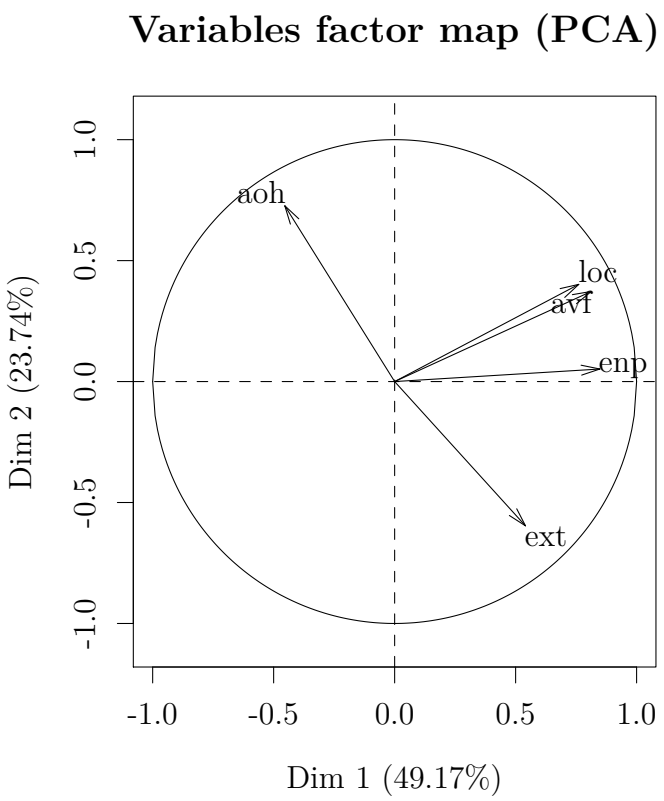
Quality of expeience:
aoh - awareness of headphones
avf - audiovisual fusion
enp - environmental plausibility
ext - externalisation
loc - localisation

Figure 4.30: PCA factor loadings for loudspeaker judgements

analysis are accounted for by the perceived origin of a stimulus, analysis was performed to determine if the independent variables in this study had any effect on the shift in responses for perceived source origin. Distances between headphone or loudspeaker source judgements were calculated using the following:

$$x_{\Delta} = x_{speakers} - x_{headphones} \tag{4.1}$$

$PC1_{\Delta}$ and $PC2_{\Delta}$ were computed and subjected to Shapiro-Wilks testing to test for normality. Both $PC1_{\Delta}$ and $PC2_{\Delta}$ were found to violate normailty assumptions ($p < 0.001$). Wilcoxon signed-rank tests were used to test for significant differences between rendering type and RT_{60} conditions. RT_{60} was found to have no significant effect ($p = 0.67$). Audio rendering type, however, was found to be a significant factor ($p = 1.196\times^{-10}$). Figure 4.33 shows $PC1_{\Delta}$ values distributed by audio rendering



	Dim.1	Dim.2
avf	0.81	0.37
loc	0.76	0.4
enp	0.85	0.052
ext	0.54	-0.6
aoh	-0.45	0.73

Quality of expeience:
aoh - awareness of headphones
avf - audiovisual fusion
enp - environmental plausibility
ext - externalisation
loc - localisation

Figure 4.31: PCA factor loadings for headphone judgements

method.

It can be seen that ambisonic rendering produces reduced differentials between headphone and loudspeaker judgements, with lower interquartile range in response differentials. This suggests that the use of the image source hybrid method produces a greater difference between perceived real and simulated sources than rendering using measured responses. $PC2_{\Delta}$ values were subjected to Wilcoxon signed rank tests for both rendering type and RT_{60} conditions. Both factors were found to have no significant effect ($p > 0.05$).

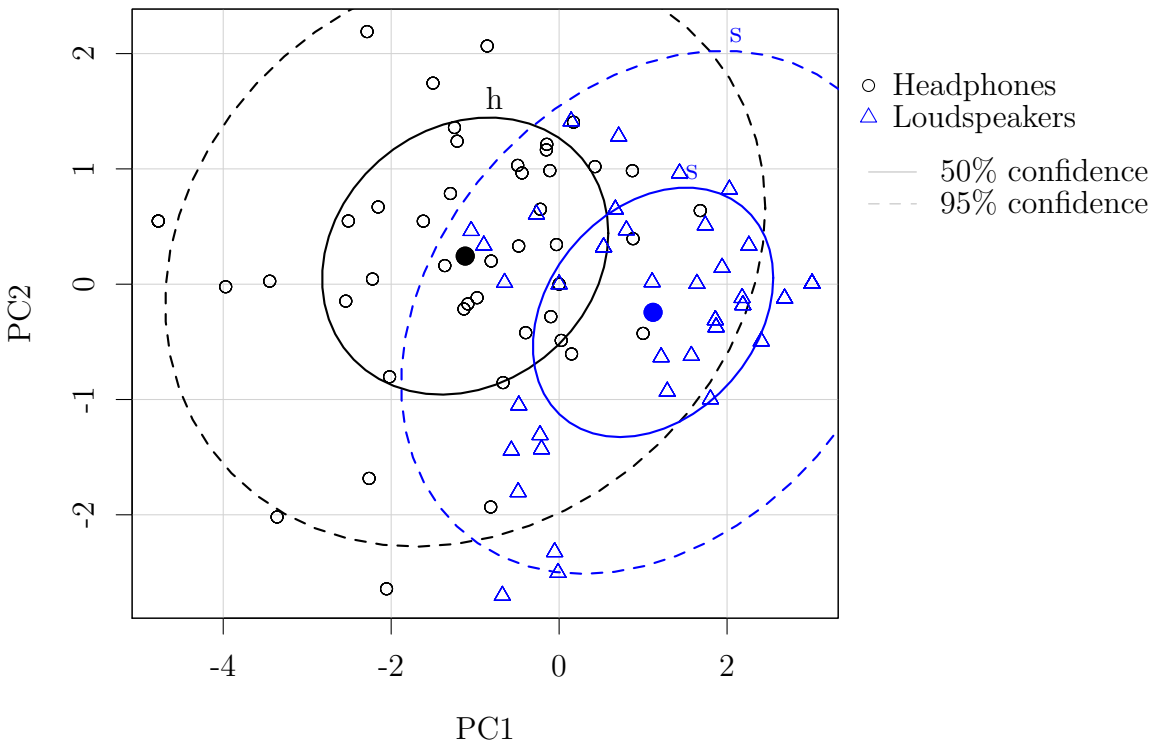


Figure 4.32: Individual responses clustered by judged stimulus origin

4.5 Summary

In this chapter, the construct of audiovisual quality of experience has been analysed in dissimilar and similar virtual environments. It can be shown that the percepts of audiovisual fusion, plausibility, and localisation are strongly related. It can also be shown that while in some cases there is some independence from these aesthetic qualities, such as in dissimilar environments and in smaller smaller sizes, the psychoacoustic response of externalisation and the awareness of listening to simulated audio over headphones is related to the former three items when conditions are sufficiently varied, particularly in comparison with similar environments. It can be shown that the assessment of audiovisual quality using these scales can be affected by impulse response content, and that this effect is more pronounced when there is similarity between real and virtual spaces. It is also possible to modulate this percept using the explicitness of the spatial relationship between audio and visual stimuli in the virtual space. It was also demonstrated that this percept has higher responses

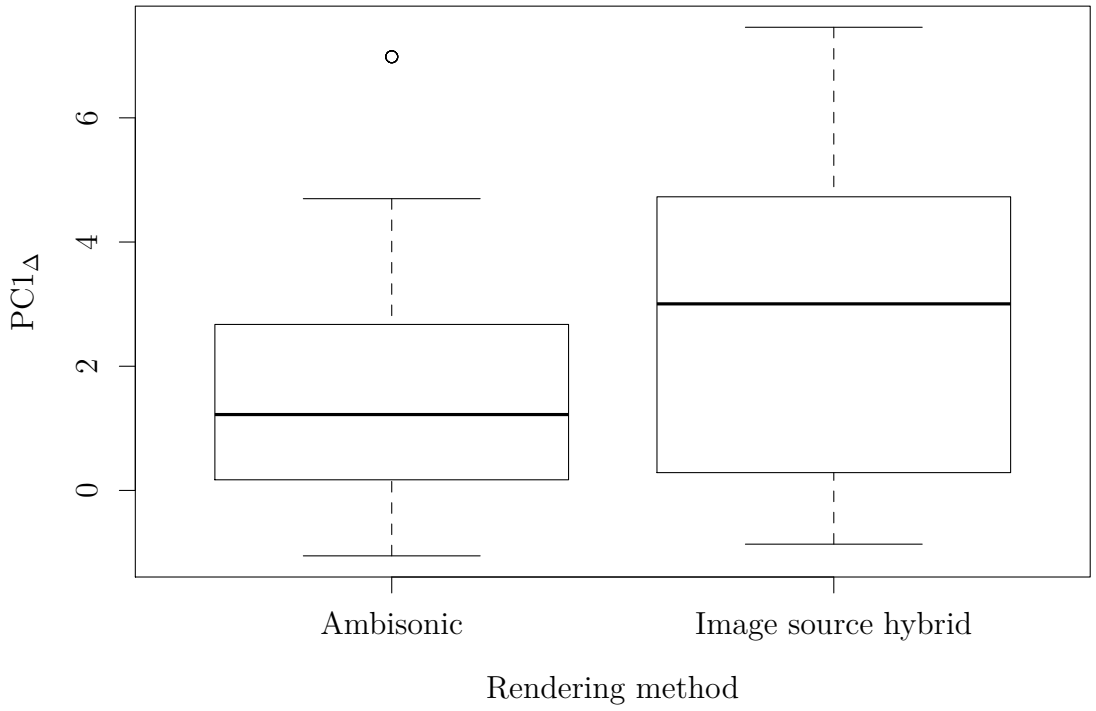


Figure 4.33: $PC1_{\Delta}$ values by rendering type condition

elicited when stimuli are perceived as real and that the responses for 'perceived as real' sources are comparable to ratings for stimuli in unreal, pre/post exposure dissimilar VEs. This points to a possible anchoring bias which promotes suspension of disbelief in unreal VEs and dampens it when VEs are similar to the pre exposure environment. In addition, it was shown that lower physical realism auditory sources produce higher differentials between QoE responses for perceived real and simulated sources.

5

QoE and Presence

Contents

5.1 Introduction

This chapter introduces measures of reported presence and investigates the dependence of these measures on the level of spatial audio rendering and stimulus co-location ambiguity. Relationships between reported presence and audiovisual quality of experience, as described in chapter 4, are investigated. In addition, the dependence of differing level of spatial audio rendering and audiovisual spatial relationship is determined.

5.2 Reported presence and manipulation of spatial models and stimulus co-location

5.2.1 Introduction

As discussed in section 2.3.1, one of the main goals of immersive virtual reality technologies is the elicitation of *presence*. This experience is defined as the sense that one is actually situated within the environment that is presented using the VR system. As a perceptual construct, presence can be considered as being formed of two components, place illusion and plausibility [21]. However, direct measurement of this experience is problematic, especially without inference from physiological or neural measurement methods. As such, self reporting on experiential phenomena has been developed as a way to estimate the extent of presence experienced while using a VR system. The development and use of constructs which are used to assess presence are reviewed in section 2.3.1. Previous work in this part has focussed on audiovisual quality of experience factors which were identified as salient from other work within the literature in the field. This section describes an investigation into the relationship between these items and two constructs of presence: the i-group presence questionnaire (IPQ) [79] and the unitary model of *presence* posited by Witmer and Singer [22]. The aim of this investigation is to identify correlation between the measures of presence as defined by the latent variables assumed by the measures of presence cited above. Explicitly, does the direct reporting of these subscales result in data which demonstrates independent components of an overall measure of presence? In addition, the effect of independent variables used in chapter 4 (varying the content of a simulated impulse response using a simplified room model and the ambiguity of the audiovisual spatial relationship of a stimulus) on responses for these measures of presence are tested.

5.2.2 Materials and methods

The virtual environments used for this study are the 'similar' low reverberation time room and the high reverberation time environments described in section 4.3. The stimuli and response tasks for this study were identical to that used in section 4.3, as were the participants who took part in that study. Acoustic modelling

conditions consisted of the inclusion or omission of the early reflections, late reverberation or both. Audiovisual co-location ambiguity conditions consisted of: obvious co-location (low ambiguity), ambiguous association (mid ambiguity), and obvious dislocation (high ambiguity). In addition to quality of experience items as described above, participants were asked to give responses to questions relating to the subscales described in the i-group presence questionnaire and the Witmer and Singer presence measure. For the igroup subscales, the terms used in the questions were defined before exposure using the following:

- Spatial presence is defined as:

  The sense of being physically present in the VE

  The sense that the environment surrounds you

  The sense that you are not just viewing pictures

- Involvement is defined as:

  Being unaware of the real world environment

  Attention focussed on the virtual environment

  Attention drawn to the virtual environment

- Realism is defined as:

  Visual imagery realism

  Experience is similar to real experience

  Realism to an imagined ideal of realism

As in section 4.3, participants were asked to give responses on a stimulus-wise basis via a textual interface using a wireless controller as a pointing device to indicate the number (1-7) corresponding to a response ranging from "the least" to "the most" in reference to a shown statement.

5.2.3 Results and discussion

Reports on components of presence were subjected to principal component analysis to determine the independence of the data collected. Figure 5.1 shows eigenvalues for the extracted feature space. The point of inflection and the small eigenvalues for components 2-4 suggest a one component model. Figure 5.2 shows factor loadings

for the input variables. Dimension 1 accounts for 61% of the variance in the data and loadings suggest that this dimension represents the unitary percept of overall presence as defined by Witmer and Singer [22]. It appears that visual realism and involvement are correlated and spatial presence is independent of these two variables. Between them, they contribute to overall presence. Therefore, a rotation of the extracted space using varimax rotation [196] shows a two factor solution with simple structure in which visual realism and involvement constitute one dimension, accounting for 40.2% of the variance in the data and spatial presence loads strongly on to dimension 2, accounting for 36% of the variance in the data (figure 5.3). Plotting the data variance in this way shows that the Witmer and Singer measure of overall presence is fully described by a combination of visual realism and involvement and spatial presence, from the igroup questionnaire. The advantage of such a representation of data is that it allows a more detailed appreciation of the factors contributing to the percept of overall presence in terms of the constituent subscales. It is notable that, as latent variables themselves, it should be expected that involvement and visual realism should be independent. However, the experimental design did not use modulation of visual realism, beyond the degree of audiovisual co-location ambiguity, as a design variable. It might be hypothesised that assessment of visual realism is correlated with involvement in cases where the level of visual detail remains constant, however this is not empirically investigated in this work.

Table 5.1: Multilevel ANOVA of rotated PCA component 1 (Involvement and visual realism) with participant as a random effect

	Model	df	AIC	logLik	Test	L.Ratio	p-value
Intercept only	1	2	1435.3	-715.6			
Participant random effect	2	3	1136.6	-565.3	1 vs 2	300.6	<.0001
Audio (modelling) fixed effect	3	6	1131.1	-559.5	2 vs 3	11.57	0.009
Co-location fixed effect	4	8	1132.6	-558.3	3 vs 4	2.5	0.287
RT_{60} fixed effect	5	9	1134.1	-558.03	4 vs 5	0.50	0.4788

Table 5.2: Multilevel ANOVA of rotated PCA component 2 (Spatial presence) with participant as a random effect

	Model	df	AIC	logLik	Test	L.Ratio	p-value
Intercept only	1	2	1390.3	-693.1			
Participant random effect	2	3	1265.1	-629.6	1 vs 2	127.1	<.0001
Audio fixed effect	3	6	1217.0	-602.5	2 vs 3	54.15	<.0001
Co-location fixed effect	4	8	1216.9	-600.4	3 vs 4	4.09	0.1291
RT_{60} fixed effect	5	9	1218.5	-600.3	4 vs 5	0.33	0.56

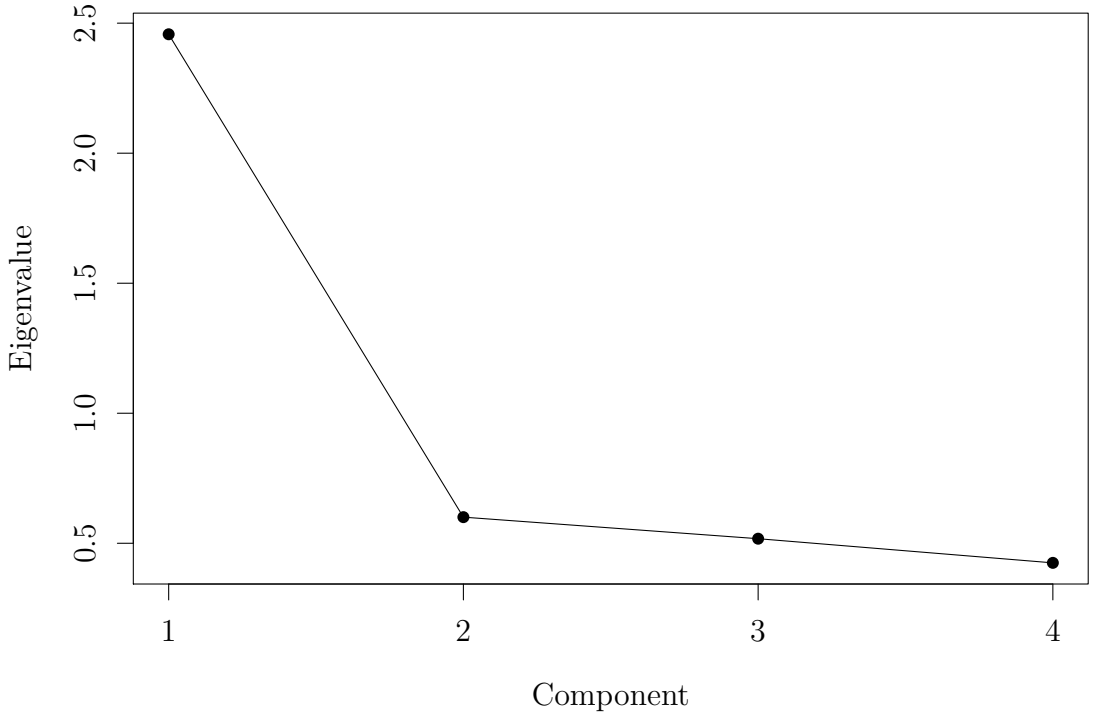
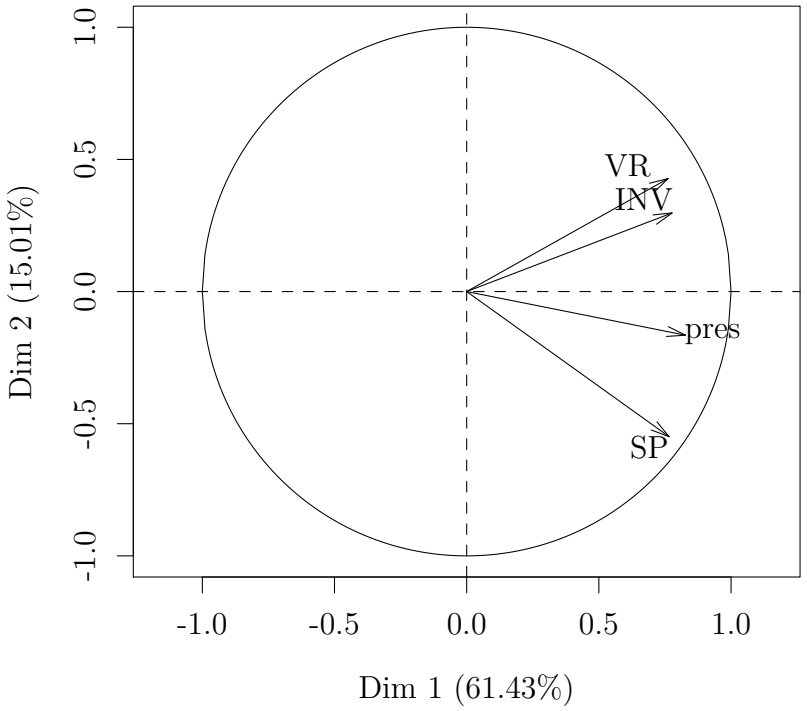


Figure 5.1: Eigenvalues for PCA of presence factors

Table 5.3: General linear hypothesis test (GLHT) statistics for difference in rotated PCA component 1 (Involvement and visual realism) by audio condition

	Estimate	Std. Error	z value	Pr(>|z|)
Intercept	-0.103	0.19822	-0.519	0.96
Early Only	0.06	0.10873	0.560	0.96
Late Only	0.24	0.10873	2.176	0.1
Full	0.33	0.10873	3.016	0.0096

Independent univariate testing for differences by independent variable factors was performed using multilevel ANOVA of mixed effects linear models on the rotated solution (table 5.1 and table 5.2). It was found that in the case of both the 'involvement/visual realism' and 'spatial presence' dimensions inclusion of random effects by participant has the largest effect in explaining the data for involvement/visual realism. Additionally, alteration of impulse response simulation was found to be a significant factor ($p = 0.009$). However, the likely size of this effect is small as the L ratio associated with the inclusion of this variable is relatively small compared with the inclusion of participant level effects. General linear hypothesis tests (GLHT)



	Dim.1	Dim.2
INV	0.78	0.3
pres	0.83	-0.17
VR	0.76	0.43
SP	0.77	-0.55

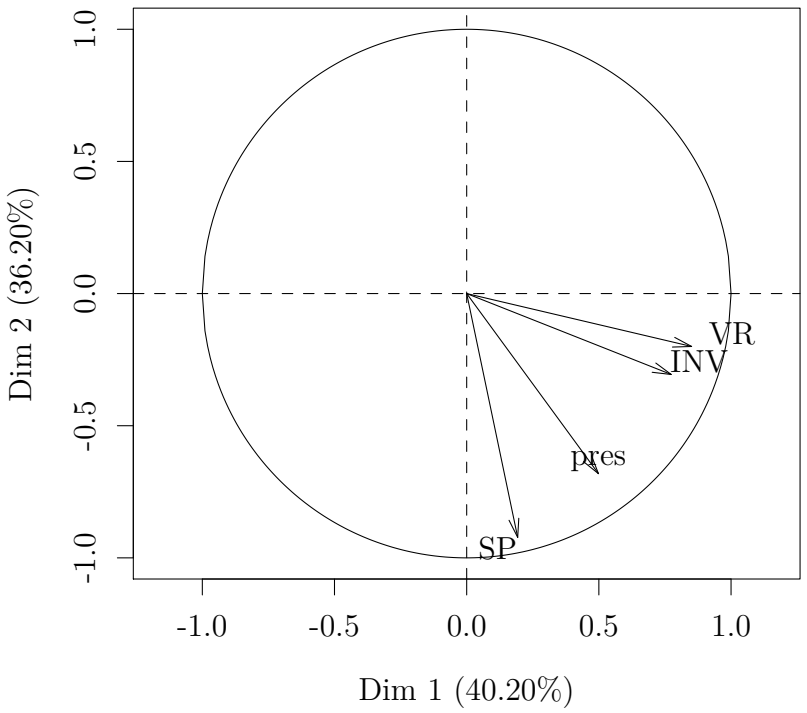
I-group questionnaire:
VR - visual realism SP - spatial presence
INV - involvment
Witmer & Singer
pres - overall presence

Figure 5.2: PCA variable factor map and factor loading table for principal component analysis of reported presence factors

Table 5.4: General linear hypothesis test (GLHT) statistics for difference in rotated PCA component 2 (Spatial presence) by audio condition

	Estimate	Std. Error	z value	Pr(>|z|)
Intercept	0.47	0.16	2.84	0.016
Early only	-0.34	0.12	-2.79	0.019
Late only	-0.67	0.12	-5.45	<0.001
Full	-0.87	0.12	-7.08	<0.001

were performed as a post-hoc procedure (table 5.3). GLHT results show that, when controlling for participant level effects, full impulse response simulation produces higher responses for involvement and visual realism (p = 0.0096). However, the magnitude of this difference is estimated at 0.33 units on the extracted feature space, which in not a large shift. Given the range of responses which are between -3.28



	Dim.1	Dim.2
INV	0.77	-0.31
pres	0.5	-0.68
VR	0.85	-0.2
SP	0.19	-0.92

I-group questionnaire:
VR - visual realism SP - spatial presence
INV - involvment
Witmer & Singer
pres - overall presence

Figure 5.3: PCA variable factor map and factor loading table for principal component analysis of reported presence factors. Varimax rotated

and 2.3 in the extracted dimension, giving a range of 5.58, the estimated shift between anechoic stimuli and full impulse response simulation is approximately 5.91% of the range of responses. This result suggests that although there is some increase in reported visual realism and involvement which can be attributed to increase in order of impulse response simulation, the effect is not large and is easily overshadowed by any intrinsic differences in responses for this dimension. Participant-level random effects were again found to be the most significant factor in accounting for variance in PC2 (reported spatial presence) (Table 5.2). However, impulse response simulation was found to have a larger effect on this component of presence than the previous analysis ($p < 0.0001$, $LR = 54$). GLHT results (Table 5.4) show that average (intercept) PC2 score (spatial presence) for all other impulse response conditions are significantly different from anechoic stimuli. Full simulation produces the

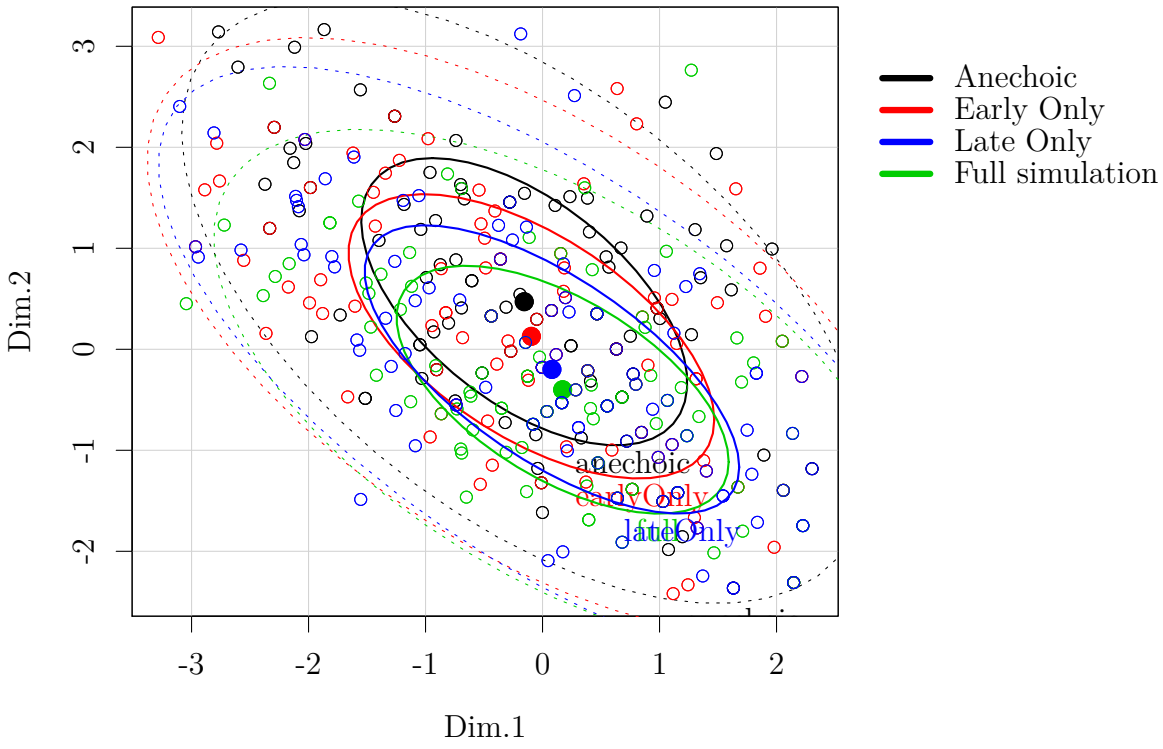


Figure 5.4: Individuals plot of PCA of reported presence subjected to varimanx rotation. Confidence ellipses are drawn for impulse response simulation level. Dotted lines are 95% confidence ellipses, solid lines are 50% confidence ellipses

largest shift in response corresponding to higher spatial presence, producing higher responses of spatial presence of approximately 15%. Analysis of the individuals plot for the rotated data grouped by audio processing condition (figure 5.4), shows movement of the barycentres for responses along the both dimensions. In addition, it is clear that variance in response which is independent of audio is aligned with the loading for the overall presence vector. The combination of these results lend support to the notion that increasing order of simulation of impulse response contributes to spatial presence and, although there is a statistically significant effect on increase in visual realism that can be attributed to audio stimuli, the effect is small.

5.2.4 Conclusions

Reported presence was assessed under exposure to audiovisual stimuli rendered with varying levels of binaural impulse response simulation and varying degrees of audiovisual co-location ambiguity. It was found that spatial presence is largely independent of levels of involvement and ratings of visual realism and, between these two dimensions, contributes towards ratings of overall experience of presence. It was found that changes to the level of spatial audio rendering of stimuli modulates ratings of spatial presence with modest effect, when inter-participant variation was taken into account. It was also found that full impulse response simulation improves ratings on the involvement/visual realism dimension. However, this effect was very small, suggesting that the greatest determinant of experienced presence in similar environments is inter-participant variation.

5.3 Quality of experience metrics and reported presence

5.3.1 Introduction

In section 5.2, measures of reported presence were introduced and investigated to determine their interrelationship and dependence on spatial audio rendering. In chapter 4, it was shown that audiovisual quality of experience was dependent on level of spatial audio rendering and audiovisual spatial co-location ambiguity. The follwing section aims to investigate the relationship between reported presence and audiovisual quality of experience and, in light of any such relationship, how these percepts are affected by changes to audio rendering and audiovisual spatial relationship.

5.3.2 Materials and methods

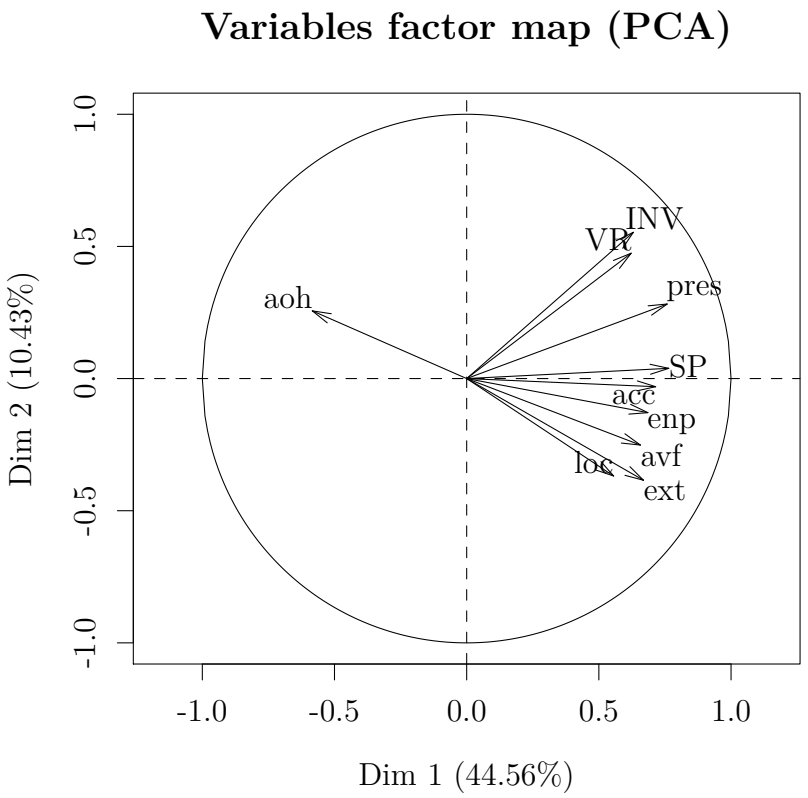
Virtual environments, tasks and stimuli used were identical to those described in section 5.2. In-VE questionnaires were used as with the i-group and Witmer and Singer items. In addition to the questions being presented as described in section 4.2.1, pre exposure briefing was given to participants with terms defined as below:

- Externalisation is defined as the experience of perceiving audio outside of the head and not 'inside' the head or 'on' the ears [17]
- Localisation is defined as the ability to pinpoint the position of the audio source [19][20]
- Audio-visual fusion is defined as the sense that audio is collocated or 'belongs' to a visual stimulus [16]
- Plausibility is defined as the acoustic response of the environment matching that of the visual scene [18]
- Awareness of headphones is defined as being conscious of audio emanating from the headphones [21]

Participants were exposed to stimuli in either the short or long reverb time similar environments as described in chapter 4.3 with audio and audiovisual conditions randomised at runtime.

5.3.3 Results and discussion

All responses were subjected to principal component analysis (PCA) for the purposes of dimensionality reduction and analysis of correlations between response variables. PCA of the response data suggests a one factor model with 9 of 10 constituent components loading with medium to large effect onto that factor. Figure 5.5 shows loadings for this PCA. Eigenvalues and parallel analysis are shown in figure 5.6. Parallel analysis shows that dimensions 2 and above are no more significant than noise in this case. This analysis suggests that responses for spatial presence and overall audio accuracy account for most of the variance observed in this sample. Overall presence is also loaded highly, but has some noise which gives 0.28 loading independence from PC1. Despite the simple structure of this result, inspection of the factor map suggests an alternative interpretation of the results. The audiovisual QoE items are clustered in the lower right quartile of the factor map, with 'awareness of headphones' showing inverse loading. The Witmer and Singer presence model and the i-group items are clustered in the top right quartile and, between the two constructs, show varying degrees of orthogonality. Further analysis was performed by subjecting the PCA to varimax rotation to obtain a two factor solution which maximises loading along input variables. Results for this analysis are shown in figure 5.7. The



	Dim.1	Dim.2
avf	0.66	-0.25
enp	0.69	-0.13
ext	0.67	-0.38
aoh	-0.58	0.26
INV	0.63	0.55
loc	0.56	-0.37
pres	0.76	0.28
VR	0.62	0.47
acc	0.71	-0.031
SP	0.76	0.039

Quality of experience:
aoh - awareness of headphones
enp - environmental plausibility
loc - localisation
avf - audiovisual fusion
ext - externalisation
acc - overall audio realism
I-group questionnaire:
VR - visual realism
INV - involvment
SP - spatial presence
Witmer & Singer
pres - overall presence

Figure 5.5: PCA loadings of all responses

rotated solution, naturally, retains 55% explanatory power for the variance in the responses. However, it can be seen that this produces two dimensions which equally contribute to the perception of spatial presence. Dimension 1 is most heavily loaded by externalisation and localisation, With audiovisual fusion, plausibility and overall rendering accuracy loading equally onto both dimensions. Conversely, dimension 2, consists of involvement in the virtual scene and visual realism. This analysis suggests that although spatial realism can be considered the most important element when assessing multimodal VR environments, in terms of variance in response, it in fact consists of two equal components: visual realism/involvement, and externalisation

of auditory stimuli which have definite localisable positions. It is curious to note that awareness of headphones is dependent on the level of reported externalisation and is independent of the level of attention given to the overall virtual environment. Other quality of experience metrics appear to correlate to varying degrees with spatial presence, however in this context it is neither possible nor appropriate to infer a causal link in either direction. Analysis was performed on dimension 1

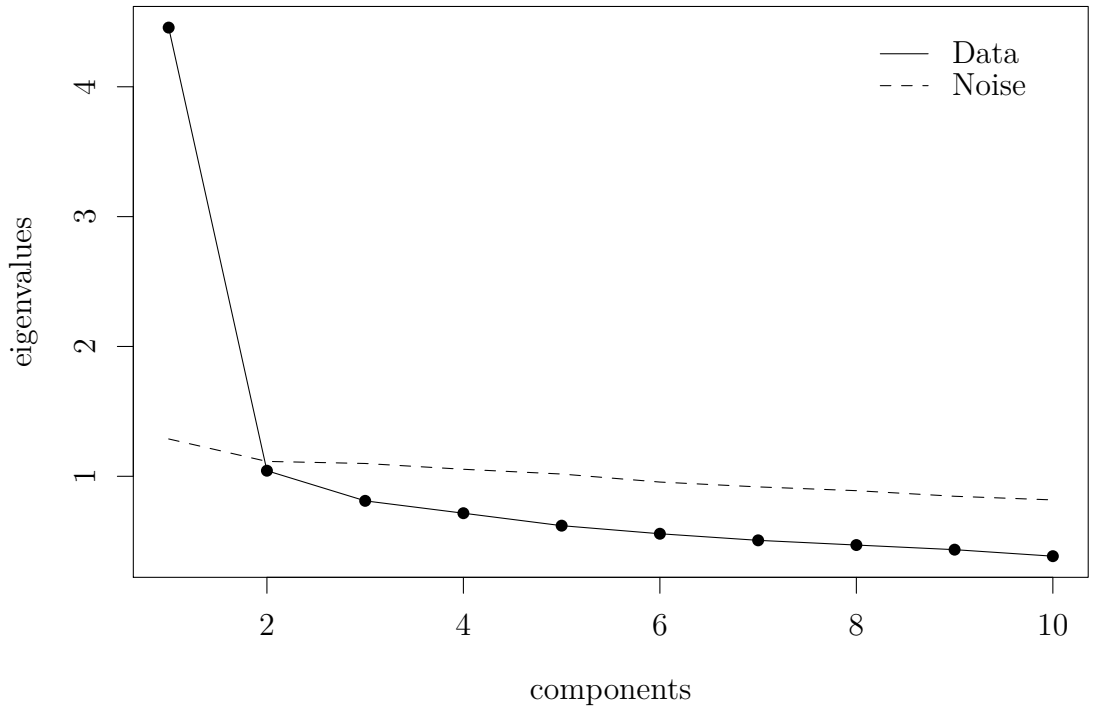
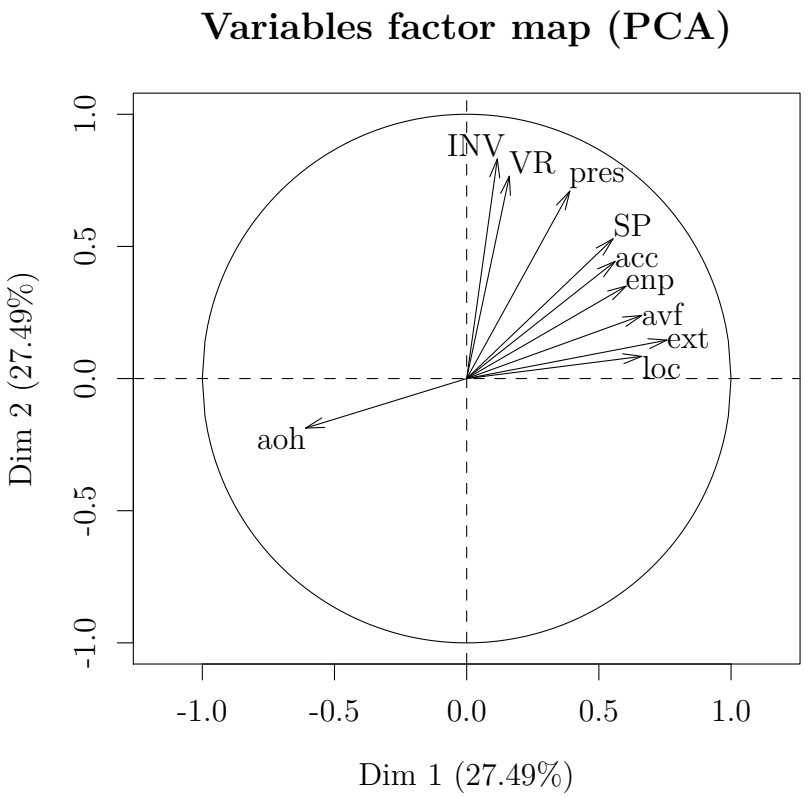


Figure 5.6: Eigenvalues and parallel analysis of PCA of all response variables

of this solution to identify significant effects from independent variables. Multilevel ANOVA of mixed linear models were performed with participant effects included as a random effect to test for significance of inter-participant responses. Table 5.5 show the results for these analyses. Random effects for participant level responses were shown to be significant in both dimensions with likelyhood ratios of 171 (PC1) and 117 (PC2) demonstrating that that the inclusion of participant level random effects significantly improves model fit. This suggests that there is a large component of variance in the responses that is due to the individual differences between subjects. While controlling for participant level effects, it was found that both modelled acoustic response and audiovisual co-location were significant predictors for the



	Dim.1	Dim.2
avf	0.66	0.24
enp	0.6	0.35
ext	0.76	0.15
aoh	-0.61	-0.19
INV	0.12	0.83
loc	0.66	0.084
pres	0.39	0.71
VR	0.16	0.77
acc	0.56	0.44
SP	0.55	0.53

Quality of experience:
aoh - awareness of headphones
avf - audiovisual fusion
enp - environmental plausibility
ext - externalisation
loc - localisation
acc - overall audio realism
I-group questionnaire:
VR - visual realism
SP - spatial presence
INV - involvment
Witmer & Singer
pres - overall presence

Figure 5.7: PCA loadings of all responses subject to varimax rotation

responses on both dimensions, with auditory spatial cues having the largest effect on dimension 1 and an equal effect to co-location ambiguity on dimension 2.

Figure 5.8 shows PCA scores subsetted by audio conditions. It can be seen that, by comparing the angle of the direction of travel in the barycentres with the factor map shown in figure 5.5, that the effect between audio modelling conditions only affects the audiovisual QoE items. The pattern of this effect is comparable to that reported in section 4.3. Figure 5.9 shows PCA scores subsetted by audiovisual co-location conditions. Again, it can be shown that the direction of difference between factors

Table 5.5: Multilevel ANOVA of PC1 of all item response PCA (unrotated)

	Model	df	AIC	logLik	Test	L.Ratio	p-value
Intercept	1	2	1875.4	-935.7335			
Participant as random effect	2	3	1705.8	-849.92	1 vs 2	171.62	<.0001
Audio (Modelling)	3	6	1626.2	-807.09	2 vs 3	85.66	<.0001
Co-location ambiguity	4	8	1587.8	-785.89	3 vs 4	42.41	<.0001
Reverb time (Environment)	5	9	1589.6	-785.81	4 vs 5	0.15	0.6938

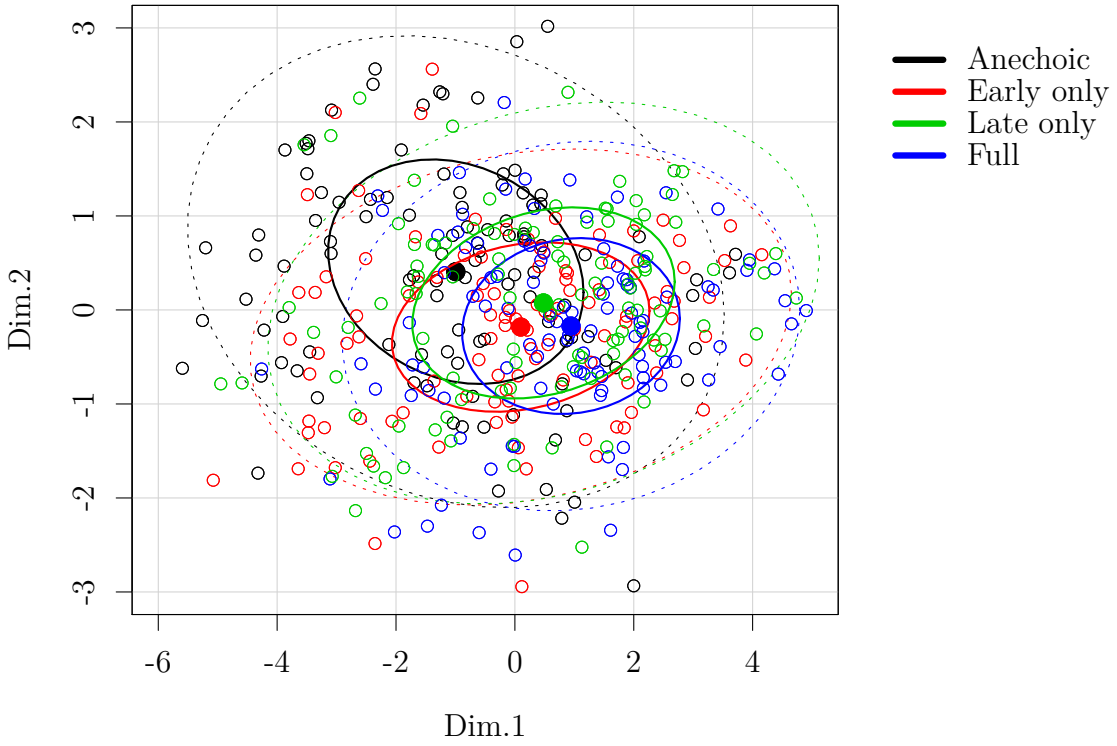


Figure 5.8: Individuals plot of unrotated PCA scores for presence and audiovisual QoE responses with confidence ellipses for audio factors. Inner ellipses (solid) are 50% confidence bounds, outer ellipses (dotted) are 95% confidence bounds

correlates with the audiovisual QoE factors in the unrotated PCA. It can be inferred, then, that although the content of an impulse response simulated using computationally efficient methods can affect the assessment of overall quality of experience of stimuli in a virtual environment which is geometrically similar with pre-exposure conditions, it, in and of itself, does not directly contribute to spatial presence, overall presence, attention and involvement or visual realism. The same can be said for the

relationship between auditory and visual events. Spatial dislocation between audio and visual stimuli lowers quality of experience responses, but semi-ambiguous and explicit co-location of audiovisual stimuli result in higher responses on these items. In contrast, audiovisual ambiguity does not significantly effect the other measures of presence investigated. Particularly, involvement and visual realism. However, it can be argued that spatial presence, which has been posited as the main component of overall presence [79], is a linear combination of two independent factors, visual realism/involvement and audiovisual quality of experience. As such acoustic modelling and ambiguity of audiovisual co-location are significant contributors to this percept.

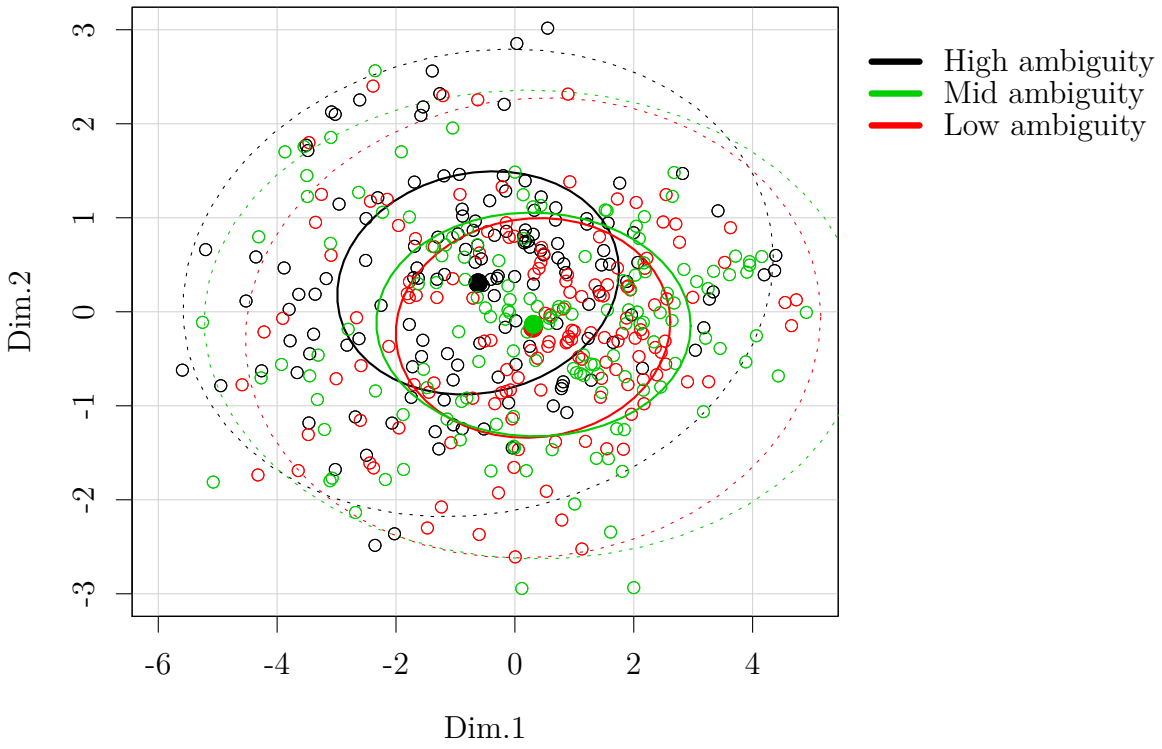


Figure 5.9: Individuals plot of PCA scores for presence and audiovisual QoE responses with confidence ellipses for audiovisual co-location factors. Inner ellipses (solid) are 50% confidence bounds, outer ellipses (dotted) are 95% confidence bounds

5.4 Summary

In this chapter, measures of reported presence have been introduced. It was found that spatial presence is independent of involvement and ratings of visual realism and that these two dimensions contribute to overall presence, with spatial presence providing the larger contribution. It was found that increased content in a modelled spatial impulse response can influence ratings of spatial presence. It was also determined that audiovisual co-location does not have a significant effect on ratings of presence. Work presented in this chapter also demonstrates the relationship between ratings of audiovisual quality and the experience of presence, particularly spatial presence, and that this percept contributes equally with visual realism and involvement to produce this percept. It was further shown that the greatest source of variation in response is due to participant level random effects and, as the explanatory models result in between 50% - 75% of explained variance between experiments, there may be individual level predictors which can further improve predictive power of these models.

6

Individual differences and the judgement of real and simulated sources

Contents

6.1 Overview

This chapter focuses on subject profiling and the extent to which this can be used to reduce the variance observed in the data which has been hitherto attributed to participant level random effects. It was shown in section 4.4 that differences in response to questions pertaining to audiovisual quality of experience, consisting of audiovisual fusion, plausibility, externalisation, localisation and awareness of headphones, can be used to predict whether a stimulus was perceived as either emitted from a loudspeaker or simulated over headphones, while participants were exposed to a virtual environment with similar geometry to the pre-exposure environment. Principal component analysis of subscales of presence and audiovisual quality is also analysed in terms of individual differences. It was found that the inclusion of participant level intercepts and variance provided a large contribution to goodness of fit in the statistical model. In this chapter, sensory cognitive and personality factors will be investigated to determine if quantifiable attributes of individual participants can be used as predictors to reduce the relative contribution of participant level random effects and account for some of the variance observed in the data, allowing better understanding of the trends within the data.

6.2 Auditory and visual sensitivity in the judgement of real and simulated sources

6.2.1 Introduction

In this section work is presented which investigates cognitive-sensory factors and their relation to the classification of perceived stimulus source as described in section 4.4. Aural sensitivity, the ability to correctly identify real or simulated stimuli, bias in classifying stimulus source, and the relative time for perceiving global and local visual structures are measured and correlated with dimensions of quality of experience (QoE) introduced in chapter 4. The effect of differing levels of accuracy of time domain information in the form of comparison between measured impulse responses and room responses simulated with hybrid image-source/artificial reverberation is investigated as well as the influence of the presence of visual components

in a multimodal secene with similar pre-exposure and virtual environment geometry. Two environments were used for the study, and the effect of changes in room size/reverberation time was tested. The purpose of this study was to quantify aural sensitivity and visual cognition within the pool of participants and to identify if this had any influence on the reporting of QoE factors.

6.2.2 Materials and Methods

6.2.2.1 Participants

Twenty-four participants aged between 19 and 37 volunteered to take part in this experiment (16 male and 8 female). Participants were not compensated for their time. After taking part in one set of conditions, the subjects were given the opportunity to volunteer to complete more trials at a later date. As such, not all participants completed the same number of trials under the same number of conditions. All data collection was undertaken in compliance with University of Salford ethical guidelines.

6.2.2.2 Virtual Environments (VEs)

Virtual environments were built in Unity Editor [206] and presented using an HTC Vive head mounted display. Environments were constructed as to be visually analogous to the space in which the experiments were taking place. Two spaces were used for the purposes of data collection. The first was medium sized ($6m \times 7m \times 3.4m$) room which is acoustically treated for the purposes of subjective audio evaluation testing and has a broadband RT_{60} of approximately 230ms. The second space was a small ($3m \times 3m \times 2.5m$) acoustically treated booth designed for speaker based spatial audio reproduction with a broadband RT_{60} of approximately 90ms. Comparisons of real and virtual spaces are given in figures 6.1, 6.2, 6.3 and 6.4. Audio from loudspeakers was emitted from either one of a pair of Genelec 8030a studio monitors positioned at ± 45 deg with respect to the listener. In the larger of the rooms, speakers were positioned at 2 metres from the central listening position, to correspond to source-receiver positions in the soundfield measurements used for ambisonic rendering. In the smaller space, loudspeakers were positioned 1.4m from the listening position.

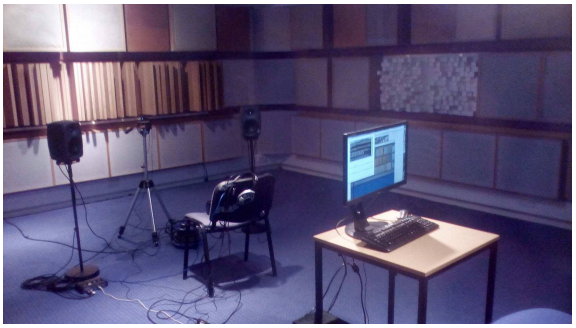


Figure 6.1: Medium ($270ms$) RT_{60} room

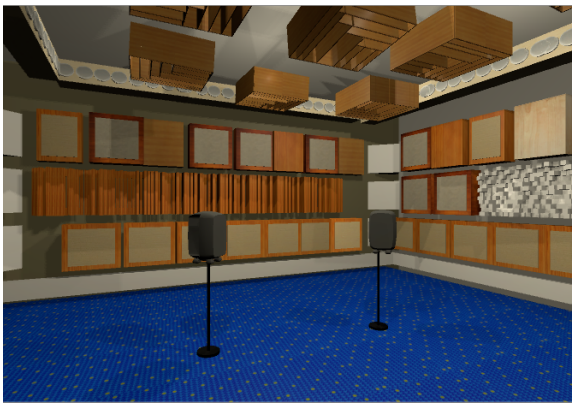


Figure 6.2: VE recreation of medium ($270ms$) RT_{60} room

Figure 6.3: Low ($90ms$) RT_{60} room

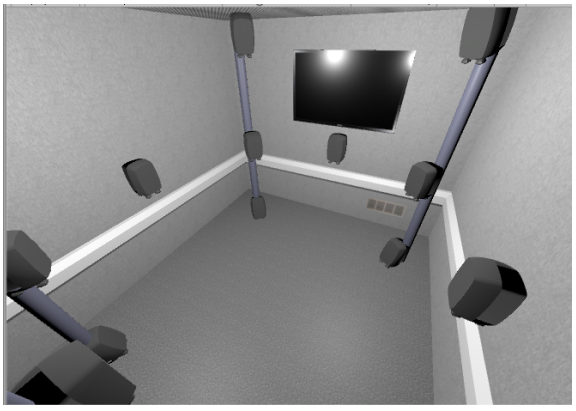


Figure 6.4: VE recreation of low ($90ms$) RT_{60} room

6.2.2.3 Auralisation

Spatial rendering of auditory stimuli was achieved in two ways intended to contrast between a high and low level of physical accuracy. This contrast was not intended to form an investigation into the effects of dissimilarity between room response parameters such as RT_{60} or direct-to-reverberant energy ratio, a subject which has been reported on within the literature [218][219]. The differences between rendering conditions constitute differing temporal and spatial information content arising from the recovery of the first-order ambisonic soundfield in the case of the high physical accuracy condition and the simplified geometric approximation of the impulse response in the low physical accuracy condition. However, in terms of energy decay envelope, the conditions were designed to be similar with the environments in which the experiments were performed. The high physical accuracy condition was achieved by convolving stimuli with ambisonic impulse responses measured at the listening position from source positions described in [214]. Decoding and rendering for playback was performed using the GoogleVR AudioSoundfield object from the GVR Unity SDK [215]. Although is has been demonstrated that B-Format directional encoding results in low levels of accuracy of response in the spatial domain [216], the time-energy content of such measurements can be considered to be a complete representation of the impulse response of the room. As such, the physical accuracy of audio rendered by this method was considered high for the purposes of this study. Low physical accuracy stimuli were processed using an image source algorithm which assumed a 'shoebox approximation' of the space for the early component of the impulse response (IR) and used the inbuilt reverb processor in Unity Editor for the diffuse component of the IR. Direct path and early reflections were convolved with HRTFs from the KEMAR compact dataset [205]. Values for the RT_{60} of the late part of the simulated response were obtained by backward integration [220] of the W channel of the B-format measurements used in the converse rendering condition. Headphone playback was achieved using Sennheiser HD800 headphones. To reduce the opportunity for consistent level differences between headphones and loudspeakers to be used as a cue, playback level was randomised between 54dBA and 68dBA at the listening position for both headphone and speaker playback. Audio that was rendered over headphones was also processed to take into account the impulse response of the loudspeakers used and the transmission of the sound through the headphone ear-cup. Loudspeaker impulse responses were obtained from [214]. Filters to account for the transmission of sound through the headphones were obtained by first recording the measurement signal at 0.5m at 90 degrees azimuth through the

Figure 6.5: Example of an incongruent Navon embedded figure

ipsilateral inner ear microphone of a Brüel & Kjær head and torso simulator (HATS). A second signal was recorded with the headphones in place and these two signals were deconvolved to produce an impulse response which approximated the occlusion effect of the headphones. Audio which was to be rendered over headphones was first filtered with these two impulse responses in MATLAB before further processing.

6.2.2.4 Auditory sensitivity (d') and bias (c)

Auditory sensitivity and bias statistics were collected using the procedure described in section 6.3.

Global Precedence - Global precendence was determined using Navon embedded figures [140]. The test consists of letter figures made up of smaller figures, in this case 'S' and 'H'. Participants were instructed to identify either the smaller (local) or larger (global) figure in a combination of congruent or incongruent conditions, where global and local figures were the same or differed. An example stimulus is shown in figure 6.5. Reaction time for both trials was recorded and the ratio between global and local response times was used as a proxy for differential cognitive load required for global and local processing.

Audiovisual quality of experience - After the externalisation task was complete and participants had removed the HMD, participants were asked to complete a questionnaire about their audiovisual experience while performing the task. Participants were asked to reference conditions in which virtual speakers were seen and to make their response to items described in section 4.2.1 for decisions about whether the audio was produced by headphones or speakers.

6.2.3 Results and discussion

6.2.3.1 Auditory sensitivity

Signal detection theory metrics were calculated on a per participant basis. Sensitivity and bias were calculated from the pooled headphone/speaker decision of participants across all factors (Figure 6.6).

- Only one participant demonstrated an overall d' lower than the critical value (p <0.55) recommended as a threshold of plausibility [18].
- 75% of participants demonstrated d' between 0.05 and 3, corresponding to range of probabilities of detection between 0.51 and 0.98
- Of these 75%, all participants demonstrated absolute bias ($|c|$) of less than 1
- 25% of participants responses demonstrated very high d', saturating the model due to truncation of small numbers (>10, p >0.999), however, this was mostly associated with very high bias (c)

Separating responses by experimental design factors does not alter the distribution of d' scores so that there is a significant change in plausibility within the definition of the statistical model used. To identify variation in task performance between experimental conditions, total rates of true positive, false positive, true negative and false negative were analysed using the χ^2 test of association. For the purpose of this analysis, results were referenced to the ability of the participant to detect the headphone signal, such that a true positive was the correct identification of simulated audio.

It was found that reverb time had no significant effect on result frequency ($\chi^2 = 3.08$, d.f. $= 3$, p $= 0.37$). There was a significant association observed between the rendering method used and misidentification rates (Table 6.1). χ^2 contributions for

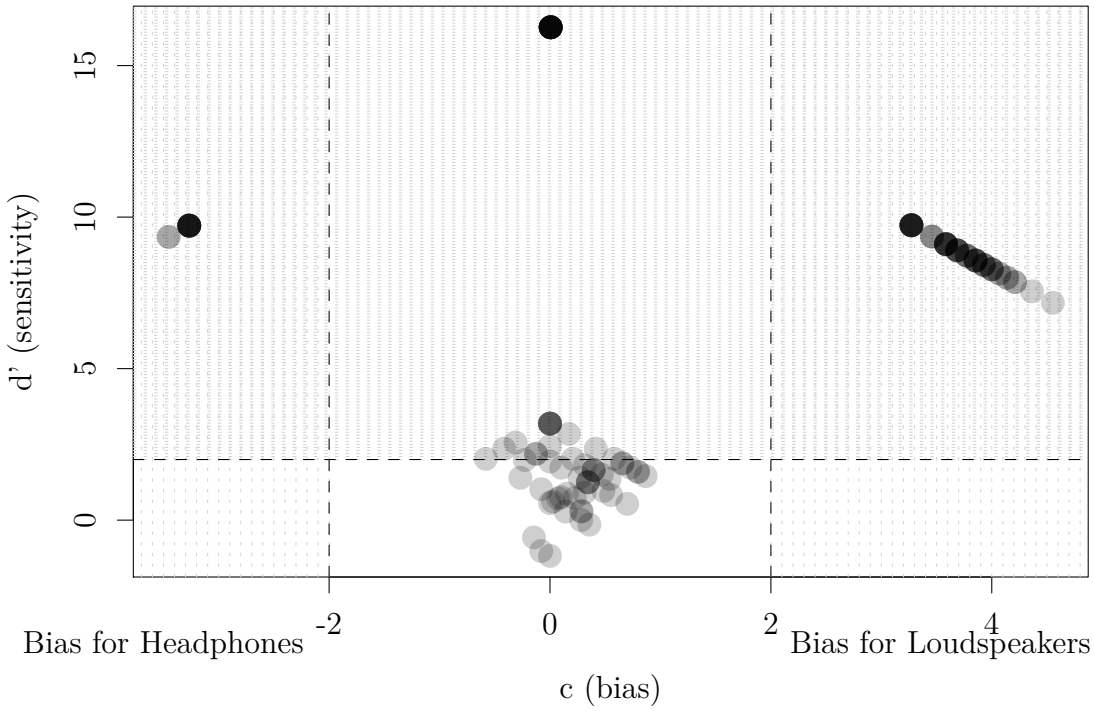


Figure 6.6: Sensitivity and bias of participants for all experimental conditions. Shaded areas indicate p >95% regions for scores. Darkness indicates repeated values for scores

false positive frequencies in both measured IRs and modelled responses suggest a significant difference in the rate of mistaking audio emitted from the loudspeakers as being emitted from the headphones, with measured responses having a greater incidence of false positives by a factor of 2.8. The Cramer's V of 0.11 indicates a medium effect size on three degrees of freedom.

Additionally, comparisons between frequencies of false positive and negatives between visual conditions show a greater incidence of false positives and a reduction of false negatives in conditions where the room was visible and the speakers were not, and an increase in attribution of headphone rendered audio as emitted by loudspeaker in conditions where there was no visual stimulus compared to conditions where both room and speakers were visible. This is suggestive of an association between visual conditions and the rate of false identifications (Table 6.2) suggesting that the presence of a visual component to the stimulus has an effect on the perception of the event. When virtual speakers were invisible, there was a small but

Table 6.1: Cross tabulation and χ^2 test of association between rendering types

	Measured responses	Image Source model
True Pos.	1068	1044
Expected	1113.2	998.7
χ^2	1.837	2.048
True Neg.	1291	1280
Expected	1355.2	1215.8
χ^2	3.038	3.386
False Pos.	**221**	**78**
Expected	**157.6**	**141.4**
χ^2	**25.50**	**28.43**
False Neg.	445	312
Expected	399	358
χ^2	5.301	5.908

$\chi^2 = 75.44$
d.f. $= 3$
$p = 2.9 \times 10^{16}$
Cramer's V $= 0.11$

significant increase in the number of stimuli incorrectly identified as being simulated, and a corresponding decrease in the number of stimuli incorrectly identified as being emitted from loudspeakers. Additionally, when no visual information was provided, the rate of stimuli incorrectly identified as loudspeakers increased. However, the Cramér's V for this analysis (Cramer's V = 0.039, d.f. = 6) is below the 'small' threshold for this number of degrees of freedom and can be considered negligible.

To test for an interaction between rendering type and visual condition, χ^2 tests were performed to look for association between visual conditions independently by rendering type. It was found that there was no significant association between visual conditions when auralisation was performed using measured responses ($\chi^2 = 6.18$, d.f. = 6, p = 0.4). However, the use of modelled responses shows a significant association between the rate of false negatives and the presence of any visual stimulus. The cells with the highest χ^2 contributions in this case indicate that judging that simulated sources are emitted from headphones is significantly less likely when the room was visible but virtual loudspeakers were not with this being significantly more likely when no visual stimulus was presented (Table 6.3).

Due to the high variation between high and low bias participants identified in the

Table 6.2: Cross tabulation and χ^2 test of association between visual conditions

	Speakers visible	Speakers invisible	Nothing visible
True Pos.	699	739	674
Expected	704	704	704
χ^2	0.036	1.74	1.27
True Neg.	869	836	866
Expected	857	857	857
χ^2	0.168	0.515	0.095
False Pos.	89	**119**	91
Expected	99.67	**99.67**	99.67
χ^2	1.142	**3.75**	0.754
False Neg.	256	**219**	**282**
Expected	252.3	**252.3**	**252.3**
χ^2	0.053	**4.403**	**3.488**

$\chi^2 = 17.42$
d.f. $= 6$
$p = 0.0079$
Cramer's V $= 0.039$

initial analysis, differences in rates of false positive and false negatives could be accounted for by differences in the biases of participants in each condition group. To discount this hypothesis, bias scores were subjected to Kruskal-Wallis test to determine if there were significant differences in the bias scores between subjects who completed the tests in different conditions. No signficant difference was found between room groups ($p = 0.98$) or rendering type groups ($p = 0.8$).

The results suggest that the use of measured B-format impulse responses rendered using commercially available tools produces results which perform better than simplified real time models, which produced a statistically significant increase in identification error. However, it cannot be said that any condition provided plausible auralisation such that participants were not able to distinguish between reality and simulation. Furthermore, it should be noted that the type of error which is more likely when using a more physically accurate representation is one of more conservative estimation. The measured impulse responses elicited a greater number of responses in which audio that was emitted from loudspeakers were perceived as being emitted by headphones. The results also suggest that the influence of visual

Table 6.3: Cross tabulation and χ^2 test of association between visual conditions for modelled responses only

	Speakers visible	Speakers invisible	Nothing visible
True Pos.	351	371	322
Expected	347.74	348.12	348.12
χ^2	0.030	1.5	1.96
True Neg.	431	418	431
Expected	426.35	426.82	426.82
χ^2	0.05	0.182	0.041
False Pos.	23	33	22
Expected	26	26	26
χ^2	1.34	1.78	0.62
False Neg.	99	**83**	**130**
Expected	104	**104**	**104**
χ^2	0.23	**4.245**	**6.478**

$\chi^2 = 17.57$
d.f. $= 6$
$p = 0.0074$
Cramer's V $= 0.054$

stimuli on source identification is dependent on the level of realism of the audio rendering. There was a small main effect due to presentation of visual cues in the virtual environment. Incorrectly identifying headphone rendered audio as a real source was less likely when a representation of the room was shown, but without virtual visual sources, but was more likely when no image was displayed to participants. This effect was not present when audio was rendered using measured responses but was marginally magnified when audio was spatialised using a simplified geometric model.

6.2.3.2 Visual sensitivity

Descriptive statistics for global precedence scores are given in table 6.4. As can be expected, there is a tendency for faster processing of global features. Figure 6.7 shows kernel densities of response times for correct responses to local and global stimuli. The majority or participants required between 9% and 25% more time to process local features than the overall structure of the stimulus (Figure 6.8).

Linear regression was used to identify any association between d' and c for auditory stimuli and global precedence. In both cases, auditory sensitivity and bias

Table 6.4: Descriptive statistics for global precedence responses

	1st Q	Median	Mean	3rd Q
Global RT (s)	0.44	0.53	0.60	0.61
Local RT (s)	0.53	0.61	0.70	0.80
Local/global ratio	0.8	0.84	0.85	0.92

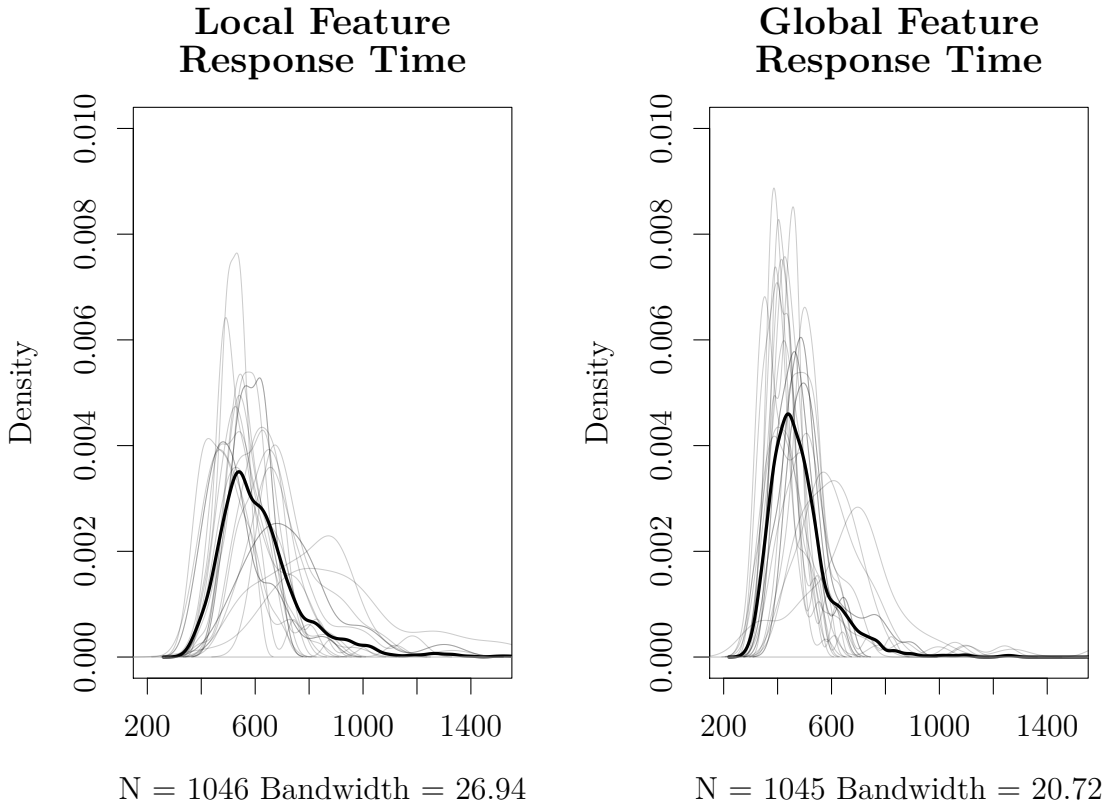


Figure 6.7: Kernel densities for local and global reaction times. Heavy lines are for all participants, grey lines are individual participants

were found to be independent of global precedence for visual stimuli (d' global precedence: p = 0.29, c global precedence: p = 0.24).

6.2.3.3 Multimodal quality of experience (QoE) and sensitivity metrics

Five point QoE questionnaire results across all conditions including both headphone and loudspeaker decisions are presented in table 6.5. Questionnaire responses were subjected to principal component analysis (PCA) for the purposes of dimensionality reduction. PCA vector maps for aggregated data, responses for 'loudspeaker' decisions and responses for 'headphone' decisions were generated (Fig. 6.9).

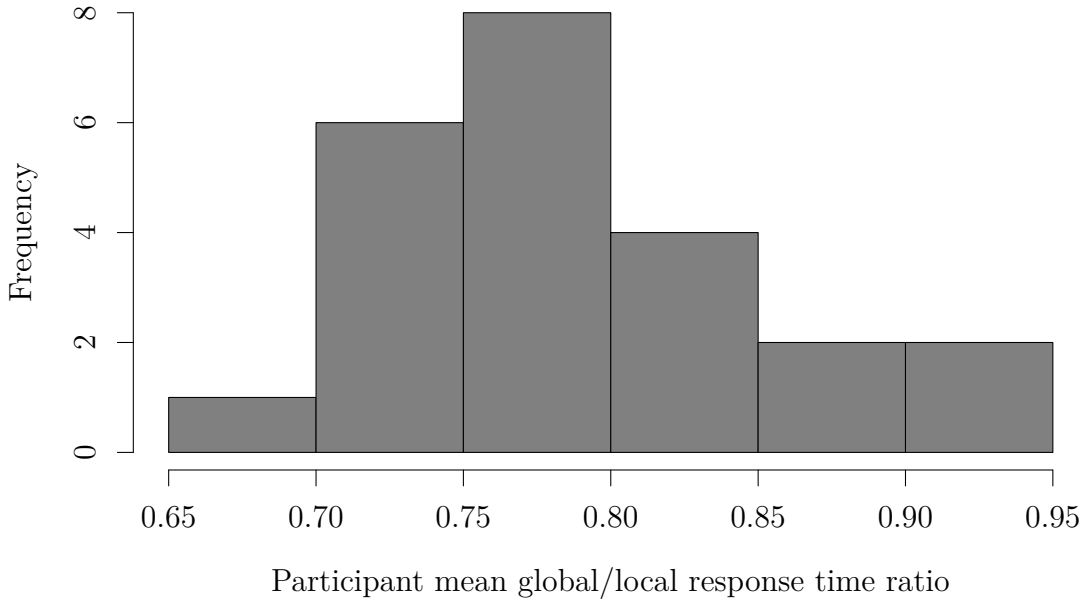
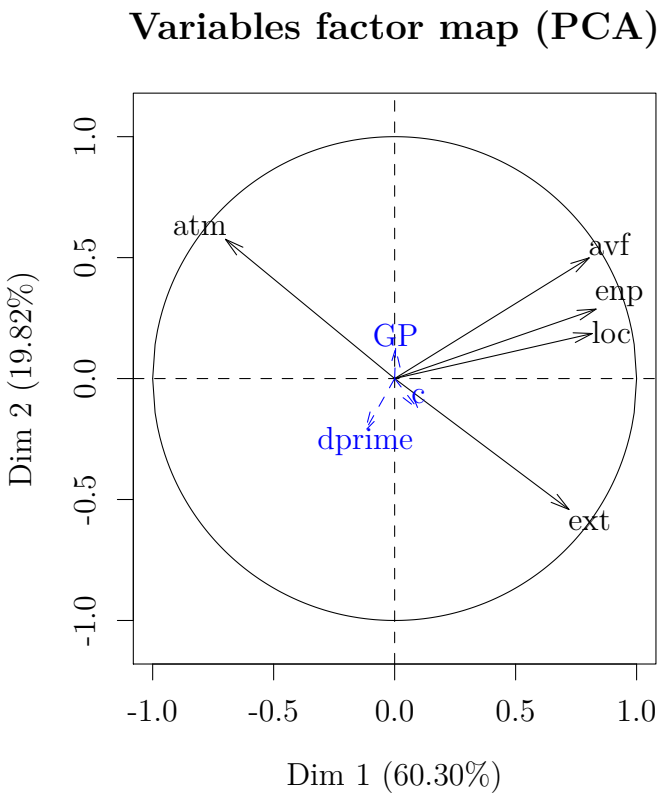


Figure 6.8: Histogram of global/local response time ratios for the Navon embedded figures task

Table 6.5: Descriptive statistics for QoE responses

	1st Q	Median	Mean	3rd Q
A/V fusion	3	5	4.62	6
Atten. to Playback	2	3	3.65	5
Env. Plausibilty	4	5	4.8	6
Externalisation	4	6	5.16	7
Localisation	4.25	6	5.2	6

In all three cases there are only two significant dimensions (Eigenvalues greater than 1), accounting for between 72% - 78% of the variation in the data. Awareness of headphones and reported externalisation are inversely correlated, with audiovisual fusion, localisation and environmental plausibility independent of these factors. Loadings reflect results that have been previously reported, with replication of the independence between perceptual and representational attributes [221]. This suggests that the five attributes used describe two percepts, the first describing physical properties of the stimulus event and the second relating to how it is perceived. When loudspeaker and headphone judgements are aggregated, there appears to be a smaller degree of independence between the representational features and externalisation. It



	Dim.1	Dim.2
avf	0.8	0.5
loc	0.82	0.19
enp	0.83	0.29
ext	0.72	-0.54
atm	-0.7	0.58

	Dim.1	Dim.2
dprime	-0.12	-0.21
c	0.098	-0.13
GP	0.0041	0.12

PCA variables (QoE):
aoh - awareness of headphones
avf - audiovisual fusion
enp - environmental plausibility
ext - externalisation
loc - localisation

Supplementary variables:
dprime - D' (auditory sensitivity)
c - bias
GP - global precedence

Figure 6.9: PCA factor loadings for all data.

can be seen in figure 6.10 that there is significant difference along the first dimension between judgements of headphone or loudspeaker playback and a smaller apparent, but opposite difference, along PC2. There is a clear distinction and separation of 95% confidence ellipses.

Stimuli judged to be originating from loudspeakers scores generally higher on PC1 corresponding to higher scores for audiovisual fusion (AVF), environmental plausibility (ENP), localisation (LOC) and externalisation (EXT). Conversely, audio judged as originating from headphones shows lower scores on these measures and higher for awareness of headphones (AoH). The difference between response profiles for loudspeaker and headphone judgements becomes evident when questionnaires for speaker and headphone responses are subjected to separate PCA analysis (Fig 6.11

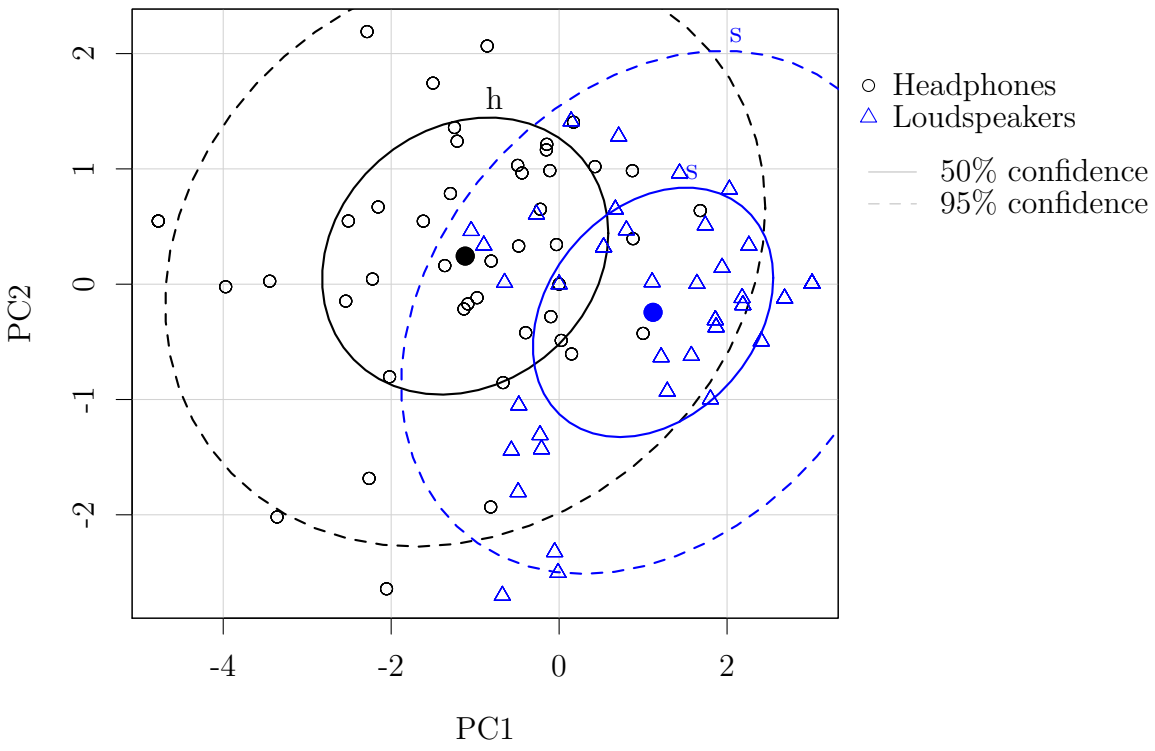
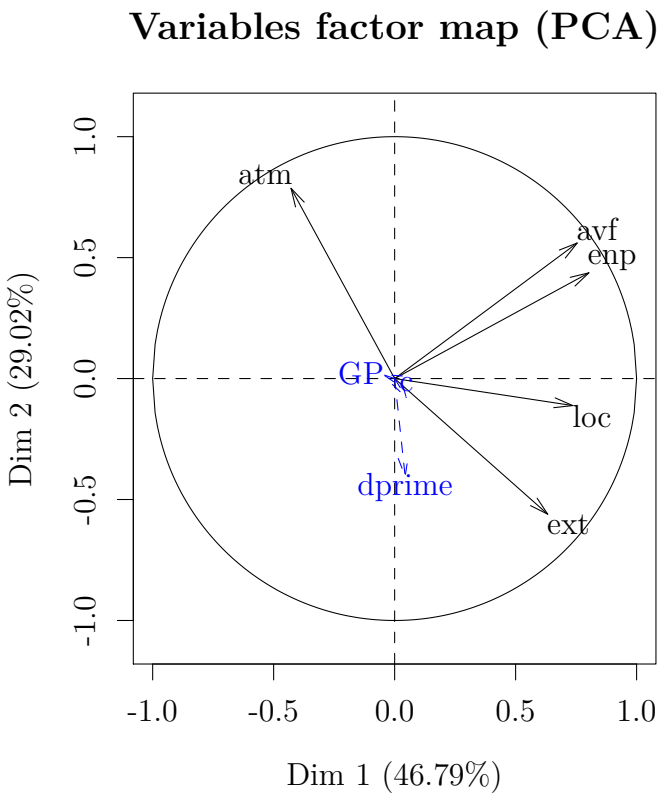


Figure 6.10: Individual responses clustered by judged stimulus origin

and Fig. 6.12). When participants believe they were presented with a loudspeaker source, there appears to be marginally greater independence between audiovisual fusion and localisation. However, when participants judged the sound to be emitted from the headphones, these factors become more closely correlated and more heavily loaded on to dimension 1. This suggests that when audio is perceived as external to the virtual world, localisation is, to some extent, associated with the externalisation of the audio source. However, when audio is perceived as being simulated, the degree of localisation experienced was more associated with the apparent co-location of the visual component of the virtual sound source.

PCA loadings of questionnaire responses associated with judgements made on the origin of audio stimuli were generated to identify correlations with psychometric and signal detection measures (Fig. 6.11 and Fig. 6.12). It was found that for 'loudspeaker' judgements (Figure 6.11), auditory sensitivity (d') was the only supplementary measure to have a correlation coefficient greater than 0.3 [$R_I = 0.264, R_{II} = -0.392$]. d' is loaded on to dimension 2, which shares loading with externalisation to



	Dim.1	Dim.2
avf	0.75	0.56
loc	0.74	-0.11
enp	0.8	0.44
ext	0.63	-0.56
atm	-0.43	0.79

	Dim.1	Dim.2
dprime	0.043	-0.4
c	0.048	-0.079
GP	-0.047	0.016

PCA variables (QoE):
aoh - awareness of headphones
avf - audiovisual fusion
enp - environmental plausibility
ext - externalisation
loc - localisation

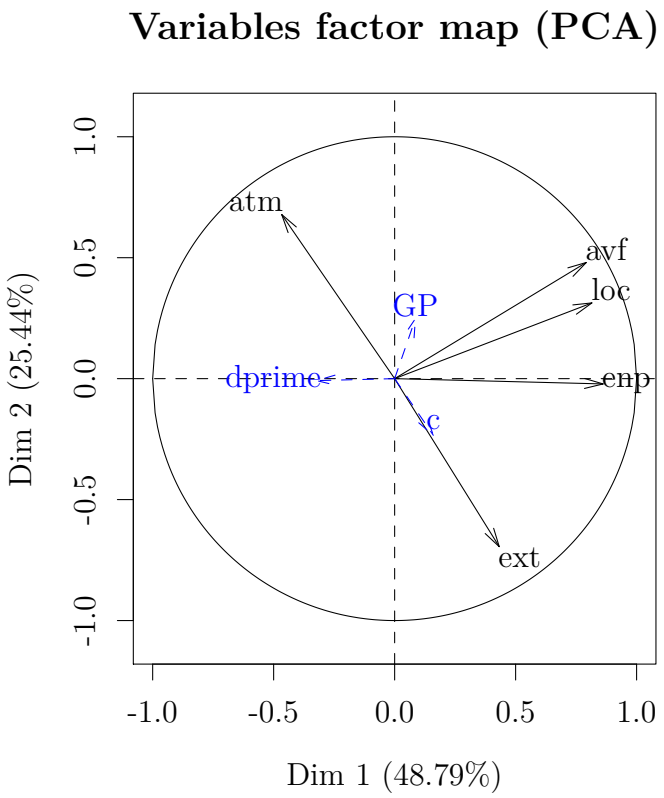
Supplementary variables:
dprime - D' (auditory sensitivity)
c - bias
GP - global precedence

Figure 6.11: PCA factor loadings for loudspeaker judgements

some extent. This indicates that ability to distinguish between real external sources and simulated stimuli is correlated with an increase in reported externalisation when those decisions are made. It may also suggest that there is an association between the degree of externalisation reported when perceiving real sources and the accuracy of that perception.

In the case of 'headphone' judgements, d' is inversely loaded on to PC1 by a factor of 0.29, suggesting a small effect where participants who were better able to distinguish between real and simulated audio rated stimuli perceived as originating from the headphones as lower on this dimension.

Analysis of clustering of individuals on the extracted dimensions from both subsetted PCAs suggested no significant difference between groups on dimensions I or



	Dim.1	Dim.2
avf	0.79	0.48
loc	0.82	0.31
enp	0.86	-0.022
ext	0.43	-0.69
atm	-0.47	0.68

	Dim.1	Dim.2
dprime	-0.32	-0.0081
c	0.16	-0.23
GP	0.086	0.25

PCA variables:
aoh - awareness of headphones
avf - audiovisual fusion
enp - environmental plausibility
ext - externalisation
loc - localisation

Supplementary variables (QoE):
dprime - D' (auditory sensitivity)
c - bias
GP - global precedence

Figure 6.12: PCA factor loadings for headphone judgements

Table 6.6: Kruskal-Wallis test results for extracted dimensions by reverb time and rendering type

	K-W χ^2	p-value
RT_{60} Dim. I	2.13	0.14
RT_{60} Dim. II	0.27	0.59
Rendering Dim. I	1.32	0.24
Rendering Dim. II	0.77	0.37

Table 6.7: Regression analyses of cognitive factors with PC1 differentials between loudspeaker and headphone stimulus origin judgements

	Coefficient	R^2	p-value
d'	0.05	0.012	0.07
c	-0.03	0.003	0.59
Global precedence	-2.97	0.018	0.035

II (Table 6.6) suggesting no significant difference in response between reverb time and rendering type conditions. As such, the variation along PC1 of the PCA of the responses to the QoE items for pooled source origin judgements can be explained by differences in perceived overall audiovisual quality; with loudspeakers perceived as being of higher audiovisual quality within the VE. It was hypothesised that either auditory sensitivity or bias could explain the differentials between responses. Distances between headphone or loudspeaker source judgements were calculated as in section 4.4.3 using the following:

$$x_{\Delta} = x_{speakers} - x_{headphones} \tag{6.1}$$

where $x_{speakers}$ is the pca individual score on the first dimension for loudspeaker judgements and $x_{headphones}$ is the equivalent value for judgement that a stimulus originated from headphones. This value was analysed using linear regression to identify whether auditory sensitivity, bias or global precedence were significant predictors.

It can be seen in table 6.7, that global precedence is a statistically significant predictor of diffecence in QoE response between perceived origin of stimulus. However, this result should be treated with caution, as the effect observed is in the negligible range (0.018), indicating a large residuals within the model. It should also be noted that neither d' or c were significant predictors of distance between ratings. This suggests that the reported quality ratings, although a measure of perceived difference of stimulus origin, are independent of the actual ability to discriminate or the tendency to favour one potential source over another.

6.2.3.4 Conclusions

An investigation into the plausibility of measured impulse responses and a simplified acoustic model of small and medium rooms in visually similar virtual environments was carried out. It was found that the overall plausibility of both reproduction techniques did not result in plausible auralisation as defined by signal detection theory

Table 6.8: Anova of random effects vs intercept only model for PC1 differentials between speaker and headphone judgements

	Model	df	AIC	logLik	Test	L.Ratio	p-value
Intercept only	1	2	1205.22	-600.61			
Participant as random effect	2	3	939.76	-466.88	1 vs 2	267.5	<.0001
RT_{60}	3	4	941.7	-466.85	2 vs 3	0.05	0.8215
Rendering method	4	5	895.82	-442.91	3 vs 4	47.87	<.0001

Table 6.9: Anova of random effects vs intercept only model for PC2 differentials between speaker and headphone judgements

	Model	df	AIC	logLik	Test	L.Ratio	p-value
Intercept only	1	2	887.7	-441.85			
Participant as random effect	2	3	658.48	-326.24	1 vs 2	231.22	<.0001
RT_{60}	3	4	632.87	-312.43	2 vs 3	27.61	<.0001
Rendering method	4	5	634.07	-312.03	3 vs 4	0.8	0.37

metrics. However, the results indicate that the absence of visual sources in the VE reduced the number of simulated stimuli being identified as being emitted from the loudspeakers, while the complete absence of visual stimuli increased the frequency of misidentification of simulated audio as being emitted from the loudspeakers. This was matched with a converse pattern in terms of false positives, where loudspeaker emitted audio was perceived as being emitted from the headphones. This effect appears to be dependent on the method of audio rendering used. However, the effect sizes for these results were small. It was also shown that although neither auralisation method produced objectively convincing results, there was significant difference in the rates of false negatives, with the measured impulse responses demonstrating significantly higher rates of loudspeaker emitted audio being identified as being emitted from the headphones.

6.3 Personality differences in the reporting of QoE in judging real or simulated sources

6.3.1 Introduction

In this section work is presented which investigates personality factors and their relation to the classification of perceived stimulus source as described in section 4.4. Empathy, systeamtisation of information, immersive tendencies and the five factor model (big five) were measured and correlated with dimensions of quality of experience (QoE) introduced in chapter 4 using the same environments and stimuli as the work described in section 6 and 6.2.

6.3.2 Materials and methods

Empathy was assessed using the EQ/SQ-short questionnaire [222], a tool designed as part of the diagnostic battery for the assessment of autistic spectrum disorders [223]. Although the use of empathy tests has been cited in the literature [177], its association with the identification of autistic traits is not always highlighted in the context of VR-based studies, which have identified empathy as a potential correlate to the experience of presence. This test aims to determine two independent and complimentary metrics, empathy and propensity for systematisation. Empathy is defined as the tendency to identify emotionally with others and systematising is taken to be the tendency to organise and categorise information. However, there is some body of work which suggests that empathy is a semantic grouping of a wide range of traits which may have a more complex structure [224][225][226]. Within the EQ/SQ model, the two quantities are assumed to be independent. This test takes the form of a 7-point Likert scale and is included in Appendix B. Propensity for the experience of presence was measured using the immersive tendencies questionnaire (ITQ) [22]. This took the form of a dichotomous forced choice test and is included in Appendix B Additionally, participants were asked to complete a five factor model, or 'big five', personality inventory [227]. This inventory classifies subjects on five continuous dimensions: Extraversion, neuroticism, contentiousness, openness, and agreeableness. This took the form of a 5 point Likert scale and is included in Appendix B.

Table 6.10: Descriptive statistics for EQ/SQ responses

	Min	1st Q	Median	Mean	3rd Q	Max
EQ	20	26	28	27	28	32
SQ	12	18	20	20.65	24.5	26

Table 6.11: Descriptive statistics for ITQ responses

	1st Q	Median	Mean	3rd Q
Focus	22	25	24.86	28
Involvement	16	20	19.34	23
Emotion	18	19	19.21	21
Gaming	11	15	14.59	17
Totals	72	79	77.35	82

6.3.3 Results and discussion

6.3.3.1 Personality factors

Empathy and systematisation quotients were tabulated using the procedures described in [222]. Items corresponding to either empathy or systematisation preference are summed independently to obtain a metric for those dimensions. Table 6.10 shows descriptive statistics for both metrics.

Shapiro-Wilks tests for normality showed that both data for EQ and SQ did not conform to a normal distribution (EQ: $p < 2.9 \times 10^{-7}$, SQ: $p < 1.4 \times 10^{-7}$). This may be related to the sample size of the study. Despite this, both mean and median empathy quotient (EQ) and systematisation quotient (SQ) scores are within one standard deviation of results reported in the literature. The validation report of this instrument (N = 1761) [222] reports a mean EQ of 22.8 with a standard deviation of 8.75 and SQ of 19.0 (sd = 10.05). It should be noted that the range of responses is quite narrow, with an interquartile range of 2 points for empathy quotient and 12 points between the minimum and maximum values.

Immersive tendencies scores were computed as described in [228]. Responses were found to be non-normal using Shapiro-Wilks testing ($p = 0.0008$). Descriptive statistics are presented in table 6.11. Again, mean and median total ITQ scores lie within one standard deviation of those reported in a previous study (Mean = 70.69, sd = 10.52) [172]. The results presented indicate that the participants in this study are representative of the general population for these metrics.

Table 6.12: Descriptive statistics for Big-5 responses

	1st Q	Median	Mean	3rd Q
Extraversion	2.25	2.375	2.44	2.72
Neuroticism	2.16	2.375	2.34	2.59
Conscientiousness	2.25	2.44	2.48	2.64
Openness	2.25	2.65	2.63	2.9
Agreeableness	2.25	2.38	2.36	2.52

Values for the big five questionnaire were computed as described in [227]. Item response scores were computed by factor association, with reversed response items inverted, and means for each factor obtained. This produced five zero-referenced values with a maximum possible value of +4. Big five responses were tested for normality for each of the factors. Data were found to be normal ($p > 0.05$). Descriptive statistics are presented in table 6.12. As can be seen, the interquartile ranges for these values are quite small, with 50% of responses falling into a range of <0.75 for all five factors.

It should be noted that in the case of a large number of measured variables, it is advantageous to perform some form of dimensionality reduction for the purposes of further analysis. However, in this case it can be argued that this approach would be inappropriate. The structure of the EQ/SQ framework and the five factor model (big 5) are such that each component within the collected computed dataset should be treated as an independent quantity [222][229]. Furthermore, it lies outside of the scope of this work to infer relationships between the ITQ and personality factors in a general due to the sample size constraints. This notwithstanding, the level of independence between the collected data was investigated to inform further analysis. Table 6.13 shows the correlation matrix for the personality items collected. It can be seen that, in the case of this dataset, there is not independence between factors. Many factor pairs exhibit large ($|R|>0.6$), medium ($|R|>0.3$), or small ($|R|>0.1$) correlation coefficients. This correlation is likely due to the size of the dataset and sampling bias inherent in the selection procedure. Sampling bias due to relying on volunteers available within faculties within universities has been observed within the literature. It has been observed that studies of this kind are vulnerable to selection bias due to the available pool of participants [230]. In this instance we can observe this in the results presented in table 6.12. Not only are the interquartile ranges small, but 50% of the data for all dimensions of the big-5 model are above the midpoints of the scale. As a five point, zero referenced scale, an unbiased sample would show central tendencies around 1.5. However, this sample as a whole displays tendency

towards extraversion, neuroticism, conscientiousness, openness and agreeableness. While it may not be possible to fully account for this bias, there may be some explanations for the tendencies observed in the data. It has been reported that there is a minor extraversion bias in the general population [231]. It has also been shown in a large scale (N = 1472) study of university students and faculty, engineering departments exhibit students with higher than average conscientiousness [232]. It has also been reported that data collected from a population of 'high achieving' was shown to have higher than average metrics across all five dimensions of the big 5 model [233]. With this in mind, the analysis undertaken in this study should be read with the understanding that the limitations described above mean that some results may not be generalisable to the general population. However, the principles of the analysis and interpretation of the data discussed remain valid.

Table 6.13: Correlation matrix for personality factors

	ITQ	EQ	SQ	Agree	Extra	Consc	Open	Neuro
ITQ	1							
EQ	-0.069	1						
SQ	-0.058	0.36	1					
Agree	-0.11	-0.032	-0.12	1				
Extra	0.48	0.072	0.45	0.26	1			
Consc	0.34	-0.056	0.27	0.13	0.52	1		
Open	0.36	-0.2	-0.05	0.39	0.68	0.36	1	
Neuro	0.38	-0.50	-0.53	0.022	-0.31	0.081	0.041	1

Legend

ITQ:	Immersive tendencies	EQ:	Empathy quotient
SQ:	Systematisation quotient	Agree:	Agreeableness
Extra:	Extraversion	Consc:	Conscientiousness
Open:	Openness to experience	Neuro:	Neuroticism

6.3.3.2 Quality of experience responses

In sections 4.4 and 6.2, data were presented showing that differences between responses on dimension 1 of the PCA of quality of experience items were associated with the judgement that audio was emitted from either a real or virtual source. Response data for these values were, as such, subjected to principal components

Table 6.14: Optimal multiple regression for predictors of $PC1_{\Delta}$ selected by stepwise model selection by AIC

	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	-20.28	10.71	-1.89	0.060
Immersive tendencies (ITQ)	0.37	0.11	3.52	0.00055
Empathy	4.23	1.26	3.36	0.00096
Systematisation	1.13	0.48	2.38	0.018
Extraversion	-43.79	7.91	-5.54	1.12×10^{-7}
Neuroticism	3.15	1.18	2.67	0.0083
Conscientiousness	-1.85	0.39	-4.76	4.13×10^{-6}
Openness	2.047	0.59	3.47	0.00066
Agreeableness	-5.48	1.71	-3.20	0.0016
ITQ×Empathy	-0.049	0.016	-3.17	0.0018
ITQ×Systematisation	-0.027	0.0062	-4.40	1.88×10^{-5}
ITQ×Extraversion	0.53	0.094	5.64	6.85×10^{-8}
Systematisation×Extraversion	0.28	0.10	2.679	0.0081

Adjusted $R^2 = 0.74$

Table 6.15: Anova of random effect of participant vs intercept only model for PC1 differentials between speaker and headphone judgements

	Model	df	AIC	logLik	Test	L.Ratio	p-value
Intercept only	1	2	783.25	-389.63			
Participant as random effect	2	3	590.97	-292.48	1 vs 2	194.28	<.0001

Nagelkerke $R^2 = 0.648$

analysis (PCA) as in section 6.2.3.3 and differences along the 1st principal component between speaker and headphone judgements were computed for individuals for each level of independent variable factor described in section 4.4.2.4. A saturated linear model containing interactions between all independent variables was subjected to bidirectional stepwise model selection with AIC as the selection criterion. The result of this procedure was a multiple regression model with predictors shown in table 6.14. It can be seen that the stepwise model selection procedure retains every independent variable as a main effects. However, the procedure has identified four significant interactions. The presence of significant interactions supersedes the presence of main effects [234]. However, initial inspection of the main effects of the stepwise selected multiple regression provides information about the higher order effects observed. It can be seen in the summary of the model in table 6.14 that extraversion is a significant predictor of $PC1_{\Delta}$ with a large estimate for $\hat{\beta}$. However, this result is only observed when extraversion is part of a multiple

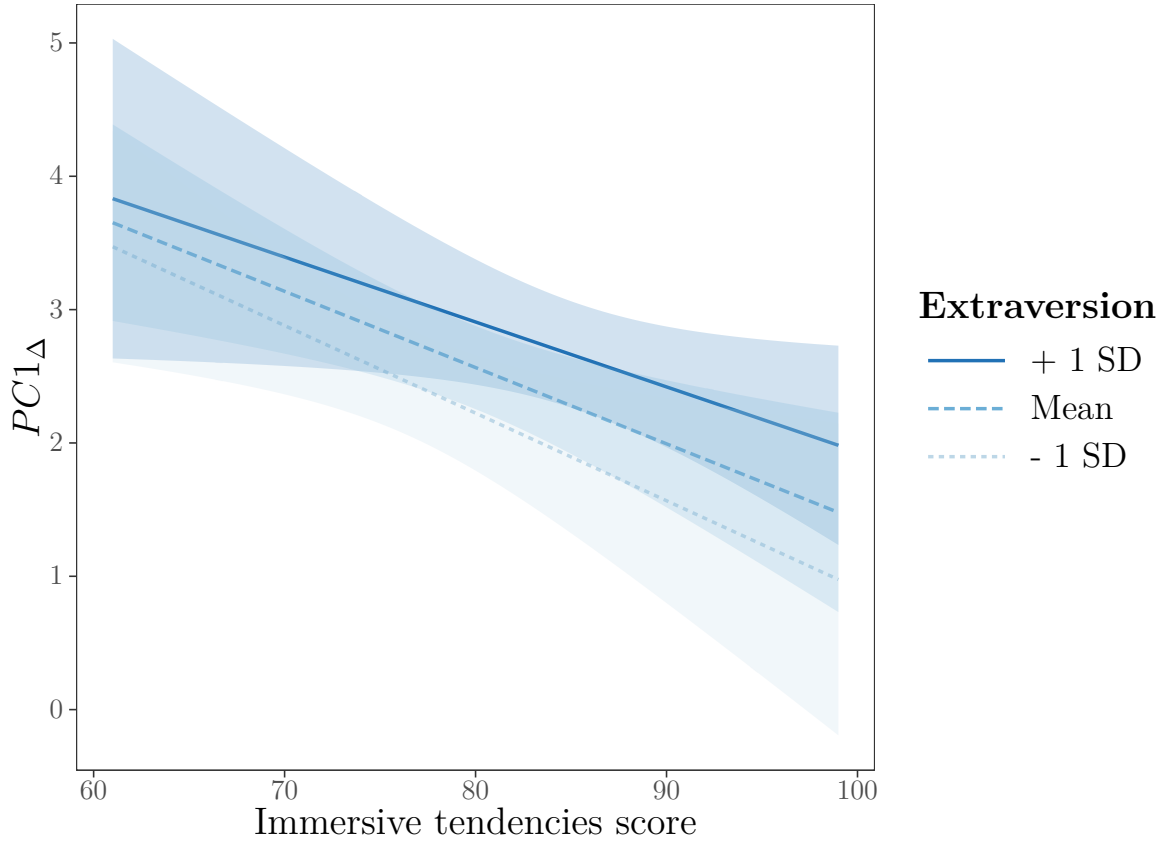


Figure 6.13: Fitted values for interaction on $PC1_\Delta$ scores between immersive tendencies and extraversion. Shaded areas show 95% confidence intervals

Table 6.16: Anova of fixed effects and mixed effects models vs intercept only model for PC1 differentials between speaker and headphone judgements

	Model	df	AIC	logLik	Test	L.Ratio	p-value
Intercept only	1	2	783.25	-389.6252			
ITQ×Extraversion	2	5	778.48	-384.24	1 vs 2	10.76	0.013
SQ×Extraversion	3	7	695.59	-340.80	2 vs 3	86.89	<.0001
ITQ×SQ	4	8	639.27	-311.63	3 vs 4	58.32	<.0001
ITQ×EQ	5	10	563.95	-271.97	4 vs 5	79.32	<.0001
Participant as random effect	6	11	563.092	-270.55	5 vs 6	2.85	0.091
Total L.Ratio						276.32	

regression. Simple single predictor regression of $PC1_\Delta$ against extraversion shows that this variable on its own is not significant ($\hat{\beta} = 0.32$, Std. error = 0.41, t = 0.76, p = 0.45). This shows that the significance of this variable is present only in the presence of other predictors, indicating modulation or suppressive effects evident in interaction with other predictors [200]. Figure 6.13 shows the interactive effect on

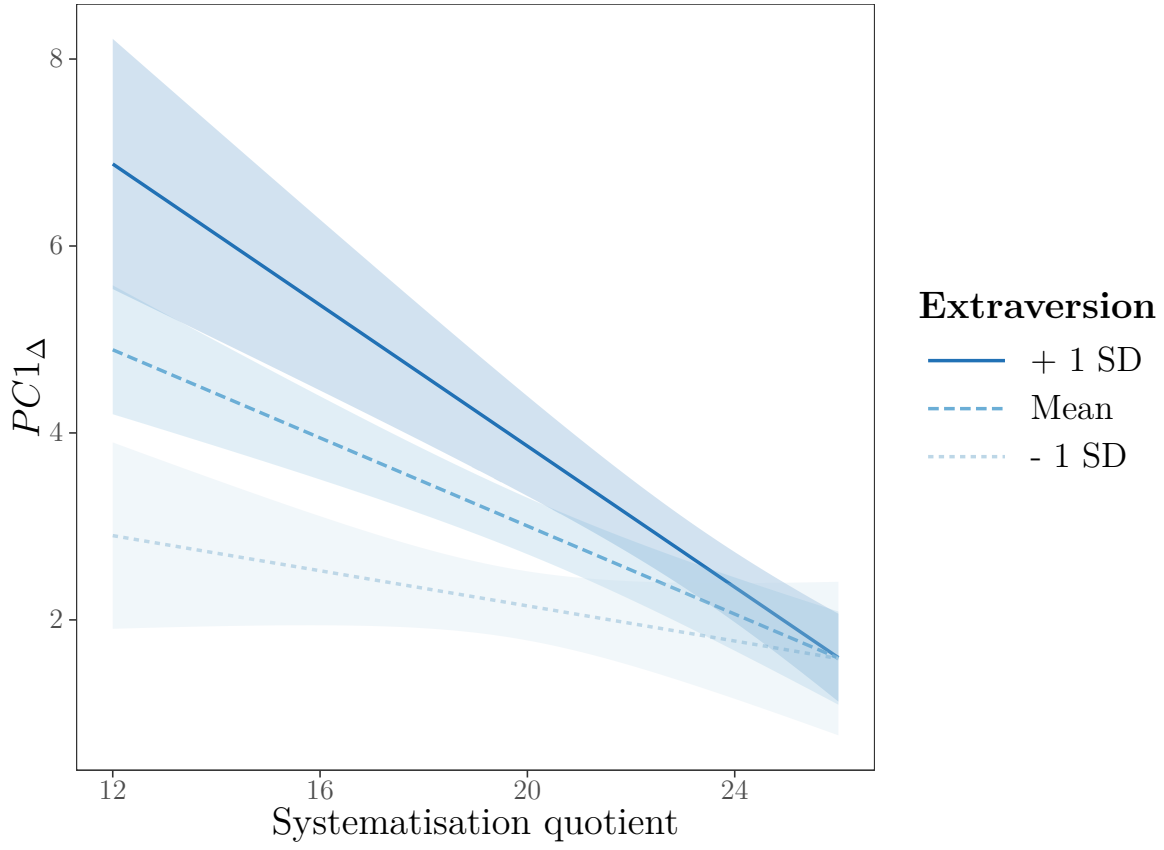


Figure 6.14: Interaction on $PC1_\Delta$ scores between systematisation and extraversion. Shaded areas show 95% confidence intervals

$PC1_\Delta$ between immersive tendencies and extraversion. It can be seen that extraversion modulates $PC1_\Delta$ scores as a function of ITQ by increasing the intercept of the function as extraversion increases. This corresponds to larger differences between responses corresponding to perceived difference between real and simulated sources for a given response for immersive tendencies. In all cases, immersive tendencies has a negative relationship with $PC1_\Delta$. Participants scoring higher on the ITQ report smaller differences in QoE factors relating to the perceived differences between real and simulated sources. Similarly, figure 6.14 shows the interactive effect on $PC1_\Delta$ between systematisation and extraversion. Once again, higher extraversion results in greater $PC1_\Delta$ scores for participants scoring low on systematisation with the association between systematisation and $PC1_\Delta$ decreasing as extraversion decreases. Figure 6.15 shows the Johnson-Neyman interval plot for the interactive effect between systematisation and extraversion on $PC1_\Delta$. This visualisation shows the change in slope of systematisation as a function of extraversion. In both of these interactions, extraversion acts as an modulatory factor which increases the

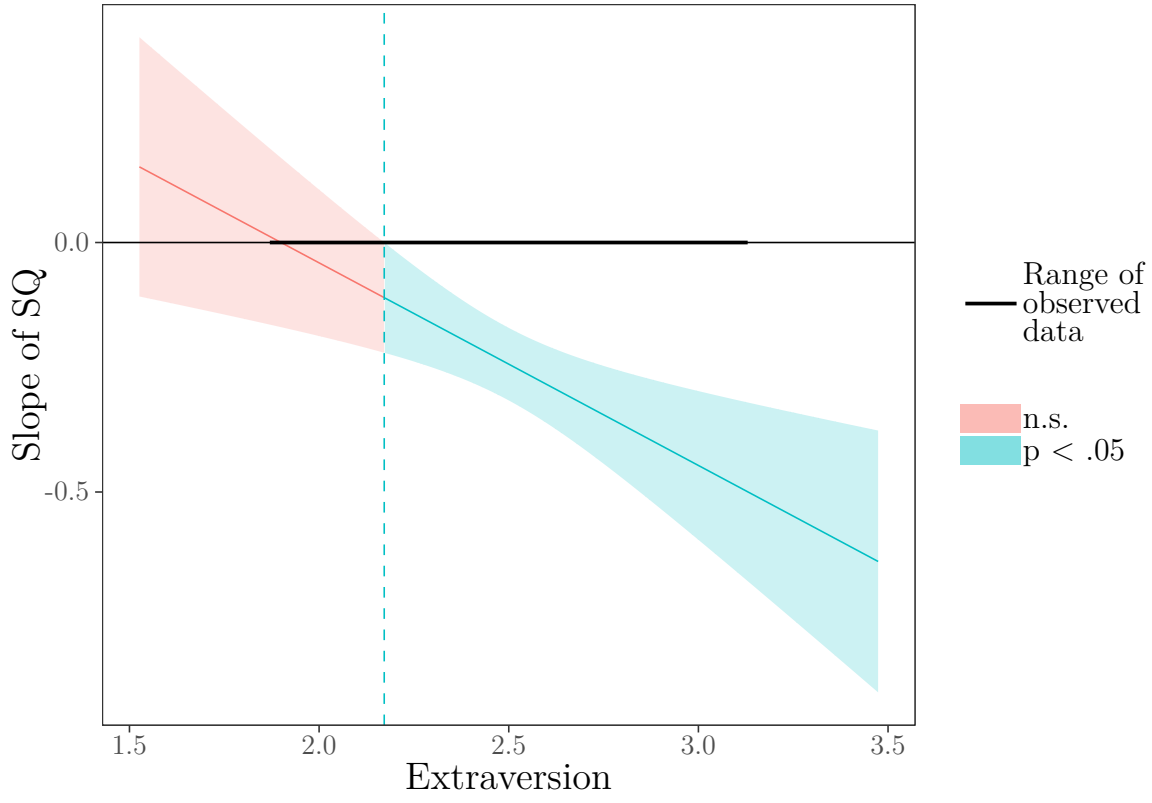


Figure 6.15: Johnson-Neyman interval plot showing interaction on $PC1_{\Delta}$ scores between systematisation and extraversion

range in the available scale which participants use for the percept of real and simulated sources. The interaction with immersive tendencies suggests this effect is to simply amplify differences in judgement ratings, while tendency toward immersion reduces perceived differences between real and simulated sources. The interaction with systematisation, however, may provide more information to the mechanisms of this effect. As systematisation is defined as a propensity for systemic categorisation of information [222], the decrease in $PC1_{\Delta}$ as a function of systematisation may reflect more nuanced or cautious responses for higher values of this factor. That this is only observed in higher extraversion subjects suggests a suppressive effect on the large response range tendency of extraverted subjects. The Johnson-Neyman interval shown in figure 6.15 indicates the cutoff for significant values of extraversion in this interaction is in the upper region of the extraversion scale. Although introversion has been implicated in the perception of presence [178], the relationship to other factors was assumed to be additive. In the case of immersive tendencies, it is plausible that this is the case. However, in the relationship between systematisation

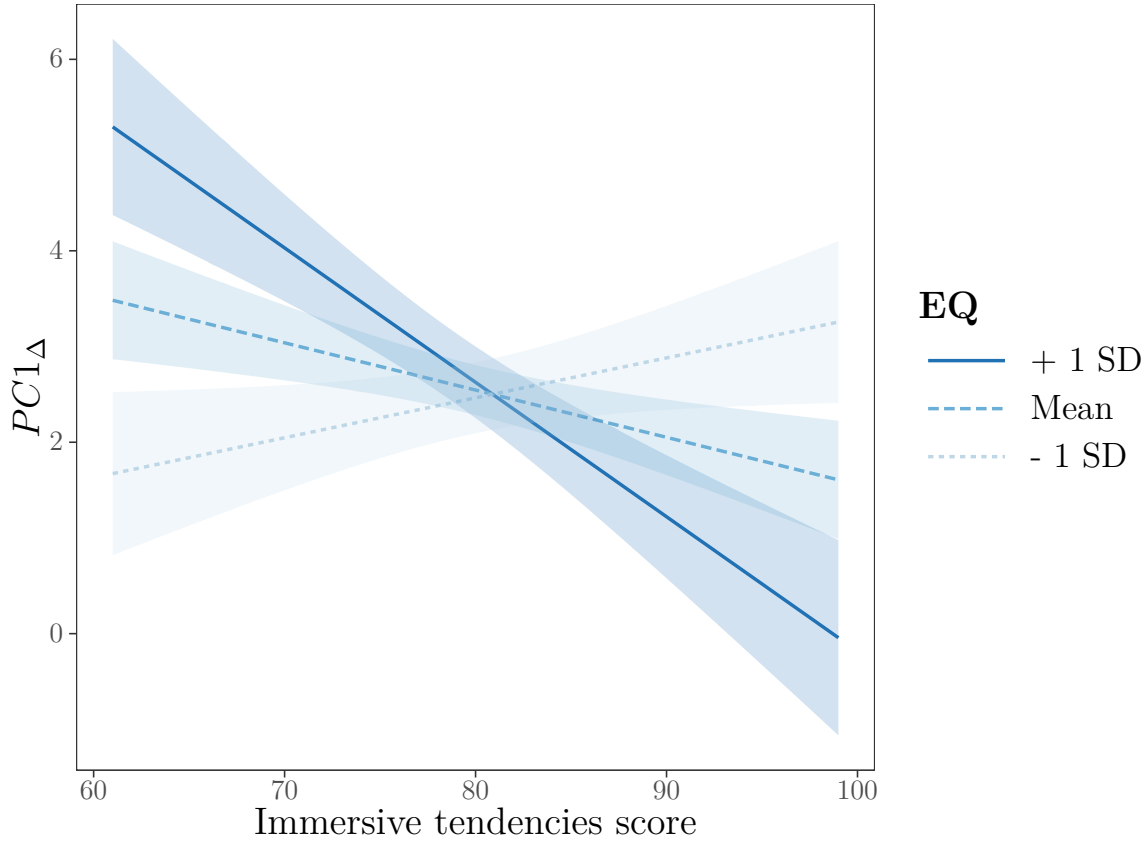


Figure 6.16: Fitted values for interaction on $PC1_\Delta$ scores between immersive tendencies and empathy quotient. Shaded areas show 95% confidence intervals

and introversion/extraversion, it is observed that there is a modulatory effect on this factor. Further work beyond the scope that is described in this document may be of interest to identify associated attitudes with this pattern of response, to identify if such response tendencies are related to such attributes as confidence or intuition. Figure 6.16 shows the interaction between immersive tendencies and empathy quotient on fitted values for $PC1_\Delta$ scores with 95% confidence intervals. It can be seen that empathy quotient modulates $PC1_\Delta$ as a function of immersive tendencies score. Participants responding lower for empathy ($<$ -1 SD from sample mean) have a positive association between $PC1_\Delta$ and immersive tendencies, corresponding to larger differences in QoE factors corresponding to the perceived difference between real and simulated sources when immersive tendencies are high. Participants responding close to the sample mean have no significant relationship between immersive tendencies and $PC1_\Delta$. Participants scoring high on empathy ($>$ 1 SD from sample mean) have a negative association between $PC1_\Delta$ and immersive tendencies. Regressions were performed on $PC1_\Delta$ scores as a function of ITQ for participants above and below

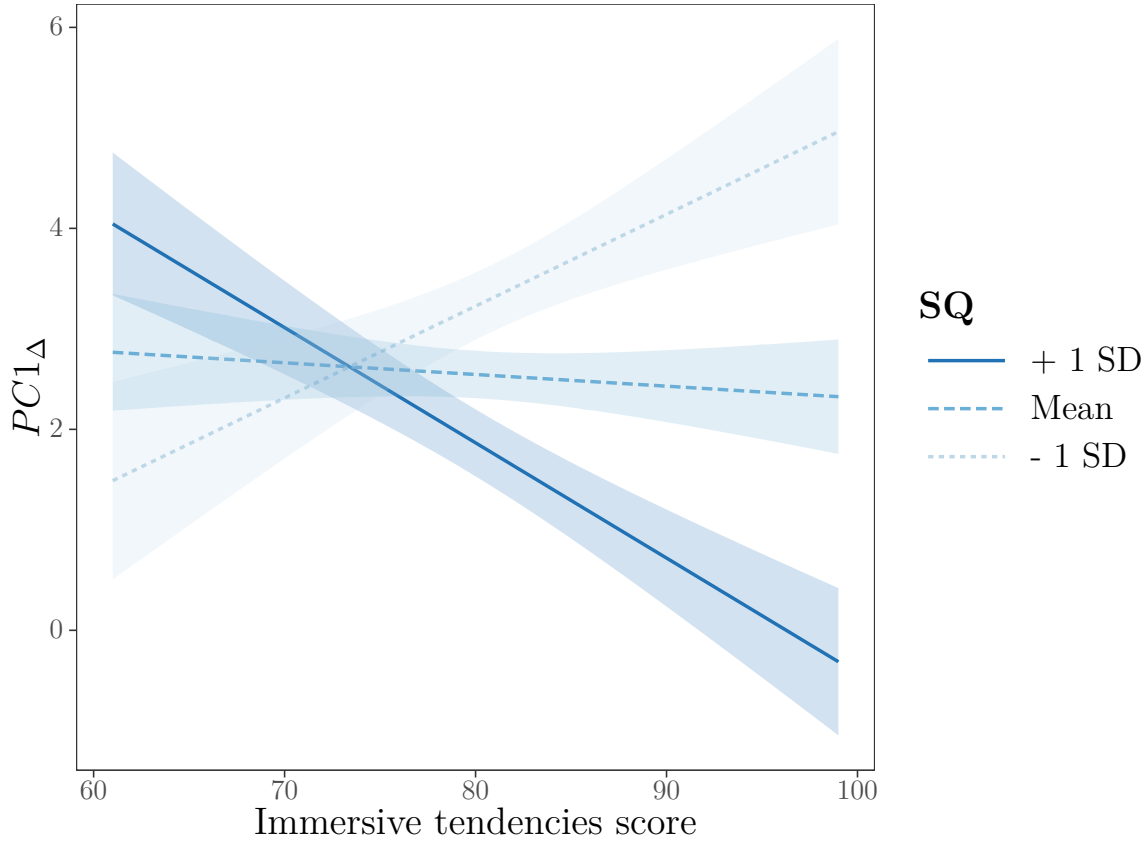


Figure 6.17: Interaction on $PC1_\Delta$ scores between immersive tendencies and systematisation. Shaded areas show 95% confidence intervals

mean EQ score. $PC1_\Delta$ was increased with ITQ in below mean participants ($\beta = 0.1$ $R^2 = 0.20$ $p = 0.0003$). $PC1_\Delta$ scores decrease in above median EQ participants ($\beta = -0.14$ $R^2 = 0.45$ $p = 2.2 \times 10^{-16}$). This indicates that immesive tendencies predicts a low range of discrimination between real and simulated sources only in participants responding higher than the sample average for empathy. This effect is also observed in the interaction between immersive tendencies and systematisation quotient. In this case, the effect appears more pronounced in figure 6.17. However, when regressions on data above and below sample mean for SQ scores were performed, below mean systematisation produced negligible positive association between $PC1_\Delta$ and ITQ scores ($\beta = 0.05$ $R^2 = 0.03$ $p = 0.022$). Above mean systematisation participants had a negative association between $PC1_\Delta$ and ITQ ($\beta = -0.1$ $R^2 = 0.34$ $p = 2.04 \times 10^{-7}$). Both analyses suggest that the EQxITQ and SQxITQ interactions have interpretable structure with The interpretation of these observations having multiple descriptive factors. As discussed in section 2.4, empathy has been implicated as a predictor of presence [172]. Given that presence is characterised

as perceiving simulated stimuli as if real, it would be plausible to hypothesise that higher empathy would result in lower $PC1_{\Delta}$ as the perceptual distance between real and simulated stimuli is reduced. What is observed, is that high empathy and high immersive tendencies produces this effect. However, the absence of one of these components results in an increase in perceptual distance. Similarly, the hypothesis of nuanced response from participants responding highly on the systematisation scale would have to be viewed in light of a dependence on immersive tendency. Nuanced, rather than impulsive, responses would only have a small perceptual distance if the subject is prone to perceiving stimuli in such a way as to require a granular and less definitive response. Interpretation of these effects, however is, mitigated by the colinearity observed in table 6.13. Further work with a larger sample size would be recommended to concretely determine if the effects observed can be replicated and support the postulated hypotheses or are anomolous due to sample colinearity, particularly as there is some degree of homogeneity in the responses for empathy and systematisation. Table 6.15 shows multilevel ANOVA statistics with likelihood test between intercept only prediction of $PC1_{\Delta}$ and the inclusion of participant as a random effect predictor. Participant random effects have a large likelihood ratio compared to the intercept only model and the Nargelkerke pseudo-R^2 for this model suggests that inter-participant variation explains 64.8% of the variance in the data. Table 6.16 shows the multilevel ANOVA statistics with likelihood tests. It can be seen that the contribution to model fit from the inclusion of participant level random effects is no longer significant in the presence of the interactions identified by the stepwise model selection process. It can therefore be posited that, within the sample of participants used in this study, the two way interactions identified in table 6.14 describe the variance contributed by inter-participant variation. Additionally, the model without the inclusion of random effects from participants has an adjusted R^2 of 0.705, which constitutes an increase of 5.7% of explained variance within the data.

6.4 Personality and visual sensitivity and the reporting of presence

6.4.1 Introduction

This section presents the results of analysis of questionnaire data from subsection 5.3 in which participants exposed to a virtual environment with similar pre-expousre environment geometry rated the experience in terms of the iGroup presence questionnaire dimensions, unitary experience of presence and audiovisual quality.

6.4.2 Materials and methods

The virtual environments and stimuli used in this chapter were the ones described in chapter 4, with pre-exposure and virtual environment geometry similarity. Participants were asked to complete psychological questionnaires and a global precedence task, as described in section 6.3, before exposure to the virtual environment with the HMD. Aural sensitivity was not measured before this test due to experimental design constraints. Participants who had completed the experiments described in subsection 4.4 were invited back to take part in this study to allow for the comparison of this factor to be investigated. However, the return rate for these participants was too low (N=4), producing a sample size too low to be considered appropriate for analysis. As such, this factor was omitted from the analysis. Personality and global precedence ratio scores were used as predictors of the combined presence and audiovisual quality of experience responses given in subsection 5.3 in order to identify the extent to which these factors accounted for individual variation in responses between participants.

6.4.3 Results and discussion

Big 5 personality scores, empathy and systematisation quotients, immersive tendencies and global precedence ratios were tabulated and the correlation matrix for these scores was obtained. Figure 6.18 shows the combined scatterplot and correlation matrix with regression lines for these scores. It can be seen that in the subset of the population that was sampled for this study, there is a high level of multicollinearity between the five factors of the big 5 model. Pearson correlation coefficients (r) between factors in the sample range between 0.47 and 0.8, suggesting that the sample

Table 6.17: Multiple regression of interactive effects between global precedence ratio, ITQ, EQ, SQ and extraversion on externalised/localised audio

	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	169.55686	24.05977	7.047	1.10e-11
GP	-175.65583	23.21391	-7.567	3.97e-13
ITQ	0.44090	0.36684	1.202	0.230273
EQ	-7.00433	0.68729	-10.191	$< 2\times10^{-16}$
SQ	-0.84219	0.74293	-1.134	0.257798
Extraversion	-2.39189	4.75239	-0.503	0.615092
GP×EQ	7.08952	1.13270	6.259	1.22×10^{-9}
SQ×Extraversion	-0.32580	0.09590	-3.397	0.000765
SQ×EQ	0.06027	0.02791	2.160	0.031538
GP×Extraversion	11.40670	4.27330	2.669	0.007983
GP×ITQ	-0.68689	0.48006	-1.431	0.153433

Table 6.18: Multilevel ANOVA statistics of significant interactive effects on ratings of externalised/localised audio

	Model	df	AIC	logLik	Test	L.Ratio	p-value
Intercept only	1	2	1375.436	-685.7179			
GP×EQ	2	5	1327.405	-658.7022	1 vs 2	54.03141	<.0001
SQ×Extraversion	3	8	1324.624	-654.3122	2 vs 3	8.78012	0.0324
GP×Extraversion	4	9	1306.342	-644.1711	3 vs 4	20.28223	<.0001
SQ×EQ	5	10	1279.509	-629.7543	4 vs 5	28.83347	<.0001
Participant as random effect	6	11	1261.60	-619.7991	5 vs 6	19.91	<.0001

does not fit the assumptions of independence of predictors for multiple regression between these factors. This notwithstanding, results presented in subsection 6.3.3.2 suggest that of the five factors in the big 5 model, only extraversion was shown to be a significant interactive factor in predicting response the profiles of interest. Due to this fact, and the lack of independence observed between factors, extraversion scores were retained and the remaining four factors from the big five model were discarded for the analysis presented here. It should be noted that this decision, therefore, limits the interpretation of results within the context of results observed in subsection 6.3.3.2 and, more generalised findings may be missed due to a paucity of variance in the psychology of the sampled population.

Figure 6.19 shows the rotated two factor solution to the principal components analysis of all presence and audiovisual quality response data as presented in subsection 5.3. Factor loadings in this analysis suggest a two latent factor model, externalised/localised sound (PC1) and visual realism and attention/involvement (PC2),

Table 6.19: Multiple regression of interactions between ITQ, EQ, SQ and extraversion on visual realism and attention

	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	-79.070	25.44	-3.108	0.0021
GP	7.76	17.02	0.456	0.65
ITQ	-0.41	0.11	-3.719	0.00024
EQ	2.03	0.81	2.504	0.013
SQ	3.93	0.75	5.251	2.75×10^{-7}
Extraversion	20.69	4.18	4.956	1.16×10^{-6}
EQ×Extraversion	-0.92	0.15	-6.259	1.23×10^{-9}
GP×SQ	-1.88	0.34	-5.574	5.24×10^{-8}
EQ×SQ	-0.085	0.019	-4.463	1.12×10^{-5}
ITQ×Extraversion	0.077	0.026	2.924	0.0037
GP×EQ	1.38	0.46	3.023	0.0027
ITQ×EQ	0.0064	0.0039	1.667	0.097

Table 6.20: Multilevel ANOVA statistics of significant interactive effects on ratings of visual realism/involvement

	Model	df	AIC	logLik	Test	L.Ratio	p-value
Intercept only	1	2	1076.7961	-536.3980			
EQ×Extraversion	2	5	1062.1457	-526.0728	1 vs 2	20.65041	0.0001
GP×SQ	3	8	909.8520	-446.9260	2 vs 3	158.29369	<.0001
EQ×SQ	4	9	906.9445	-444.4723	3 vs 4	4.90750	0.0267
ITQ×Extraversion	5	11	896.3172	-437.1586	4 vs 5	14.62733	0.0007
GP×EQ	6	12	887.3979	-431.6990	5 vs 6	10.91925	0.0010
Participant as random effect	7	13	866.6759	-420.3379	6 vs 7	22.72206	<.0001

with other factor loading suggesting a mix of loading based on the semantics of these underlying factors. Personality factors identified as significant in the analysis in subsection 6.3.3.2 were entered as terms in a saturated interactive multiple regression and subjected to bi-directional stepwise model selection by AIC. Model selection was performed with personality factors as predictors for both externalisation/localisation and realism/involvement. Table 6.17 shows the output of the stepwise model selection procedure for externalisation/localisation. The results of this analysis suggest that there are four interactive terms which are significant predictors of responses along the dimension relating to externalised audio. The results suggest that extraversion and empathy have modulatory effects on externalisation predicted by global precedence ratio and systematisation quotient. To identify if the identified interactions explain variance in the data that is attributable to inter-participant variation, these interactive terms were subjected to multilevel ANOVA

comparing the addition of fixed effects and participant level random effects (Table 6.18). Direct comparison between the null model and participant level random effects shows that taking in to account inter-participant variation signficiantly improves model fit (L.Ratio = 108.59, $p < 0.0001$). The analysis shown in table 6.18 suggests that the inclusion of the interactions identified above accounts of a large portion of this variance. However, it can be seen that participant level random effects are still a significant contributor to goodness of fit. This notwithstanding, the Nagelkerke R^2 of the random effects only model is 0.27. The fixed effects multiple regression described in table 6.17 is 0.34, indicating that interactions between empathy and extraversion and systematisation and global precedence ratio are better predictors of externalisation responses than simply accounting for individual variation in response. Figures 6.20, 6.21, 6.22 and 6.23 show the interactive effects identified. Figure 6.20 and 6.21 suggest that in the case of participants responding highly on empathy scales, a higher global to local processing ratio results in higher externalisation and sense of localistion, with this effect being reversed in the case of less empathic individuals. A similar trend can be observed in the case of the interaction between global precedence ratio and extraversion. In the case of participants indicating higher extraversion, visual processing speed is associated with higher externalisation and localisation. Whereas the slope of the model fit is reversed in the case of empathy, introverted participants show no association between visual processing speed and reported externalisation and localisation. Figure 6.22 shows that the relationship between systematisation and empathy is similar to that of global precedence and empathy. High systematisers with high empathy report high externalisation and localisation, while in low empathy participants, externalisation and localisation is low for high systematisation individuals. Figure 6.23 shows that the effect of extraversion on the relationship between systematisation and externalisation and localisation is the reverse of that of empathy. Responses on this quality of experience scale are higher for introverted systematisers and low for extraverted systematisers. It should be noted that the large confidence intervals seen in this interaction support the observation in table 6.18 that this effect is not as strong as the others identified and is a less good predictor of the outcome in question.

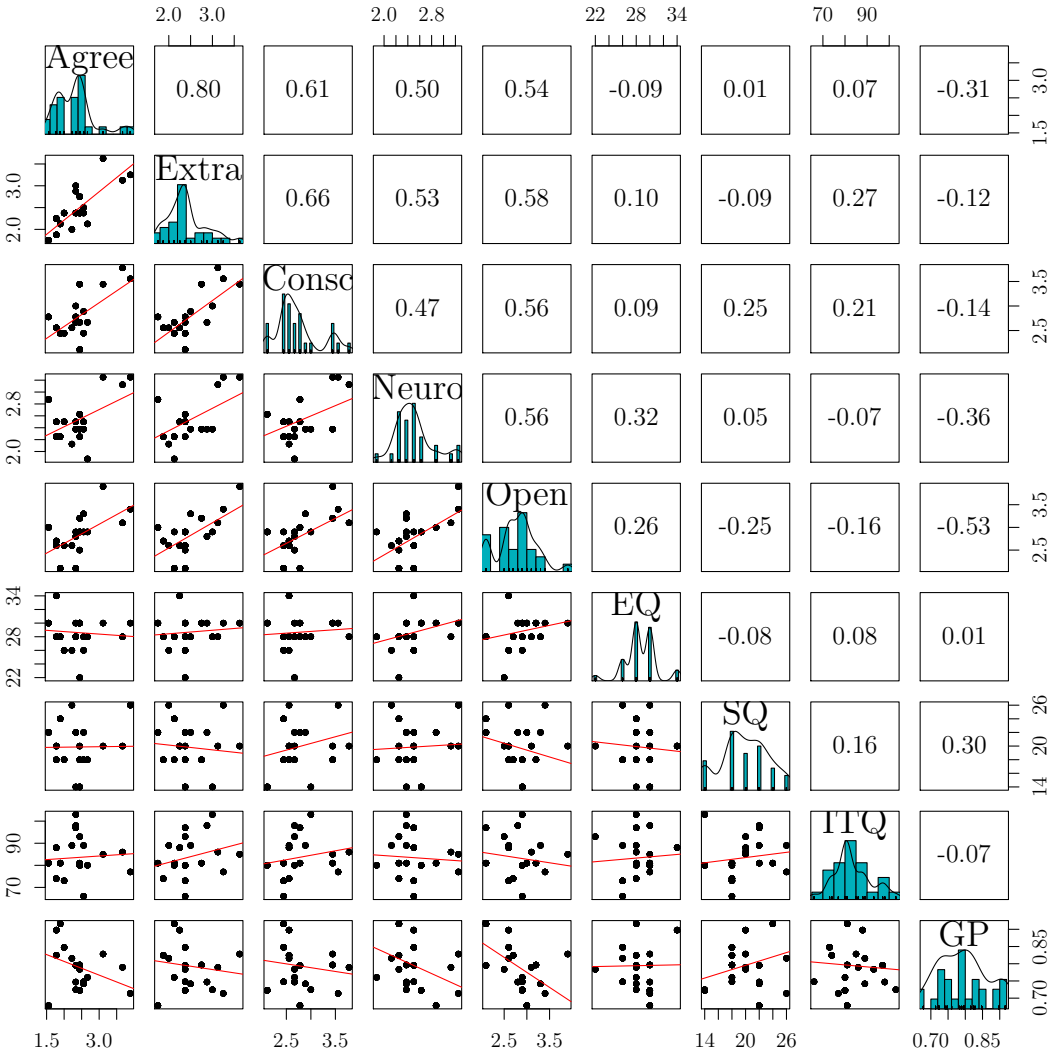
Table 6.19 shows the statistic of the optimal model identified by stepwise model selection for predictors of involvement and visual realism. The results suggest that there are five significant interaction terms which were retained. Table 6.20 shows the multilevel ANOVA to determine the contribution to model fit of each term, in comparison with the mixed effects model including participant level random effects.

It can be seen that the interaction between global precedence and systematisation had the large contribution to overall model fit, with an L. ratio of 158. The L ratio for the comparison between the intercept only model and the inclusion of participant level variance is 195.54 with a Nagelkerke R^2 of 0.44. The R^2 of the fixed effects model is 0.46, demonstrating a small improvement over simply including participant level effects. The reduction in model fit contribution from participant level random effects suggests that a portion of the variance explained by individual variation is explained by the interaction terms, but that accounting for individual variation between participants is still a significant contributor to model fit. Both figures 6.24 and 6.25 show that there is a positive relationship between global precedence ratio and reporting of realism and involvement. Participants who score higher on the global precedence task, indicating better resolution of local visual structures report higher visual realism and report higher involvement and attention on the virtual environment. This appears to be modulated by both empathy and systematisation, however, in different ways. Systematisation amplifies this relationship, with high systematisers with better visual detail resolution reporting even higher realism and involvement. Conversely, empathy serves to raise the baseline level of responses on this dimension for those who scored less well on the global precedence task. Immersive tendencies scores have an inverse relationship with responses for visual realism/involvement (Figure 6.26) This is somewhat modulated by extraversion, however the effect is small, with $\pm 1SD$ regressions cohabiting 95% confidence intervals within the region observed. This relationship is the reverse expected given the intention of the ITQ, designed as a predictor of presence. Similarly, the relationship observed in figure 6.27 confounds the expectation set by the literature that empathy should be a positive predictor of a component of presence. When modulated by extraversion, the association between empathy and reported realism and involvement in the virtual environment is negative for most of the range of values of extraversion. It must be noted, however, that the effect has a limited contribution to overall model fit in comparison with the global precedence ratio and systematisation interaction, which dominates in this model.

6.5 Summary

In this chapter, individual differences in cognitive sensory and personality factors have been investigated as explanatory variables for the variation seen in responses

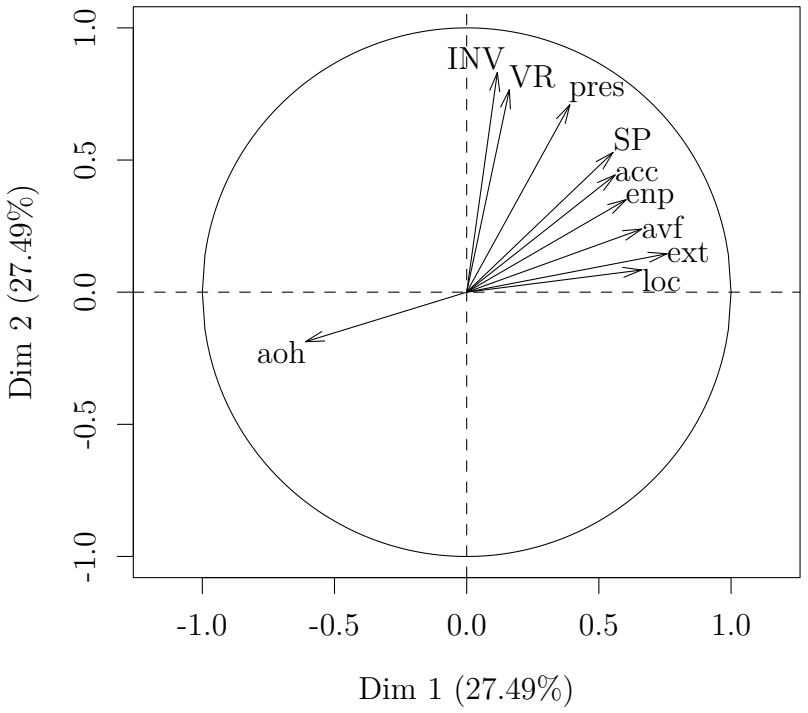
to quality of experience scales when judging whether stimuli were real or simulated. It was shown that neither ability to accurately discern real and simulated sources nor bias for choosing either loudspeakers or headphones as a stimulus source was a significant predictor of QoE responses or distance between responses for stimulus classifications. Neither was ability to cognitively resolve visual detail a predictor of QoE or differential responses between source classifications. However, independent analysis of responses for stimulus source classifications showed that ability to accurately determine stimulus source was associated with different percepts, dependent on the perceived stimulus origin. Sensitivity to aural stimuli was associated with higher reported externalisation in the case of perceived loudspeaker sources and lower aesthetic factors in the case of perceived simulated sources. It was shown that personality has a measurable and significant effect on QoE responses between source classifications. It was also demonstrated that personality factors not only account for differences in the distance between the percepts used as a decision criterion in classifying real and simulated sources, but were shown to be better predictors of perceptual distance than raw inter-participant variation. It was also shown that individual differences in extraversion, systematisation, empathy and immersive tendencies are interactive within the sample analysed and that, in the case of assessment of 'closeness to reality' of auditory sources in virtual reality, further models which attempt to provide predictors which may be generalised to the wider population should adopt a interactive or modulatory approach to the prediction of data, rather than a simple additive approach to factor structure. This was also found to be true for accounting for variation in responses relating to externalisation and localistion, and involvement and visual realism, in virtual environments with similar geometry to the pre-exposure environment. It was shown that the associations between personality and cognitive predictors of these dimensions are interactive and have less predictive power when assumed to be a simple additive component of a fixed effects model. In summary, it was demonstrated that accounting for interactive effects between personality factors in the analysis of subjective response data reduces the magnitude of model fit improvement contributed by the inclusion of participant-level random effects, and allows for accounting for variance in a dataset which might otherwise be unexplained.



Legend

ITQ:	Immersive tendencies	EQ:	Empathy quotient
SQ:	Systematisation quotient	Agree:	Agreeableness
Extra:	Extraversion	Consc:	Conentiousness
Open:	Openness to experience	Neuro:	Neuroticism
GP:	Global precedence ratio		

Figure 6.18: Correlation and scatterplot matrix for personality and global precedence ratio scores



	Dim.1	Dim.2
avf	0.66	0.24
enp	0.6	0.35
ext	0.76	0.15
aoh	-0.61	-0.19
INV	0.12	0.83
loc	0.66	0.084
pres	0.39	0.71
VR	0.16	0.77
acc	0.56	0.44
SP	0.55	0.53

Quality of expeience:

aoh - awareness of headphones
avf - audiovisual fusion
enp - environmental plausibility
ext - externalisation
loc - localisation
acc - overall audio realism

I-group questionnaire:

VR - visual realism
SP - spatial presence
inv - involvment

Witmer & Singer

pres - overall presence

Figure 6.19: PCA loadings of all presence and audiovisual quality responses subject to varimax rotation

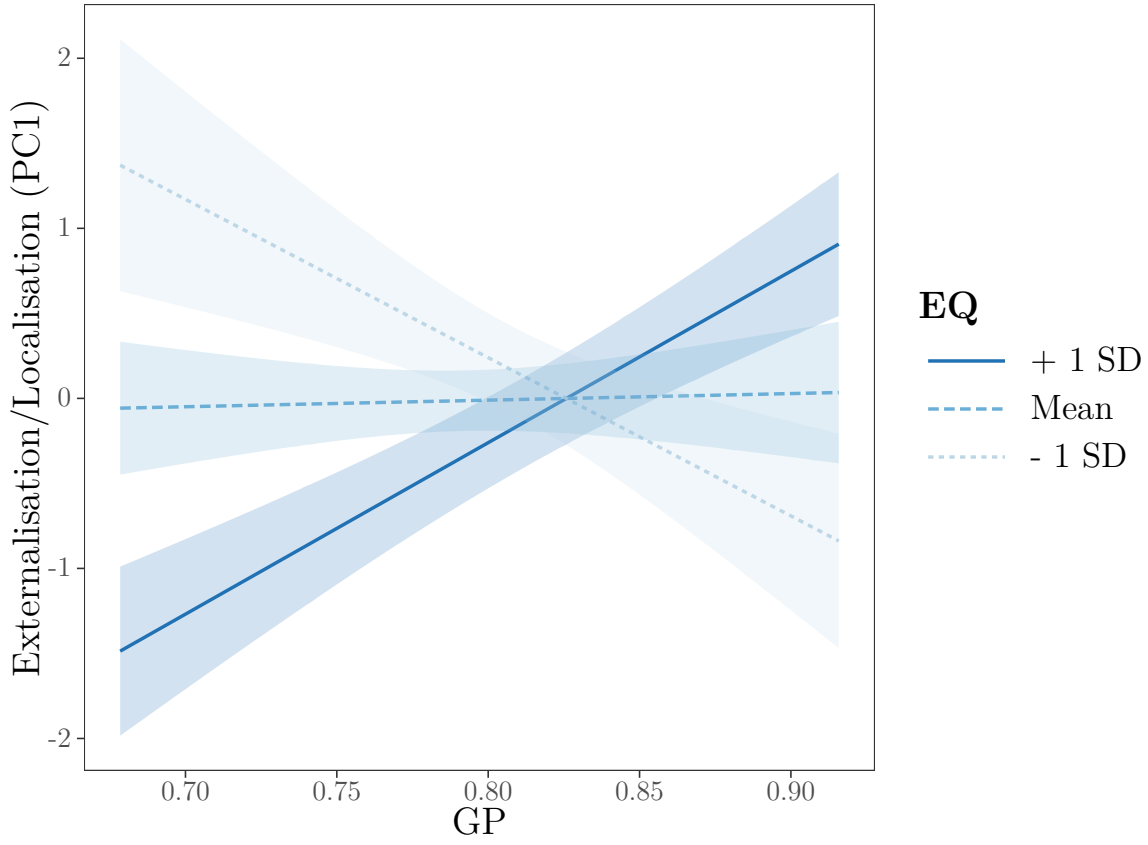


Figure 6.20: Interaction between global precedence ratio and empathy quotient on externalisation/localisation responses

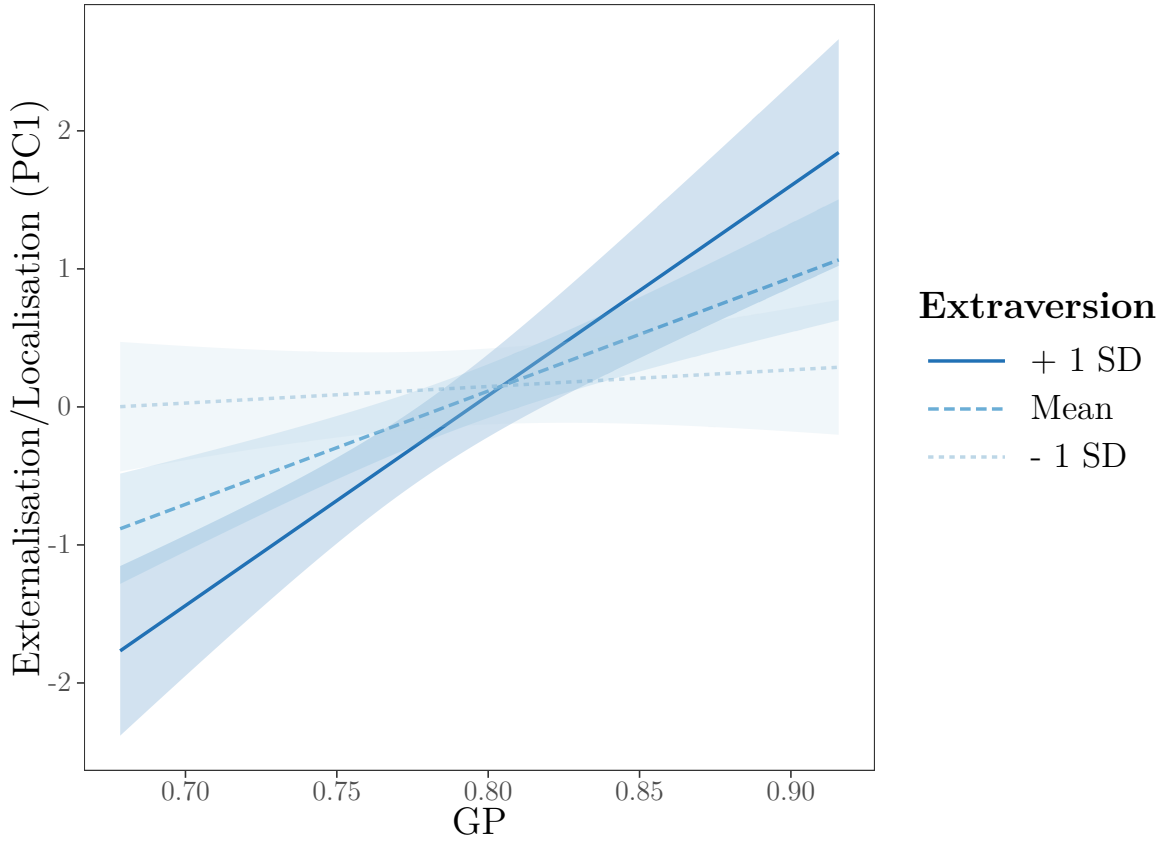


Figure 6.21: Interaction between global precedence ratio and extraversion on externalisation/localisation responses

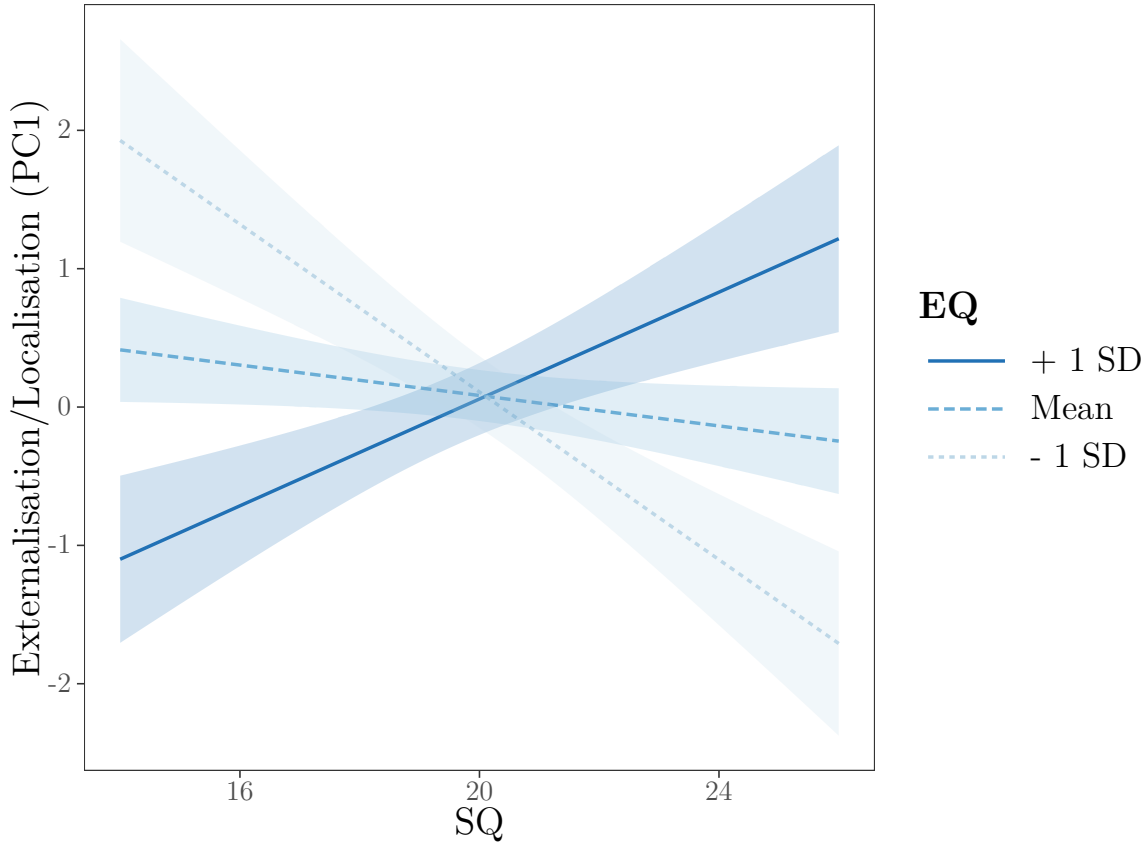


Figure 6.22: Interaction between systematisation quotient and empathy quotient on externalisation/localisation responses

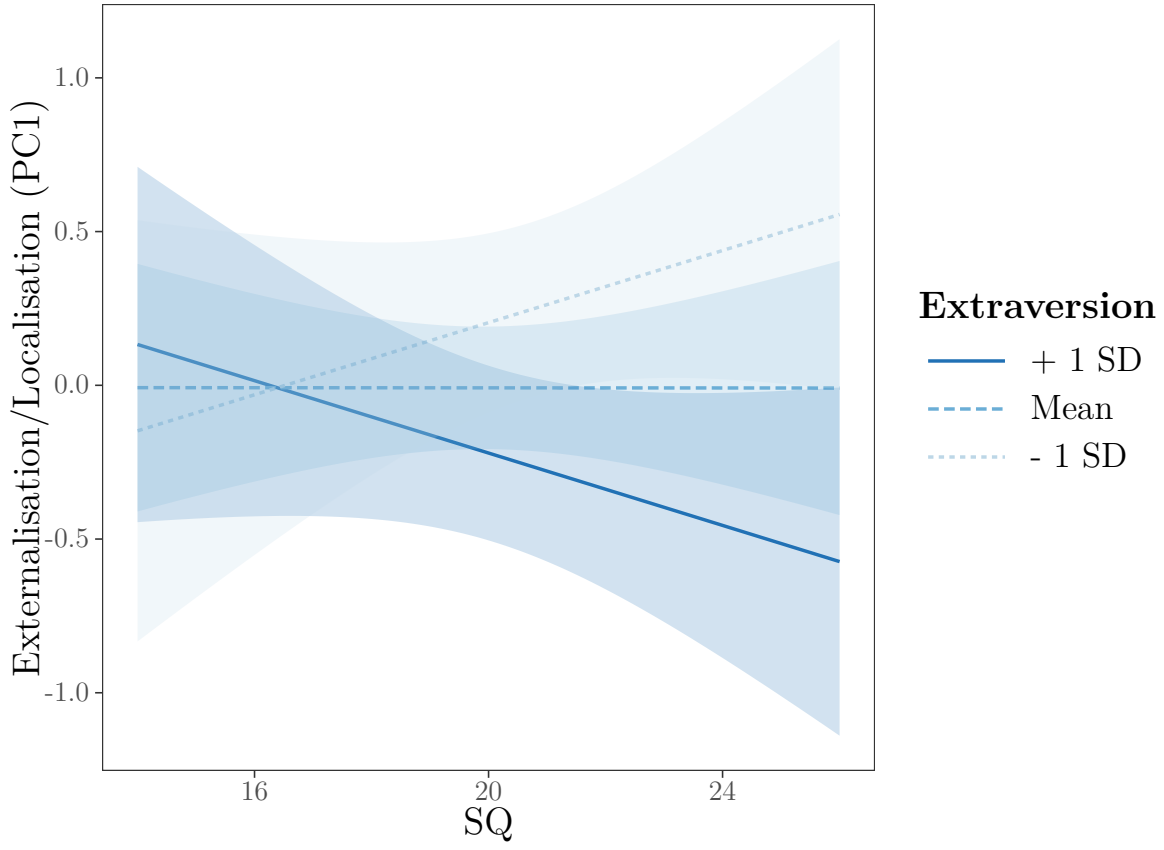


Figure 6.23: Interaction between systematisation quotient and extraversion on externalisation/localisation responses

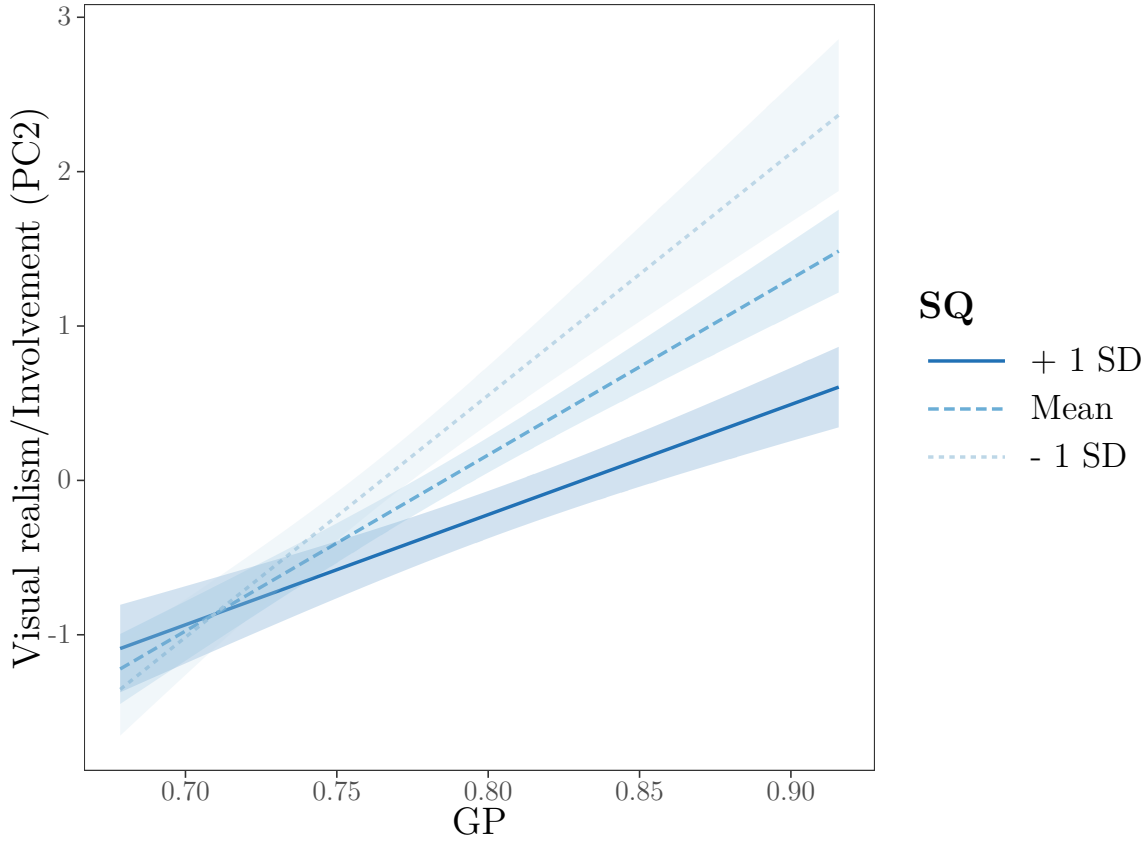


Figure 6.24: Interaction between global precedence ratio and systematisation quotient on visual realism and involvement responses

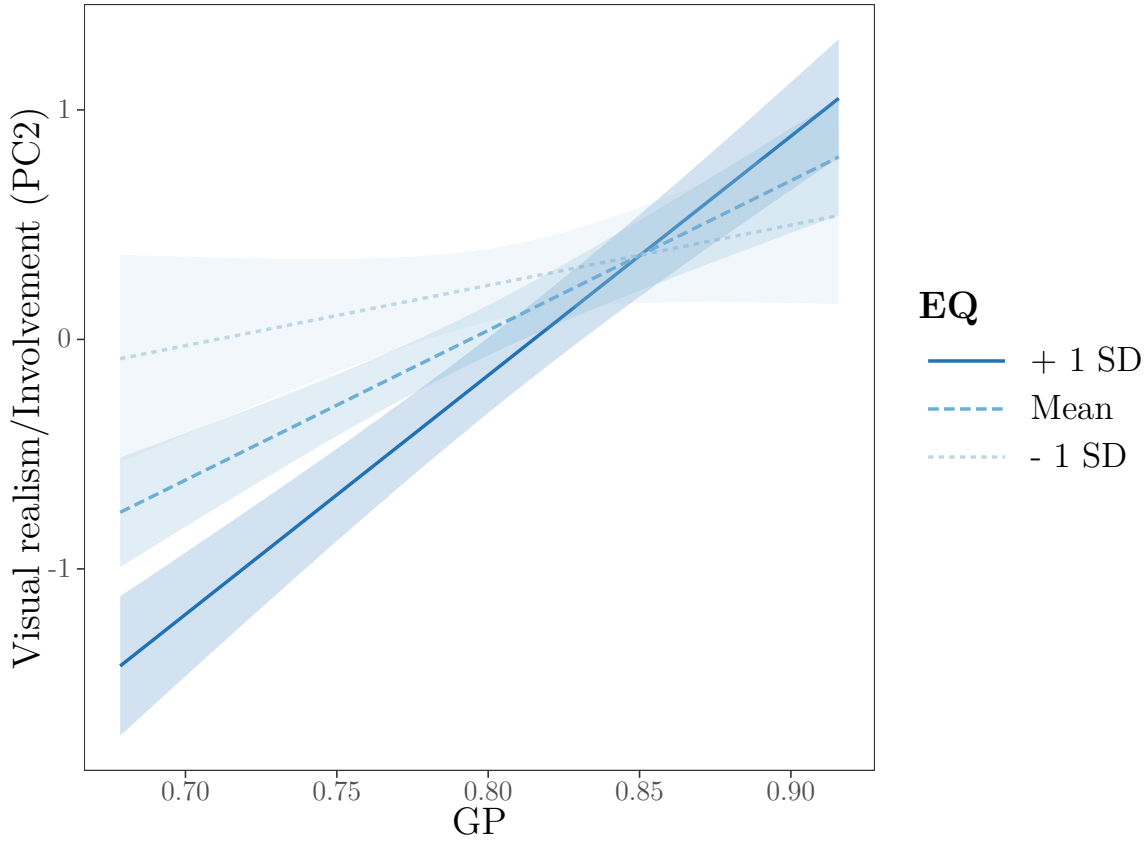


Figure 6.25: Interaction between global precedence ratio and empathy quotient on visual realism and involvement responses

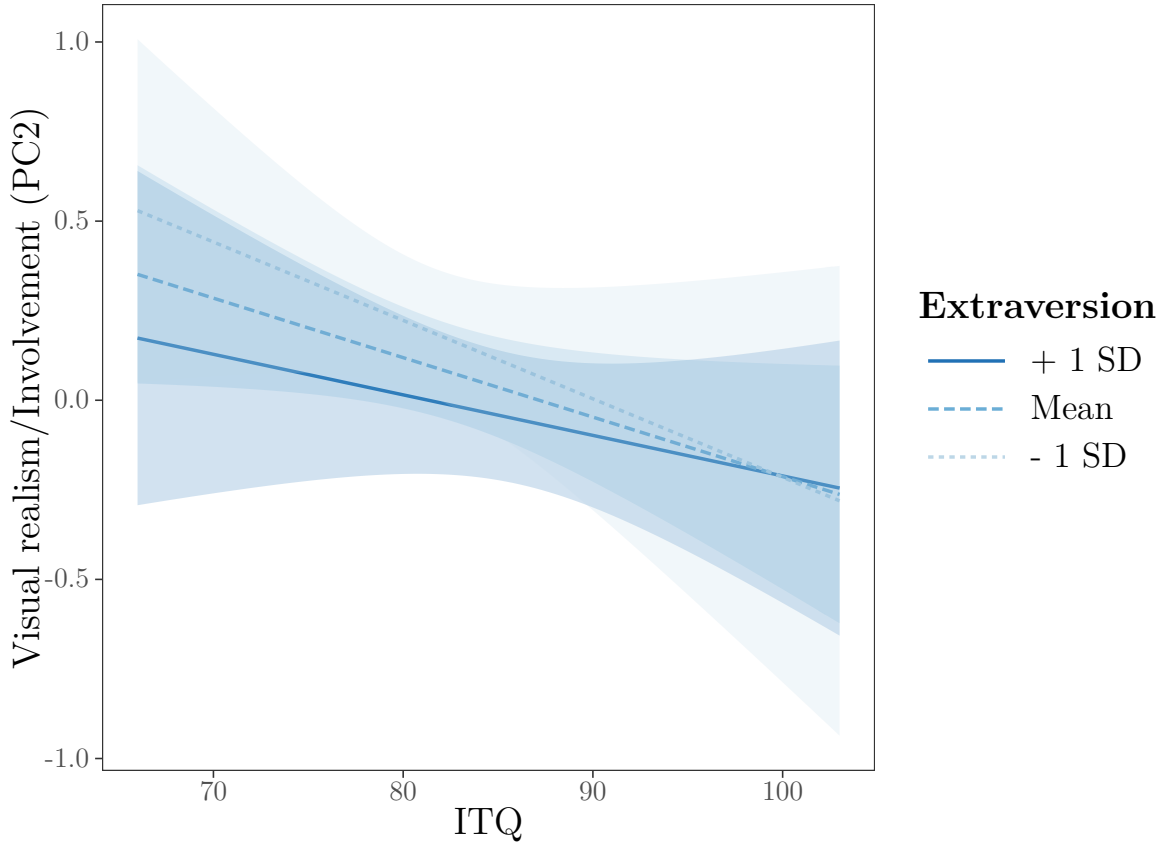


Figure 6.26: Interaction between immersive tendencies and extraversion on visual realism and involvement responses

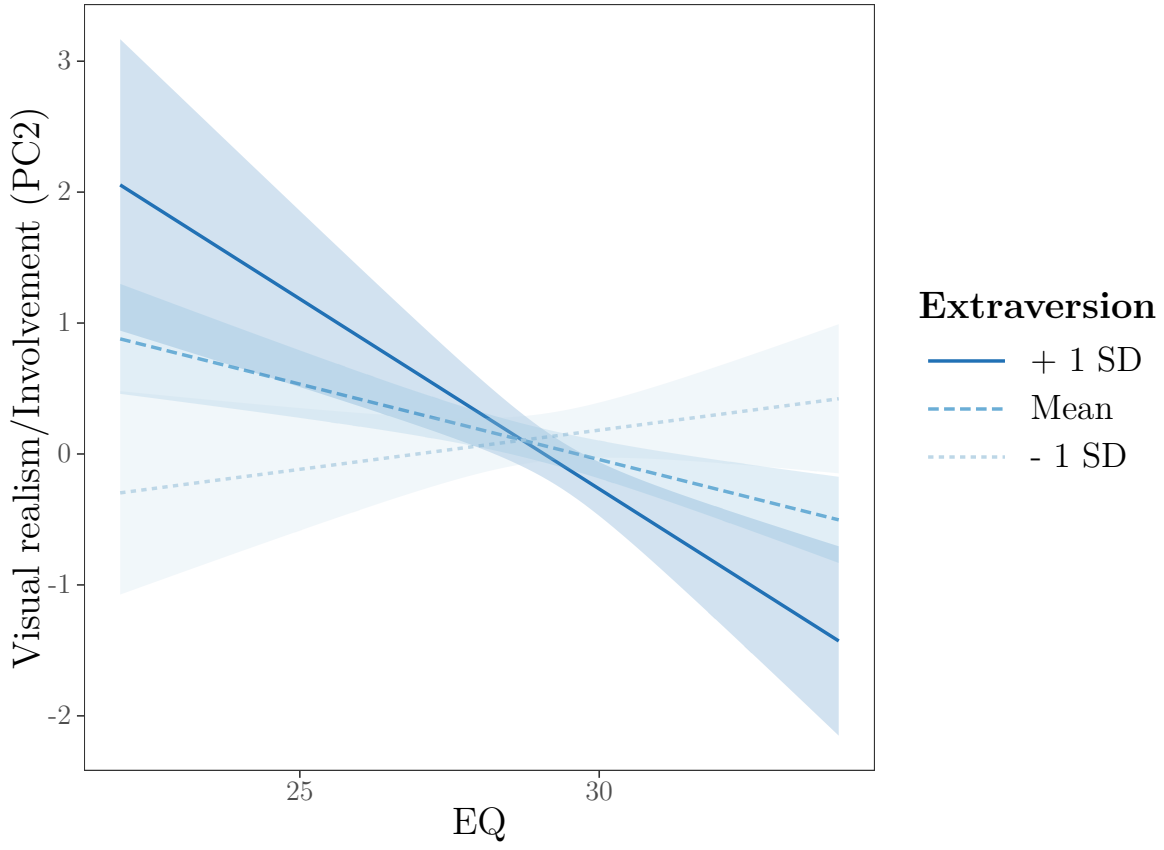


Figure 6.27: Interaction between empathy quotient and extraversion on visual realism and involvement responses

7

Discussion of experimental results and further work

7.1 Pre-exposure anchoring effects

As presented in chapter 4.3, it is suggested that there is a tendency to rate stimuli as more externalised, plausible, localisable and with better audiovisual fusion when there is no relationship between the geometry of the VE and the pre-exposure environment. This result can be viewed as a generalised anchoring bias, an effect that has been long known within the field of psychology. It can be demonstrated that for many basic perceptual stimuli, such as weight, anchoring to a reference can influence the judgements of degree on an internal scale of evaluation [235].

Anchoring is commonly used in perceptual testing in audio, and forms the basis of tests like the multiple stimulus hidden reference (MUSHRA) protocol for the assessment of signal degradation [11]. It has also been demonstrated in terms of judgement of realistic playback levels differing in real-world and laboratory environments [236]. However, this phenomenon has not yet been shown to be a salient contributor to subjective responses in the assessment of virtual reality environments and as such has not been considered as a constituent of presence. Particularly with reference to the immediate pre-exposure conditions of a subject. Comparisons between subjective responses to real and matching simulated spaces have been performed in the past [82][237]. However, these studies aimed to compare the real world experience with the corresponding VE. Additionally, they do not control for the pre-exposure/post-exposure similarity analysed in this work, as the locations of the real and simulated environments were different in the case of these studies. As this is an area of research that has not been systematically explored and, although not investigated in

great depth within this document, identifies an area of research which may produce interesting premises for further work.

Despite the implicit need for all findings to be replicated and have the benefit of increased sample size, there are questions of degree which could be explored within this area. In the work described here, a large unrealistic space was used as the geometrically dissimilar space. In addition, the pre-exposure environments were uncontrolled when subjecting participants to this set of conditions. Further work in this area may investigate the existence of a granular relationship between pre-exposure and VE geometries in terms of similarity of scale or shape. Although referred to as 'similarity of geometry' in this document, design of the VEs was such that simulacra of surface textures were attempted to within a reasonable degree at the discretion of the author. This was not controlled as an independent variable, however, with the results observed in section 5.3 suggesting that spatial presence is related to both the audiovisual/spatial audio dimension and the visual realism/involvement dimension, it may be of interest to determine the contributions of high quality realistic textures to a model of audio quality rating which uses pre-exposure anchoring as a predictor variable.

7.2 The relationship between audiovisual quality and presence

The findings discussed above, pertaining to differences between similar and dissimilar environments, define the scope of the interpretations derived from the remainder of the body of work described in this thesis. It is shown that results obtained with similar and dissimilar pre-exposure environments are distinguishable. Therefore, it is assumed that until the relationships identified between audio quality and presence are investigated, contrasting the loadings of those constructs in terms of pre-exposure similarity, care should be taken when assuming that the relationships observed are valid in dissimilar VE experiences. However, within these limitations, the relationship between the experience of presence and spatial audiovisual quality can be shown to comprise of two latent factors that can be broadly interpreted as externalisation/perceived localisation of audio and perceived visual realism/reported involvement in the virtual scene. Other responses appear to be loaded as linear combinations of these two dimensions which are easily interpreted.

It may be hypothesised, for instance, that spatial presence is reported when participants experience both high visual realism and externalised audio. It may also be posited that audiovisual fusion or high plausibility are phenomena related to the experience of spatial presence in conjunction with externalised and localised sound, when the visual stimuli in the VE have an anchoring reference in the form of the pre-exposure environment.

Within the similar environments, it has been demonstrated in this body of work that audiovisual quality, externalisation of audio and sense of localisation in particular, is independent of visual realism/involvement and together contribute to spatial presence. Due to the changes in loadings within the construct of audiovisual quality due to the various experimental conditions (i.e. spatial co-location and similarity) observed throughout the experimental work presented here, it can only be asserted that this construct remains valid for the conditions presented above. The independence of environmental geometry and the attendant acoustic responses suggest that the results seen are consistent when pre-exposure and VE geometry are similar. However, further work would be useful to see if the distribution of loadings observed between the audiovisual quality and the construct of presence are comparable in dissimilar VEs and to investigate the existence of any dependence on the relationship between the pre-exposure environment and the virtual environment.

7.3 Range of responses used in assessing plausibility

The experiments described in section 4.4 suggest that audiovisual quality is associated with real sources. However, given the loadings of audiovisual fusion and plausibility on to the extracted dimensions associated with higher response, these stimuli were perceived as belonging to the virtual environments, rather than as external stimuli. In these experiments, loudspeaker sources were discrete emitters, with no phantom sources rendered. This result, although nominally used as a metric for quantifying inter-participant variability in quality of experience begs further questions about the reproduction systems used in immersive VR, and the relationship between loudspeaker reproduction systems and head mounted displays.

The findings that, in the case of convergent environments, real sources are attributed to emission within the visually rendered virtual space and not necessarily perceived as an stimulus external to the virtual world is one of interest. As with the conditions discussed above, it would be of interest to compare this to dissimilar environments. However, the use of convergent environments introduces a built-in 'congruence' of expected acoustic response, and with it design challenges for any experimental set up that would be used to identify any effect. If the virtual environment presented is the same as the pre-exposure environment, there is some reasonable expectation as to why this may be the case. However, it also presents complications when implementing such an experimental apparatus in the case of dissimilar environments.

This notwithstanding, it has been suggested in the literature that in cases where spatial reproductions of environments are rendered in non-anechoic environments, the more reverberant space takes precedence when eliciting a sense of environment size [214]. As such, it may be reasonable to investigate the effect of headphone based soundfield rendering versus speaker based soundfield rendering based on this premise. To maintain continuity and to make results comparable with those presented in this work, assessment would be in terms of the audiovisual quality constructs described, with the constraint that the virtual environments would necessarily be more or equally reverberant than the pre exposure geometry.

7.4 Individual differences and inter-participant variability

It was consistently shown in the results presented over the course of this document that although it was possible to detect effects attributed to the manipulation of stimulus signals, the majority of the variance observed in the data were a result of inter-participant variability.

This variability manifested in changes in both intercept and variance. This is evidenced by the dependence of individual differences accounting for variance in both central tendency and range of responses in the subjective tests. This indicates that both parameters of the profile of responses given by a participant is more dependent on predisposition of the subject than the properties of a stimulus which may constitute an independent variable.

This notwithstanding, the variables which were identified and shown to explain this variance, both in terms of overall response and deltas of ratings between source types, as indicated elsewhere in this document, must be treated with care. This work has included high level constructs such as immersive tendencies, empathy, systematisation and extraversion as possible explanatory factors which have significantly reduced standard error and increase likelihood ratio in models derived from datasets which have a large amount of inter-participant variability.

The inclusion of these factors was informed by literature which had, in the past, used combinations of these to produce additive predictors of the experience of presence [178][238]. The inclusion of these factors in this work was to identify if these differences could be used as predictors for the auditory/audiovisual component of the experience presence, as this relationship has not hitherto been identified.

Although it was demonstrated that these high level trait constructs have predictive power in terms of model fit improvement, it was shown that the use of these traits in a multiple regression analysis produces anomalously large main effects which are only explicable when interpreting the first-order interactions present between constructs.

The interactions observed are complex and do not offer a simple interpretation which would contribute to a neat predictive metric for the experience of presence or the subjective response to audiovisual stimuli in immersive VEs. The conditional nature of the modulation of the coefficients in these interactions may go some way to explaining the inconsistency observed in the associations between personality and presence in the literature where only main effects are considered[179][180][181][182][183][177].

7.5 Information processing interpretations of observed results

It is also possible that the complexity of the interactions observed in the work presented in the context of the inconsistencies in the wider literature are indicative of the possibility that the use of high-level constructs such as those used here may be inappropriate for the characterisation of the inter-participant variation observed. There is something of an inherent imprecision that is introduced by the use of categories which comprise of a wide range of sub-traits which are semantically grouped by behavioural consistency. Although the categorisations used here have been shown

to *statistically* improve modelling of reported perception, the power of the factors to give insight into reasons *why* such responses should be expected is, on a superficial level, limited.

Despite this, it may be possible to construct a falsifiable hypothesis which may be tested to produce a meaningful construct which both predicts and explains some part of the perceptual heuristics which result in a given report of audiovisual quality of experience or presence. This might be achieved through closer inspection of what is implied by the behavioural constructs that were shown to significantly interact in this work and the theoretical and conceptual commonalities which may underlie those interactions.

Empathy is recognised as a multidimensional construct [239] and there have been shown to be distinct neural correlates associated with different components of this trait [240]. The distinctions made between the subscales relate primarily to differentiation between theory of mind, heuristic estimation of the thoughts of others and the mimetic experience of the emotional states of others. These properties are formed of an object-subject relationship which is mediated as a sensory stimulus. Stimulus-response processes can be understood as an information to motor transformation using the perception-action model of response [241].

This framework has framed the notion that empathy should be treated as a perception-action process whereby information produces an mimetic response which is analagous to the internal state of the object [225]. The perception-action model is, in essence, an information-processing theory of cognition in which stimuli constitute information which are processed and appropriate responses are formulated. Within this framework, the trait of empathy can be considered a a behavioural expression of the level of automaticity of a response which triggers in the subject its internal representation of the state of the object [225]. This process, however, is shown to comprise of conceptually distinct facets [239] and anatomically discrete locations [240].

In the experimental conditions presented in this thesis, it is unlikely that the stimuli presented elicited mimetic representations of the internal states of the stimuli within participants scoring high for empathy. However, there is a component to empathy related to the automatic experience of somatic response given a stimulus containing information corresponding to a bodily state [225].

The systematisation quotient, the complimentary metric to empathy quotient used in this work, is a high level concept that describes the tendency to organise information [222]. Despite this, its relationship to its counterpart in terms of information processing theory is somewhat easier to conceive, being the automaticity of analysis of incoming information within a stimulus [223]. Similarly the global precedence effect is a behavioural measure of an inherently information theoretical premise, that a multipass process of information processing is required to generate the internal representation of the corresponding objects which the subject perceives through the stimulus of focused light [140]. Finally, extraversion is often characterised as a sensitivity to the intensity of sensory inputs and the magnitude of physiological response to stimuli [242]. Although the validity of this construct is disputed [243], neuroimaging studies have shown structural differences associated with this dimension indicating that it is a useful behavioural metric with a demonstrable neural basis [159][244]. Of the significant factors identified in this thesis, only immersive tendencies is unable to be understood in an information processing context.

Given the discussion above, it is possible to reinterpret the results presented in chapter 5 in terms of an information processing paradigm:

- Extraversion - sensitivity to the magnitude of information content within a stimulus
- Empathy - sensitivity to stimuli with somatic information which elicits automatic somatic response
- Systematisation - organisation of stimulus information
- Global precedence - sensitivity to high resolution stimulus information in the visual domain

When parameters are redefined in this way, the range of responses relating to audio visual quality of experience when judging whether auditory sources are real or simulated can be said to be affected by intrinsic qualities of participants in the following ways:

- The effect of immersive tendencies (ITQ) score is modulated by the sensitivity to the magnitude of information content within a stimulus (ITQ $\times$ Extraversion interaction):

  - Immersive tendencies scores are associated with smaller differences between judgements of real and simulated sources. As stimulus information sensitivity decreases, there is a greater difference between QoE scores for real and simulated stimuli.

- The effect of organisation of stimulus information is modulated by sensitivity to the magnitude of information content within a stimulus (Systematisation × Extraversion interaction):

  - Greater organisation of stimulus information is associated with smaller differences between ratings for real and simulated sources. This association is accentuated by a reduction in stimulus information sensitivity.

- The effect of ITQ score is modulated by sensitivity to stimuli with somatic information which elicits automatic somatic response (ITQ × Empathy interaction):

  - Immersive tendencies scores are associated with smaller differences between judgements of real and simulated sources in subjects with a high sensitivity to information which provokes an automatic somatic response. In subjects with a low somatic information sensitivity, ITQ scores are associated with a modest increase in differences between QoE scores for real and simulated stimuli.

- The effect of ITQ score is modulated by organisation of stimulus information (ITQ × Systematisation interaction):

  - Immersive tendencies scores are associated with smaller differences between judgements of real and simulated sources in subjects who tend to organise information. Immersive tendencies scores are associated with greater differences between judgements of real and simulated sources in subjects with less tendency to organise information.

Similarly, the interpretation of responses for the components of spatial presence may be interpreted as below:

For responses relating to externalistion and localisaton of auditory stimuli:

- The effect of sensitivity to high resolution stimulus information in the visual domain is modulated negatively by sensitivity to the magnitude of information content within a stimulus (Global precedence × Extraversion interaction):

  - Resolution of information sensitivity predicts externalisation only when overall information sensitivity is low.

- The effect sensitivity to high resolution stimulus information in the visual domain is modulated by sensitivity to stimuli with somatic information which elicits automatic somatic response response (Global precedence × Empathy interaction):

  - In subjects with high sensitivity to information which elicits a somatic response, greater resolution of information sensitivity has a positive association. In subjects with low sensitivity to information which elicits a somatic response, greater resolution of information sensitivity predicts lower externalisation and localisation.

- The effect of organisation of stimulus information is modulated by sensitivity to stimuli with somatic information which elicits automatic somatic response (Systematisation × Empathy interaction):

  - In subjects with high sensitivity to information which elicits a somatic response, higher organisation of stimulus information has a positive association. In subjects with low sensitivity to information which elicits a somatic response the association is negative.

For responses relating to involvement and the subjective assessment of visual realism:

- The effect of sensitivity to high resolution stimulus information in the visual domain is modulated by organisation of stimulus information (Global precedence × Systematisation interaction):

  - Higher information resolution results in higher responses on this scale. This effect becomes more pronounced as information organisation is reduced. This is the strongest effect on this percept.

- The effect of sensitivity to high resolution stimulus information in the visual domain is modulated by sensitivity to stimuli with somatic information which elicits automatic somatic response (Global precedence × Empathy interaction):

  - Higher information resolution results in higher responses on this scale. This effect becomes more pronounced as sensitivity to stimuli with somatic information is increased.

- The effect of sensitivity to stimuli with somatic information which elicits automatic somatic response is modulated by sensitivity to the magnitude of information content within a stimulus (Empathy × Extraversion interaction):
  - Greater sensitivity to somatic information is associated with lower responses for detail and involvement. However, this is only in participants with lower overall information sensitivity.
- The effect of immersive tendencies is modulated by sensitivity to the magnitude of information content within a stimulus (ITQ × Extraversion interaction):
  - Higher scores on the ITQ result in lower assessments of visual realism and involvement. As sensory information sensitivity decreases, low scoring subjects report higher levels of visual realism and involvement.

These interpretations, derived from a information processing theory viewpoint, provide some insight that may be used to form a basis of a set of hypotheses which may be tested using behavioural tasks specifically designed to directly test these inferences using behavioural and physiological methods.

It can be seen that consistencies exist within the reported results. The modulation effect of extraversion suggests that increases in the thresholds of sensory information has characteristic effects on the domain in which information is restricted. This takes the form or either damping the effect of granularity of classification seen in systematisation, or by increasing the importance of the effect of greater visual information resolution in ratings of externalisation. It has been shown that the strength of the ventriloquist effect is affected by inter-participant variation. In addition, brain regions associated with spatial perception which are only activated when congruent sensory inputs in more than one modality are detected are implicated in the neural correlates of the ventriloquist [245]. The association of externalisation being dependent on visual information resolution when overall information sensitivity is low can be seen to fit this mechanism.

Empathy quotient, as interpreted as an inference of the automaticity of somatic response to stimuli, is associated with increased responses on components of presence and smaller differences between QoE ratings of real and simulated sources. In terms of externalisation of audio, this may be explained as an automatic reaction to a stimulus relating to spatial information eliciting the bodily feeling of externalised

sound or being situated within a space. The distinguishing effect of this information processing property is that it predominantly either amplifies an existing effect which contributes to this response or to override an effect which would otherwise produce a reduction in response. This property might, conversely, be characterised as a low specificity to somatic information stimuli with accompanied automatic response, in the case of the work presented here. The auralisations and environments used in this work are objectively unrealistic when compared to their real world counterparts. This is further supported by the association with lower visual realism only being observed in low stimulus information sensitivity participants. If this were true it would be necessarily to characterise both dimensions of the receiver operating characteristic for this information channel and determine the physiological response characteristics of subjects to describe the effect fully.

Systematisation as an organisation process appears to have several effects. The first is to reduce the distance in QoE ratings between objects perceived as real or simulated, with low information organisers using a wider range of scores to rate the stimuli. This may be indicative of a more granular approach to stimulus classification, with low information organisers operating under the rubric of binomial classification. This effect is amplified by a reduction in information sensitivity, suggesting that as received information decreases, the heuristics used by low organisation subjects must rely on a smaller number of salient features, in turn resulting in more definite classification decisions. Similarly, greater ability to resolve visual information is generally associated with higher ratings of visual realism. This might be considered surprising if reported realism were not associated with involvement in this dataset. As such, it is suggested that the more visual information that can be perceived by a subject who is attending to the virtual environment, the greater the subjective assessment of realism. As information organisation reduces, the use of the extremes of the scales increases due to the tendency to binomial classification. As visual information resolution is positively correlated with realism and involvement, this effect is magnified.

8

Conclusions

This thesis has presented work which constitutes novel contributions to the field of human perception of spatial audio in immersive virtual environments. The literature review in section 2.1 attempts to demonstrate that, despite active research in the fields of spatial audio reproduction and human response to immersive virtual environments, gaps in knowledge remain within the field. The literature discussed, particularly in section 2.3, illustrates the importance of presence as a dimension of response to stimulus in VR. It is also widely recognised that spatial audio reproduction can be used to dramatic effect when placing a listener within a scene in terms of unimodal presentation of programme material. However, beyond simple recognition that spatial audio should contribute to an increase in presence in a multimodal immersive environment, such a relationship had not hitherto been systematically explored in the way described in this work. Similarly, taxonomies of spatial quality for multimodal stimuli in immersive VR are not represented in the literature. Finally, although there are some studies which compare real environments to simulacra in terms of response to stimuli, there are no studies which make these comparisons directly, to determine if differences between immediate pre-exposure and VE geometry influence subjective responses.

To address these gaps in knowledge, an assessment instrument was constructed using a small number of assessment parameters from studies which have aimed to quantify the perception spatial audio and immersive media. The factors selected were audiovisual fusion, environmental plausibility, sense of localisation, externalisation and awareness of headphones. It was found throughout this work that a two factor model can be extracted from responses on these scales, explaining the variance in the data. However, the interpretation of these extracted dimensions is heavily context dependent. It was shown that the structure of the extracted dimensions were more explicable in conditions where pre-exposure geometry was dissimilar to that of

the virtual environment and when the spatial relationship between audio and visual stimuli was clear and obvious, with a low degree of ambiguity. This model consisted of a dimension associated with spatial scales (audiovisual fusion, environmental plausibility and sense of localisation) and a dimension associated with physical response (externalisation and awareness of headphones). However, when pre-exposure room geometry was similar to that of the VE, the strength of loading onto this simple structure model breaks down to one of 'overall' quality of experience, and externalisation becoming more associated with audiovisual fusion. A similar effect was also observed when the degree of ambiguity in the spatial relationship between audio and visual stimulus was increased. These context dependent effects contributed to an averaging of the loadings of the data into a univariate model of audiovisual quality of experience. It was found that pre-exposure similarity produced anchoring effects on this dimension, with similar environments assessed more critically than a novel environment with comparable acoustic properties.

These scales were also used to assess response when participants believed whether an auditory stimulus was reproduced in the real space over loudspeaker or simulated to be reproduced over headphones. It was found that higher responses were associated with believing sources to be emitted over loudspeaker independent of the ability of the subject to correctly identify loudspeaker emitted audio. Although time domain accuracy of the impulse response used in auralisation was shown to effect the difference between judgements of real or simulated, the largest effect was individual differences in response between individuals. This was also the case with responses in experiments rating quality of experience with simulated impulse response length as an experimental design variable. Although manipulation of IR content affected responses on spatial subscales of QoE, the largest effect was that introduced by differences between individual subjects.

Audiovisual quality of experience scales were correlated with ratings on subscales of presence as defined by the iGroup presence questionnaire (IPQ), a tool identified in the literature review as the best performing model of presence in terms of describing the variance in responses for presence. The scales used were spatial presence, visual realism and involvement. The extracted feature space suggested that subjective rating of visual realism and involvement were largely independent of audiovisual qualities, particularly externalisation and sense of localisation. However, these two dimensions (visual realism/involvement and AV QoE) contributed together to form spatial presence. That visual attention and audiovisual spatial information combine

to form the sense that one is spatially located within the virtual space. In the IPQ, which does not take auditory cues into account, this quantity is assumed to be an independent dimension. Overall presence scores were found to load equally between the visual/involvement and spatial presence vectors. This result describes the relationships between audiovisual quality of experience and presence and its component factors to an existing model of presence in a way which makes a new contribution to the understanding of the importance of spatial audio for the design and implementation of successful immersive virtual reality systems. Additionally, it was found that manipulation of impulse response and spatial co-location of audio and visual objects only affected responses which loaded onto the avQoE dimension. Visual realism and involvement were not impacted. The effects of IR content and spatial relationships were reflected in their being identified as predictors of spatial presence, which was the single significant dimension in the unrotated solution for the principal component analysis of the response data. As with the AV QoE data, extracted dimensions were significantly affected by individual differences between participants.

Identifying individual differences as being a large effect predictor of presence and its constituent factors provided an opportunity to make a new contribution to the understanding of subjective testing in VR, with wider potential implications for subjective testing in audio. The literature reviewed in section 2.4 covers models of personality and cognition which have been used to characterise individuals in terms of behavioural preferences and cognitive processing differences. Existing models of response to immersive VR environments are also addressed, where personality factors are attempted to be used to predict reported presence to VR scenes. It is noted, however, that the literature on this subject is not conclusive, even when the same factors are included in models and similar statistical methodologies are employed. It was noted that this inconsistency in the reported models within the data may stem from an assumption of an additive relationship between predictors and ignores the possibility of multidimensionality or interactivity between such predictors. Using a hierarchical statistical approach with mixed effects modelling it was found that interactions did occur between predictors and that when taking into account the modulatory relationships between extraversion, empathy, systematisation, and visual response speed (global precedence), in addition to responses on the immersive tendencies questionnaire (ITQ), it was possible to fit a fixed-effects model predicting the independent constituents of spatial presence (visual realism/involvement and externalisation/localisation of sound) with better goodness-of-fit than a random-effects

model grouping responses by individual participant. This portion of the work presented demonstrates the new knowledge that simple additive models of presence to predict complex perceptual responses are insufficient and a small number of interactive predictors can perform better than the naïve assumption that all subjects will respond differently to each other. Further to this, the identified predictors are related to each other in such a way as to allow for the construction of a model of presence which does not rely on disparate semantic groupings but on a model of information processing which, while posited here as an explanatory device to describe the patterns observed in the data, can be rigorously and falsifiably tested in further work. The proposed model of presence is described fully in section 7.5. However, it can be summarised as a set of interactions between sensitivity thresholds to sensory stimuli, automaticity of constructing internal models of object states, automaticity of stimulus information organisation and cognitive effort required to resolve local detail in stimuli.

The work presented in this thesis can be thought of as falling into three main categories. The first is the support of the assertion that perception of audiovisual stimuli in immersive virtual environments, where pre-exposure geometry is dissimilar from the VE and those where pre-exposure geometry is similar, is distinguishable. This is underpinned by objectives 1a, 1b and 1c identified in chapter 1. It was determined that the audiovisual and spatial metrics of audiovisual fusion, plausibility, localisation, externalisation and awareness of headphones can be generalised into a construct of generalised audiovisual quality of experience. This underlying latent variable was shown to be reported as lower in similar environments with a contrasting dissimilar environment case showing higher responses and lower variance with ceiling effects. This was interpreted as a result of perceptual anchoring in the case of similar environments, allowing subjects to better distinguish inaccurate spatial cues and resolve audiovisual inconsistencies. This effect was observed in the interactions between the similarity and spatial room response simulation and spatial co-location of auditory and visual events.

Secondly, this thesis investigates the relationship between the construct of audiovisual quality, latent dimensions of presence identified in the literature and judgements of source realism. This is underpinned by objectives 2a, 2b and 2c identified in chapter 1. It was shown that spatial presence can be thought of as a combination of visual realism and involvement and audiovisual quality, with audiovisual stimuli relaying information pertaining to the spatial quality of the objects and events within the

virtual environment. Manipulation of acoustic room response modelling and audiovisual spatial co-location of stimuli results in changes to responses on the audiovisual quality component of this construct. Changes to the stimuli made in this way resulted in an overall change in spatial presence, but only as a result of the effect on audiovisual quality. Responses on the visual realism/involvement component were not affected by either manipulation of acoustic response or audiovisual co-location.

Thirdly, this thesis attempts to account for the variation in *presence* and responses of audiovisual spatial quality observed in the data. This is in line with objective 3a identified in chapter 1. Although some of the variance can be explained by the manipulation of independent variables related to the stimuli presented, the majority of the variance is attributable to inter-participant variation. The results presented here support the formation of a hypothesis that such variation is due to intrinsic properties of the subject as a receiver of information. Using information processing theory and perception-action models of concepts already identified in the literature as salient to the experience of *presence*, allows for the formulation of a model which would attempt to explain observed results rather than simply describing them. It is hypothesised that it is possible to account for differences in the data on an inter-participant level. This can be understood as a set of interactions between organisation, sensitivity and specificity to information within multimodal stimuli that elicit the internal representation of a somatic state that is recognised as spatial presence. Although such a model would require further systematic investigation to confirm, the identification of an hypothetical explanatory model of this type is a contribution to an area of study which has hitherto relied on semantic constructs which imply no level of theoretical causation, and have had limited success in replication.

PART III. BIBLIOGRAPHY AND APPENDICES

Bibliography

[1] F. L. Wightman and D. J. Kistler. A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction. *The Journal of the Acoustical Society of America*, 91(1992):1637–1647, 1992. ISSN 0001-4966. doi: 10.1121/1.402444.

[2] M. Wenzel, Elizabeth, M. Arruda, J. Kistler, Doris, and F. L. Wightman. Localization using nonindividualized head-related transfer functions. *Journal of the Acoustical Society of America.*, 94(1):111–123, 1993.

[3] C. Müller-Tomfelde. Low-Latency Convolution for Real-Time Applications. In *Audio Engineering Society Conference: 16th International Conference: Spatial Sound Reproduction*, mar 1999.

[4] D. Satongar, Y. W. Lam, and C. Pike. Measurement and analysis of a spatially sampled binaural room impulse response dataset. In *21st International Congress on Sound and Vibration, Beijing*, 2014.

[5] D. Poirier-quinot, B. N. J. Postma, B. F. G. Katz, A. A. Group, and U. Paris-saclay. Simulation and Auralization of Concert Halls / Opera Houses : Augmented auralization : Complementing auralizations with immersive virtual reality technologies Augmented auralization. *International Symposium on Musical and Room Acoustics*, (Kinect 2), 2016.

[6] B. N. J. Postma and B. F. G. Katz. The Influence of Visual Distance on the Room-acoustic Experience of Auralizations. *The Journal of the Acoustical Society of America*, 142(5):3035–3046, 2017. ISSN 00014966. doi: 10.1121/1.5009554. URL `https://doi.org/10.1121/1.5009554{%}0Ahttp://asa.scitation.org/toc/jas/142/5`.

[7] A. Andreopoulou and B. F. G. Katz. Investigation on Subjective HRTF Rating Repeatability. *140th Audio Engineering Society Convention*, 2016.

[8] A. Andreopoulou and B. F. G. Katz. Subjective HRTF evaluations for obtaining global similarity metrics of assessors and assessees. *Journal on Multimodal User Interfaces*, 10(3):259–271, 2016. ISSN 17838738. doi: 10.1007/s12193-016-0214-y.

[9] J. J. Gibson. *The Ecological Approach to Visual Perception.* Houghton Miffin, Boston, Massachusetts, USA, 1979. ISBN 9781848725775.

[10] D. M. Green and J. A. Swets. *Signal Detection Theory and Psychophysics.* Wiley, New York, NY, USA, 1966.

[11] The ITU Radiocommunication Assembly. RECOMMENDATION ITU-R BS . 1534-1 Method for the subjective assessment of intermediate quality level of coding systems Annex 1. Technical report, The ITU Radiocommunication Assembly, 2003.

[12] The ITU Radiocommunication Assembly. Recommendation ITU-R BS.116-3: Methods for the subjective assessment of small impairments in audio systems. Technical report, The ITU Radiocommunication Assembly, 2015.

[13] J. Berg. How do we determine the attribute scales and questions that we should ask of subjects when evaluating spatial audio quality. *Int. Workshop on Spatial Audio and Sensory Evaluation*, (1):1–5, 2006. URL `http://www.surrey.ac.uk/soundrec/ias/papers/Berg.pdf`.

[14] N. Zacharov and T. H. Pedersen. Spatial sound attributes - development of a common lexicon. *AES 139th Convention*, 2015.

[15] A. Wilson, T. Cox, N. Zacharov, and C. Pike. Perceptual Audio Evaluation of Media Device Orchestration Using the Multi-Stimulus Ideal Profile Method. In *Audio Engineering Society Convention 145*, oct 2018. URL `http://www.aes.org/e-lib/browse.cfm?elib=19793`.

[16] P. Bertelson and J. O. N. Driver. The ventriloquist effect does not depend on the direction of deliberate visual attention. *Perception & Psychophysics*, 62(2):321–332, 2000.

[17] W. M. Hartmann and A. Wittenberg. On the externalization of sound images, 1996. ISSN 00014966.

[18] A. Lindau and S. Weinzierl. Assessing the plausibility of virtual acoustic environments. *Acta Acustica united with Acustica*, 98(5):804–810, 2012. ISSN 16101928. doi: 10.3813/AAA.918562.

[19] A. Lindau, V. Erbes, S. Lepa, H. J. Maempel, F. Brinkman, and S. Weinzierl. A spatial audio quality inventory (SAQI). *Acta Acustica united with Acustica*, 100(5): 984–994, 2014. ISSN 16101928. doi: 10.3813/AAA.918778.

[20] J. Berg. Evaluation of perceived spatial audio quality. *The 9th World Multi-Conference on Systemics, Cybernetics and Informatics*, 4(2):10–14, 2005. URL http://www.iiisci.org/journal/CV{\protect\T1\textdollar}/sci/pdfs/P146141.pdf.

[21] M. Slater. 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, 2009. ISSN 0962-8436. doi: 10.1098/rstb.2009.0138. URL http://rstb.royalsocietypublishing.org/cgi/doi/10.1098/rstb.2009.0138.

[22] B. G. Witmer and M. J. Singer. Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoper. Virtual Environ.*, 7(3):225–240, 1998. ISSN 1054-7460. doi: 10.1162/105474698565686. URL http://www.mitpressjournals.org/doi/abs/10.1162/105474698565686.

[23] M. Sierra and G. E. Berrios. Depersonalization: neurobiological perspectives. *Biological Psychiatry*, 44(9):898–908, 1998. ISSN 0006-3223. doi: https://doi.org/10.1016/S0006-3223(98)00015-8. URL http://www.sciencedirect.com/science/article/pii/S0006322398000158.

[24] B. G. Witmer and M. J. Singer. Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoper. Virtual Environ.*, 7(3):225–240, 1998. ISSN 1054-7460. doi: 10.1162/105474698565686.

[25] F. Rumsey. *Spatial Audio*. Music technology series. Focal Press, 2001. ISBN 9780240516233. URL https://books.google.co.uk/books?id=0kVx{_}JjFIKIC.

[26] ITU. Multichannel sound technology in home and broadcasting applications BS Series. Technical report, 2012.

[27] V. Pulkki. Virtual Sound Source Positioning Using Vector Base Amplitude Panning. *Journal of the Audio Engeenering Society*, 45(6):456–466, 1997.

[28] M. A. Gerzon. Ambisonics in Multichannel Broadcasting and Video. *Journal of the Audio Engineering Society.*, 33(11):859–871, 1985.

[29] A. D. Blumlein. Improvements Relating to Sound-transmission, Sound-recording and Sound-reproducing Systems, 1933. URL papers3://publication/uuid/7B6533D1-0F08-4115-ACDD-2B829216F2F4.

[30] J. Trevino, T. Okamoto, Y. Iwaya, and Y. Suzuki. High order Ambisonic decoding method for irregular loudspeaker arrays. *International Congress on Acoustics*, (August):1–8, 2010.

[31] J. Daniel, R. Nicol, and S. Moreau. Further Investigations of High Order Ambisonics and Wavefield Synthesis for Holophonic Sound Imaging. In *114th Convention of the Audio Engineering Society .2003 March 22-25*, pages 1–12, Amsterdam, The Netherlands, 2003.

[32] M. Kronlachner and F. Zotter. Spatial transformations for the enhancement of Ambisonic recordings. *2nd International Conference on Spatial Audio*, (2):1–5, 2014.

[33] A. J. Berkhout, D. de Vries, and P. Vogel. Acoustic control by wave field synthesis. *The Journal of the Acoustical Society of America*, 93(5):2764–2778, 1993. ISSN 0001-4966. doi: 10.1121/1.405852. URL `http://asa.scitation.org/doi/10.1121/1.405852`.

[34] T. Caulkins, E. Corteel, and O. Warusfel. Wave Field Synthesis Interaction With the Listening Environment , Improvements in the Reproduction of Virtual Sources Situated Inside the Listening Room. *Audio*, pages 6–9, 2003. URL `http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.9.6790{&}rep=rep1{&}type=pdf`.

[35] D. de Vries. Sound Reinforcement by Wavefield Synthesis: Adaptation of the Synthesis Operator to the Loudspeaker Directivity Characteristics. *Journal of the Audio Engineering Society*, 44(12):1120—-1131, 1996. ISSN 00047554. URL `http://www.aes.org/e-lib/browse.cfm?elib=7872`.

[36] M. Noisternig, T. Musil, A. Sontacchi, and R. Höldrich. 3D binaural sound reproduction using a virtual ambisonic approach. *VECIMS 2003 - 2003 International Symposium on Virtual Environments, Human-Computer Interfaces and Measurement Systems*, (July):174–178, 2003. doi: 10.1109/VECIMS.2003.1227050.

[37] D. R. Begault, M. Wenzel, Elizabeth, and M. R. Anderson. Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source. *Journal of the Audio Engineering Society. Audio Engineering Society*, 49(10):904–916, 2001. ISSN 0004-7554.

[38] H. Møller, M. F. Sørensen, C. B. Jensen, and D. Hammershøi. Binaural technique: Do we need individual recordings? *Journal of the Audio Engineering Society*, 44(6): 451–469, 1996. ISSN 0004-7554. URL `http://www.aes.org/e-lib/browse.cfm?elib=7897`.

[39] C. Mendonça, G. Campos, P. Dias, J. Vieira, J. Ferreira, and J. Santos. On the improvement of localization accuracy with non- individualized HRTF-based sounds. *Journal of the Audio Engeenering Society*, 60(10):821–830, 2012.

[40] G. Parseihian and B. F. G. Katz. Rapid head-related transfer function adaptation using a virtual auditory environment. *The Journal of the Acoustical Society of America*, 131(4):2948, 2012. ISSN 00014966. doi: 10.1121/1.3687448.

[41] C. C. Berger, M. Gonzalez-Franco, A. Tajadura-Jiménez, D. Florencio, and Z. Zhang. Generic HRTFs may be good enough in virtual reality. Improving source localization through cross-modal plasticity. *Frontiers in Neuroscience*, 12(FEB), 2018. ISSN 1662453X. doi: 10.3389/fnins.2018.00021.

[42] K. Iida, M. Itoh, A. Itagaki, and M. Morimoto. Median plane localization using a parametric model of the head-related transfer function based on spectral cues. *Applied Acoustics*, 68(8):835–850, 2007. ISSN 0003682X. doi: 10.1016/j.apacoust.2006.07.016.

[43] A. Lindau. On the extraction of interaural time differences from binaural room impulse responses. *Ak.Tu-Berlin.De*, (September 2010), 2010. URL `http://www2.ak.tu-berlin.de/{~}akgroup/ak{_}pub/abschlussarbeiten/2011/EstrellaJorgos{_}StudA.pdf`.

[44] C. Schissler, A. Nicholls, and R. Mehra. Efficient HRTF-based Spatial Audio for Area and Volumetric Sources. *IEEE Transactions on Visualization and Computer Graphics*, 22(4):1356–1366, 2016. ISSN 10772626. doi: 10.1109/TVCG.2016.2518134.

[45] S. Mehrotra, W.-g. Chen, and Z. Zhang. Interpolation of combined head and room impulse response for audio spatialization. *2011 IEEE 13th International Workshop on Multimedia Signal Processing*, pages 1–6, 2011. doi: 10.1109/MMSP.2011.6093794. URL `http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6093794`.

[46] R. Stewart and M. Sandler. Generating a Spatial Average Reverberation Tail Across Multiple Impulse Responses. *Electronic Engineering*, pages 1–6, 2009.

[47] A. Lindau. On the extraction of interaural time differences from binaural room impulse responses. *Ak.Tu-Berlin.De*, (September 2010), 2010. URL `http://www2.ak.tu-berlin.de/{%}7B{~}{%}7Dakgroup/ak{_}pub/abschlussarbeiten/2011/EstrellaJorgos{_}StudA.pdf`.

[48] J. S. Abel and P. Huang. A Simple, Robust Measure of Reverberation Echo Density. *Audio Engineering Society*, page 6985, 2006. ISSN 08957061. doi: 10.1016/S0895-7061(01)01478-9. URL `http://www.aes.org/e-lib/browse.cfm?elib=13819`.

[49] M. R. Schroeder. Digital Simulation of Sound Transmission in Reverberant Spaces. *The Journal of the Acoustical Society of America*, 47(2A):424–431, 1970. ISSN 0001-4966. doi: 10.1121/1.1911541. URL `http://asa.scitation.org/doi/10.1121/1.1911541`.

[50] J. A. Moorer. About this Reverberation Business. *Computer Music Journal*, 3(2): 13 – 28, 1979.

[51] D. Rocchesso and J. O. Smith. Circulant and elliptic feedback delay networks for artificial reverberation. *IEEE Transactions on Speech and Audio Processing*, 5(1): 51–63, 1997. ISSN 10636676. doi: 10.1109/89.554269.

[52] Y. W. Cheng and M. C. Ku. Apparatus for generating stereo sound and method for the same, 2005. URL `https://www.google.com/patents/US20050213770`.

[53] S. Heise, M. Hlatky, and J. Loviscach. Automatic Adjustment of Off-the-Shelf Reverberation Effects. In *Audio Engineering Society Convention 126*, pages 1–8, Munich, Germany, 2009.

[54] I. Smith, Julius O. A New Approach to Digital Reverberation Using Closed Waveguide Networks.pdf, 1985.

[55] M. Vorländer. *Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*. RWTHedition. Springer Berlin Heidelberg, 2007. ISBN 9783540488309. URL `https://books.google.co.uk/books?id=CuXF3JkTuhAC`.

[56] A. J. Burton and G. F. Miller. The Application of Integral Equation Methods to the Numerical Solution of Some Exterior Boundary Value Problems. *Proceedings of the Royal Society of London*, 323(1553):201–210, 1970.

[57] O. C. Zienkiewicz and R. L. Taylor. The Finite Element Method Volume 1 : The Basis. *Methods*, 1:708, 2000.

[58] L. Yang and B. Sheild. Development of a Ray Tracing Computer Model for the Prediction of the Sound Field in Long Enclosures. *Journal of Sound and Vibration*, 229(1):133–146, 2000. ISSN 0022460X. doi: 10.1006/jsvi.1999.2477. URL `http://linkinghub.elsevier.com/retrieve/pii/S0022460X9992477X`.

[59] J. B. Allen. Image method for efficiently simulating small-room acoustics. *The Journal of the Acoustical Society of America*, 65(4):943, 1979. ISSN 00014966. doi: 10.1121/1.382599.

[60] J. Borish. Extension of the image model to arbitrary polyhedra. *The Journal of the Acoustical Society of America*, 75(6):1827, 1984. ISSN 00014966. doi: 10.1121/1.390983.

[61] F. P. Mechel. Improved mirror source method in roomacoustics. *Journal of Sound and Vibration*, 256(5):873–940, 2002. ISSN 0022460X. doi: 10.1006/jsvi.5025. URL http://linkinghub.elsevier.com/retrieve/pii/S0022460X0295025X.

[62] R. Heinz. Binaural room simulation based on an image source model with addition of statistical methods to include the diffuse sound scattering of walls and to predict the reverberant tail. *Applied Acoustics*, 38(2-4):145–159, 1993. ISSN 0003682X. doi: 10.1016/0003-682X(93)90048-B.

[63] N. Raghuvanshi and J. Snyder. Triton: Practical pre-computed sound propagation for games and virtual reality. *The Journal of the Acoustical Society of America*, 141 (5), 2017.

[64] M. Gorzel, F. Boland, B. O'Toole, and I. Kelly. 3D Immersive Spatial Audio Systems and Methods, 2016.

[65] I. E. Sutherland. The ultimate display. *Proceedings of the Congress of the Internation Federation of Information Processing (IFIP)*, 21(3):506–508, 1965. doi: 10.1109/MC.2005.274. URL http://www.wired.com/beyond{_}the{_}beyond/2009/09/augmented-reality-the-ultimate-display-by-ivan-sutherland-1965/.

[66] I. E. Sutherland. A Head-Mounted Three-Dimensional Display. In *AFIPS Conference Proceedings (1968) 33, I*, pages 757–764, 1968.

[67] G. E. Moore. Moore's Law at Forty. In D. C. Brock, editor, *Understanding Moore's Law: Four Decades of Innovation*, number Figure 2, pages 67–84. Chemical Heritage Foundation, Philadelphia, PA, 2006.

[68] R. A. Robison. Moore's law: Predictor and driver of the silicon era. *World Neurosurgery*, 78(5):399–403, 2012. ISSN 18788750. doi: 10.1016/j.wneu.2012.08.019. URL http://dx.doi.org/10.1016/j.wneu.2012.08.019.

[69] Facebook. No Title, 2016. URL www.oculus.com.

[70] Sony. PlayStation VR, 2016.

[71] HTC. No Title, 2016.

[72] Y. Boger and R. A. Pavlik. An Introduction to OSVR, 2015. URL http://osvr.github.io/whitepapers/introduction{_}to{_}osvr/.

[73] M. Hall. Position tracking system that exploits arbitrary configurations to determine loop closure, 2018. URL https://patents.justia.com/patent/9983665.

[74] C. Cruz-Neira, D. J. Sandin, T. A. DeFanti, R. V. Kenyon, and J. C. Hart. The CAVE: audio visual experience automatic virtual environment. *Communications of the ACM*, 35(6):64–72, 1992. ISSN 00010782. doi: 10.1145/129888.129892. URL http://portal.acm.org/citation.cfm?doid=129888.129892.

[75] M. C. Juan and D. Perez. Comparison of the Levels of Presence and Anxiety in an Acrophobic Environment Viewed via HMD or CAVE. *Presence: Teleoperators and Virtual Environments*, 18(3):232–248, 2009.

[76] M. V. Sanchez-Vives and M. Slater. From presence to consciousness through virtual reality. *Nat Rev Neurosci*, 6(4):332–339, apr 2005. ISSN 1471-003X. URL http://dx.doi.org/10.1038/nrn1651.

[77] M. Slater and M. Usoh. Body Centred Interaction in Immersive Virtual Environments. pages 1–22, 1968.

[78] M. Slater and S. Wilbur. A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments. *Presence: Teleoperators and Virtual Environments*, 6(6):603–616, 1997. ISSN 10547460. doi: 10.1007/s10750-008-9541-7. URL http://discovery.ucl.ac.uk/79956/.

[79] T. Schubert, F. Friedmann, and H. Regenbrecht. The Experience of Presence: Factor Analytic Insights. *Presence: Teleoperators and Virtual Environments*, 10 (3):266–281, 2001. ISSN 1054-7460. doi: 10.1162/105474601300343603. URL http://www.mitpressjournals.org/doi/abs/10.1162/105474601300343603.

[80] J. M. Flach and J. G. Holden. The Reality of Experience: Gibson's Way. *Presence: Teleoperators and Virtual Environments*, 7(1):90–95, 1998. ISSN 1054-7460. doi: 10.1162/105474698565550.

[81] F. Biocca. The Cyborg's Dilemma: Progressive Embodiment in Virtual Environments [1]. *Journal of Computer-Mediated Communication*, 3(2):0, 1997. ISSN 1083-6101. doi: 10.1111/j.1083-6101.1997.tb00070.x. URL http://dx.doi.org/10.1111/j.1083-6101.1997.tb00070.x.

[82] J. L. Higuera-Trujillo, J. López-Tarruella Maldonado, and C. Llinares Millán. Psychological and physiological human responses to simulated and real environments: A comparison between Photographs, 360Â° Panoramas, and Virtual Reality. *Applied Ergonomics*, 65:398–409, 2017. ISSN 18729126. doi: 10.1016/j.apergo.2017.05.006.

[83] A. Spagnolli, C. C. Bracken, and U. Padova. Do you feel as if you are there? Measuring presence in cybertherapy. *ISPR 2011: The International Society for Presence Research Annual Conference*, pages 26–28, 2011.

[84] J. Lessiter, J. Freeman, E. Keogh, and J. Davidoff. A Cross-Media Presence Questionnaire: The ITC-Sense of Presence Inventory. *Presence: Teleoperators and Virtual Environments*, 10(3):282–297, 2001. ISSN 1054-7460. doi: 10.1162/105474601300343612. URL `http://www.mitpressjournals.org/doi/abs/10.1162/105474601300343612`.

[85] M. Slater, M. Usoh, and A. Steed. Depth of presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, 3(2):130–144, 1994. ISSN 10547460. doi: 10.1371/journal.pone.0013904. URL `http://s3.amazonaws.com/publicationslist.org/data/melslater/ref-24/depthofpresence.pdf`.

[86] R. M. Baños, C. Botella, A. Garcia-Palacios, H. Villa, C. Perpiña, and M. Alcañiz. Presence and Reality Judgment in Virtual Environments: A Unitary Construct? *CyberPsychology & Behavior*, 3(3):327–335, 2000. ISSN 1094-9313. doi: 10.1089/10949310050078760. URL `http://www.liebertonline.com/doi/abs/10.1089/10949310050078760`.

[87] G. Fontaine. The Experience of a Sense of Presence in Intercultural and International Encounters. *Presence: Teleoperators & Virtual Environments*, 1(4):482–490, 1992.

[88] A. M. Treisman. Verbal Cues , Language , and Meaning in Selective Attention. *The American Journal of Psychology*, 77(2):206–219, 1964.

[89] M. Chion. *Audio-Vision: Sound on Screen*. Columbia University Press, New York, NY, USA, 1994.

[90] Z. Whalen. Play along - An approach to videogame music. *Game Studies*, 4(1):http://www.gamestudies.org/0401/whalen/, 2004. URL `http://www.gamestudies.org/0401/whalen/`.

[91] E. Weis and J. Belton. *Film Sound: Theory and Practice*. Columbia University Press, 1985. ISBN 9780231056373. URL `https://books.google.co.uk/books?id=Dz-1LBWOod0C`.

[92] M. Gröhn, T. Lokki, and T. Takala. Comparison of auditory, visual, and audiovisual navigation in a 3D space. *ACM Transactions on Applied Perception*, 2(4):564–570, 2005. ISSN 1544-3558. doi: 10.1145/1101530.1101558.

[93] R. Shumaker and L. Stephanie. Virtual , Augmented and Mixed Reality: Designing and Developing Augmented Environments. In R. Shumaker and S. Lackey, editors, *Virtual, Augmented and Mixed Reality Designing and Developing Virtual and Augmented Environments 6th International Conference, VAMR 2014 Held as Part of HCI International*, Heraklion, Crete, Greece, 2014. Springer. ISBN 9783319074573. doi: 10.1007/978-3-319-07464-1.

[94] I. Drumm and J. O'Hare. Workflow Automations and Optimisations To Facilitate Room Acoustics Prediction Within Multimodal Virtual Environments. (July):12–16, 2015.

[95] M. Rébillat, X. Boutillon, É. Corteel, and B. F. G. Katz. Audio, visual, and audio-visual egocentric distance perception in virtual environments. *Forum Acusticum*, 1 (1):0–5, 2011. ISSN 22213767.

[96] M. Rebillat, B. F. G. Katz, and C. Oberglatt. Smart-I 2 : Spatial Multi-User Audio-Visual Real-Time Interactive Interface, a Broadcast Application Context. *Design*, pages 2–5, 2009.

[97] S. Pelzer, B. Masiero, and M. Vorländer. 3D Reproduction of Room Auralizations by Combining Intensity Panning, Crosstalk Cancellation and Ambisonics. *Proceedings of the EAA Joint Symposium on Auralization and Ambisonics*, (April):3–5, 2014. doi: 10.14279/depositonce-33.

[98] D. R. Begault. 3-D Sound for Virtual Reality and Multimedia. Technical report, NASA, Moffett Field, California, USA, 1995. URL `http://www.jstor.org/stable/3680997?origin=crossref`.

[99] B. E. Riecke, J. Schulte-Pelkum, F. Caniard, and H. H. B Ülthoff. Spatialized auditory cues enhance the visually-induced self-motion illusion (circular vection) in Virtual Reality. *Max Planck Institute for Biological Cybernetics Technical Report*, 138(138):1–11, 2005. doi: http://www.twk.tuebingen.mpg.de/twk05/abstract.php?_load_id=riecke01. URL `http://www.kyb.mpg.de/publication.html?publ=4187`.

[100] S. Dargar, R. Kennedy, W. Lai, V. Arikatla, and S. De. Towards immersive virtual reality (iVR): a route to surgical expertise. *Journal of Computational Surgery*, 2 (1):2, 2015. ISSN 2194-3990. doi: 10.1186/s40244-015-0015-8. URL `http://www.computationalsurgery.com/content/2/1/2`.

[101] K. Asakawa and H. Ishikawa. Binocular Vison and Depth Perception: Development and Disorders. In J. McCoun and L. Reeves, editors, *Binocular Vision: Development, Depth Perception and Disorders*, pages 139–153. Nova Science Publishers, 2010. ISBN 978-1-61761-957-1.

[102] J. M. Foley. Binocular distance perception. *Psychological Review*, 87(5):411–434, 1980. ISSN 0003-066X. doi: 10.1037/h0021465.

[103] L. A. Remington. Uvea. *Clinical Anatomy and Physiology of the Visual System*, (1): 40–60, 2012. doi: 10.1016/B978-1-4377-1926-0.10003-7. URL `http://linkinghub.elsevier.com/retrieve/pii/B9781437719260100037`.

[104] V. W. Grant. Accomodation and convergence in visual space perception. *Journal of Experimental Psychology*, 31(2):89–104, 1942.

[105] L. Beck, M. Wolter, N. F. Mungard, R. Vohn, M. Staedtgen, T. Kuhlen, and W. Sturm. Evaluation of spatial processing in virtual reality using functional magnetic resonance imaging (fMRI). *Cyberpsychology, behavior and social networking*, 13(2):211–5, 2010. ISSN 2152-2723. doi: 10.1089/cyber.2008.0343. URL `http://www.ncbi.nlm.nih.gov/pubmed/20528281`.

[106] J. E. Cutting and P. M. Vishton. Perceiving Layout and Knowing Distances: The Integration, Relative Potency and Contextual Use of Different Information about Depth. In W. Epstien and S. Rogers, editors, *Perception of space and motion*, pages 69 – 117. San Diego, CA., 1995.

[107] D. Rojas, B. Kapralos, A. Hogue, K. Collins, L. E. Nacke, S. Cristancho, C. Conati, and A. Dubrowski. The effect of sound on visual fidelity perception in stereoscopic 3-D. *IEEE Transactions on Cybernetics*, 43(6):1572–1583, 2013. ISSN 2168-2275. doi: 10.1109/TCYB.2013.2269712. URL `http://www.researchgate.net/publication/257755463{_}The{_}Effect{_}of{_}Sound{_}on{_}Visual{_}Fidelity{_}Perception{_}in/file/50463526b9e3a3b860.pdf`.

[108] J. Byun and C. S. Loh. Audial engagement: Effects of game sound on learner engagement in digital game-based learning environments. *Computers in Human Behavior*, 46:129–138, 2015. ISSN 07475632. doi: 10.1016/j.chb.2014.12.052. URL `http://linkinghub.elsevier.com/retrieve/pii/S0747563215000084`.

[109] N. Tsingos, E. Gallo, and G. Drettakis. Perceptual audio rendering of complex virtual environments. *ACM Transactions on Graphics*, 23(3):249, 2004. ISSN 07300301. doi: 10.1145/1015706.1015710. URL `http://portal.acm.org/citation.cfm?doid=1015706.1015710`.

[110] B. E. Riecke. Moving Sounds Enhance the Visually-Induced Self-Motion Illusion (Circular Vection) in Virtual Reality. *ACM Transactions on Applied Perception*, 6 (2):7–27, 2009.

[111] A. Väljamäe. *Self-motion and Presence in the Perceptual Optimization of a Multisensory Virtual Reality Environment*. PhD thesis, Chalmers University of Technology, Göteborg, 2005.

[112] H. McGurk and J. MacDonald. Hearing lips and seeing voices. *Nature*, 1976.

[113] D. Alais and D. Burr. The Ventriloquist Effect Results from Near-Optimal Bimodal Integration. *Current Biology*, 14(3):257–262, 2004. ISSN 09609822. doi: 10.1016/j.cub.2004.01.029. URL `http://linkinghub.elsevier.com/retrieve/pii/S0960982204000430`.

[114] J. Prado and D. H. Weissman. Spatial attention influences trial-by-trial relationships between response time and functional connectivity in the visual cortex. *NeuroImage*, 54(1):465–473, 2011. ISSN 10538119. doi: 10.1016/j.neuroimage.2010.08.038. URL `http://dx.doi.org/10.1016/j.neuroimage.2010.08.038`.

[115] A. Reeves. Temporal Resolution. In W. Prinz and B. Bridgeman, editors, *Handbook of Perception and Action*, chapter 1, pages 11–24. Elsevier, 1996.

[116] M. P. Zwiers, A. J. Van Opstal, J. R. M. Cruysberg, A. J. V. Opstal, and J. R. M. Cruysberg. A spatial hearing deficit in early-blind humans. *J.Neurosci.*, 21(1529-2401):RC142—-RC145, 2001. ISSN 1529-2401. doi: 20015180[pii].

[117] M. Turatto, V. Mazza, and C. Umiltà. Crossmodal object-based attention: Auditory objects affect visual processing. *Cognition*, 96(2):B55–B64, 2005. ISSN 00100277. doi: 10.1016/j.cognition.2004.12.001. URL `http://linkinghub.elsevier.com/retrieve/pii/S0010027704002264`.

[118] W. A. Teder-Salejarvi, F. Di Russo, J. J. Mcdonald, and S. S. Hillyard. Effects of Spatial Congruity on Audio-Visual Multimodal Integration. *Journal of Cognitive Neuroscience*, 17(9):1396–1409, 2005.

[119] C. Keysers, E. Kohler, M. A. Umiltà, L. Nanetti, L. Fogassi, and V. Gallese. Audio-visual mirror neurons and action recognition. *Experimental Brain Research*, 153(4): 628–636, 2003. ISSN 00144819. doi: 10.1007/s00221-003-1603-5.

[120] K. Alaerts, S. P. Swinnen, and N. Wenderoth. Interaction of sound and sight during action perception: Evidence for shared modality-dependent action representations. *Neuropsychologia*, 47(12):2593–2599, 2009. ISSN 00283932. doi: 10.1016/j.neuropsychologia.2009.05.006.

[121] J. H. Brockmyer, C. M. Fox, K. A. Curtiss, E. Mcbroom, K. M. Burkhart, and J. N. Pidruzny. Journal of Experimental Social Psychology The development of the Game Engagement Questionnaire : A measure of engagement in video game-playing. *Journal of Experimental Social Psychology*, 45(4):624–634, 2009. ISSN 0022-1031. doi: 10.1016/j.jesp.2009.02.016. URL `http://apps.webofknowledge.com.ejgw.nul.nagoya-u.ac.jp/full{_}record.do?product=UA{&}search{_}mode=CitingArticles{&}qid=14{&}SID=X1FaVKxvVSnqR7jUQEs{&}page=1{&}doc=2{%}5Cnhttp://dx.doi.org/10.1016/j.jesp.2009.02.016`.

[122] Y. K. Choi, S. M. Lee, and H. Li. Audio and Visual Distractions and Implicit Brand Memory: A Study of Video Game Players. *Journal of Advertising*, 42(2-3): 219–227, 2013. ISSN 0091-3367. doi: 10.1080/00913367.2013.775798. URL `http://www.tandfonline.com/doi/abs/10.1080/00913367.2013.775798`.

[123] S. E. Kober and C. Neuper. Using auditory event-related EEG potentials to assess presence in virtual reality. *International Journal of Human Computer Studies*, 70 (9):577–587, 2012. ISSN 10715819. doi: 10.1016/j.ijhcs.2012.03.004. URL `http://dx.doi.org/10.1016/j.ijhcs.2012.03.004`.

[124] C. Hendrix and W. Barfield. Presence in virtual environments as a function of visual and auditory cues. *Proceedings Virtual Reality Annual International Symposium '95*, pages 74–82, 1995. doi: 10.1109/VRAIS.1995.512482.

[125] P. Larsson, D. Västfjäll, and M. Kleiner. Better Presence and Performance in Virtual Environments By Improved Binaural Sound Rendering. *Journal of the Audio Engineering Society*, pages 1–8, 2002.

[126] K. Bormann. Presence and the Utility of Audio Spatialization. *Presence: Teleoperators and Virtual Environments*, 14(3):278–297, 2005. ISSN 1054-7460. doi: 10.1162/105474605323384645. URL `http://www.mitpressjournals.org/doi/10.1162/105474605323384645`.

[127] M. Gröhn, T. Lokki, and T. Takala. Localizing sound sources in a CAVE-like virtual environment with loudspeaker array reproduction. *Presence: Teleoperators and Virtual Environments*, 16(2):157–171, 2007. ISSN 10547460. doi: 10.1162/pres.16.2.157. URL `https://www.scopus.com/inward/record.uri?eid=2-s2.0-34247095953{&}partnerID=40{&}md5=2bfe1bf8dcede0a835f1b3f3b29d8f5d`.

[128] B. Titchener, Edward. *A Textbook of Psychology*. Macmillan, New York, NY, USA, revised ed edition, 1928. ISBN 041505740X. doi: 10.1136/jnnp.30.1.89.

[129] W. James. The Principles of Psychology, 1890.

[130] F. Perls, R. Hefferline, and P. Goodman. *Gestalt Therapy: Excitement and Growth in the Human Personality*. Gestalt Journal Press, Highland, NY, 1951. ISBN 0-939266-24-5.

[131] F. Perls. *The Gestalt Approach & Eye Witness to Therapy*. Science and Behaviour Books, Willmette, IL, USA, 1973. ISBN 0-8314-0034-X.

[132] W. Bevan. Perception: evolution of a concept. *Psychological review*, 65(1):34–55, 1958. ISSN 0033-295X. doi: 10.1037/h0045496. URL `http://www.ncbi.nlm.nih.gov/pubmed/13505980`.

[133] C. H. Graham. Behavior and the Psychophysical Methods - an Analysis of Some Recent Experiments. *Psychological Review*, 59(1):62–70, 1952. ISSN 0033295X. doi: 10.1037/h0054020.

[134] R. W. Gardner. Cognitive Styles in Categorizing Behavior. *Journal of Personality*, 22(2):214–233, 1953. ISSN 14676494. doi: 10.1111/j.1467-6494.1953.tb01807.x.

[135] G. Klein and H. J. Schlesinger. Perceptual Attitudes Toward Instability: I. Prediction of Apparent Movement Experiences From Rorschach Responses. *Journal of Personality*, 19(3):289–302, 1951. ISSN 14676494. doi: 10.1111/j.1467-6494.1951.tb01103.x.

[136] D. N. Jackson. Independence and Resistance to Perceptual Field Forces. *Journal of abnormal and social psychology*, 56(2):279–281, 1957. ISSN 0096851X. doi: 10.1037/h0042998.

[137] F. Pettigrew, T. The Measureman and Correlates of Category Width as a Cognitive Variable. *Journal of Personality*, 26(4):532–544, 1958.

[138] M. Kozhevnikov. Cognitive styles in the context of modern psychology: Toward an integrated framework of cognitive style. *Psychological Bulletin*, 133(3):464–481, 2007. ISSN 1939-1455. doi: 10.1037/0033-2909.133.3.464. URL `http://doi.apa.org/getdoi.cfm?doi=10.1037/0033-2909.133.3.464`.

[139] D. M. Broverman. Cognitive style and intra-individual variation in abilities. *Journal of Personality*, 28(2):240–255, 1960. ISSN 14676494. doi: 10.1111/j.1467-6494.1960.tb01616.x.

[140] D. Navon. Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, 9:353–383, 1977.

[141] S. Messick and F. J. Fritzky. Dimensions of analytic attitude in cognition and personality. *Journal of Personality*, 31(3):346–370, 1963.

[142] J. Kagan. REFLECTION-IMPULSIVITY : THE GENERALITY AND DYNAMICS OF CONCEPTUAL TEMPO 1. *Journal of Abnormal Psychology*, 71(1):17–24, 1966.

[143] S. H. Osipow. Cognitive styles and educational-vocational preferences and selection. *Journal of Counseling Psychology*, 16(6):534–546, 1969. ISSN 0022-0167. doi: 10.1037/h0028490. URL https://login.proxy.library.emory.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true{&}db=pdh{&}AN=1970-04204-001{&}site=ehost-live.

[144] H. K. Loomis and S. Moskowitz. Cognitive style and stimulus ambiguity. *Journal of Personality*, 26(3):349–364, 1958. ISSN 14676494. doi: 10.1111/j.1467-6494.1958.tb01591.x.

[145] D. L. McLain. Evidence of the Properties of an Ambiguity Tolerance Measure: The Multiple Stimulus Types Ambiguity Tolerance Scale-II (MSTAT-II) 1. *Psychological Reports*, 105(3):975–988, 2009. ISSN 0033-2941. doi: 10.2466/PR0.105.3.975-988. URL http://prx.sagepub.com/lookup/doi/10.2466/PR0.105.3.975-988.

[146] C. Lomberg, T. Kollmann, and C. Stöckmann. Different Styles for Different Needs - The Effect of Cognitive Styles on Idea Generation. *Creativity and Innovation Management*, 26(1):49–59, 2017. ISSN 14678691. doi: 10.1111/caim.12188.

[147] A. K. Jain and H. Jeppe Jeppesen. Knowledge management practices in a public sector organisation: the role of leaders' cognitive styles. *Journal of Knowledge Management*, 17(3):347–362, 2013. ISSN 1367-3270. doi: 10.1108/JKM-11-2012-0358. URL http://www.emeraldinsight.com/doi/10.1108/JKM-11-2012-0358.

[148] S. Hilbert, M. Bühner, N. Sarubin, K. Koschutnig, E. Weiss, I. Papousek, G. Reishofer, M. Magg, and A. Fink. The influence of cognitive styles and strategies in the digit span backwards task: Effects on performance and neuronal activity. *Personality and Individual Differences*, 87:242–247, 2015. ISSN 01918869. doi: 10.1016/j.paid.2015.08.012. URL http://dx.doi.org/10.1016/j.paid.2015.08.012.

[149] L. Lugli, M. Ragni, L. Piccardi, and R. Nori. Hypermedia navigation: Differences between spatial cognitive styles. *Computers in Human Behavior*, 66:191–200, 2017. ISSN 07475632. doi: 10.1016/j.chb.2016.09.038. URL http://dx.doi.org/10.1016/j.chb.2016.09.038.

[150] P. Honey and A. Mumford. *Manual of Learning Styles*. P. Honey, London, UK, 1982.

[151] A. Furnham. Personality and Learning Style - a Study of 3 Instruments. *Personality and Individual Differences*, 13(4):429–438, 1992. ISSN 0191-8869. doi: 10.1016/0191-8869(92)90071-V.

[152] R. R. McCrae and P. T. Costa. Validation of the Five-Factor Model of Personality Across Instruments and Observers. *Journal of Personality and Social Psychology*, 52(1):81–90, 1987.

[153] E. R. Peterson, I. J. Deary, and E. J. Austin. Are intelligence and personality related to verbal-imagery and wholistic-analytic cognitive styles? *Personality and Individual Differences*, 39(1):201–213, 2005. ISSN 01918869. doi: 10.1016/j.paid.2005.01.009.

[154] C. M. MacLeod, R. A. Jackson, and J. Palmer. On the relation between spatial ability and field dependence. *Intelligence*, 10(2):141–151, 1986. ISSN 01602896. doi: 10.1016/0160-2896(86)90011-5.

[155] F. P. McKenna. Measures of field dependence: Cognitive style or cognitive ability? *Journal of Personality and Social Psychology*, 47(3):593–603, 1984. ISSN 0022-3514. doi: 10.1037/0022-3514.47.3.593. URL `http://content.apa.org/journals/psp/47/3/593`.

[156] M. A. Guisande, M. F. Páramo, C. Tinajero, and L. S. Almeida. Field dependence-independence (FDI) cognitive style: An analysis of attentional functioning. *Psicothema*, 19(4):572–577, 2007. ISSN 02149915.

[157] E. R. Peterson and K. Meissel. The effect of Cognitive Style Analysis (CSA) test on achievement: A meta-analytic review. *Learning and Individual Differences*, 38:115–122, 2015. ISSN 18733425. doi: 10.1016/j.lindif.2015.01.011. URL `http://dx.doi.org/10.1016/j.lindif.2015.01.011`.

[158] R. L. C. Mitchell and V. Kumari. Hans Eysenck's interface between the brain and personality: Modern evidence on the cognitive neuroscience of personality. *Personality and Individual Differences*, 103:74–81, 2016. ISSN 01918869. doi: 10.1016/j.paid.2016.04.009. URL `http://dx.doi.org/10.1016/j.paid.2016.04.009`.

[159] M. Kennis, A. R. Rademaker, and E. Geuze. Neural correlates of personality: An integrative review. *Neuroscience and Biobehavioral Reviews*, 37(1):73–95, 2013. ISSN 01497634. doi: 10.1016/j.neubiorev.2012.10.012. URL `http://dx.doi.org/10.1016/j.neubiorev.2012.10.012`.

[160] C. G. Jung. *Psychological Types*. Princeton University Press, Princeton, New Jersey, USA, 1971.

[161] A. Tymula, L. A. Rosenberg Belmaker, A. K. Roy, L. Ruderman, K. Manson, P. W. Glimcher, and I. Levy. Adolescents' risk-taking behavior is driven by tolerance to ambiguity. *Proceedings of the National Academy of Sciences of the United States of America*, 109(42):17135–40, 2012. ISSN 1091-6490. doi: 10.1073/pnas.1207144109. URL /pmc/articles/PMC3479478/?report=abstract.

[162] A. L. Krain, S. Hefton, D. S. Pine, M. Ernst, F. Xavier Castellanos, R. G. Klein, and M. P. Milham. An fMRI examination of developmental differences in the neural correlates of uncertainty and decision-making. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 47(10):1023–1030, 2006. ISSN 00219630. doi: 10.1111/j.1469-7610.2006.01677.x.

[163] A. Simmons, S. C. Matthews, M. P. Paulus, and M. B. Stein. Intolerance of uncertainty correlates with insula activation during affective ambiguity. *Neuroscience Letters*, 430(2):92–97, 2008. ISSN 03043940. doi: 10.1016/j.neulet.2007.10.030.

[164] D. J. Kraemer, L. M. Rosenberg, and S. L. Thompson-Schill. The neural correlates of visual and verbal cognitive styles. *J Neurosci*, 29(12):3792–3798, 2009. ISSN 0270-6474. doi: 29/12/3792[pii]\r10.1523/JNEUROSCI.4635-08.2009. URL http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve{&}db=PubMed{&}dopt=Citation{&}list{_}uids=19321775.

[165] S. Gardini, C. R. Cloninger, and A. Venneri. Individual differences in personality traits reflect structural variance in specific brain regions. *Brain Research Bulletin*, 79(5):265–270, 2009. ISSN 03619230. doi: 10.1016/j.brainresbull.2009.03.005.

[166] M.-C. Lai, M. V. Lombardo, B. Chakrabarti, C. Ecker, S. A. Sadek, S. J. Wheelwright, D. G. M. Murphy, J. Suckling, E. T. Bullmore, and S. Baron-Cohen. Individual differences in brain structure underpin empathizing-systemizing cognitive styles in male adults. *NeuroImage*, 61(4):1347–1354, 2012. ISSN 10538119. doi: 10.1016/j.neuroimage.2012.03.018. URL http://dx.doi.org/10.1016/j.neuroimage.2012.03.018.

[167] N. Ford. Cognitive styles and virtual environments. *Journal of the American Society for Information Science*, 51(6):543–557, 2000. ISSN 1097-4571. doi: 10.1002/(SICI)1097-4571(2000)51:6<543::AID-ASI6>3.0.CO;2-S. URL http://onlinelibrary.wiley.com/doi/10.1002/(SICI)1097-4571(2000)51:6{%}3C543::AID-ASI6{%}3E3.0.CO;2-S/abstract{%}5Cnhttp://onlinelibrary.wiley.com.ezproxy.liv.ac.uk/doi/10.1002/(SICI)1097-4571(2000)51:6{%}3C543::AID-ASI6{%}3E3.0.CO;2-S/abstract;jsessionid=C970C31F9.

[168] G. Pask. Learning strategies, teaching strategies, and conceptual or learning style. In R. R. Schmeck, editor, *Learning strategies and individual competence*, volume Ch4, pages 217–253. New York, New York, USA, 1988.

[169] S. Gulliver. Cognitive style and personality: Impact on multimedia perception. *Online Information Review*, 34(1):39–58, 2010. ISSN 14684527. doi: 10.1108/14684521011024119.

[170] T. Ogle, J. Burton, and G. Worley. *The Effects of Virtual Environments on Recall in Participants of Differing Levels of Field Dependence*. PhD thesis, Virginia Polytechnic and State University, 2002.

[171] T. R. Cutmore, T. J. Hine, K. J. Maberly, N. M. Langford, and G. Hawgood. Cognitive and gender factors influencing navigation in a virtual environment. *International Journal of Human-Computer Studies*, 53(2):223–249, 2000. ISSN 10715819. doi: 10.1006/ijhc.2000.0389. URL `http://linkinghub.elsevier.com/retrieve/pii/S1071581900903896`.

[172] R. Samana, H. S. Wallach, and M. P. Safir. The impact of personality traits on the experience of presence. *2009 Virtual Rehabilitation International Conference, VR 2009*, pages 1–7, 2009. doi: 10.1109/ICVR.2009.5174197.

[173] D. R. Begault, E. M. Wenzel, and M. R. Anderson. Direct Comparison of the Impact of Head Tracking, Reverberation and Individualized Head-Related-Transfer Functions on the Spatial Perception of a Virtual Speech Source. *Journal of the Audio Engineering Society*, 49(10):904 – 916, 2001.

[174] P. Zahorik. Perceptually relevant parameters for virtual listening simulation of small room acoustics. *The Journal of the Acoustical Society of America*, 126(2):776–791, 2009. ISSN 1520-8524. doi: 10.1121/1.3167842. URL `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2730711{&}tool=pmcentrez{&}rendertype=abstract`.

[175] S. E. Kober, J. Kurzmann, and C. Neuper. Cortical correlate of spatial presence in 2D and 3D interactive virtual reality: An EEG study. *International Journal of Psychophysiology*, 83(3):365–374, 2012. ISSN 01678760. doi: 10.1016/j.ijpsycho.2011.12.003. URL `http://dx.doi.org/10.1016/j.ijpsycho.2011.12.003`.

[176] B. M. Fazenda and J. Newton. The psychoacoustic effects of stimuli plausibility on headphone externalisation. *Proceedings of the Institute of Acoustics*, 36(2), oct 2014. URL `http://usir.salford.ac.uk/39007/`.

[177] S. E. Kober and C. Neuper. Personality and Presence in Virtual Reality: Does Their Relationship Depend on the Used Presence Measure? *International Journal of Human-Computer Interaction*, 29(1):13–25, 2013. ISSN 10447318. doi: 10.1080/10447318.2012.668131.

[178] C. A. Thornson, B. F. Goldiez, and H. Le. Predicting presence: Constructing the Tendency toward Presence Inventory. *International Journal of Human Computer Studies*, 67(1):62–78, 2009. ISSN 10715819. doi: 10.1016/j.ijhcs.2008.08.006.

[179] S. Walkowiak, T. Foulsham, and A. F. Eardley. Individual differences and personality correlates of navigational performance in the virtual route learning task. *Computers in Human Behavior*, 45:402–410, 2015. ISSN 07475632. doi: 10.1016/j.chb.2014.12.041. URL `http://dx.doi.org/10.1016/j.chb.2014.12.041`.

[180] Y. Ling, H. T. Nefs, W. P. Brinkman, C. Qu, and I. Heynderickx. The relationship between individual characteristics and experienced presence. *Computers in Human Behavior*, 29(4):1519–1530, 2013. ISSN 07475632. doi: 10.1016/j.chb.2012.12.010. URL `http://dx.doi.org/10.1016/j.chb.2012.12.010`.

[181] C. Johns, D. Nu, M. Daya, D. Sellars, J. Casanueva, and E. Blake. The Interaction Between Individuals ' Immersive Tendencies and the Sensation of Presence in a Virtual Environment. *Virtual Environments Proceedings of the Eurographics Workshop*, pages 65–74, 2000.

[182] J. Hou, Y. Nam, W. Peng, and K. M. Lee. Effects of screen size, viewing angle, and players' immersion tendencies on game experience. *Computers in Human Behavior*, 28(2):617–623, 2012. ISSN 07475632. doi: 10.1016/j.chb.2011.11.007. URL `http://dx.doi.org/10.1016/j.chb.2011.11.007`.

[183] L. Phillips, V. Interrante, M. Kaeding, B. Ries, and L. Anderson. Correlations Between Physiological Response, Gait, Personality, and Presence in Immersive Virtual Environments. *Presence: Teleoperators and Virtual Environments*, 21(2):119–141, 2012. ISSN 1054-7460. doi: 10.1162/PRES_a_00100. URL `http://www.mitpressjournals.org/doi/10.1162/PRES{_}a{_}00100`.

[184] A. S. Byrk and S. W. Raudenbush. *Hieracrchical Linear Models: Applications and Data Analysis Methods*. Sage Publications, London, UK, 1992.

[185] H. Woltman, A. Feldstain, C. MacKay, and M. Rocchi. An introduction to hierarchical linear modeling. *Tutorials in Quantitative Methods for Psychology*, 8(1):52–69, 2012. ISSN 0003-1224. doi: 10.2307/2095731.

[186] H. Akaike. A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, 1974. ISSN 15582523. doi: 10.1109/TAC.1974.1100705.

[187] J. Fox and S. Weisberg. Mixed-Effects Models in R. *An R Companion to Applied Regression*, pages 1–54, 2011.

[188] T. Hothorn, F. Bretz, P. Westfall, R. M. Heiberger, and A. Schuetzenmeister. multcomp: Simultaneous Inference in General Parametric Models. *Biometrical Journal*, 50(3):346–363, 2013. ISSN 1521-4036. doi: https://cran.r-project.org/web/packages/multcomp/multcomp.pdf.

[189] A. Field, J. Miles, and Z. Field. *Discovering statistics using R*. SAGE Publications, 2012. ISBN 9781446258460. URL `https://books.google.co.uk/books?id=wd2K2zC3swIC`.

[190] S. Siegel. *Non-Parametric Statistics for the Behavioural Sciences*. McGraw Hill, New York, New York, USA, 1956.

[191] H. Cramér. *Mathematical Methods of Statistics*. Asia Publishing House, Princeton, New Jersey, USA, 1946.

[192] F. J. Gravetter and L. B. Wallnau. *Statistics for the Behavioral Sciences*. Cengage Learning, New York, NY, USA, 9th edition, 2010. ISBN 9781111830991.

[193] J. Cohen. *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum Associates, New York, NY, USA, 2nd edition, 1988. ISBN 0805802835.

[194] R. Bakeman. Recommended effect size statistics for repeated measures designs. *Behavior Research Methods*, 37(3):379–384, 2005. ISSN 1554351X. doi: 10.3758/BF03192707.

[195] L. C. Freeman. A still further note on freeman's measure of association. *Psychometrika*, 41(2):273–275, 1976. ISSN 00333123. doi: 10.1007/BF02294111.

[196] I. T. Jolliffe. Principal Component Analysis, Second Edition. *Encyclopedia of Statistics in Behavioral Science*, 30(3):487, 2002. ISSN 00401706. doi: 10.2307/1270093. URL `http://onlinelibrary.wiley.com/doi/10.1002/0470013192.bsa501/full`.

[197] T. F. Cox and M. A. A. Cox. *Multidimensional Scaling, Second Edition*. Chapman and Hall/CRC, 2 edition, 2000. ISBN 1584880945. URL `http://www.amazon.com/Multidimensional-Scaling-Second-Trevor-Cox/dp/1584880945`.

[198] N. Zacharov and K. Koivuniemi. Unravelling the Perception of Spatial Sound Reproduction: Analysis & External Preference Mapping. *Proceedings of the 111th Convention of the Audio Engineering Society (2001)*, 2001.

[199] R. Conetta, T. Brookes, F. Rumsey, S. Zielinski, M. Dewhirst, P. Jackson, S. Bech, D. Meares, and S. George. Spatial Audio Quality Perception (Part 2): A Linear Regression Model. *Journal of the Audio Engineering Society*, 62(12):847–860, 2015. ISSN 15494950. URL `http://www.aes.org/e-lib/browse.cfm?elib=17558{%}5Cnpapers3://publication/uuid/4A0E0CF8-5D42-4EAA-8A29-9087CFB68C56`.

[200] J. Cohen and P. Cohen. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. Lawrence Erlbaum Associates, Hillsdale, New Jersey, USA, second edi edition, 1983.

[201] E. Sussman, M. Steinschneider, W. Lee, and K. Lawson. Auditory scene analysis in school-aged children with developmental language disorders. *International journal of psychophysiology : official journal of the International Organization of Psychophysiology*, 95(2):113–124, 2015. ISSN 1872-7697. doi: 10.1016/j.ijpsycho.2014.02.002. URL `http://www.sciencedirect.com/science/article/pii/S0167876014000555`.

[202] B. M. Fazenda, P. Kendrick, T. J. Cox, F. Li, and I. Jackson. Perception and automated assessment of audio quality in user generated content: An improved model. *2016 8th International Conference on Quality of Multimedia Experience, QoMEX 2016*, 2016. doi: 10.1109/QoMEX.2016.7498974.

[203] H. Stanislaw and N. Todorov. Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, 31(1):137–149, 1999. ISSN 0743-3808. doi: 10.3758/BF03207704. URL `http://www.springerlink.com/index/10.3758/BF03207704`.

[204] J. K. Witt, J. E. T. Taylor, M. Sugovic, and J. T. Wixted. Signal detection measures cannot distinguish perceptual biases from response biases. *Perception*, 44(3):289–300, 2015. ISSN 03010066. doi: 10.1068/p7908.

[205] B. Gardner and K. Martin. HRTF Measurements of a KEMAR Dummy-Head Microphone. Technical report, MIT Media Lab Perceptual Computing, 1994.

[206] Unity Technologies. Unity 3D, 2016. URL `madewith.unity.com`.

[207] H. Kuttruff. *Room acoustics*. Spon Press, 2009. ISBN 9780415480215. doi: 10.1002/1521-3773(20010316)40:6<9823::AID-ANIE9823>3.3.CO;2-C.

[208] A. Lindau, L. Kasanke, and S. Weinzierl. Perceptual evaluation of physical predictors of the mixing time in binaural room impulse responses. *128th Audio Engineering Society Convention*, 2010.

[209] G.-B. Stan, J.-J. Embrechts, and D. Archambeau. Comparison of different impulse response measurement techniques. *Journal of the Audio Engineering Society*, 50(4): 249–262, 2002. ISSN 0004-7554.

[210] R. V. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano. The CIPIC HRTF Database. In *IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*, 2001. URL `http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.19.363`.

[211] M. Gardner. Proximity Image Effect in Sound Localization. *Journal of the Acoustical Society of America*, 43(1968):163, 1968. ISSN 0001-4966. doi: 10.1121/1.1910747.

[212] P. Zahorik. Auditory Distance Perception in Humans : A Summary of Past and Present Research. *Acta Acustica united with Acustica*, 91(February 2003): 409–420, 2005. URL `http://www.ingentaconnect.com/content/dav/aaua/2005/00000091/00000003/art00003`.

[213] Blender Online Community. Blender - a 3D modelling and rendering package, 2018. URL `www.blender.org`.

[214] R. J. Hughes, T. J. Cox, B. Shirley, and P. Power. The room-in-room effect and its influence on perceived room size in spatial audio reproduction. In *141st AES Convention*, Los Angeles, USA, sep 2016.

[215] Google Inc. Spatial Audio API, 2017. URL `https://developers.google.com/vr/ios/ndk/reference/group/audio`.

[216] A. Avni, J. Ahrens, M. Geier, S. Spors, H. Wierstorf, and B. Rafaely. Spatial perception of sound fields recorded by spherical microphone arrays with varying spatial resolution. *The Journal of the Acoustical Society of America*, 133(5):2711–2721, 2013. ISSN 0001-4966. doi: 10.1121/1.4795780. URL `http://asa.scitation.org/doi/10.1121/1.4795780`.

[217] R. J. Hughes, T. J. Cox, B. G. Shirley, and P. Power. Data and supporting information for AES 141 convention paper "The room-in-room effect and its influence on perceived room size in spatial audio reproduction", sep 2016.

[218] S. Werner, F. Klein, T. Mayenfels, and K. Brandenburg. A summary on acoustic room divergence and its effect on externalization of auditory events. *2016 8th International Conference on Quality of Multimedia Experience, QoMEX 2016*, 2016. ISSN 14349957. doi: 10.1109/QoMEX.2016.7498973.

[219] F. Klein, S. Werner, and T. Mayenfels. Influences of training on externalization of binaural synthesis in situations of room divergence. *AES: Journal of the Audio Engineering Society*, 65(3):178–187, 2017. ISSN 15494950. doi: 10.17743/jaes.2016.0072.

[220] M. R. Schroeder. New Method of Measuring Reverberation Time. *The Journal of the Acoustical Society of America*, 37:409–412, 1965. doi: 10.1121/1.1909343.

[221] W. Bailey and B. M. Fazenda. The effect of reverberation and audio spatialization on egocentric distance estimation of objects in stereoscopic virtual reality. *The Journal of the Acoustical Society of America*, 141(5):3510, 2017. doi: 10.1121/1.4987362. URL `https://doi.org/10.1121/1.4987362`.

[222] A. Wakabayashi, S. Baron-Cohen, S. Wheelwright, N. Goldenfeld, J. Delaney, D. Fine, R. Smith, and L. Weil. Development of short forms of the Empathy Quotient (EQ-Short) and the Systemizing Quotient (SQ-Short). *Personality and Individual Differences*, 41(5):929–940, 2006. ISSN 01918869. doi: 10.1016/j.paid.2006.03.017.

[223] S. Baron-Cohen. The empathising-systemising theory of autism: Implications for education. *Tizard Learning Disability Review*, 14(3):4–13, 2009. ISSN 13595474. doi: 10.1108/13595474200900022.

[224] D. I. Falkenberg, F. Schneider, B. Derntl, T. Kellermann, S. Eickhoff, U. Habel, and A. Finkelmeyer. Multidimensional assessment of empathic abilities: Neural correlates and gender differences. *Psychoneuroendocrinology*, 35(1):67–82, 2009. ISSN 03064530. doi: 10.1016/j.psyneuen.2009.10.006.

[225] S. D. Preston and F. B. M. D. Waal. Empathy: Its ultimate and proximate bases. *Behavioral and Brain Sciences*, (25):1–71, 2002.

[226] S. D. Preston. A perception-action model for empathy. *Empathy in Mental Illness*, pages 428–447, 2007. doi: 10.1017/CBO9780511543753.024.

[227] L. R. Goldberg. The development of markers for the Big-Five factor structure. *Psychological Assessment*, 4(1):26–42, 1992.

[228] B. G. Witmer, C. J. Jerome, and M. J. Singer. The Factor Structure of the Presence Questionnaire. *Presence: Teleoperators and Virtual Environments*, 14

(3):298–312, 2005. ISSN 1054-7460. doi: 10.1162/105474605323384654. URL http://link.springer.com/10.1007/978-3-642-02806-9{_}12{%}5Cnhttp://www.mitpressjournals.org/doi/abs/10.1162/105474605323384654.

[229] J. Digman. Personality Structure: Emergence Of The 5-Factor Model. *Annual Review of Psychology*, 41(1):417–440, 1990. ISSN 00664308. doi: 10.1146/annurev.psych.41.1.417. URL http://psych.annualreviews.org/cgi/doi/10.1146/annurev.psych.41.1.417.

[230] R. G. Smart. Subject selection bias in psychological research. *Canadian Psychologist/Psychologie canadienne*, 7a(2):115–121, 1966. ISSN 0008-4832(Print). doi: 10.1037/h0083096.

[231] P. Jylhä, O. Mantere, T. Melartin, K. Suominen, and M. Vuorilehto. Differences in neuroticism and extraversion between patients with bipolar I or II and general population subjects or major depressive disorder patients. *Journal of Affective Disorders*, 125(1-3):42–52, 2010. ISSN 0165-0327. doi: 10.1016/j.jad.2010.01.068. URL http://dx.doi.org/10.1016/j.jad.2010.01.068.

[232] P. Kline and S. L. Lapham. Personality and faculty in British universities. *Personality ad Individual Differences*, 13(7):855–857, 1992.

[233] J. B. Peterson and S. Carson. Latent Inhibition and Openness to Experience in a high- achieving student population. *Personality and Individual Differences*, 28: 323–332, 2000.

[234] D. R. Cox. Interaction. *International Statistical Review*, 52(1):1–24, 1984.

[235] D. R. Brown. Stimulus-Similarity and the Anchoring of Subjective Scales. *The American Journal of Psychology*, 66(2):199–214, 1953.

[236] A. S. Sudarsono, Y. W. Lam, and W. J. Davies. The effect of sound level on perception of reproduced soundscapes. *Applied Acoustics*, 110:53–60, 2016. ISSN 0003-682X. doi: https://doi.org/10.1016/j.apacoust.2016.03.011. URL http://www.sciencedirect.com/science/article/pii/S0003682X16300482.

[237] M. Schoeffler, J. L. Gernert, M. Neumayer, S. Westphal, and J. Herre. 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, 2015. ISSN 14349957. doi: 10.1007/s10055-015-0270-8.

[238] Y. Ling, H. T. Nefs, N. Morina, I. Heynderickx, and W. P. Brinkman. A meta-analysis on the relationship between self-reported presence and anxiety in virtual

reality exposure therapy for anxiety disorders. *PLoS ONE*, 9(5):1–12, 2014. ISSN 19326203. doi: 10.1371/journal.pone.0096144.

[239] D. Baker, A. S. David, P. Shaw, E. J. Lawrence, and S. Baron-Cohen. Measuring empathy: reliability and validity of the Empathy Quotient. *Psychological Medicine*, 34(5):911–919, 2004. ISSN 0033-2917. doi: 10.1017/s0033291703001624.

[240] R. Saxe and L. J. Powell. It's the Thought That Counts. *Psychological Science*, 17 (8):692–699, 2006. ISSN 0956-7976. doi: 10.1111/j.1467-9280.2006.01768.x. URL `http://csufresno-dspace.calstate.edu/handle/10211.3/118892{%}0Ahttp://journals.sagepub.com/doi/10.1111/j.1467-9280.2006.01768.x`.

[241] W. Prinz. Perception and Action Planning. *European Journal of Cognitive Psychology*, 9(2):129–154, 1997. ISSN 09541446. doi: 10.1080/713752551.

[242] L. Plouffe. Introversion-Extraversion: The Bell-Magendie Law Revisited. *Personality and Individual Differences*, 4(4):421–427, 1983.

[243] G. Matthews. Traits, cognitive processes and adaptation: An elegy for Hans Eysenck's personality theory. *Personality and Individual Differences*, 103:61–67, 2016. ISSN 01918869. doi: 10.1016/j.paid.2016.04.037. URL `http://dx.doi.org/10.1016/j.paid.2016.04.037`.

[244] B. C. Dickerson, E. Feczko, C. E. Schwartz, L. F. Barrett, M. M. Wedig, D. Williams, and C. I. Wright. Neuroanatomical Correlates of Extraversion and Neuroticism. *Cerebral Cortex*, 16(12):1809–1819, 2006. ISSN 1047-3211. doi: 10.1093/cercor/bhj118.

[245] M. Bischoff, B. Walter, C. R. Blecker, K. Morgen, D. Vaitl, and G. Sammer. Utilizing the ventriloquism-effect to investigate audio-visual binding. *Neuropsychologia*, 45(3): 578–586, 2007. ISSN 00283932. doi: 10.1016/j.neuropsychologia.2006.03.008.

[246] S. Azevedo, P. Campos, and J. A. Jorge. Combining EEG Data with Place and Plausibility Responses as an Approach to Measuring Presence in Outdoor Virtual Environments. *Presence: Teleoperators & Virtual Environments*, 23(4):354–368, 2014. ISSN 15313263. doi: 10.1162/PRES. URL `http://www.mitpressjournals.org/doi/pdf/10.1162/PRES{_}a{_}00135`.

[247] I. Bergstrom, S. Azevedo, P. Papiotis, N. Saldanha, and M. Slater. The Plausibility of a String Quartet Performance in Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics*, 23(4):1332–1339, 2017. ISSN 10772626. doi: 10.1109/TVCG.2017.2657138.

[248] J. G. Bolaños and V. Pulkki. Immersive audiovisual environment with 3D audio playback. *132nd Audio Engineering Society Convention 2012*, pages 404–412, 2012.

[249] K. Debattista, T. Bashford-Rogers, C. Harvey, B. Waterfield, and A. Chalmers. Subjective Evaluation of High-Fidelity Virtual Environments for Driving Simulations. *IEEE Transactions on Human-Machine Systems*, 48(1):30–40, 2017. ISSN 21682291. doi: 10.1109/THMS.2017.2762632.

[250] D. J. Finnegan, E. O'Neill, and M. J. Proulx. An approach to reducing distance compression in audiovisual virtual environments, 2017.

[251] D. J. Finnegan, b. O. Eamonn, and M. J. Proulx. Compensating for Distance Compression in Audiovisual Virtual Environments Using Incongruence. *Conference of Human-Computer Interaction 2016*, pages 200–212, 2016. doi: 10.1145/2858036.2858065.

[252] C. Greenhalgh and S. Benford. MASSIVE: a collaborative virtual environment for teleconferencing. *ACM Transactions on Computer-Human Interaction*, 2(3): 239–261, 1995. ISSN 10730516. doi: 10.1145/210079.210088. URL `http://portal.acm.org/citation.cfm?doid=210079.210088`.

[253] E. R. Hoeg, L. J. Gerry, L. Thomsen, N. C. Nilsson, and S. Serafin. Binaural sound reduces reaction time in a virtual reality search task. *2017 IEEE 3rd VR Workshop on Sonic Interactions for Virtual Environments, SIVE 2017*, 37, 2017. doi: 10.1109/SIVE.2017.7901610.

[254] T. Iachini, L. Maffei, F. Ruotolo, V. P. Senese, G. Ruggiero, M. Masullo, and N. Alekseeva. Multisensory assessment of acoustic comfort aboard metros: A virtual reality study. *Applied Cognitive Psychology*, 26(5):757–767, 2012. ISSN 08884080. doi: 10.1002/acp.2856.

[255] G. Kearney, X. Liu, A. Manns, and M. Gorzel. Auditory Distance Perception with Static and Dynamic Binaural Rendering. *Proceedings of the 57th AES International Conference, Hollywood, CA, USA*, pages 1–8, 2015.

[256] D. M. Krum, E. A. Suma, and M. Bolas. Spatial misregistration of virtual human audio: Implications of the precedence effect. *Intelligent Virtual Agents (Lecture Notes in Computer Science)*, 7502:139–145, 2012. ISSN 03029743. doi: 10.1109/3DUI.2012.6184204. URL `http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6184204`.

[257] H.-J. Maempel and M. Jentsch. Audio-visual interaction of size and distance perception in concert halls - a preliminary study. *International Symposium on Room Acoustics*, page P124, 2013.

[258] C. Pike, R. Taylor, T. Parnell, and F. Melchior. Object-based Spatial Audio Production for Virtual Reality using the Audio Definition Model. *Proceedings of the AES International Conference on Audio for Augmented and Virtual Reality*, pages 1–7, 2016.

[259] S. Poeschl, K. Wall, and N. Doering. Integration of spatial sound in immersive virtual environments an experimental study on effects of spatial sound on presence. *Virtual Reality (VR), 2013 IEEE*, pages 129–130, 2013. ISSN 1087-8270. doi: 10.1109/VR.2013.6549396.

[260] B. E. Riecke, A. Väljamäe, and J. Schulte-Pelkum. Moving sounds enhance the visually-induced self-motion illusion (circular vection) in virtual reality. *ACM Transactions on Applied Perception*, 6(2):1–27, 2009. ISSN 15443558. doi: 10.1145/1498700.1498701. URL `http://portal.acm.org/citation.cfm?doid=1498700.1498701`.

[261] O. Rummukainen, S. Schlecht, A. Plinge, and E. A. Habets. Evaluation of binaural reproduction systems from behavioral patterns in a six-degrees-of-freedom wayfinding task. *2017 9th International Conference on Quality of Multimedia Experience, QoMEX 2017*, pages 4–6, 2017. doi: 10.1109/QoMEX.2017.7965680.

[262] F. Ruotolo, L. Maffei, M. Di Gabriele, T. Iachini, M. Masullo, G. Ruggiero, and V. P. Senese. Immersive virtual reality and environmental noise assessment: An innovative audio-visual approach. *Environmental Impact Assessment Review*, 41:10–20, 2013. ISSN 01959255. doi: 10.1016/j.eiar.2013.01.007. URL `http://dx.doi.org/10.1016/j.eiar.2013.01.007`.

[263] D. Thery, D. Poirier-Quinot, B. N. J. Postma, and B. F. G. Katz. Impact of the Visual Rendering System on Subjective Auralization Assessment in VR. In J. Barbic, M. D'Cruz, M. E. Latoschik, M. Slater, and P. Bourdot, editors, *Virtual Reality and Augmented Reality*, pages 105–118, Cham, 2017. Springer International Publishing. ISBN 978-3-319-72323-5.

[264] M. Vinnikov, R. S. Allison, and S. Fernandes. Gaze-Contingent Auditory Displays for Improved Spatial Attention in Virtual Reality. *ACM Transactions on Computer-Human Interaction*, 24(3):1–38, 2017. ISSN 10730516. doi: 10.1145/3067822. URL `http://dl.acm.org/citation.cfm?doid=3086563.3067822`.

[265] N. Averbukh. Subjective-Situational Study of Presence. In R. Shumaker and S. Lackey, editors, *Virtual, Augmented and Mixed Reality. Designing and Developing Virtual and Augmented Environments. VAMR 2014. Lecture Notes in Computer Science*, volume 8525. Springer, Cham, 2014.

[266] A. C. Beall, J. M. Loomis, J. W. Philbeck, and T. G. Fikes. Absolute Motion Parallax Weakly Determines Visual Scale in Real and Virtual Environments. *Human Vision, Visual Processing, and Digital Display VI*, 2411:288–297, 1995. ISSN 0277786X. doi: 10.1117/12.207547.

[267] M. Bellani, L. Fornasari, L. Chittaro, and P. Brambilla. Virtual reality in autism: State of the art. *Epidemiology and Psychiatric Sciences*, 20(3):235–238, 2011. ISSN 20457960. doi: 10.1017/S2045796011000448.

[268] G. P. Bingham, A. Bradley, M. Bailey, and R. Vinner. Accomodation, Occlusion, and Disparity Matching Are Used to Guide Reaching: A Comparison of Actual Versus Virtual Environments. *Journal of Experimental Psychology: Human Perception and Performance*, 27(6):1314–1334, 2001.

[269] G. Bruder, F. Steinicke, P. Wieland, and M. Lappe. Tuning self-motion perception in virtual reality with visual illusions. *IEEE Transactions on Visualization and Computer Graphics*, 18(7):1068–1078, 2012. ISSN 10772626. doi: 10.1109/TVCG.2011.274.

[270] E. Combe, J. Posselt, and A. Kemeny. Virtual Prototype Visualization: A Size Perception Study. In *Proceedings of the 5th Nordic Conference on Human-computer Interaction: Building Bridges*, NordiCHI '08, pages 581–582, New York, NY, USA, 2008. ACM. ISBN 978-1-59593-704-9. doi: 10.1145/1463160.1463253. URL `http://doi.acm.org/10.1145/1463160.1463253`.

[271] T. J. Dodds, B. J. Mohler, and H. H. Bülthoff. Talk to the virtual hands: Self-animated avatars improve communication in head-mounted display virtual environments. *PLoS ONE*, 6(10), 2011. ISSN 19326203. doi: 10.1371/journal.pone.0025759.

[272] J. Fröhlich and I. Wachsmuth. Combining multi-sensory stimuli in virtual worlds - A progress report. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8525 LNCS (PART 1):44–54, 2014. ISSN 16113349. doi: 10.1007/978-3-319-07458-0_5.

[273] A. Gaggioli and R. Breining. Perception and cognition in immersive Virtual Reality. *Emerging Communication: Studies on New Technologies and Practices in*, (FEBRUARY 2004):71–86, 2001.

[274] T. Y. Grechkin, T. D. Nguyen, J. M. Plumert, J. F. Cremer, and J. K. Kearney. How does presentation method and measurement protocol affect distance estimation in real and virtual environments? *ACM Transactions on Applied Perception*, 7(4): 1–18, 2010. ISSN 15443558. doi: 10.1145/1823738.1823744.

[275] J. A. Jones, J. E. Swan, G. Singh, S. Reddy, K. Moser, C. Hua, and S. R. Ellis. Improvements in visually directed walking in virtual environments cannot be explained by changes in gait alone. In *Proceedings of ACM Symposium on Applied Perception (SAP) 2012*, pages 11–16, 2012. doi: 10.1145/2338676.2338679.

[276] F. Kellner, B. Bolte, G. Bruder, U. Rautenberg, F. Steinicke, M. Lappe, and R. Koch. Geometric Calibration of Head-Mounted Displays and its Effects on Distance Estimation. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 18 (4):589–596, 2012.

[277] R. V. Kenyon, D. Sandin, R. C. Smith, R. Pawlicki, and T. Defanti. Size-constancy in the CAVE. *Presence: Teleoperators and Virtual Environments*, 16(2):172–187, 2007. ISSN 10547460. doi: 10.1162/pres.16.2.172.

[278] E. Klein, J. E. Swan, G. S. Schmidt, M. A. Livingston, and O. G. Staadt. Measurement protocols for Medium-Field distance perception in Large-Screen immersive displays. *Proceedings - IEEE Virtual Reality*, pages 107–113, 2009. ISSN 1087-8270. doi: 10.1109/VR.2009.4811007.

[279] S. A. Kuhl, W. B. Thompson, and S. H. Creem-Regehr. Minification Influences Spatial Judgements in Virtual Environments. In *APGV '06 Proceedings of the 3rd symposium on Applied perception in graphics and visualization*, pages 15–19, Boston, Massachusetts, USA, jul 2006. ACM.

[280] S. Kuhl, K. Hinrichs, F. Steinicke, P. Willemsen, M. Lappe, and G. Bruder. Judgment of natural perspective projections in head-mounted display environments. page 35, 2009. doi: 10.1145/1643928.1643940.

[281] B. R. Kunz, L. Wouters, D. Smith, W. B. Thompson, and S. H. Creem-Regehr. Revisiting the effect of quality of graphics on distance judgments in virtual environments: A comparison of verbal reports and blind walking. *Attention, Perception and Psychophysics*, 71(6):1284–1293, 2009. doi: 10.3758/APP.

[282] N. H. Lehment, D. Merget, and G. Rigoll. Creating automatically aligned consensus realities for AR videoconferencing. *ISMAR 2014 - IEEE International Symposium on Mixed and Augmented Reality - Science and Technology 2014, Proceedings*, pages 201–206, 2014. doi: 10.1109/ISMAR.2014.6948428.

[283] T. Lentz, D. Schröder, M. Vorländer, and I. Assenmach. Virtual reality system with integrated sound field simulation and reproduction. *Eurasip Journal on Advances in Signal Processing*, 2007. ISSN 11108657. doi: 10.1155/2007/70540.

[284] P. Lubos, G. Bruder, and F. Steinicke. Analysis of direct selection in head-mounted display environments. *IEEE Symposium on 3D User Interfaces 2014, 3DUI 2014 - Proceedings*, pages 11–18, 2014. doi: 10.1109/3DUI.2014.6798834.

[285] X. Luo. From augmented reality to augmented computing: A look at cloud-mobile convergence. *Proceedings - 2009 International Symposium on Ubiquitous Virtual Reality, ISUVR 2009*, (November 2007):29–32, 2009. doi: 10.1109/ISUVR.2009.13.

[286] M. R. Marner, R. T. Smith, J. A. Walsh, and B. H. Thomas. Spatial user interfaces for large-scale projector-based augmented reality. *IEEE Computer Graphics and Applications*, 34(6):74–82, 2014. ISSN 02721716. doi: 10.1109/MCG.2014.117.

[287] R. Messing and F. H. Durgin. Distance Perception and the Visual Horizon in Head-Mounted Displays. *ACM Transactions on Applied Perception*, 2(3):234–250, 2005. ISSN 15443558. doi: 10.1145/1077399.1077403.

[288] A. Mossel, M. Froeschl, C. Schoenauer, A. Peer, J. Goellner, and H. Kaufmann. VROnSite: Towards immersive training of first responder squad leaders in untethered virtual reality. *Proceedings - IEEE Virtual Reality*, pages 357–358, 2017. doi: 10.1109/VR.2017.7892324.

[289] A. Naceri, R. Chellali, F. Dionnet, and S. Toma. Depth perception within virtual environments: A comparative study between wide screen stereoscopic displays and head mounted devices. *Computation World: Future Computing, Service Computation, Adaptive, Content, Cognitive, Patterns, ComputationWorld 2009*, pages 460–466, 2009. doi: 10.1109/ComputationWorld.2009.91.

[290] H. M. Peperkorn, J. Diemer, and A. Mühlberger. Temporal dynamics in the relation between presence and fear in virtual reality. *Computers in Human Behavior*, 48: 542–547, 2015. ISSN 07475632. doi: 10.1016/j.chb.2015.02.028. URL `http://www.sciencedirect.com/science/article/pii/S0747563215001260`.

[291] J. D. Pfautz. Depth perception in computer graphics. Technical Report 546, 2002. URL `http://www.cl.cam.ac.uk/TechReports/`.

[292] B. Ries, V. Interrante, M. Kaeding, and L. Phillips. Analyzing the effect of a virtual avatar's geometric and motion fidelity on ego-centric spatial perception in immersive virtual environments. (September 2014):59, 2009. doi: 10.1145/1643928.1643943.

[293] P. Willemsen, M. B. Colton, S. H. Creem-Regehr, and W. B. Thompson. The effects of head-mounted display mechanics on distance judgments in virtual environments. page 35, 2004. doi: 10.1145/1012551.1012558.

[294] P. Tiefenbacher, N. H. Lehment, and G. Rigoll. Don't Walk into Walls: Creating and Visualizing Consensus Realities for Next Generation Videoconferencing. In R. Shumaker and S. Lackey, editors, *Virtual, Augmented and Mixed Reality. Designing and Developing Virtual and Augmented Environments*, pages 170–180, Cham, 2014. Springer International Publishing. ISBN 978-3-319-07458-0.

[295] J. Bailey, J. N. Bailenson, A. S. Won, J. Flora, and K. C. Armel. Presence and Memory: Immersive Virtual Reality Effects on Cued Recall Jakki. In *Proceedings of the International Society for Presence Research Annual Conference, Philadelphia, PA*, pages 24–26, 2012. URL `papers2://publication/uuid/D4532D07-A0C9-4FEC-A194-B9B7B7CB9A57`.

[296] P. Willemsen, A. A. Gooch, W. B. Thompson, and S. H. Creem-Regehr. Effects of stereo viewing conditions on distance perception in virtual environments. *Presence: Teleoperators and Virtual Environments*, 17(1):91–101, 2008. ISSN 10547460. doi: 10.1162/pres.17.1.91.

[297] P. Willemsen and A. A. Gooch. Perceived egocentric distances in real, image-based, and traditional virtual environments. *IEEE Proceedings of the Virtual Reality (VR)*, 2002:275–276, 2002.

Appendices

A

Works included in analysis of VE and spatial audio technologies

A.1 Included studies in analysis of spatial audio techniques used in virtual environment studies

- S. Azevedo, P. Campos, and J. A. Jorge. Combining EEG Data with Place and Plausibility Responses as an Approach to Measuring Presence in Outdoor Virtual Environments. *Presence: Teleoperators & Virtual Environments*, 23(4):354–368, 2014. ISSN 15313263. doi: 10.1162/PRES. URL `http://www.mitpressjournals.org/doi/pdf/10.1162/PRES{_}a{_}00135`

- W. Bailey and B. M. Fazenda. The effect of reverberation and audio spatialization on egocentric distance estimation of objects in stereoscopic virtual reality. *The Journal of the Acoustical Society of America*, 141(5):3510, 2017. doi: 10.1121/1.4987362. URL `https://doi.org/10.1121/1.4987362`

- I. Bergstrom, S. Azevedo, P. Papiotis, N. Saldanha, and M. Slater. The Plausibility of a String Quartet Performance in Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics*, 23(4):1332–1339, 2017. ISSN 10772626. doi: 10.1109/TVCG.2017.2657138

- J. G. Bolaños and V. Pulkki. Immersive audiovisual environment with 3D audio playback. *132nd Audio Engineering Society Convention 2012*, pages 404–412, 2012

- K. Bormann. Presence and the Utility of Audio Spatialization. *Presence: Teleoperators and Virtual Environments*, 14(3):278–297, 2005. ISSN 1054-7460. doi: 10.1162/105474605323384645. URL http://www.mitpressjournals.org/doi/10.1162/105474605323384645

- K. Debattista, T. Bashford-Rogers, C. Harvey, B. Waterfield, and A. Chalmers. Subjective Evaluation of High-Fidelity Virtual Environments for Driving Simulations. *IEEE Transactions on Human-Machine Systems*, 48(1):30–40, 2017. ISSN 21682291. doi: 10.1109/THMS.2017.2762632

- D. J. Finnegan, E. O'Neill, and M. J. Proulx. An approach to reducing distance compression in audiovisual virtual environments, 2017

- D. J. Finnegan, b. O. Eamonn, and M. J. Proulx. Compensating for Distance Compression in Audiovisual Virtual Environments Using Incongruence. *Conference of Human-Computer Interaction 2016*, pages 200–212, 2016. doi: 10.1145/2858036.2858065

- C. Greenhalgh and S. Benford. MASSIVE: a collaborative virtual environment for teleconferencing. *ACM Transactions on Computer-Human Interaction*, 2(3):239–261, 1995. ISSN 10730516. doi: 10.1145/210079.210088. URL http://portal.acm.org/citation.cfm?doid=210079.210088

- M. Gröhn, T. Lokki, and T. Takala. Localizing sound sources in a CAVE-like virtual environment with loudspeaker array reproduction. *Presence: Teleoperators and Virtual Environments*, 16(2):157–171, 2007. ISSN 10547460. doi: 10.1162/pres.16.2.157. URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-34247095953{&}partnerID=40{&}md5=2bfe1bf8dcede0a835f1b3f3b29d8f5d

- M. Gröhn, T. Lokki, and T. Takala. Comparison of auditory, visual, and audiovisual navigation in a 3D space. *ACM Transactions on Applied Perception*, 2(4):564–570, 2005. ISSN 1544-3558. doi: 10.1145/1101530.1101558

- C. Hendrix and W. Barfield. Presence in virtual environments as a function of visual and auditory cues. *Proceedings Virtual Reality Annual International Symposium '95*, pages 74–82, 1995. doi: 10.1109/VRAIS.1995.512482

- E. R. Hoeg, L. J. Gerry, L. Thomsen, N. C. Nilsson, and S. Serafin. Binaural sound reduces reaction time in a virtual reality search task. *2017 IEEE 3rd VR Workshop on Sonic Interactions for Virtual Environments, SIVE 2017*, 37, 2017. doi: 10.1109/SIVE.2017.7901610

- T. Iachini, L. Maffei, F. Ruotolo, V. P. Senese, G. Ruggiero, M. Masullo, and N. Alekseeva. Multisensory assessment of acoustic comfort aboard metros: A virtual reality

study. *Applied Cognitive Psychology*, 26(5):757–767, 2012. ISSN 08884080. doi: 10.1002/acp.2856

- G. Kearney, X. Liu, A. Manns, and M. Gorzel. Auditory Distance Perception with Static and Dynamic Binaural Rendering. *Proceedings of the 57th AES International Conference, Hollywood, CA, USA*, pages 1–8, 2015

- D. M. Krum, E. A. Suma, and M. Bolas. Spatial misregistration of virtual human audio: Implications of the precedence effect. *Intelligent Virtual Agents (Lecture Notes in Computer Science)*, 7502:139–145, 2012. ISSN 03029743. doi: 10.1109/3DUI.2012.6184204. URL `http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6184204`

- P. Larsson, D. Västfjäll, and M. Kleiner. Better Presence and Performance in Virtual Environments By Improved Binaural Sound Rendering. *Journal of the Audio Engineering Society*, pages 1–8, 2002

- H.-J. Maempel and M. Jentsch. Audio-visual interaction of size and distance perception in concert halls - a preliminary study. *International Symposium on Room Acoustics*, page P124, 2013

- S. Pelzer, B. Masiero, and M. Vorländer. 3D Reproduction of Room Auralizations by Combining Intensity Panning, Crosstalk Cancellation and Ambisonics. *Proceedings of the EAA Joint Symposium on Auralization and Ambisonics*, (April):3–5, 2014. doi: 10.14279/depositonce-33

- C. Pike, R. Taylor, T. Parnell, and F. Melchior. Object-based Spatial Audio Production for Virtual Reality using the Audio Definition Model. *Proceedings of the AES International Conference on Audio for Augmented and Virtual Reality*, pages 1–7, 2016

- S. Poeschl, K. Wall, and N. Doering. Integration of spatial sound in immersive virtual environments an experimental study on effects of spatial sound on presence. *Virtual Reality (VR), 2013 IEEE*, pages 129–130, 2013. ISSN 1087-8270. doi: 10.1109/VR.2013.6549396

- B. N. J. Postma and B. F. G. Katz. The Influence of Visual Distance on the Room-acoustic Experience of Auralizations. *The Journal of the Acoustical Society of America*, 142(5):3035–3046, 2017. ISSN 00014966. doi: 10.1121/1.5009554. URL `https://doi.org/10.1121/1.5009554{%}0Ahttp://asa.scitation.org/toc/jas/142/5`

- M. Rebillat, B. F. G. Katz, and C. Oberglatt. Smart-I 2 : Spatial Multi-User Audio-Visual Real-Time Interactive Interface, a Broadcast Application Context. *Design*, pages 2–5, 2009

- M. Rébillat, X. Boutillon, É. Corteel, and B. F. G. Katz. Audio, visual, and audio-visual egocentric distance perception in virtual environments. *Forum Acusticum*, 1 (1):0–5, 2011. ISSN 22213767

- B. E. Riecke, A. Väljamäe, and J. Schulte-Pelkum. Moving sounds enhance the visually-induced self-motion illusion (circular vection) in virtual reality. *ACM Transactions on Applied Perception*, 6(2):1–27, 2009. ISSN 15443558. doi: 10.1145/1498700.1498701. URL `http://portal.acm.org/citation.cfm?doid=1498700.1498701`

- B. E. Riecke, J. Schulte-Pelkum, F. Caniard, and H. H. B Ülthoff. Spatialized auditory cues enhance the visually-induced self-motion illusion (circular vection) in Virtual Reality. *Max Planck Institute for Biological Cybernetics Technical Report*, 138(138):1–11, 2005. doi: http://www.twk.tuebingen.mpg.de/twk05/abstract.php?_load_id=riecke01. URL `http://www.kyb.mpg.de/publication.html?publ=4187`

- O. Rummukainen, S. Schlecht, A. Plinge, and E. A. Habets. Evaluation of binaural reproduction systems from behavioral patterns in a six-degrees-of-freedom wayfinding task. *2017 9th International Conference on Quality of Multimedia Experience, QoMEX 2017*, pages 4–6, 2017. doi: 10.1109/QoMEX.2017.7965680

- O. Rummukainen, S. Schlecht, A. Plinge, and E. A. Habets. Evaluation of binaural reproduction systems from behavioral patterns in a six-degrees-of-freedom wayfinding task. *2017 9th International Conference on Quality of Multimedia Experience, QoMEX 2017*, pages 4–6, 2017. doi: 10.1109/QoMEX.2017.7965680

- F. Ruotolo, L. Maffei, M. Di Gabriele, T. Iachini, M. Masullo, G. Ruggiero, and V. P. Senese. Immersive virtual reality and environmental noise assessment: An innovative audio-visual approach. *Environmental Impact Assessment Review*, 41: 10–20, 2013. ISSN 01959255. doi: 10.1016/j.eiar.2013.01.007. URL `http://dx.doi.org/10.1016/j.eiar.2013.01.007`

- M. Schoeffler, J. L. Gernert, M. Neumayer, S. Westphal, and J. Herre. 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, 2015. ISSN 14349957. doi: 10.1007/s10055-015-0270-8

- D. Thery, D. Poirier-Quinot, B. N. J. Postma, and B. F. G. Katz. Impact of the Visual Rendering System on Subjective Auralization Assessment in VR. In J. Barbic, M. D'Cruz, M. E. Latoschik, M. Slater, and P. Bourdot, editors, *Virtual Reality and Augmented Reality*, pages 105–118, Cham, 2017. Springer International Publishing. ISBN 978-3-319-72323-5

- N. Tsingos, E. Gallo, and G. Drettakis. Perceptual audio rendering of complex virtual environments. *ACM Transactions on Graphics*, 23(3):249, 2004. ISSN 07300301. doi: 10.1145/1015706.1015710. URL `http://portal.acm.org/citation.cfm?doid=1015706.1015710`

- M. Vinnikov, R. S. Allison, and S. Fernandes. Gaze-Contingent Auditory Displays for Improved Spatial Attention in Virtual Reality. *ACM Transactions on Computer-Human Interaction*, 24(3):1–38, 2017. ISSN 10730516. doi: 10.1145/3067822. URL `http://dl.acm.org/citation.cfm?doid=3086563.3067822`

A.2 Included studies in analysis of visual presentation media used in virtual environment studies

- N. Averbukh. Subjective-Situational Study of Presence. In R. Shumaker and S. Lackey, editors, *Virtual, Augmented and Mixed Reality. Designing and Developing Virtual and Augmented Environments. VAMR 2014. Lecture Notes in Computer Science*, volume 8525. Springer, Cham, 2014

- A. C. Beall, J. M. Loomis, J. W. Philbeck, and T. G. Fikes. Absolute Motion Parallax Weakly Determines Visual Scale in Real and Virtual Environments. *Human Vision, Visual Processing, and Digital Display VI*, 2411:288–297, 1995. ISSN 0277786X. doi: 10.1117/12.207547

- M. Bellani, L. Fornasari, L. Chittaro, and P. Brambilla. Virtual reality in autism: State of the art. *Epidemiology and Psychiatric Sciences*, 20(3):235–238, 2011. ISSN 20457960. doi: 10.1017/S2045796011000448

- G. P. Bingham, A. Bradley, M. Bailey, and R. Vinner. Accomodation, Occlusion, and Disparity Matching Are Used to Guide Reaching: A Comparison of Actual Versus Virtual Environments. *Journal of Experimental Psychology: Human Perception and Performance*, 27(6):1314–1334, 2001

- G. Bruder, F. Steinicke, P. Wieland, and M. Lappe. Tuning self-motion perception in virtual reality with visual illusions. *IEEE Transactions on Visualization and Computer Graphics*, 18(7):1068–1078, 2012. ISSN 10772626. doi: 10.1109/TVCG.2011.274

- E. Combe, J. Posselt, and A. Kemeny. Virtual Prototype Visualization: A Size Perception Study. In *Proceedings of the 5th Nordic Conference on Human-computer Interaction: Building Bridges*, NordiCHI '08, pages 581–582, New York, NY, USA, 2008. ACM. ISBN 978-1-59593-704-9. doi: 10.1145/1463160.1463253. URL `http://doi.acm.org/10.1145/1463160.1463253`

- E. Combe, J. Posselt, and A. Kemeny. Virtual Prototype Visualization: A Size Perception Study. In *Proceedings of the 5th Nordic Conference on Human-computer Interaction: Building Bridges*, NordiCHI '08, pages 581–582, New York, NY, USA, 2008. ACM. ISBN 978-1-59593-704-9. doi: 10.1145/1463160.1463253. URL `http://doi.acm.org/10.1145/1463160.1463253`

- T. J. Dodds, B. J. Mohler, and H. H. Bülthoff. Talk to the virtual hands: Self-animated avatars improve communication in head-mounted display virtual environments. *PLoS ONE*, 6(10), 2011. ISSN 19326203. doi: 10.1371/journal.pone.0025759

- J. Fröhlich and I. Wachsmuth. Combining multi-sensory stimuli in virtual worlds - A progress report. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8525 LNCS (PART 1):44–54, 2014. ISSN 16113349. doi: 10.1007/978-3-319-07458-0_5

- A. Gaggioli and R. Breining. Perception and cognition in immersive Virtual Reality. *Emerging Communication: Studies on New Technologies and Practices in*, (FEBRUARY 2004):71–86, 2001

- T. Y. Grechkin, T. D. Nguyen, J. M. Plumert, J. F. Cremer, and J. K. Kearney. How does presentation method and measurement protocol affect distance estimation in real and virtual environments? *ACM Transactions on Applied Perception*, 7(4): 1–18, 2010. ISSN 15443558. doi: 10.1145/1823738.1823744

- M. Gröhn, T. Lokki, and T. Takala. Localizing sound sources in a CAVE-like virtual environment with loudspeaker array reproduction. *Presence: Teleoperators and Virtual Environments*, 16(2):157–171, 2007. ISSN 10547460. doi: 10.1162/pres.16.2.157. URL `https://www.scopus.com/inward/record.uri?eid=2-s2.0-34247095953{&}partnerID=40{&}md5=2bfe1bf8dcede0a835f1b3f3b29d8f5d`

- J. A. Jones, J. E. Swan, G. Singh, S. Reddy, K. Moser, C. Hua, and S. R. Ellis. Improvements in visually directed walking in virtual environments cannot be explained by changes in gait alone. In *Proceedings of ACM Symposium on Applied Perception (SAP) 2012*, pages 11–16, 2012. doi: 10.1145/2338676.2338679

- F. Kellner, B. Bolte, G. Bruder, U. Rautenberg, F. Steinicke, M. Lappe, and R. Koch. Geometric Calibration of Head-Mounted Displays and its Effects on Distance Estimation. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 18 (4):589–596, 2012

- R. V. Kenyon, D. Sandin, R. C. Smith, R. Pawlicki, and T. Defanti. Size-constancy in the CAVE. *Presence: Teleoperators and Virtual Environments*, 16(2):172–187, 2007. ISSN 10547460. doi: 10.1162/pres.16.2.172

- E. Klein, J. E. Swan, G. S. Schmidt, M. A. Livingston, and O. G. Staadt. Measurement protocols for Medium-Field distance perception in Large-Screen immersive displays. *Proceedings - IEEE Virtual Reality*, pages 107–113, 2009. ISSN 1087-8270. doi: 10.1109/VR.2009.4811007

- E. Klein, J. E. Swan, G. S. Schmidt, M. A. Livingston, and O. G. Staadt. Measurement protocols for Medium-Field distance perception in Large-Screen immersive displays. *Proceedings - IEEE Virtual Reality*, pages 107–113, 2009. ISSN 1087-8270. doi: 10.1109/VR.2009.4811007

- S. E. Kober and C. Neuper. Using auditory event-related EEG potentials to assess presence in virtual reality. *International Journal of Human Computer Studies*, 70 (9):577–587, 2012. ISSN 10715819. doi: 10.1016/j.ijhcs.2012.03.004. URL `http://dx.doi.org/10.1016/j.ijhcs.2012.03.004`

- S. A. Kuhl, W. B. Thompson, and S. H. Creem-Regehr. Minification Influences Spatial Judgements in Virtual Environments. In *APGV '06 Proceedings of the 3rd symposium on Applied perception in graphics and visualization*, pages 15–19, Boston, Massachusetts, USA, jul 2006. ACM

- S. Kuhl, K. Hinrichs, F. Steinicke, P. Willemsen, M. Lappe, and G. Bruder. Judgment of natural perspective projections in head-mounted display environments. page 35, 2009. doi: 10.1145/1643928.1643940

- B. R. Kunz, L. Wouters, D. Smith, W. B. Thompson, and S. H. Creem-Regehr. Revisiting the effect of quality of graphics on distance judgments in virtual environments: A comparison of verbal reports and blind walking. *Attention, Perception and Psychophysics*, 71(6):1284–1293, 2009. doi: 10.3758/APP

- N. H. Lehment, D. Merget, and G. Rigoll. Creating automatically aligned consensus realities for AR videoconferencing. *ISMAR 2014 - IEEE International Symposium on Mixed and Augmented Reality - Science and Technology 2014, Proceedings*, pages 201–206, 2014. doi: 10.1109/ISMAR.2014.6948428

- T. Lentz, D. Schröder, M. Vorländer, and I. Assenmach. Virtual reality system with integrated sound field simulation and reproduction. *Eurasip Journal on Advances in Signal Processing*, 2007. ISSN 11108657. doi: 10.1155/2007/70540

- P. Lubos, G. Bruder, and F. Steinicke. Analysis of direct selection in head-mounted display environments. *IEEE Symposium on 3D User Interfaces 2014, 3DUI 2014 - Proceedings*, pages 11–18, 2014. doi: 10.1109/3DUI.2014.6798834

- X. Luo. From augmented reality to augmented computing: A look at cloud-mobile convergence. *Proceedings - 2009 International Symposium on Ubiquitous Virtual Reality, ISUVR 2009*, (November 2007):29–32, 2009. doi: 10.1109/ISUVR.2009.13

- M. R. Marner, R. T. Smith, J. A. Walsh, and B. H. Thomas. Spatial user interfaces for large-scale projector-based augmented reality. *IEEE Computer Graphics and Applications*, 34(6):74–82, 2014. ISSN 02721716. doi: 10.1109/MCG.2014.117

- R. Messing and F. H. Durgin. Distance Perception and the Visual Horizon in Head-Mounted Displays. *ACM Transactions on Applied Perception*, 2(3):234–250, 2005. ISSN 15443558. doi: 10.1145/1077399.1077403

- A. Mossel, M. Froeschl, C. Schoenauer, A. Peer, J. Goellner, and H. Kaufmann. VROnSite: Towards immersive training of first responder squad leaders in untethered virtual reality. *Proceedings - IEEE Virtual Reality*, pages 357–358, 2017. doi: 10.1109/VR.2017.7892324

- A. Naceri, R. Chellali, F. Dionnet, and S. Toma. Depth perception within virtual environments: A comparative study between wide screen stereoscopic displays and head mounted devices. *Computation World: Future Computing, Service Computation, Adaptive, Content, Cognitive, Patterns, ComputationWorld 2009*, pages 460–466, 2009. doi: 10.1109/ComputationWorld.2009.91

- H. M. Peperkorn, J. Diemer, and A. Mühlberger. Temporal dynamics in the relation between presence and fear in virtual reality. *Computers in Human Behavior*, 48: 542–547, 2015. ISSN 07475632. doi: 10.1016/j.chb.2015.02.028. URL `http://www.sciencedirect.com/science/article/pii/S0747563215001260`

- J. D. Pfautz. Depth perception in computer graphics. Technical Report 546, 2002. URL `http://www.cl.cam.ac.uk/TechReports/`

- B. E. Riecke, A. Väljamäe, and J. Schulte-Pelkum. Moving sounds enhance the visually-induced self-motion illusion (circular vection) in virtual reality. *ACM Transactions on Applied Perception*, 6(2):1–27, 2009. ISSN 15443558. doi: 10.1145/1498700.1498701. URL `http://portal.acm.org/citation.cfm?doid=1498700.1498701`

- B. Ries, V. Interrante, M. Kaeding, and L. Phillips. Analyzing the effect of a virtual avatar's geometric and motion fidelity on ego-centric spatial perception in immersive virtual environments. (September 2014):59, 2009. doi: 10.1145/1643928.1643943

-

- P. Willemsen, M. B. Colton, S. H. Creem-Regehr, and W. B. Thompson. The effects of head-mounted display mechanics on distance judgments in virtual environments. page 35, 2004. doi: 10.1145/1012551.1012558

- P. Tiefenbacher, N. H. Lehment, and G. Rigoll. Don't Walk into Walls: Creating and Visualizing Consensus Realities for Next Generation Videoconferencing. In R. Shumaker and S. Lackey, editors, *Virtual, Augmented and Mixed Reality. Designing and Developing Virtual and Augmented Environments*, pages 170–180, Cham, 2014. Springer International Publishing. ISBN 978-3-319-07458-0

- M. Vinnikov, R. S. Allison, and S. Fernandes. Gaze-Contingent Auditory Displays for Improved Spatial Attention in Virtual Reality. *ACM Transactions on Computer-Human Interaction*, 24(3):1–38, 2017. ISSN 10730516. doi: 10.1145/3067822. URL `http://dl.acm.org/citation.cfm?doid=3086563.3067822`

- J. Bailey, J. N. Bailenson, A. S. Won, J. Flora, and K. C. Armel. Presence and Memory: Immersive Virtual Reality Effects on Cued Recall Jakki. In *Proceedings of the International Society for Presence Research Annual Conference, Philadelphia, PA*, pages 24–26, 2012. URL `papers2://publication/uuid/D4532D07-A0C9-4FEC-A194-B9B7B7CE`

- P. Willemsen, A. A. Gooch, W. B. Thompson, and S. H. Creem-Regehr. Effects of stereo viewing conditions on distance perception in virtual environments. *Presence: Teleoperators and Virtual Environments*, 17(1):91–101, 2008. ISSN 10547460. doi: 10.1162/pres.17.1.91

- P. Willemsen and A. A. Gooch. Perceived egocentric distances in real, image-based, and traditional virtual environments. *IEEE Proceedings of the Virtual Reality (VR)*, 2002:275–276, 2002

B

Questionnaires

B.1 EQ-SQ Short Form

A. Wakabayashi, S. Baron-Cohen, S. Wheelwright, N. Goldenfeld, J. Delaney, D. Fine, R. Smith, and L. Weil. Development of short forms of the Empathy Quotient (EQ-Short) and the Systemizing Quotient (SQ-Short). *Personality and Individual Differences*, 41(5):929–940, 2006. ISSN 01918869. doi: 10.1016/j.paid.2006.03.017

Empathy Quotient

1. I can easily tell if someone else wants to enter a conversation
2. I really enjoy caring for other people
3. I find it hard to know what to do in a social situation - *Reversal*
4. I often find it difficult to judge if something is rude or polite - *Reversal*
5. In a conversation I tend to focus on my own thoughts rather than on what my listener might be thinking - *Reversal*
6. I can pick up quickly if someone says one thing but means another
7. It is hard for me to see why some things upset people so much - *Reversal*
8. I find it easy to put myself in somebody else's shoes
9. I am good at predicting how someone will feel
10. I am quick to spot when someone in a group is feeling awkward or uncomfortable

11. I can't always see why someone should have felt offended by a remark - *Reversal*
12. I don't tend to find social situations confusing
13. Other people tell me I am good at understanding how they are feeling and what they are thinking
14. I can easily tell if someone else is interested or bored with what I am saying
15. Friends usually talk to me about their problems as they say that I am very understanding
16. I can sense if I am intruding, even if the other person doesn't tell me
17. Other people often say that I am insensitive, though I don't always see why - *Reversal*
18. I can tune into how someone else feels rapidly and intuitively
19. I can easily work out what another person might want to talk about
20. I can tell if someone is masking their true emotion
21. I am good at predicting what someone will do
22. I tend to get emotionally involved with a friend's problems

Systemizing Quotient

1. If I were buying a car, I would want to obtain specific information about its engine capacity
2. If there was a problem with the electrical wiring in my home, I'd be able to fix it myself
3. I rarely read articles or web pages about new technology - *Reversal*
4. I do not enjoy games that involve a high degree of strategy - *Reversal*
5. I am fascinated by how machines work
6. In math, I am intrigued by the rules and patterns governing numbers,
7. I find it difficult to understand instruction manuals for putting appliances together - *Reversal*

8. If I were buying a computer, I would want to know exact details about its hard disc drive capacity and processor speed
9. I find it difficult to read and understand maps - *Reversal*
10. When I look at a piece of furniture, I do not notice the details of how it was constructed - *Reversal*
11. I find it difficult to learn my way around a new city - *Reversal*
12. I do not tend to watch science documentaries on television or read articles about science and nature - *Reversal*
13. If I were buying a stereo, I would want to know about its precise technical features
14. I find it easy to grasp exactly how odds work in betting
15. I am not very meticulous when I carry out D.I.Y - *Reversal*
16. When I look at a building, I am curious about the precise way it was constructed
17. I find it difficult to understand information the bank sends me on different investment and saving systems - *Reversal*
18. When travelling by train, I often wonder exactly how the rail networks are coordinated,
19. If I were buying a camera, I would not look carefully into the quality of the lens - *Reversal*
20. When I hear the weather forecast, I am not very interested in the meteorological patterns - *Reversal*
21. When I look at a mountain, I think about how precisely it was formed
22. I can easily visualize how the motorways in my region link up
23. When I'm in a plane, I do not think about the aerodynamics - *Reversal*
24. I am interested in knowing the path a river takes from its source to the sea
25. I am not interested in understanding how wireless communication works

B.2 Immersive Tendencies Questionnaire

B. G. Witmer and M. J. Singer. Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoper. Virtual Environ.*, 7(3):225–240, 1998. ISSN 1054-7460. doi: 10.1162/105474698565686

1. Do you easily become deeply involved in movies or tv dramas?
2. Do you ever become so involved in a television program or book that people have problems getting your attention? How mentally alert do you feel at the present time?
3. Do you ever become so involved in a movie that you are not aware of things happening around you?
4. How frequently do you find yourself closely identifying with the characters in a story line?
5. Do you ever become so involved in a video game that it is as if you are inside the game rather than moving a joystick and watching the screen?
6. How physically fit do you feel today?
7. How good are you at blocking out external distractions when you are involved in something?
8. When watching sports, do you ever become so involved in the game that you react as if you were one of the players?
9. Do you ever become so involved in a daydream that you are not aware of things happening around you?
10. Do you ever have dreams that are so real that you feel disoriented when you awake?
11. When playing sports, do you become so involved in the game that you lose track of time?
12. How well do you concentrate on enjoyable activities?
13. How often do you play arcade or video games? (OFTEN should be taken to mean every day or every two days, on average.)

14. Have you ever gotten excited during a chase or fight scene on TV or in the movies?

15. Have you ever gotten scared by something happening on a TV show or in a movie?

16. Have you ever remained apprehensive or fearful long after watching a scary movie?

17. Do you ever become so involved in doing something that you lose all track of time?

B.3 Five Factor Personality Test

L. R. Goldberg. The development of markers for the Big-Five factor structure. *Psychological Assessment*, 4(1):26–42, 1992

Respond with one of the following:

- Strongly disagree
- Disagree a little
- Neither agree or disagree
- Agree a little
- Strongly agree

I am someone who...

1. Is talkative - **Extraversion**
2. Tends to find fault with others - *Reversal* - **Agreeableness**
3. Does a thorough job - **Conscientiousness**
4. Is depressed, blue - **Neuroticism**
5. Is original, comes up with new ideas - **Openness**

6. Is reserved - *Reversal* - **Extraversion**
7. Is helpful and unselfish with others - **Agreeableness**
8. Can be somewhat careless - *Reversal* - **Conscientiousness**
9. Is relaxed, handles stress well. - *Reversal* - **Neuroticism**
10. Is curious about many different things - **Openness**
11. Is full of energy - **Extraversion**
12. Starts quarrels with others - *Reversal* - **Agreeableness**
13. Is a reliable worker - **Conscientiousness**
14. Can be tense - **Neuroticism**
15. Is ingenious, a deep thinker - **Openness**
16. Generates a lot of enthusiasm - **Extraversion**
17. Has a forgiving nature - **Agreeableness**
18. Tends to be disorganized - *Reversal* - **Conscientiousness**
19. Worries a lot - **Neuroticism**
20. Has an active imagination - **Openness**
21. Tends to be quiet - *Reversal* - **Extraversion**
22. Is generally trusting - **Agreeableness**
23. Tends to be lazy - *Reversal* - **Conscientiousness**
24. Is emotionally stable, not easily upset - *Reversal* - **Neuroticism**
25. Is inventive - **Openness**
26. Has an assertive personality - **Extraversion**
27. Can be cold and aloof - *Reversal* - **Agreeableness**
28. Perseveres until the task is finished - **Conscientiousness**
29. Can be moody - **Neuroticism**

30. Values artistic, aesthetic experiences - **Openness**
31. Is sometimes shy, inhibited - *Reversal* - **Extraversion**
32. Is considerate and kind to almost everyone - **Agreeableness**
33. Does things efficiently - **Conscientiousness**
34. Remains calm in tense situations - *Reversal* - **Neuroticism**
35. Prefers work that is routine - *Reversal* - **Openness**
36. Is outgoing, sociable - **Extraversion**
37. Is sometimes rude to others - *Reversal* - **Agreeableness**
38. Makes plans and follows through with them - **Conscientiousness**
39. Gets nervous easily - **Neuroticism**
40. Likes to reflect, play with ideas - **Openness**
41. Has few artistic interests - *Reversal* - **Openness**
42. Likes to cooperate with others - **Agreeableness**
43. Is easily distracted - *Reversal* - **Conscientiousness**
44. Is sophisticated in art, music, or literature - **Openness**