NUI Galway
OÉ Gaillimh

Effect of Spatial Audio and Movement Control on User Performance, Awareness, and Preference

James Broderick, B.Sc.

Research Supervisors
Prof Jim Duggan, Dr Sam Redfern

School of Computer Science
College of Science and Engineering
National University of Ireland Galway

A thesis submitted for the degree of Doctor of Philosophy

Submitted September 2021

Table of Contents

Table of Contents 1
List of Tables 4
List of Figures 5
Declaration of Originality 7
James Broderick, B.Sc., 2021 7
Acknowledgements 8
James Broderick 8
Glossary/Abbreviations 9
1 Introduction 10
1.1 Overview 10
1.2 Contributions 12
1.3 Publications 13
1.4 Thesis Roadmap 14
2 Background, Literature Survey and Meta Analysis 15
2.1 Introduction 15
2.2 Spatial Audio 18
2.2.1 Defining HRTF and Spatial Audio 18
2.2.2 Audio Searching Techniques 21
2.2.3 Audio in other fields of study 23
2.2.4 Spatial Audio Technologies 25
2.3 Virtual Reality 26
2.3.1 What is Virtual Reality? 26
2.3.2 Presence in Virtual Reality 28
2.3.3 Head Tracking 29
2.4 Spatial Audio and Virtual Reality 31
2.4.1 Spatial Audio and Navigation 31
2.4.2 Spatial Audio and Sonification 32
2.4.3 Spatial Audio in Other Task Management 35
2.4.4 Immersion and Presence 36

Table of Contents

2.5 Applications of Spatial Audio in Virtual Spaces 38
2.5.1 Serious Games and Game Engines 38
2.5.2 Serious Games for Training and Education 41
2.5.3 Examples of Spatial Audio in modern games 44
2.5.4 Collaborative Virtual Environments and Decision Support Systems 47
2.6 Meta Analysis 54
2.7 Research Gaps 60
2.8 Conclusion 61
3. Methodology and Experimental Design 63
3.1 Introduction 63
3.2 Research Questions 63
3.3 Summary of the experiments 66
3.3.1 Experiment 1: Locating sound sources between HRTF Spatial Audio and Panning Stereo Audio 67
3.3.2 Experiment 2: Navigating a virtual maze using audio-visual and pure visual 74
3.3.3 Experiment 3: Locating sound sources between Mouse + Keyboard and Head Mounted Virtual Reality 81
3.4 Research Constraints 95
3.5 Technologies chosen 99
3.5.1 Unity for building environments/game controls 99
3.5.2 HTC Vive for Virtual Reality headset 101
3.5.3 Steam Audio for audio spatialisation 102
3.6 Conclusion 103
4 Experimental Results 104
4.1 Introduction 104
4.2 Subject Cohort 105
4.3 Experiment 2: Maze Navigation Experiment 109
4.3.1 Overview 109
4.3.2 Quantitative Results Discussion 110
4.4 Experiment 2: Varying Control Scheme Sound Location Experiment Results 125
4.4.1 Overview 125
4.4.2 Research Questions 125
4.4.3 Quantitative Results Discussion 128
4.4.4 Qualitative Results Discussion 146

Table of Contents

4.5 Conclusion....154
5 Conclusions and Further Research....155
5.1 Summary of Contributions....155
5.2 Research Questions....156
5.3 Future Research....163
5.4 Research Challenges Due to COVID-19....165
References and Bibliography....167

List of Tables

Table 2.1 Table of Reviews Examined 59
Table 3.1 Research questions and relevant experiments 64
Table 3.2 Experimental Summary 66
Table 3.3 Feedback Questionnaire 92
Table 4.1 Research Question Table 105
Table 4.2 Maze Experiment Subject Details 106
Table 4.3 Sound Location Experiment Subject Details 106
Table 4.4 Maze Experiment Results by Audio Presence 111
Table 4.5 Individual Performance Table 115
Table 4.6 Group Mistakes Results 135
Table 4.7 Group Time Results 135
Table 4.8 Group Results for VR 137
Table 4.9 Group Results for Non VR 138
Table 4.10 User Questionnaire Responses 147
Table 4.11 User NASA-TLX Responses 153
Table 5.1 Research Questions 156

List of Figures

Fig 2.1 HRTF Diagram by Soumyasch, 2007 20
Fig 2.3 Triage Trainer screenshot, (Knight et al, 2010)....42
Fig 2.4 Diagram of the CDS-based flood management procedure, (Ivanov et al 2012) 48
Fig 2.5, Galway Bay Current Visualisation screenshot (Broderick, 2016) 53
Fig 2.6 3D modelled Galway Bay Seabed in Unity 54
Fig 3.2 Experimental environment overview, with cubes coloured by group 68
Fig 3.3 Unity Audio Source 70
Fig 3.4 Experimental Flow....73
Fig 3.10 Experimental flow and results tracking 82
Fig 3.11 Steam Audio settings 86
Fig 4.1 Research Question Table 105
Fig 4.1 Participant Experience with virtual environments/games 108
Fig 4.2 Maze Example 109
Fig 4.3 Audio vs No Audio Time Analysis....111
Fig 4.4 Audio vs No Audio Mistake Analysis 111
Fig 4.5 Average Mistakes Made by Maze....112
Fig 4.6 Average Time Taken by Maze....112
Fig 4.7 Average Mistakes By Audio Presence....113
Fig 4.8 Average time By Audio Presence 113
Fig 4.9.2 Low-Medium Experience vs High Experience F Test....116
Fig 4.9.3 Low-Medium Experience vs High Experience Mistakes....116
Fig 4.10 Average Mistakes by Experience 117
Fig 4.11 High Experience Mistakes Comparison 118
Fig 4.14 Maze 3 user paths by experience level....121
Fig 4.15 Maze 3 mistake locations by user....122
Fig 4.16 Maze 3 mistake locations by experience level 123
Fig 4.17 Maze 3 mistake locations by audio presence 124
Fig 4.19 Proportion of Measurements VR/Non-VR 130
Fig 4.20 Distribution of Sound Locations....131
Fig 4.21 Total Mistakes By Group 133
Fig 4.22 Average Mistakes By Group....134
Fig 4.24 Average Mistakes by Sound Location....138
Fig 4.25 Average Time Taken by Grouping VR vs Non-VR 139
Fig 4.26 Average Time Taken by Sound Locations 140
Fig 4.27 Individual Time Taken Comparison T-Test 141
Fig 4.27 Average Mistakes by Experience Level 142
Fig 4.28 Comparison of time taken between experience level groups 142
Fig 4.29 Comparison of mistakes made between experience level groups....143
Fig 4.30 Average Time Taken by Experience Level 143
Fig 4.31 User viewpoint in VR 144
Fig 4.32 User viewpoint in VR 145

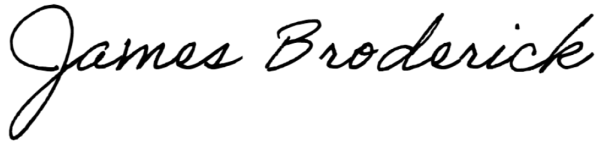
List of Figures

Fig 4.33 Was either method more difficult feedback.....148
Fig 4.34 Was either method more natural feedback.....149
Fig 4.35 Was it difficult to tell where the sounds were coming from feedback 150

Declaration of Originality

I, James Broderick, declare that this thesis is entirely composed of my own work, except where explicitly stated otherwise, and that this work has not been submitted for any other degree or professional qualification. I acknowledge that I have read the NUI Galway Code of Practice regarding Plagiarism and the University Code of Conduct, and that I am bound by them.

James Broderick, B.Sc., 2021



Acknowledgements

A doctorate is a hugely taxing undertaking, and one I did not fully expect when I started. I could not have kept going and reached this stage without the support and assistance of so many people.

To my supervisors, Prof Jim Duggan and Dr Sam Redfern, thank you for your years of countless patience, advice, and understanding. I could not have kept the course without your guidance, and none of this would be possible without you both.

Thanks to my graduate research committee Dr Enda Howley and Dr Karen Young, and all the staff in the IT Department at NUI Galway for the support and opportunities over my many years at NUIG.

I would also like to acknowledge the support of NUI Galway and the MaREI Research group for the funding and support for my work.

I also could not have done this without the support of my friends, who helped keep me sane and provided sounding boards to talk to.

And finally, I couldn't have gotten here without the unending support and love of my parents, Brid and Gerry Broderick. I wouldn't be here without the choices they provided to me, and throughout the overly long process of my education, they were a constant source of encouragement and help. Thank you both so much for being there for me.

James Broderick

Glossary/Abbreviations

- **API** Application Programming Interface
- **CDS** Collaborative Decision Support
- **CSV** Comma Separated Values
- **CVE** Collaborative Virtual Environment
- **DSS** Decision Support System
- **HMD** Head Mounted Display
- **HRTF** Head Related Transfer Function
- **LIDAR** Light Detection and Ranging
- **NASA-TLX** NASA Task Load Index
- **VR** Virtual Reality

1 Introduction

1.1 Overview

In recent years, more people are engaging with virtual environments, both recreationally and professionally. Game engines have become more accessible, making it easier to craft these virtual spaces, and home hardware improvements have made it easier for people to partake in these environments without large financial investment. There has also been a new growth in virtual reality (VR) technologies, with home VR setups becoming available to consumers easily, and more commercial and academic research being performed in these spaces. However, despite the recent improvements in these areas, the importance of high-quality spatial audio is often overlooked, with most focus being on the visual fidelity of these environments. There is also a lack of research into how different control types for these environments change how users will interact with the space and may increase the importance of factors such as spatial sound. The impact of user experience with video games is something that can also impact performance in these virtual environments, even if that experience isn't specifically with VR.

Spatial audio is sound which "is perceived by a listener as emanating from a certain location in space" (Microsoft, 2015). Spatial audio has many useful applications, from locating targets in video games, drawing attention to specific locations in a virtual space, or enhancing feelings of immersion and presence in virtual environments. However, research into its use has been relatively limited, and it's use in industry was uncommon until the last few years. Due to its computational cost and specialised knowledge requirement, it was usually left to experts in the field. However, it's importance to mixed reality experiences and it's growth in industry has led to creation of many useful tools, which allow for non-experts to implement spatial audio easily in their projects. The

constant growth of computing power and game efficiency has also made it easier for games to make use of the more realistic audio type. As more research has been done, especially in the realms of mixed reality and non-gaming environments, it is more important than ever to examine how this audio type can affect people's experience in virtual environments, and how it interacts with more realistic control types.

The goal of this thesis project is to investigate the impact of audio in virtual environments, how proper spatialisation of audio within these virtual environments can improve users' awareness and ability to locate things within these environments, and the way different control schemes for virtual environments lend themselves to working with audio virtually. Control scheme refers to the way with which a user interacts with a virtual environment, such as using a mouse and keyboard with a monitor in a traditional computer setup, or the use of a head mounted VR kit for head tracking control. Different types of control schemes require different equipment and design, and some may be more suited to particular individuals or scenarios. The three main factors being examined for this are user performance, a user's awareness of their virtual surroundings and changes in those surroundings, and the user's quality of experience or intuitive understanding of the system. This was achieved through a set of experiments looking at a user's ability to locate a source of audio in a virtual space using both mouse and keyboard controls, and a head tracking VR control scheme, as well as navigating a maze while avoiding obstacles either through pure visual awareness or supplemented with spatial audio.

From our research, while it can be difficult to find significant performance improvements in short term experiments, we have observed the way people search and interact with an environment changes when moving between traditional mouse and keyboard control schemes and head mounted VR control schemes. This change more closely relies on natural forms of sound location, which only work in virtual spaces through properly spatialised audio. We also see that people overwhelmingly feel a preference for these natural, virtual reality control schemes, and felt that finding the location of sounds was much easier and more intuitive in this control type. Finally, we discovered that the

most influential factor on differences in user performance between the experiment types was their experience with games and virtual environments prior to the experiment.

1.2 Contributions

Throughout this thesis, we discuss our work in analysing research gaps in the crossover of spatial audio, VR, and virtual environments, as well as the experiments we performed while attempting to fill these gaps. A brief overview of our contributions are as follows:

1. We reviewed historical and cutting-edge research in the areas of Virtual Reality, Spatial Audio, Serious Games, Collaborative Virtual Environments, and Decision Support Systems, as well as the overlap between these fields. We also provide background information on each of these fields. This is presented in Chapter 2.
2. We performed a meta-analysis of papers and reviews in the areas of Serious Games and Virtual Environments, highlighting the lack of research and focus on the importance of audio and spatial audio in these fields. This is presented in Chapter 2.
3. We designed and developed experiments examining the impact of Spatial Audio on user performance, awareness, and preference in virtual environments, as well as how the control scheme being used affects these factors. This involves analysis of other similar experiments and examining some of the potential design decisions not previously implemented, as well as some analysis of the tools available to make experiments in these domains. This is covered in Chapters 2 and 3.
4. We analyse the results of our navigation experiment, showing while user performance does not appear to be greatly altered in short-term experiments when using spatial audio for navigation, there is evidence that the impact of audio is greater when including user experience in virtual environments, and that user movement patterns change depending on the presence of audio. This is presented in Chapter 4.
5. We analyse the results of our sound location experiment, showing users overwhelmingly prefer VR control schemes for this type of virtual task compared

to traditional mouse and keyboard schemes, regardless of perceived performance. Users also felt that VR controls were much more natural and intuitive for this type of task. Regarding performance, while there wasn't an overall improvement of user performance, changes to performance seem to tie to user experience more closely in virtual environments. This is presented in Chapter 4.

6. We highlight some areas for expanded research and provide some potential research experiments growing off the research covered in this thesis. This is covered in Chapter 5.

1.3 Publications

Some of the research covered in this thesis has been published and presented at international conferences as follows:

- J. Broderick, J. Duggan, and S. Redfern, "Using game engines for marine visualisation and collaboration," 2016 International Conference on Image, Vision and Computing (ICIVC), 2016, pp. 96-101, doi: 10.1109/ICIVC.2016.7571280.
- Broderick, J., Duggan, J., Redfern, S. "Using Auditory Display Techniques to Enhance Decision Making And Perceive Changing Environmental Data Within a 3D Virtual Game Environment" Presented at the 23rd International Conference on Auditory Display (ICAD2017), June 20-23, 2017, Pennsylvania State University, State College, PA, USA.
- J. Broderick, J. Duggan, and S. Redfern, "The Importance of Spatial Audio in Modern Games and Virtual Environments," 2018 IEEE Games, Entertainment, Media Conference (GEM), 2018, pp. 1-9, doi: 10.1109/GEM.2018.8516445.

Additional to this, we have presented several posters covering the work performed over the course of this thesis and have prepared a journal paper detailing recent work covered in Chapter 4.

1.4 Thesis Roadmap

In this section we will provide a brief overview of each upcoming chapter of this thesis.

Chapter 2 serves as a literature review examining the relevant research topics to the work completed in this thesis. This includes Spatial Audio, Virtual Reality (VR), and the overlapping of those two fields. It also discusses some broader topics in the areas of Serious/Entertainment Games, Collaborative Virtual Environments (CVEs) and Decision Support Systems (DSS), and how research into those topics address (or in some cases ignore) spatial audio implementation. Finally, the chapter presents a meta-analysis of systematic review collections in related fields with regard to their coverage of audio, before highlighting the existing research gaps in all of the discussed areas.

Chapter 3 covers the Methodology and Experimental Design of the different experiments performed as part of this thesis, as well as the Research Questions examined, the technologies used, and the obstacles faced. We cover how the experiments were designed, the logic behind their purpose and design, and what type of data we were hoping to extract from each set of experiments.

In Chapter 4 we cover the results gathered from the experiments performed and discuss the potential implications of those results. This includes review of user movement patterns, statistical analysis of the quantitative performance data gathered within the experiments, and examination of the qualitative data gathered from users through questionnaires and workload surveys.

Finally, in Chapter 5 we summarise the findings of our research, the potential contributions, and examine some of the possible future research work. We discuss what conclusions we can draw from the analysis of qualitative and quantitative data gathered, new research areas and experiments to be performed, and some of the research constraints imposed by the COVID-19 pandemic.

2 Background, Literature Survey and Meta Analysis

2.1 Introduction

The goal of this thesis is to investigate (1) the impact of audio in virtual environments, (2) how proper spatialisation of audio within these virtual environments can improve users' awareness and ability to locate things within these environments, and (3) the way different control schemes for virtual environments lend themselves to working with audio virtually. The three main factors being examined for this are user performance, a user's awareness of their virtual surroundings and changes in those surroundings, and the user's quality of experience or intuitive understanding of the system. This chapter offers an overview of research in the fields of spatial audio, VR, and game engines, as well as coverage of some research areas that have used or could benefit from spatial audio. We examine not only relevant existing research to provide necessary knowledge of the field, but also highlight gaps in the existing fields listed above regarding their use of audio. We also performed a meta-analysis of review papers in the fields of VR and serious games and discuss the lack of audio focus within these reviews.

Spatial audio is sound which "is perceived by a listener as emanating from a certain location in space" (Microsoft, 2015). As a technology it has been having a resurgence in recent times (Cohen et al, 2015; Rajguru et al, 2020), and its importance has grown with the improvements and increased availability of VR. High quality audio makes users feel more immersed in VR, and when it comes to performing tasks or playing games in virtual spaces, properly spatialized audio can also make users more aware of their surroundings and perform better in competitive environments (Blizzard, 2016; Valve 2016; Semionov et al, 2020). While spatial audio still has technical hurdles with its implementation, new technologies and techniques are making it more accessible for researchers and for creators.

2 Background, Literature Survey and Meta Analysis

Research into uses of *Virtual Reality* has been greatly aided by the growing availability of VR technologies such as those from Oculus, HTC, and Sony, as well as the accessibility of game engines such as Unity and Unreal, which offer VR support built in. The growing market for commercial games and experiences in VR has allowed for more technologies to become available in research fields. This varies from psychological studies of immersion and presence, accessibility to non-technology fields such as cinema, as well as more fundamental studies in computer science, examining how VR can be improved and how new ways of interacting with virtual environments are needed when using VR (Martel et al, 2017).

An area which requires further study and investigation is the *crossover of Virtual Reality and Spatial Audio*. Modern games are the main application of spatial audio technology, leading to the development of audio packages that can be used within game development tools. Spatial audio in games can lead to players being more aware of their surroundings, feeling more immersed, and aid them in locating sound sources compared to traditional forms of game audio. As more games and experiences make use of VR, it is important to augment these situations with properly spatialised audio to create the best experience for users (Kern et al, 2020; Freeman et al, 2020, Rajguru et al, 2020). As virtual environments and technology improve, the way users interact with the virtual world more closely replicates how humans interact with the real world, making it important that the audio aspects of our virtual world also grow more realistic.

Over the course of this research, there were a variety of areas examined for either improvement using spatial audio, or areas where spatial audio could improve performance, awareness, or enjoyment. Some of the areas examined for this are serious and entertainment games, decision support systems, and collaborative virtual environments.

2 Background, Literature Survey and Meta Analysis

Serious gaming are games which are not focused on entertainment, and instead are aimed towards training, teaching or other non-leisure concepts (Abt, 1970). As a field it has grown significantly in recent times, with game environments being used for training (Knight et al, 2010; Chittaro et al, 2020), simulation (Sharma et al, 2014; Mayer et al, 2014; Yamin et al, 2021) and urban planning (Bishop et al, 2010; Greenwood et al, 2009; Robinson et al, 2021; Bathke et al, 2019). Game engines were originally developed for internal use by game development companies but starting in the 90s with the Unreal Engine and Id Software's Doom and Quake, these companies licensed out their engines to third parties for game creation. Previous to this, some development environments existed, but were highly specialised (e.g., Pinball Construction Set, RPG Maker; Softline, 1982). Since those early days, we now have powerful and highly customisable development kits such as Unity, CryEngine and Unreal, which not only have made game development more accessible and simpler but have also opened the technology behind games for use in non-entertainment sectors. As serious games often try to replicate real world situations, being able to make an environment feel realistic and natural is of high importance, as well as making users feel immersed and present in the environment (Freeman et al 2020; Kern et al 2020).

Decision Support Systems aid individuals and groups of people in making decisions by providing relevant information and data concisely. They can also be expanded to provide simulations of potential events and suggest courses of action based on the situation at hand. They can be used in a wide variety of cases, including disaster reaction (Ivanov et al, 2012), ecological impact (Schluter et al, 2001; Ramsey, 2009), medical support (Prange et al, 2017; Niederriter et al, 2020) and urban planning (Bishop et al, 2010; Greenwood et al, 2009). Many of these systems are collaborative, as the context behind information can be important for decision making and many decisions can affect a significant number of groups and individuals who may all have a say in the decision-making process. While there is some work looking at audio for assisting in decision support or providing information in 3D spaces, it is still a limited section of the overall research area.

Collaborative virtual environments create a virtual space inside of which users can meet each other, share, and analyse data. They can be used for decision making and planning (Bishop et al, 2010; Ramsey, 2009) and teaching (Scheucher et al, 2009), and are quite often a part of decision support systems (Greenwood et al, 2009; Ivanov et al, 2012) due to these support systems requiring collaboration between multiple sources.

2.2 Spatial Audio

2.2.1 Defining HRTF and Spatial Audio

The way people perceive and locate sounds in the real world is a complex and personalised physiological process. Humans can locate sources of sound in a 3D space despite only having two ears. This is achieved by the brain taking the sounds that reach the ears and comparing them, using differences in time, frequency, etc, as well as changes to the sound caused by body geometry, from head shape, ear shape, pinnae, shoulders, etc. This information allows us to locate the sounds elevation, distance, and azimuth within a few degrees of accuracy (Middlebrooks, 2015; Yost et al, 2013; Carlile et al, 1997) and with faster reaction times compared to visual stimuli (Jain et al., 2015). This ability to locate sound so accurately and quickly is not only physical, but something that develops as we use it, improving with our lived experiences. This is something we use unconsciously but is an important way in which we interact with and stay aware of the world around us. Being a key factor of our everyday lives, being able to use our hearing in a similar manner while in a virtual space is something that can be very impactful.

A Head-Related Transfer Function (HRTF) is a measurement which can be used to digitally replicate sound locations (Cheng et al., 2001). By examining how a sound is affected by all the factors listed above for an individual, a function is created which can be used to digitally place a sound at a location in space. The function artificially changes the sound so that it will sound as if it has travelled to the listener from the chosen location.

The measurements vary from person to person and are often gathered by placing microphones in a person's ear canal and recording a set of sounds played from various distances, azimuths, and elevation around the user. A pair of HRTF measurements, one for each ear, are used to create a binaural sound which a person can then locate in an environment exactly as they would in real life. Compared to simple stereo sound, HRTF adjusted sound can reduce problems with the *front-back problem*, where sounds directly in front of and behind the user sound quite similar. It also helps with determining elevation differences, whereas with stereo setups, sounds above, below, or at the user's elevation all sound similar. We can see an example demonstrating the different paths sound takes from a source to each ear in Fig 2.1. This highlights why sounds are different in each ear, and what HRTF aims to replicate in a digital way. In the diagram, x(t) is the sound source, hR(t) and hL(t) are the different paths the sound takes to get to our right and left ears, and XL(t) and XR(t) are where the sounds enter our ears.

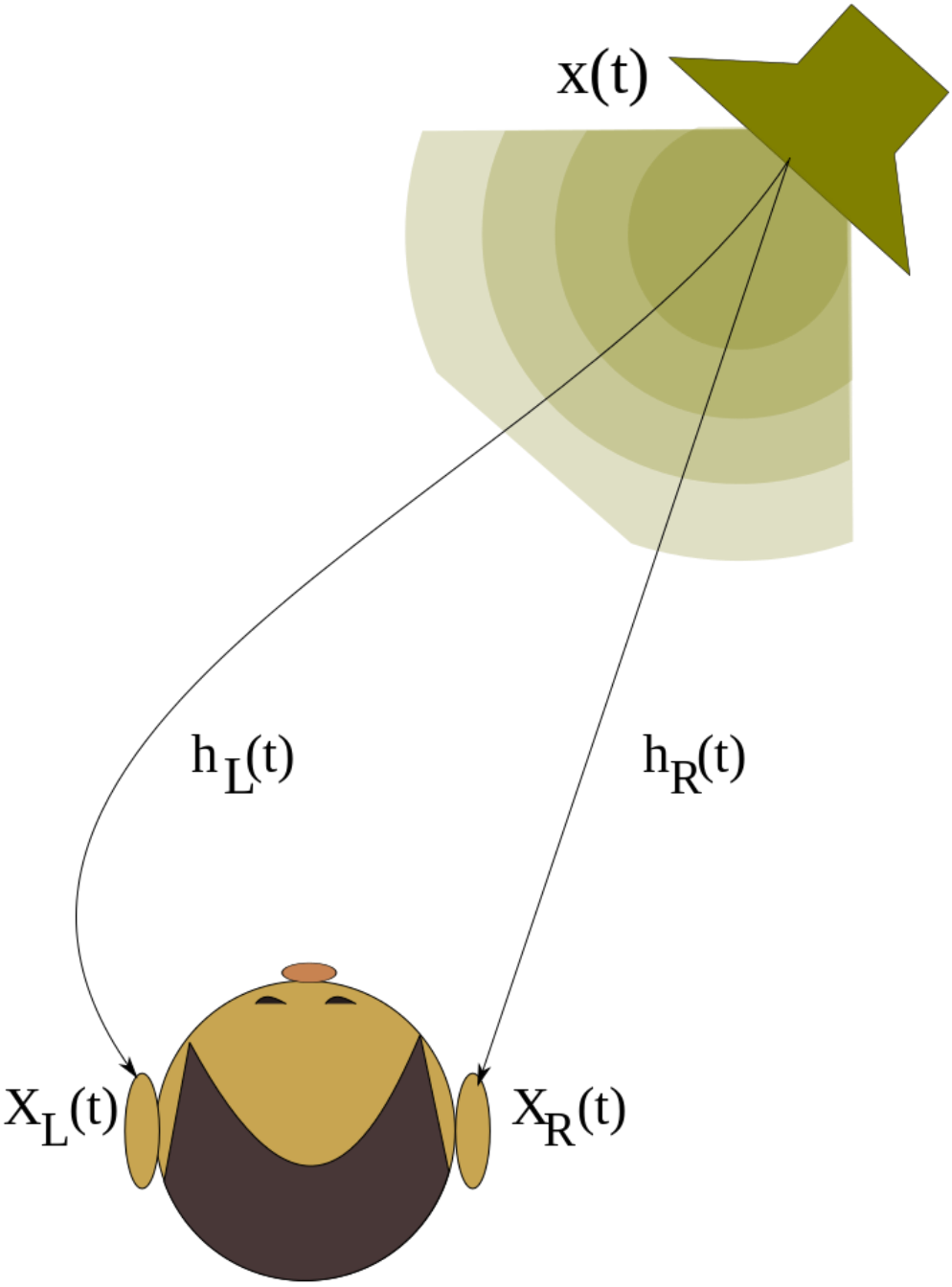


Fig 2.1 HRTF Diagram, (Soumyasch, 2007)

Because HRTFs are a completely unique measurement to an individual, one of the biggest issues with creating spatial audio for general public use is trying to craft a “best fit” generic HRTF. Being able to measure an individual's HRTF leads to the most accurate replication of how they hear real world sounds, but it is an involved process that isn’t feasible for large scale use (Anderson et al, 2021). The current methodology is to use generic HRTF libraries that are made to be the best fit for the average user. While using a HRTF created from a different individual can cause issues (Wenzel et al, 1993), a perfect HRTF is not required to benefit from sound location or realism(Kulkarni et al., 1998). Research on HRTFs has included the differences between symmetric and asymmetric HRTF datasets (Wu et al, 2016) and the effect of training on performance with non-individualised HRTF datasets (Campos et al, 2012), but otherwise the core concept of HRTF usage in spatial audio has remained quite similar.

Spatial audio leads to reduction of issues with sound location (Letowski, 2016; Kato et al, 2003; Iwaya et al., 2003), but it also creates a more natural feel to an environment when compared to simpler forms of digital audio. The sounds act more as we would expect from a real environment, and, depending on the control scheme, it can lead to a more natural way of staying aware of our surroundings or pinpointing locations.

2.2.2 Audio Searching Techniques

The way humans locate sounds around them is quite complex. As we discussed previously when discussing HRTF, as sound reaches us, it is affected by the shape of our body, head, ears, and environment. Our brain processes a wide array of factors, from time and intensity differences between each ear, the frequency of the sound, etc, and uses our learned experience of our entire lives to approximate the direction and location of the sound source, to within even a few degrees of accuracy (Middlebrooks, 2015; Yost et al, 2013; Carlile et al, 1997).

One of the most important factors when locating sounds is head movement, both allowing the user to move their head freely, and having the sound accurately adjust based on the user's head movements. Moving our head is key for reducing front-back audio issues and getting more accurate locational information (Letowski, 2016; Link and Lehnhardt, 1966; Van Soest, 1929; Majdak et al., 2010; Kato et al, 2003). This is especially true when given a longer audio clip to locate, compared to very short sounds (Iwaya et al., 2003; Lambert, 1974). Fig 2.2 shows an example of the front-back problem, where sounds to the sides of a user are easier to identify, while sounds directly behind or in front of the user sound similar. The yellow nodes signify that the sounds to the front and back of the user sound almost the same, whereas the sounds to the right and left are more easily locatable.

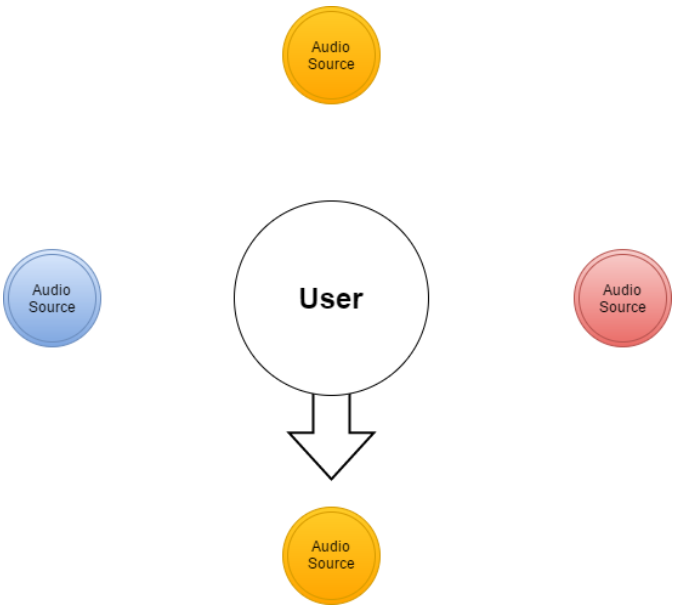


Fig 2.2 Front-Back Problem, (Broderick, 2021)

Different head movements are used to determine different information about the sound source's location. Horizontal movements can help to reduce issues with front-back confusion (Wallach, 1939, 1940; Kato et al, 2003), while tilting the chin or head towards the shoulders can help with determining the elevation of a sound source (Perrett and

Noble, 1997). We see this happening naturally in our everyday lives where people attempt to locate a sound, and it's something we do subconsciously. It can be seen while gaming, where even without spatialised audio, people will move and tilt their head while trying to determine a sound's location. This trained behaviour can extend into VR, where users can accurately move their camera using their head movements. This both leads to a user more naturally being able to work with a virtual environment and can also lead to a greater disconnect from the environment if things such as sound don't react as naturally as they would expect (Kern et al, 2020).

2.2.3 Audio in other fields of study

In the real world, we use hearing to stay aware of our surroundings in 360 degrees and while we are focusing our vision on a specific task. Being able to accurately identify the spatial location of sounds in virtual environments is important for games, VR, and non-entertainment virtual environments (Murphy et al, 2011). Being able to use a headphone and PC setup compared to a full surround sound system is far more achievable for both end users and academic research. In recent times, more openly available options for HRTF libraries or plugins have become available and are even seeing integration into game engines, which are already ideal tools for creating virtual environments. Some recent specific examples will be looked at later and how this industry response to spatial sound can have a hugely positive shared effect on academia in the areas of VR, Virtual Environments, Auditory Display, and game research. There is some research into spatial audio with augmented reality, such as automatically adding spatial audio to panoramic images (Huang et al., 2019), finding targets in augmented reality using spatial audio (Ruminski et al., 2015), or creation of audio-only augmented reality games(Rovithis et al., 2019; Moustakas et al., 2019).

Some older research looked at spatial audio for real world GPS navigation (Holland et al., 2002; Bellotti et al., 2002) or in target detection for aircraft (Nelson et al.,

1998), all of which showed that spatialised audio to provide locational information to the user provided some advantage. This allowed users to make faster and more accurate decisions about their surroundings outside of their vision, while adding very little mental load compared to purely visual target detection and navigation. This highlights how decision making can be added with spatialised audio waypoints, cues, and information, like some virtual environment navigation experiments performed later.

Even early iterations of limited virtual environments considered the ability of spatial audio to aid user interaction and navigation (Wenzel et al., 1990). There is often an advantage to spatial audio when the user needs to focus their vision on something different, allowing for information about their surroundings or some alert to be sent to the user without requiring vision changes. This is useful in navigation, tasks such as driving, and with some forms of feedback, such as correcting motion gestures (Morrison-Smith et al., 2016). Allowing a user to maintain their visual focus while still providing them with information is something seen in many video games as well through sound effects and voice lines, such as Overwatch and Counter Strike which are discussed in detail later.

In general, humans react to audio very quickly, much quicker than they can react to visual stimulus (Jain et al., 2015). Average reaction times to audio stimulus are between 140-160ms, whereas visual stimulus reaction times are between 180-200ms. We can also experience a sound while not being aware of it (Simmons et al., 2016). A sound such as an air conditioner can be constantly humming in the background, but when it stops, we can immediately feel the change. Thinking back, we know it was running, but it wasn't a conscious awareness. This makes awareness of sounds and changes in sound something that is very important to our lives, hence it is something that should be taken advantage of in our virtual lives, both for performance benefits, and for the unconscious feelings of immersion. Study of earcons[1] and auditory display techniques can be combined with spatial audio to create more identifiable audio cues in a 3D virtual space, such as while monitoring network traffic in the background (Debashi et al, 2018; Debashi et al, 2018).

[1] A brief, distinctive sound that represents an event or conveys information

However, the addition of many sound sources in a virtual environment can be overwhelming to a user, similar to real world situations. There has been research in countering this, such as highlighting sounds in the user's vision (Vinnikov et al, 2017), but it is an issue to be aware of in more complex environments.

2.2.4 Spatial Audio Technologies

The use of spatial audio in virtual environments and games is currently going through a resurgence of interest, both academically and in industry (Lawlor et al, 2016; Valve, 2016; Cohen et al, 2015; Rajguru et al, 2020). In the 1990s when dedicated sound cards were more common, some games used spatial audio. However, as sound cards were replaced with software handling sound, spatial audio fell to the wayside. Stereo audio was the norm, especially with many games being developed for a living room setup. In recent years, there has been more focus on spatial audio from the games industry, especially with the growing popularity and accessibility of VR headsets. Companies such as Oculus, Valve, and HTC are creating and releasing tools such as Steam Audio (Valve, 2016) for handling spatialised audio in games, allowing modern developers and academics to work with this type of audio without needing to build a spatial audio system from scratch. This in turn allows people who are not involved with audio technologies to use this technology in new ways, or to add spatial audio to projects where audio isn't the focus. This should further normalise the presence of spatial audio.

Working with spatial audio can be daunting at first, as there is a lot of terminology that can be confusing, as well as a wide variety of ways to implement spatial audio, including tools such as Steam Audio. Users unfamiliar with audio need to consider how the environment can occlude and reflect audio, how the elevation and azimuth may change the ability to locate sound, what the difference between generalised and individualised HRTFs are. There are valuable overviews of the concepts behind spatial audio (Cheng et al, 2001; Kapralos et al, 2008; Beig et al., 2019; Serafin et al., 2018),

and when it comes to the technologies coming from industry, there is often quite extensive documentation regarding implementation.

2.3 Virtual Reality

2.3.1 What is Virtual Reality?

Virtual Reality is defined as “a medium composed of interactive computer simulations that sense the participant’s position and actions and replace or augment the feedback of one or more senses, giving the feeling of being mentally immersed or present in the simulation (a virtual world)” (Sherman et al, 2002). While not inherently dependent on any type of technology, modern forms of VR often focus on the use of a head mounted display and headphones which allow the user's vision and hearing to be replaced by the virtual environment.

VR is a technology that has become highly popular in the last decade for several reasons. Previously, VR was expensive, required large and powerful equipment, and it was difficult to work with for creating experiences, with one of the first commercial VR kits being the EyePhone costing $9000 without a computer to run it (Teitel, 1990). Now, companies such as Oculus and HTC have made VR equipment much more accessible to the masses, with costs under €1000, and with computers capable of running VR experiences becoming more mainstream (Coburn et al., 2017). Game engines such as Unity and Unreal Engine have also made it easier than ever for people to make games at home, even with relatively low technical skills. With an increased amount of people making games and environments, and these engines integrating support for VR equipment, there is a wealth of content available for users, from highly crafted games made by huge companies to simpler experiments and projects made by individuals. This has made people more aware of VR and made it more accessible for researchers to work with, with more than 50 million people in the US using VR once a month (Petrock, 2021).

2 Background, Literature Survey and Meta Analysis

As it has become more available and more people have worked with the technology, the applications of VR have grown widely. While the primary commercialisation of VR has been focused on entertainment, there has also been research which has looked at its application in education and training (Klippel et al, 2020; Freina et al., 2015; Moldoveanu et al., 2017), psychological therapy (Jensen et al, 2004; Lohse et al, 2014), and studies of the growing community of people socialising in VR (Maloney et al, 2020; Maloney et al, 2020). In addition to how virtual environments allow people to enter environments and perform tasks that would be too costly or complicated to do in the real world, VR can enhance our interactions with these environments and allows for finer control and more immersion. It allows people to put themselves in stressful situations while being completely safe, lets people practice skills or visit locations that may be inaccessible to them, and creates a social outlet for people who might not easily be able to enjoy or attend more commonplace social events (Maloney et al, 2020; Lohse et al, 2014). As this enhanced virtual world becomes more normalised and more integrated with our lives, it is important to constantly aim to improve and to create the best environments we can. If this virtual system is being used to educate children (Freina et al, 2015), to train doctors in surgery (Dargar et al, 2015) or first responders, or if it is someone's primary form of socialising with the world, it is important to make it as effective and realistic as possible(Maloney et al, 2020).

Graphical improvements are constantly being driven by the commercial games sector, and since 2010 we have seen full body control (such as Kinect) and headtracking becoming available to everyday people, with more home VR kits such as Oculus Quest 2 at lower price points and working as standalone devices. However, there is a lack of research focused on the importance of properly spatialised audio in these VR environments, as we shall discuss. While the game engines used to make many of these environments provide simple audio spatialising (such as implementation in Unity and Unreal Engine, as well as integration with Oculus), we need to try and keep innovating and using new technologies to improve this, especially as the use of head mounted displays and full body tracking is changing the way we interact with virtual spaces (Seibert et al, 2018). With the growth of the metaverse, more and more people who use VR are

non-gamers, making it more important to be able to interact and control experiences in a natural and intuitive way.

There are some disadvantages to VR of course. Even though it is much more accessible than previously, it still requires both specific equipment and powerful computers or games consoles to be able to properly enjoy or implement. One major issue still plaguing VR is VR sickness *(*Chang et al, 2020). VR sickness usually has symptoms of nausea or headache and can affect between 30% and 80% of users (Rebenitsch et al, 2016). Compared to traditional desktop setups, long term usage of VR can cause feelings of sickness in users (Cao et al., 2020), and this can be aggravated by the way users move in the virtual space (Munafo et al, 2017). While some of this can be reduced through using different in-game movement systems such as teleportation(Tanaka et al., 2004; Moghadam et al., 2020; Russell et al, 2014; Liu et al., 2019) or by keeping the player stationary, it is still something that people need to be aware of when creating or participating in VR experiments and experiences.

2.3.2 Presence in Virtual Reality

Presence has been linked to virtual environments and VR from their inception. Presence is a complex concept, but in terms of VR it can be defined as “the sensation ‘you are there’”, or the “sensation of being in the virtual world”. (Schuemie et al., 2001). Even in the early days of telecommunication research, presence was considered an important aspect of the idea, and that has continued into the field of virtual environments, and further to VR. As people aim to become more immersed in virtual environments and use them for socialising, being able to accurately portray themselves can increase their feeling of presence (Freeman et al., 2020). Just simply using a VR headset rather than a traditional monitor makes users more present (Schuemie et al., 2001). When dealing with some of the downsides of VR such as VR sickness, research tries to balance reducing sickness, through changes to movement such as teleportation, with maintaining immersion and feeling present in the environment (Tanaka et al., 2004).

Users' feelings of presence can be improved by making their experience and environment feel more realistic. Interacting with the environment in the most natural way and moving their body accurately increases this feeling (Freeman et al, 2020), while things that differ greatly from what they expect can remind them it is a virtual experience. Being able to interact with the environment in a predictable way, as well as high quality and realistic sensory information, are also key factors for presence (Schuemie et al, 2001). These factors all point to the importance of using high quality and realistic audio in these virtual spaces. Having audio that isn't spatialised at all, or that is only spatialised in a simple way, can cause users to feel less present by having the audio react in a way that isn't natural or expected. If users need to use different techniques to the real world to accomplish tasks in the virtual environment, it can remind them they are in a virtual space. Users being able to interact with environments in a natural way, and having the environment respond as they would expect, creates greater immersion and feeling of presence, which can increase their enjoyment (Freeman et al, 2020; Schuemie et al, 2001).

2.3.3 Head Tracking

One key difference when using VR (as opposed to traditional monitor-based display) is the use of a head mounted display unit (HMD). While the primary purpose is to provide the visual part of the experience for the user, most modern HMDs provide head tracking capability for use in virtual experiences and games. This greatly alters the way users control the virtual camera and engage with their environment.

Users can feel more immersed and present in a VR situation compared to a traditional computer setting. There is evidence that the way the camera is controlled through head tracking can be the greatest influence on this, with the movement and interaction controls not having as significant an impact (Seibert et al, 2018). Users generally tend to prefer head tracking systems compared to traditional control schemes

despite performance reduction (Martel et al., 2015; Ashdown et al., 2005). Of course, the task being performed can impact the enjoyability of and performance with head tracking. A significant amount of research that examines head tracking in games looks at first person shooters, which can be affected positively with the addition of head tracking, while other genres such as racing can be negatively impacted (Kulshreshth et al., 2013). Performance increases are more likely to be affected by a player's experience with a game rather than the control scheme used, but as discussed, people often prefer the more intuitive and natural control scheme without a performance increase, and even with a performance decrease (Kulshreshth et al., 2013; Martel et al., 2015; Ashdown et al., 2005).

Outside of VR, head tracking also has uses in computer control. There has been research examining how it can be a benefit to people with disabilities (Sahadat et al, 2018; Honye et al., 2012). While traditional control schemes may be difficult or impossible to control for certain types of disabilities, head tracking is one area that can potentially open up accessibility for these people. It can reduce future ailments by reducing risks of repetitive strain injury compared to traditional mouse controls (Dernoncourt, 2012). For people with disabilities, it can be used with other technologies such as tongue controls to provide a varied control scheme for users (Sahadat et al. 2018; Honye et al., 2012). This opens up access to people who may have difficulty with existing controls for technology. In areas such as image control, head tracking can be used instead of mouse control, avoiding repetitive strain injury, aiding workers with disabilities, and avoiding contamination (Alcaraz-Mateos et al., 2020).

However, to the best of our knowledge after our review of the literature, there hasn’t been any examination found of how intuitive and natural control schemes can affect the importance of environmental factors in VR, such as sound realism.

2.4 Spatial Audio and Virtual Reality

2.4.1 Spatial Audio and Navigation

It has already been shown that adding audio cues, such as sounds attached to important objects, to virtual environments aids users' navigational ability. There have been several studies of using auditory navigation cues at specific goals or locations to aid users in being able to move through the environment (Grohn et al, 2005; Walker et al, 2006). Grohn and Lokki (2005) measured improvements in users finding objects within the environment using visual, audio, and audio visual cues (referred to as gates) to represent the target goals. Users were in a 3D virtual space and were asked to locate as many gates as possible within a time limit. While audio on its own was the least successful method of locating goal objects, combining visual and audio led to users finding many more objects within a set time constraint. The combined audio/visual method found each new gate more quickly (median of 5.7 seconds compared to visual only median of 7.5 seconds) and found more gates in the time limit (between 24 and 31, compared to 16 to 24 for visual only). It was also found that users would use audio cues first to roughly locate an object before using visual stimuli for the final approach.

Walker and Lindsay (2006) also examined using auditory cues (referred to as waypoints) for navigation of an environment, and more specifically, how users are able to follow a path of waypoints with primarily auditory stimuli. It was found that users were well able to find and follow these waypoints, with the only common issue being the overshooting of waypoints upon getting too close to them. People use sound as a way to roughly approximate a destination, especially over longer distances or when visual cues may not be available. It was found that as a person got closer to their target, they switched to visual searching to locate the final target. While audio on its own isn't always an effective method of locating destinations or navigating, it greatly enhances user performance when added to a visual element (Grohn et al, 2005; Walker et al, 2006).

2.4.2 Spatial Audio and Sonification

As well as spatial audio allowing for users to be more aware of changes around them, it is also an important aid to sonification and auditory display. Sonification is “the technique of rendering sound in response to data and interactions” (Hermann et al, 2011, pg. 1). It works to represent some data as sound in a way that is algorithmic and reproducible. This can then aid users with pattern recognition or monitoring. Related to this is the concept of auditory icons and earcons. Auditory icons are “caricatures of naturally occurring sounds such as bumps, scrapes, or even files hitting mailboxes” (Gaver, 1986, pg. 167). They rely on a user's mental connections between a sound effect and an event. Earcons are “abstract, synthetic tones that can be used in structured combinations to create auditory messages” (Hermann et al, 2011, pg. 339), with no direct relation between the sound and the information it represents. Games already use these concepts to give users information within virtual environments, both spatially and not. Virtual environment research can benefit from more research into this, as up to now, most focus has been on graphical quality and user interactions with the environment (Hamidi et al, 2009).

Even within the areas of auditory display research, VR is a topic looked at more theoretically rather than with substantial empirical research, with some work looking into evaluating sonification in VR (Kuppanda et al, 2015), and recent auditory display work in the area of games looks at augmented reality and VR (Jäger et al, 2017; Wu et al, 2017; Heller et al, 2017). Ideally, with VR technology becoming much cheaper and more accessible to the average person, more research work not focused specifically on the VR research space will use VR in interesting and new ways, such as studies of socialisation (Maloney et al, 2020). In fact, it was shown by Schoeffler, (Schoeffler et al, 2015) that performing audio-based experiments in VR and in the real world have negligible differences in user response and rating if the quality of the experiment is high enough. This is important to acknowledge as any systems for training and education that focus strongly on audio must be able to provide the same or similar benefit to a real-world system.

Looking specifically at the work of Kuppanda et al (2015), a system was built to evaluate sonification techniques in VR, using Unity3D and Oculus Rift. The goal was to evaluate different types of sonification to compare suitability for use in an auditory display, with the experimental values being different types of sounds, such as white noise, pure tones, or plucked sounds somewhere in-between. Such a system could be used in a car for driving assistance. Users were placed in a virtual environment built in Unity3D and controlled their camera using an Oculus Rift head mounted display unit. Users were asked to centre the sound source within the Field of View (FOV) of the camera and then activate a mechanical clicker. After clicking, the system would reset, and the task would be repeated with a different sound source. Subjects performed two versions of the test: a sighted version where they could see the objects within the environment, and a blind version where they could not see the objects. The evaluation measures recorded were user response time, subjective questions on whether or not the sound was pleasant, and whether or not the mapping of location to sound was intuitive. It was found that a sighted test with both visual and audio elements had the fastest reaction time for most sonification types, as well having a smaller range of minimum and maximum times compared to purely auditory tests. There was also evidence that some of the sound types were rated subjectively higher than objectively better performing sounds and vice versa, showing that users preferred a pleasant sound and would rather use it over a better performing sound. This is similar to what we've seen regarding user preference for VR control schemes (Kulshreshth et al., 2013; Martel et al., 2015; Ashdown et al., 2005). The most negatively reviewed sound was noise with jitter, with noise generally being less pleasant compared to pluck sounds. Overall, the application of jitter, a special effect created by slight oscillation of the sound source, was found to be more pleasant to the users.

This experiment shows similarities with the work conducted in the current thesis, though some aspects are different due to a different emphasis. As the aim of Kuppanda's work was evaluation for usage in an auditory display, some factors are less important when compared to use within a fully 3D environment. Within Kuppanda's experiment, all sound source locations were at the same elevation as the user and were all contained

within a 180-degree stereo sound field. There was no need to focus on the effect that elevation can have on sound locating and using a 180 degree sound field reduces the effect of front-back problems. Since the goal of an auditory display system is usually to bring a user's attention to something specific very quickly, response time is the most important factor to consider, as there will usually be a visual element involved as well. There is also a smaller number of subjective values analysed, as intuitiveness and user preference are key factors.

Comparatively, while looking to evaluate something similar in a 3D space, certain aspects greatly increase in importance. Sounds can come from any direction around a user; from in front of, behind, above, or below, which can be the case in entertainment games, training simulations, or collaborative environments. It needs to be considered how these factors can affect a user's ability to locate a source. There are also other important factors beyond reaction time when considering a virtual environment. While speed to react is useful in many cases, it is important to gather information on accuracy, search methods, and related measures. This is particularly with limited visual elements, people are going to search and react differently depending on if they are relying on only auditory information, or a mixture of audio-visual. It can also be useful to expand on the subjective questions asked of users to cover as many areas as possible. Within Kuppanda's study the users were only asked if the sound was pleasant and if the mapping of location to sound was intuitive. It could also be worth asking about perceived difficulty, both between sighted and blind tests performed, and between the different sounds. Perceived difficulty can be different to measured performance, and users may feel a higher workload when doing something that feels more difficult. While the mapping is the same, there could still be a perceived difference to the users, which can still be useful information.

Many of these considerations are different as the goal of what is being measured is greatly changed. VR evaluation of sound is something that has not been widely researched previously, so we need to look at examples that align with the work being examined in this thesis.

2.4.3 Spatial Audio in Other Task Management

Both in terms of navigation and in more complex areas such as awareness of surroundings and alerts to changes in the environment, sound is key for aiding user decision making. Beyond examples of navigational decision making using auditory signals, there is also evidence of sonification aiding in other types of more complex decision making. Whether it be monitoring network traffic (Worrall, 2015) or keeping track of multiple task prioritization (Donmez et al, 2009; Brock et al, 2002), auditory display enhances the user experience as a background process. Brock noted a faster response time for audio compared to no audio in dual task experiments (improvement of 245ms, $p < .05$), and also a reduction in head movements in two window setups (343.2 movements with sound compared to 461.2 for no sound). Donmez found that sonification alerts compared to traditional discrete auditory alerts caused a 1.90 second faster reaction time ($p=.006$), with 85% of users reporting that the audio worked well in conjunction with standard visual displays. 66% also stated that traditional discrete audio alerts were harder to identify against background noise. Rather than distracting from the task at hand, users are simply listening for changes in their auditory background to indicate their attention is needed elsewhere. In situations where a purely visual display would hinder task completion and decision making, users can instead focus on the task at hand. Users were then not only able to handle their tasks to the best of their ability but were better able to react to changing situations within multiple tasks simultaneously.

Of course, games use audio for all these tasks. Whether it be locating where incoming gunfire is coming from, recognising the difference in types of footsteps, being made aware of important events taking place off-screen and warning the player when their health is low, games provide players with large amounts of information through audio. This is why sound design is so important in games: to not drown the user in information, but instead teach them to automatically react to key sound cues (Westerberg et al, 2015). Rather than focusing on the types of sound information being provided to the player, it is important to examine why the sound *source* is relevant to the player (Blizzard, 2016; Lawlor et al, 2016). This informs how the sound should be created and

changed, such as increasing or decreasing the volume, adding a looming effect, or even removing the sound entirely.

The type of sound we hear can also change how quickly and accurately we can locate it or become aware of it. A “looming” sound, one growing louder, replicates that of a sound source getting closer to us. It was shown by McCarthy (McCarthy et al, 2017) that users could locate a “looming” noise much more quickly and accurately than a receding sound source. This can be used to highlight specific types of sounds in a game or environment or make them easier to locate for a user.

2.4.4 Immersion and Presence

Beyond navigation and decision making, audio in virtual environments can be a key component of increasing user immersion and understanding of their environment, something that becomes even more key when the user is experiencing the virtual environment through VR. It has been shown that having higher quality audio can lead to a user perceiving a higher overall quality of a visual object (Bonneel et al, 2010). The quality of the information in the environment can also outweigh the quality of the technology, meaning a focus on high quality audio and information could be the most important factor in some cases (Baños et al, 2008). Being able to improve your audio quality is generally computationally cheaper compared to improving your visual quality, so this can be quite a positive trade-off. Audio can also potentially create a sense of pseudo-haptic feedback when interacting with virtual objects and environments (Collins et al, 2019). Haptic can be used as an umbrella term for anything relating to touch (Oakley et al, 2000), though there are a variety of distinct categories within it. Pseudo-haptic specifically is about making a user feel a haptic sensation using other sensory information. Collins and Kapralos (2019) look at several examples of how audio can change users' perception of an object’s material and weight, which can especially help in VR where you might have to touch and move an object but may not have a true haptic feedback device. There is a lot more research of visuo-haptic feedback for virtual

environments (Ebrahimi et al, 2016), but auditory feedback to provide a pseudo-haptic response can also provide a lot of information to a user.

We use all of our senses to engage with and perceive our environment, so by improving the quality of information provided through any one sense, we can potentially improve the user's overall perception and depth of understanding. "Increasing the modalities of sensory input in a virtual environment can increase both the sense of presence and memory for objects in the environment" (Dinh et al., 1999). Dinh showed that by adding more sensory information, including sound, to a user's engagement with a virtual environment led to users remembering more about the environment, the objects in the environment, and feeling more present in the environment, and that increases in visual fidelity did not increase these factors in a similar way. This is also examined by Kapralos et. al (Kapralos et al., 2017). Having these modalities change with your movements and having a certain amount of control over the environment is also key to increasing users' perception of presence (Sheridan et al, 1992; Held et al, 1992). Control of your virtual camera gives users an ability to move and look around their environment but having audio that breaks your expected understanding compared to the real world can cause a break in immersion for a user. Having properly spatialised audio that will react as closely to the real world as possible will stop the user having this break in immersion, allowing them to focus on their environment and the task at hand.

Increasing presence and immersion is not only useful for entertainment, but also for learning and training in virtual environments and VR (Dargar et al, 2015; Dragos et al, 2017). An issue with presence is that it is a highly individual experience, meaning one virtual environment could elicit a completely different level of presence from two different users (Moghimi et al, 2016). Because of this, there is evidence supporting the positive impact of creating an environment that can increase a user's sense of presence in more generalised ways, such as improving sensory information across touch, sight, and sound, without overcommitting resources attempting to reach complete immersion. VR also increases a user's sense of immersion and presence compared to traditional virtual environments. People tend to prefer head mounted VR setups as they feel more realistic

and natural, as seen by Shu (Shu et al, 2019). While PC monitors are still the most common setup for individuals, people will often choose a VR if possible.

2.5 Applications of Spatial Audio in Virtual Spaces

2.5.1 Serious Games and Game Engines

Video games have quickly grown to be an important form of entertainment media, with the Global Gaming Market valued at $173.70 billion USD (Mordor Intelligence, 2021) and 67% of Americans playing games regularly as of 2021 (ESA, 2021). People of all ages are playing games, and this leads to a unique opportunity. Compared to many other entertainment media, games are about involvement, with the player controlling the game and interacting with the environment created. This direct connection between the player and the medium has led to the use of games for education and training purposes, hence the discipline of “serious games”.

The definition of a game engine can be broadly defined as “that collection of modules of simulation code that do not directly specify the game’s behaviour (game logic) or game’s environment (level data)” (Lewis et al, 2002). Game engines provide a platform upon which games and applications can be built, providing basic tools and functionality needed to make modern games. By using a game engine for the basic setup of a system, more focus can be put on the actual functionality being created, rather than spending time recreating a rendering or physics engine.

The Unity3D game engine has been the engine of choice for the research work described later in this thesis, the reasons for which will be covered now. Most game engines provide a similar baseline functionality, with systems for rendering/visualization, sound, physics, and scripting for functionality. Unity has significant flexibility with its scripting, as it allows for use of Unity versions of C#, JavaScript, and Boo. Scripts of different languages can be used in the same project and on the same game objects with

little conflict. This implies that there is a smaller learning curve and gives more room for coders to work in their preferred language. Unity is also highly portable, easily able to deploy on platforms such as PC, OS X, Linux, Mobile, Web, and a variety of game consoles. This portability means programs can be built to work on a variety of systems so that the maximum number of users can be supported. Unity has a user base of over 2 billion users and 1.5 million monthly active creators (Unity, 2020), with people constantly writing tutorials, creating assets, and answering questions. Additionally, projects and companies have used Unity for serious games in the past, including virtual environments (Wang et al, 2010), urban planning (Indraprastha et al, 2009) and disaster simulation (Sharma et al, 2015).

Unity3D is also ideal for projects focusing on VR or HRTF work. Oculus and Unity worked together very closely to make the workflow for implementing Oculus features in a Unity project as simple and painless as possible. There is a wealth of tutorials and resources available, allowing less technology or gaming-oriented researchers to use Unity for VR projects and virtual environment creation. Unity has also added support for non-programmers with the Timeline and Cinemachine assets now available in Unity (Unity, 2017). The primary objective of these assets is to make it easier for artists and creatives to work with Unity when creating movies, clips, installations, or simple games, by allowing animations and camera work to be controlled and created in a manner like video editing. It is likely that this will make it easier for non-programmers to work with Unity in a research, academic, or individual project. This is especially applicable in areas similar to auditory display and sonification, where the field is not only audio engineers and programmers, but also artists, musicians, scientists, etc.

Additionally, Unity’s later versions have both Microsoft HRTF and Oculus HRTF directly integrated into the audio spatialisation options of its editor, which makes it even easier for games or research projects to use this higher quality audio system (Microsoft, 2018; Unity, 2020). This integration not only saves time with setup of resources and plugins, but also guarantees a quality of integration with Unity3D. Unity 3D is also one of the supported platforms for Steam Audio.

There are some limitations to game engines. For example, depending on the engine used, the source code may or may not be available (Pranckevičius, 2018). Not being able to access the entirety of the engine can make it harder to have the system work exactly as needed for certain projects, or it can make it more difficult to figure out why bugs or interaction of objects happens a certain way. Related to this is that game engines usually have a specific language that they work with for scripting new functions into the game. Depending on the coder, this can require learning of a new language that they may not have experience with. Many game engines used to use custom versions of languages which meant even with experience in a language, there may be changes or additions that need to be learned. The same issue can be found in any new project though, as many must eventually make use of new technology. However, in recent times, more game engines have moved away from these custom scripting languages (Unity moved from UnityScript to C#, Unreal Engine uses C++).

There can also be the issue of licensing. Game engines can charge for a license to use their software for game and program creation, or to unlock the full features available. This cost can also increase with the amount of people working on a project. Those that don't have a licensing fee usually have a charge when the game sells over a certain amount (Unity, 2021). Some engines, such as Unreal, remove this cost for film projects or consulting/contracting projects related to visualisation, simulation, and architecture (Unreal, 2015). This shows that the companies behind these engines are seeing the increasing use of serious games and are trying to support these new ideas.

2.5.2 Serious Games for Training and Education

Serious games are games which are not focused on entertainment, and instead are aimed towards training, teaching or other non-leisure concepts (Abt, 1970). There are a huge number of advantages to using games in this way:

- User interaction: The ability for learners to engage with a virtual environment and see some reaction to their input.
- Cost effectiveness: Compared to some training and educational scenarios that require either expensive components or paid participants, there is usually a once-off cost for development and equipment
- Visualisations: Being able to visualise information or objects in great detail, such as for graphing of results, display of biological materials and processes
- More in depth simulations for complicated scenarios, such as disaster training, medical emergency training, compared to traditional training methods.
- Distributed training: Being able to train groups of people without having them all in one place.

By having users directly involved and interacting with an exercise, it is much easier to learn and for information to be internalised. Rather than simply having to read about a potential subject, the user can instead work with a virtual simulation, adding a level of realism and practicality to the topic being learned. For some types of learners, it is much easier to understand a topic when it has a level of tangibility compared to just text, and this can be provided by things such as serious games and VR (Bell et al, 1995).

Triage Trainer is an example of a serious game training application (Knight et al, 2010) (Fig 2.3). Knight was researching the advantages of using serious games as a training method compared to more traditional means. Traditionally triage is taught in workshops with instructors and live actors, which require significant planning and resources, especially on larger scales. They are also limited in their realism, due to having to use mannequins or actors to simulate situations involving triage.

Fig 2.3 Triage Trainer screenshot, (Knight et al, 2010)

In Triage Trainer, the user moves between virtual patients, judging their conditions using the triage methods taught to them, and marking the patient's urgency before moving on. The system then rates the user based on the speed and accuracy of their judgement, giving them feedback on where mistakes were made. In a live test, those trained using Triage Trainer performed noticeably better in their assessments of the patients compared to those trained using traditional methods. When it came to accuracy of tagging, Triage Trainer users scored higher perfect scores than the traditional card sorting training group (34 compared to 24, p=.02). They also had a higher accuracy of performing the correct medical steps, with 13 receiving a perfect score compared to 3 in the card sort group (p=.0196).

Another example of serious games as training applications is in Sharma's creation of an immersive subway evacuation program (Sharma et al, 2014). This system was made to show the possibilities of a collaborative immersive VR evacuation drill and the advantages this can have over real-world drills. The users would traverse a virtual

replication of a subway train and stations while using an Oculus Rift headset to heighten user immersion. User and computer-controlled agents would be populated in the environment and would have to evacuate the subway station due to a simulated disaster. This system allows for mass training of staff in environments that cannot be available for training regularly and can create scenarios not feasible or safe in a real world training system. It is unlikely that systems such as this will replace real-world drills, as they can currently not offer perfect immersion. However, they can be a useful supplement to traditional training methods.

Persson created a game-based system that simulates a Swedish city and allows for the replication of disasters for decision-making training (Persson et al, 2007). Compared to previous training methods which relied on either archived images/video or simply didn't have visual information, user comments were that this system was typically preferable and provided clearer and more accurate information. While most users don't directly interact with the game environment, the information it creates and how it reacts to the actions of the training personnel are of great value. It is easy for the system to create varied and new simulations, and it was used to test new equipment and technology that was not yet prepared for field testing.

However, with all these examples of serious games, audio was not the primary focus of the research or the tools. Many systems such as these will implement at least simple panning audio as most games have, but rarely has the focus been on how audio can greatly improve a user's awareness of their surroundings or improve their feeling of immersion and presence in the environment. This lack of audio research in serious games is examined in our meta-analysis later in this chapter. Recent research has shown that higher fidelity serious games don't lead to enhanced learning when compared to lower fidelity games (Norman et al, 2021), but users' experience with modern games has led to an expectation of higher fidelity (Visschedijk et al, 2012). Usually when there is a focus on audio in serious games, it is aimed at sonification and auditory learning (Radecki et al, 2020) rather than spatial audio. However, using audio cues can improve people's perception of visual fidelity in virtual environments (Kapralos et al, 2017), as well as

creating a more natural feeling environment. As serious games continue to become more common, especially with the availability of VR and with our current world moving more towards online tools, we need research to look further at these techniques that can improve the quality and effect of these tools.

2.5.3 Examples of Spatial Audio in modern games

From personal experience and conversation with both friends and other players online, one of the most frustrating experiences when playing competitive first-person games is the panic of knowing you're in danger, but not being able to find where the threat is coming from. As the player swings their camera from right to left, trying to locate the source of danger, they can already find themselves defeated. That crucial time spent trying to better determine the location of the threat can be the difference between winning and losing. Games support stereo audio by default, and while knowing if the enemy is to the left or right is a huge help, users may not have an accurate idea of where the opponent actually is. There are also issues tied to the *front-back problem* or knowing if the sound source is on similar elevation or above, as well as the distance of the target. However, some modern game companies are starting to see the value of high-quality 3D audio and how it can benefit their players, such as the ones we will now discuss.

2.5.3.1 Overwatch

Overwatch, a competitive, team-based first-person shooter released in 2016, was advertised as being the first game to support Dolby Atmos over headphones (Blizzard, 2016). Dolby Atmos for headphones aims to create a richer sound experience by creating a virtual sphere of sound, rather than using a set of predefined channels. This allows sounds to come from any position around the user. The goal of this was to allow users to more accurately locate enemies around them, since many characters can appear from above or below, making fast and accurate sound location an important aspect of the

game. Blizzard spoke at the Game Developers Conference in 2016 (Lawlor et al, 2016), discussing specifically the sound design for the game, looking at both the importance of the type of sounds (dialog, sound effects), sound layering, and 3D sound.

One major aspect which Blizzard focused on was occlusion of sound. Sound bounces around environments, it can be muffled by distance and objects it passes. In Overwatch, the path between the player and the source of a sound is calculated, both as a raycast and as a traversable path. The difference in length is used to calculate an occlusion value, which then affects how the user hears the sounds. This combined with the 3D positional sound provided by Dolby Atmos allows users to tell not only the direction of the sound, but how far away in the environment. It also allows players to be more tactical, knowing their sounds are muffled by objects and distance. Overwatch has been very open about their sound design process through presentations, which is uncommon for major games, and this may push more companies to improve their audio design.

2.5.3.2 CSGO

Counter Strike: Global Offensive (CSGO) by Valve is another competitive first person shooter, though it focuses more on realism and being tactical than something as fast paced as Overwatch. In 2016, Valve released a change to the game which allowed spatial audio to be used by headphone users (Valve, 2016). This was implemented using generalised HRTFs which would aid users in sound location by azimuth and elevation. This change added all the benefits of HRTF audio for players, allowing them to better locate sounds around them, especially since they would now be able to tell the elevation of a sound source, where previously sound was identical regardless of elevation. In 2017, Valve also added sound occlusion to the game, causing sound to be distorted and changed as it passed through or around objects in the environment. The goal of all these changes was to create a much more realistic experience, and ideally help players be more aware of their surroundings.

However, many fans did not react positively to these changes. Some complained of a drop in sound quality, or that sound effects sounded odd, especially when moving while the sound effect played (Malimbo, 2017). Since the release of the first HRTF patch, Valve has been actively taking user feedback and attempting to make positive changes to their sound system, and the community has started to have more positive feedback regarding these changes, such as improving the sound system for specific sounds. While early reactions create worry for the quality of HRTF audio in virtual environments, there is also hope. This is the largest scale testing of HRTF audio available, on a scale researchers would dream of. CSGO had an average player amount of 354,270.3 for March 2018 (SteamCharts, 2018). If even a small percentage of that player base use the HRTF audio and give feedback, Valve is going to have a valuable amount of information available to them to help improve the sound systems they have created, as well as having their own significant financial backing. And with this information being funnelled into Steam Audio, everyone should hopefully be able to benefit from these advances.

Audio-only games, games which use purely audio without a visual component, are quite rare at the current time and are mostly the area of indie titles and artistic projects. While there is a lot possible in these games, for most people the experience is quite difficult to navigate, especially as it is difficult to have tactile feedback for navigation of a completely virtual environment. There has been some research of these styles of games (Jäger et al, 2017), but it is currently quite limited, and the genre needs more consideration, though the possibilities are hopeful, especially for visually impaired players.

2.5.4 Collaborative Virtual Environments and Decision Support Systems

2.5.4.1 Decision Support Systems

Decision support systems are defined as “interactive computer-based systems, which help decision makers utilize data and models to solve unstructured problems”, which is expanded to include “any system that makes some contribution to decision making” (Sprague, 1980). Decision support systems can either provide information in a way to users that simplifies and improves the decision-making process, or they can supplement this with generated suggestions of what decisions can be made, similar to the St Petersburg Flood System (Ivanov et al, 2012), an example of which is seen in Fig 2.4). Whether for training, immediate emergency response or long-term planning, these support systems must provide information clearly to users of varying technical skill and ability, allow for networked connection to external datastores or systems and take input from the users.

Collaborative Decision Support is an area, where multiple users can view a situation’s data, potentially in different formats, and discuss how best to proceed. People of various skill levels and from multiple knowledge backgrounds can all work with and collaborate in the same system, making it much easier to demonstrate project ideas to stakeholders or to have a team of varying individuals come to a group decision together. By having a collaborative system that users can access over the internet, experts from all over the world can communicate and analyse these virtual scenarios, providing more information to whoever is in charge of decision making, as well as displaying it in such a way that they do not need to have an expert’s knowledge of the area to understand the problem, or the solutions suggested.

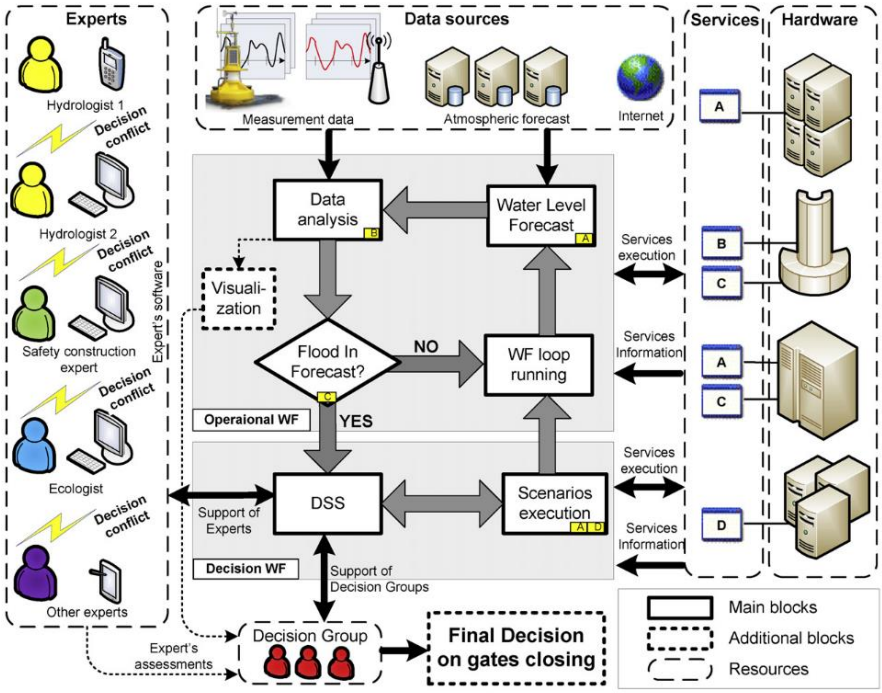


Fig 2.4 Diagram of the CDS-based flood management procedure, (Ivanov et al 2012)

Audio in decision support can allow users to maintain awareness of new or urgent information, while focusing their visual attention elsewhere. We saw earlier how this can be useful for pilots, but it can also be useful for Unmanned Aerial Vehicle (UAV) control (Nehme et al, 2006) in a variety of capacities. In these scenarios, the spatialisation of audio is highly important as the location of the information is often important as well. Machines can use a combination of audio and vibration for fault detection (Henriquez et al, 2014), aiding passive awareness of when decisions on replacement or repair need to be made. There is additionally work looking at sonifying network traffic for network monitoring and searching for botnet activity (Debashi et al, 2018; Debashi et al, 2018; Ballora et al, 2011). This can not only improve user's performance in decision making by leading to less false positives, but it also allows users to remain aware of this information while focusing elsewhere.

2.5.4.2 Collaborative Virtual Environments

A collaborative virtual environment (CVE) is “a computer-based, distributed, virtual space or set of places. In such places, people can meet and interact with others, with agents or with virtual objects” (Churchill et al, 2001). Whether these systems are graphically represented in 2D or 3D, or are text-based, they provide an environment for interaction between multiple users and possibly some data or functions. They can be run from a computer, a mobile device, or VR equipment. CVEs aim to fulfil 4 main goals (Snowdon et al, 2001):

- Shared context: Seeing the data and environment in the same way as the other users.
- Awareness of others: Having a way to tell who else is in the CVE, where they are, and what they are doing.
- Negotiation and communication: Being able to communicate with the other users to work together on changes, discussions, and creations.
- Flexible and multiple viewpoints: Having multiple ways to view the CVE and being able to move that viewpoint as needed.

Shared context is about shared understandings. By allowing users to have access to historical actions, information, and environments, they can better understand each other and the current context. This is also highlighted by allowing users to direct the attention of others, either directly or indirectly. This can include passing viewpoints to other users, highlighting changes being made to all users or allowing users to leave annotations viewable by other users. The type of context sharing depends on the environment (3D, text based) and if the collaboration is asynchronous or not (Annotations and messages are even more important when users are interacting with a scene at different times to each other).

Awareness of others has some similarities to shared context, but there are some key differences when directly comparing the two. Where shared context focuses on

understanding others' work and actively drawing attention to it, awareness of others is focused on peripheral awareness. This allows users to work on separate tasks in the collaborative environment, while remaining aware of what other users are working on in the environment. Both things are key facets of CVEs, but they require different approaches, as they share different goals.

2.5.4.3 Examples of Audio in Decision Support Systems and CVEs

Communication is arguably one of the most important parts of collaboration, as it is near impossible for users to work together on a single project without some level of communication (Bjørn et al, 2014). CVEs must incorporate some way for users in the environment to communicate, whether it be by text, video and/or audio. By using a more human shaped avatar rather than simpler or faceless graphics, basic "emotes" can be added to work off of text prompts or buttons (Bishop et al, 2010). The avatar not only shows where the user is and where they are looking, the avatar points in front of them if the user says "look", or nods when the user says "yes". While these aren't nearly at the depth humans use non-verbal cues for, it is a step towards a more intricate and intimate communication style.

Earlier research into VR collaborative environments recognised the importance of audio (Benford et al, 1994). Audio in VR collaborative environments can help greatly with navigation (Khalid et al, 2019), awareness of other users (Benford et al, 1994), and are necessary for high quality communication (Salinas et al, 2005). A combination of spatial audio, sound cues, and haptic response can aid visually impaired users' interactions with virtual CVEs, and with sighted users in those environments (Moll et al, 2010). Being able to create a meaningful environment and allowing users to communicate and interact in a way that is natural to them increases their feelings of presence and can lead to them more quickly adapting to the system, similar to the other work on presence in VR discussed earlier.

Bishop constructed a system called SIEVE, based on the Torque Game Engine (Bishop et al, 2010). The construction of wind farms effects wildlife, and the aesthetics of the area. This addressed the problem by aiding in the decision-making process for where wind farms should be built, by allowing residents of an area to see a full visual and audio representation of possible wind farms in their area. Compared to animations or still images which are usually designed to give the best possible impression, SIEVE allowed users to view the area from any position or angle in a real-world accurate model. Users could also adjust placement of the wind farms throughout the area and place down shielding trees to help planners find the best possible layout that suited all residents. Being able to make these changes meant that both planners and residents could work together. This system allowed for fast turnaround on feedback between planners and residents and involved residents with the process more directly. Unfortunately, the system was a proof of concept and hadn't received any public feedback as of publication.

CVEs and DSSs can be enhanced with the addition of spatial audio (Sahai et al, 2016). Being able to apply spatial audio to the situation can give a more accurate and convincing example of real-world effects, similar to the airplane noises generated by Sahai. This could be expanded for council planning services, by showing residents the impact of construction, roads in an area or school and playground noises. Being able to give people a realistic visual and auditory environment replicating their own can make them more open to projects such as these.

Nehme (2016) reviewed literature on audio decision support related to supervisory control of unmanned vehicles was performed. While the research covered is from a time before many modern technologies, it still shows some useful examples of how spatial audio can improve many aspects of unmanned vehicle control, as well as the tasks that surround this job. The general themes examined were spatial audio and target acquisition, 3D audio displays to enhance alerts, workload, real vs virtual sound sources, speech communications, and adaptive interfaces. Reflecting on these themes, aspects of the information provided remain relevant in modern research, and in applications outside of the military. Regarding target acquisition, it was found that the addition of

spatial audio can improve search times and reaction times compared to purely visual stimuli (100-200ms reduction, Perrott et al 1996; 15% improvement, Bronkhorst et al, 1996). This is particularly useful when additional visual information cannot be provided due to limited visual real estate, and it was also found that there were statistically significant results in subjective measurement of user enjoyment of spatial audio. With regards to alerts and warnings, it was found an audio element could improve response times (2.63s compared to 2.13s for audio, Begault et al, 1996; 4.7s to 2.5s for audio, Begault et al, 1993) and avoid change blindness when moving between multiple visual elements. It was also shown that overall workload was not increased with the addition of audio elements, while also increasing user performance. Finally, it was seen that equivalent accuracy was achievable with both real sound sources and virtually sounds created using HRTFs. All these tests were using sound in addition to an existing visual element.

2.5.4.4 Previous Related Research Work

In a study related to the research focus of this thesis, the author examined the usage of game engines as decision support systems and collaborative virtual environments (Broderick et al, 2016). An early concept was for a 3D model of a marine environment which could be populated with ocean data, allowing users to examine and work with the data collaboratively. The environment would need to visualise both a real world 3D environment and data which would be rendered both as visual objects and as text. It would need to be accessible by groups of users from various locations and should support communication between all of these users, ideally through voice chat and text. Users should easily be able to navigate the environment and available data, move through timelines of data visualisation, and show other users specific visualisations and viewpoints that are relevant to current discussion. Annotations should be addable to datasets, either for specific data points or to larger sets of data. All of this should persist through uses of the system, allowing further groups of users to be shown previously discussed data or bring in additional data. More specific tools and functionality would be

added as more datasets were found and user input was given. Using a game engine as a basis for this system greatly simplified much of the basic functionality, such as the rendering system, user interactivity and networking.

The project prototype was built in Unity 3D using LIDAR and bathymetry data of Galway Bay, as well as using ocean current data gathered from a variety of points throughout the bay. Fig 2.6 shows an example of the LIDAR data after being converted into a piece of terrain in Unity3D. While the early version used 2D surface data (seen in Fig 2.5), the goal was to have layers of data from the ocean surface to the floor. However, while examining other work similar and general implementation of large amounts of data in a 3D virtual space, there was more examination done on how an audio element could make working within this space simpler and more efficient. The research focus shifted away from this initial concept and instead examined the role of audio within these sorts of systems - since we had identified this as a fundamental component worthy of more in-depth study.

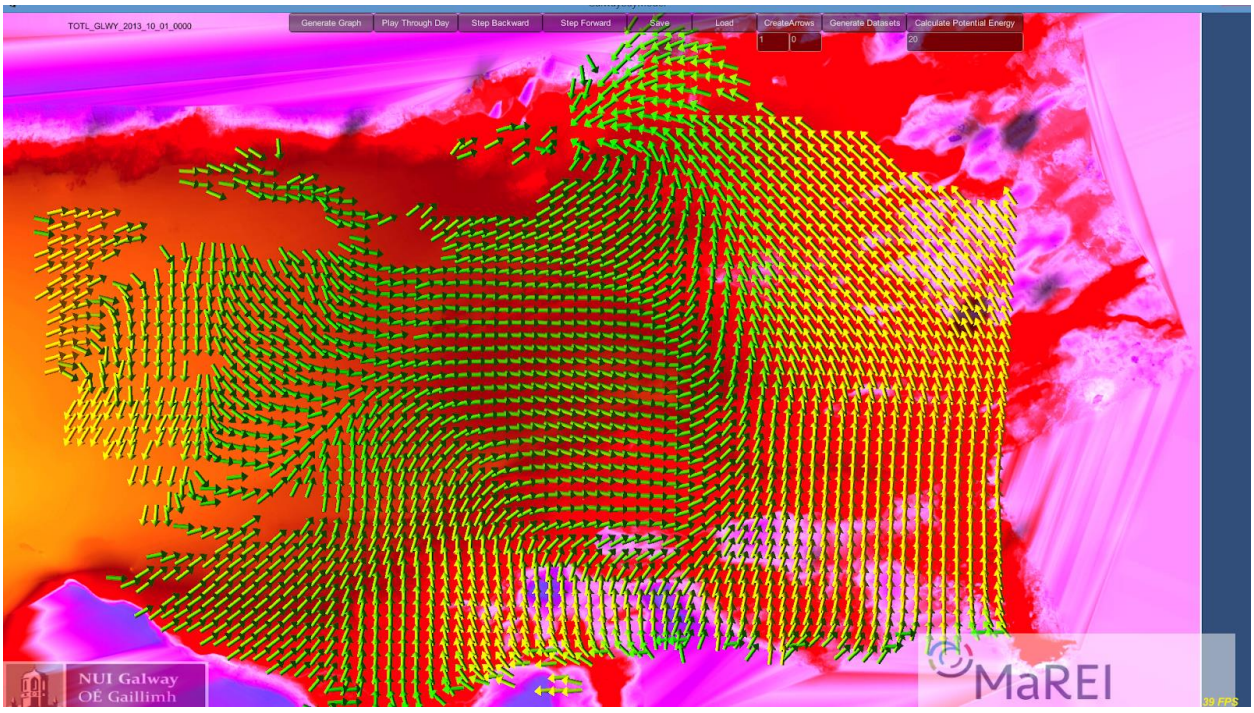


Fig 2.5, Galway Bay Current Visualisation screenshot, (Broderick, 2016)

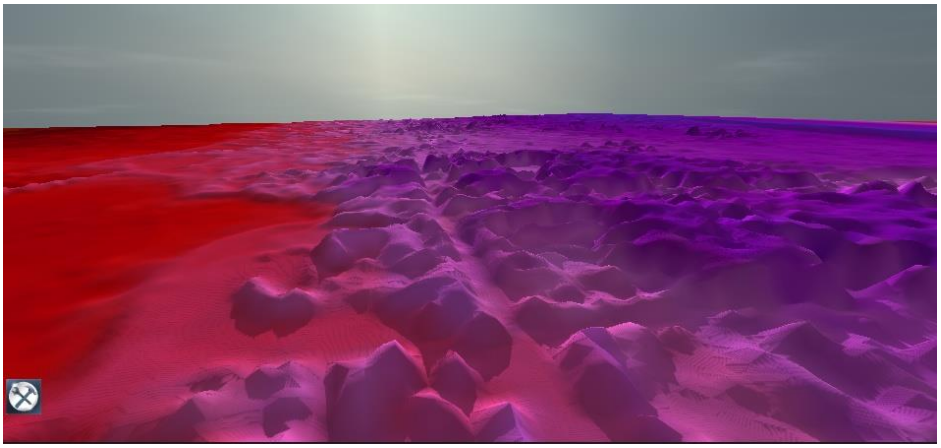


Fig 2.6 3D modelled Galway Bay Seabed in Unity, (Broderick, 2016)

2.6 Meta Analysis

While examining papers related to VR and serious games, a pattern is evident. Audio was rarely a focus within the majority of papers (approx. 120 total papers reviewed in those areas, as well as review papers examining a further 150+ papers), with usually just a brief mention of an auditory element, such as background noise, a voiced narrative, or something similar. Several papers acknowledge the important role audio has to play in virtual environments, but if audio wasn't the overall focus of the research, it was often just a side note. Considering the importance of immersion and presence in virtual environments (Freeman et al., 2020; Schuemie et al., 2001), and how many serious games aim to recreate real world situations for training or education (Knight et al, 2010; Sharma et al, 2014), the lack of focus on audio in these research areas is a deficit. This is especially true considering the proven advantages brought by audio as discussed throughout this literature review. A meta-analysis was performed on a series of review papers covering different areas of serious gaming and VR to examine how often these field reviews examined audio, and how many papers within these reviews address audio as a major focus. Our theory was that the majority of papers and review papers in these fields would either not cover the importance of audio in the experiments performed or its

possible impact or would only give a brief mention of audio. For the meta-analysis, we primarily examined review papers which focused on either serious games or virtual reality. Within these reviews, we examined their criteria for paper inclusion and discussion, as well as the names and keywords for all the papers covered within these reviews.

These review papers investigated a variety of areas, such as medical training, general education and learning, tourism, presence, decision making, and more generalised reviews of the fields of serious gaming and VR. They ranged in year from 2001 to 2020 and were gathered through Scopus and Google Scholar. Across these reviews, the majority of reviews examined (87%) had barely any mention of audio within the content of the review, if any mention at all. While defining the criteria for review and discussing specific papers, there was often a single mention of audio, oftentimes stating an “audio-visual element”, or how audio can improve these environments. Only three reviews (Nehme et al, 2006; Ravyse et al, 2017; Laamarti et al, 2014) made any focus on audio or sound as an important factor to consider, either as a form of environmental fidelity, or providing benefit to research outcomes. One of these reviews (Nehme et al, 2006) focused specifically on audio decision support, while the other two papers looked at serious games or VR for learning, tourism or provided a more general overview of serious games. Even papers looking at measuring engagement or presence and perception didn’t focus on audio in any meaningful way. Table 2.1 summarises all the review papers examined, and what mention of audio they contained, as well as how many papers within those reviews contained reference to audio

This lack of focus is also present in the papers examined within the reviews. Reviewing the titles, abstracts, and keywords of the reviewed papers, only a small percentage had mention of audio, sound, or multi-modality. From the papers within these reviews examined by the author, it did not appear to be that audio was overlooked but was generally either not present or not a focus of the paper. It appears that audio isn’t seen as an important factor within the paper’s description, and from looking at other research within the field of VR and serious gaming, this is often the case. While audio is considered a nice bonus, or something that comes part and parcel with games and VR,

it typically isn't something considered of importance in discussion. The few papers that do mention audio often have it as a reference to "audio visual", or in reference to specific medical work such as "ultrasound guided needles" or "heart sounds". The review papers would make mention of these technologies, but not discuss how sound was used in the papers or any advantage it could give. As many of these reviews were critiquing the difference between digital training vs traditional training, it seems like audio would be something of import to those cases.

Paper Title	Paper Authors	Review Audio Focus	Papers with Audio Focus	Context of Audio Papers
Developing serious games for cultural heritage: A state-of-the-art Review	Anderson, E. F., McLoughlin, L., Liarokapis, F., Peters, C., Petridis, P., & de Freitas, S.	No	2	1 audio-visual
A Comprehensive Review of Serious Games in Health Professions	Ricciardi, F., & De Paolis, L. T.	No	2	Ultrasound guided needles + audio-visual
Immersive virtual reality serious games for evacuation training and research: A systematic literature review	Feng, Z., González, V. A., Amor, R., Lovreglio, R., & Cabrera-Guerrero, G.	No	0	
A review of immersive virtual reality serious games to enhance learning and training.	Checa, D., & Bustillo, A.	No	1	Ultrasound guided needles
A systematic review of serious games in training: Health care professionals. Simulation in Healthcare	Wang, R., DeMaria, S., Goldberg, A., & Katz, D.	No	1	Ultrasound guided needles
A systematic literature review	Connolly, T. M., Boyle,	No	1	

of empirical evidence on computer games and serious games	E. A., MacArthur, E., Hainey, T., & Boyle, J. M.			
A systematic literature review on serious games evaluation: An application to software project management.	Calderón, A., & Ruiz, M.	No	1	Ultrasound guided needles
Trends on the application of serious games to neuropsychological evaluation: A scoping review	Valladares-Rodríguez, S., Pérez-Rodríguez, R., Anido-Rifón, L., & Fernández-Iglesias, M	No	2	
VR to OR: A review of the evidence that virtual reality simulation improves operating room performance	Seymour, N. E.	No	0	
The impact of perception and presence on emotional reactions: A review of research in virtual reality.	Diemer, J., Alpers, G. W., Peperkorn, H. M., Shiban, Y., & Mühlberger, A.	No	0	
The minimal relationship between simulation fidelity and transfer of learning.	Norman, G., Dore, K., & Grierson, L.	No	5	heart sounds
A literature review on immersive virtual reality in education: State of the art and perspectives	Freina, L., & Ott, M.	No	0	
Immersive Environments and Virtual Reality : Systematic Review and Advances in	Rubio-tamayo, J. L.	No	2	audio-visual

Communication , Interaction and Simulation				
Virtual Reality Goes to War : A Brief Review of the Future of Military Behavioral Healthcare	Rizzo, A., Parsons, T. D., Lange, B., Kenny, P., Buckwalter, J., Rothbaum, B.,Reger,G	No	0	
Current Issues in Tourism New realities : a systematic literature review on virtual reality and augmented reality in tourism research.	Yung, R., & Khoo-lattimore, C.	No	1	audio-visual
A systematic review of serious games in medical education : quality of evidence and pedagogical strategy	Gorbanev, I., Agudelo-londoño, S., González, R. A., Cortes, A., Pomares, A., Delgadillo, V., Cortés, A.	No	0	
Why so serious ? On the relation of serious games and learning	Breuer, J., & Bente, G.	No	0	
Serious Games For Treatment or Prevention of Depression: A Systematic Review	Ridgman, H. E. B., Tasiak, K. A. S., Hepherd, M. A. S., & Erry, Y. A. E. L. P.	No	0	
Adaptive and interactive climate futures: Systematic review of “serious games” for engagement and decision-making.	Flood, S., Cradock-Henry, N. A., Blackett, P., & Edwards, P.	No	0	
An Overview of Serious Games	Laamarti, F., Eid, M., & Saddik, A. El.	Partial Yes	0	
Success factors for serious games to enhance learning : a systematic review	Ravyse, W., Blignaut, S., & Leendertz, V.	Yes	12	

Audio Decision Support for Supervisory Control of Unmanned Vehicles: Literature Review	Nehme, C. E., & Cummings, M. L.	Yes	29	
A Systematic Review of the Definition and Measurement of Engagement in Serious Games	Hookham, G., & Nesbitt, K.	No	1	

Table 2.1 Table of Reviews Examined

Evaluating the reviews which mentioned audio as an important factor and contained information about reviewed papers audio content, some conclusions that can be drawn about the perceived importance of audio in non-audio areas of research. In the review by Nehme et al. (2006), the focus was on audio decision support specifically, which explains why every paper analysed was aimed at audio. This was the only review found which explicitly examined audio in relation to decision support systems, serious games, or VR.

The other paper that provided information on papers with audio is the review by Ravyse, et al. (2017). This review acknowledged the importance audio can play within serious games and made note of how audio was presented in each of the examined papers. Most of the papers simply had a reference to audio-visual components, with only a couple using audio for something greater (combined with haptic feedback for blind players, informing players of incorrect choices), and with only one paper targeting audio as a core area of study. Even in a review paper which acknowledged the importance of audio, there are very few papers available which focused on audio.

2.7 Research Gaps

Until recently, most work in virtual environments and game engines was focused on the visual possibilities of these technologies (Hamidi et al, 2009). Most audio research in these fields was quite limited, specialised to specific areas, or came from before recent technological improvements in game engines and VR. In more recent times, especially with the growth of the games industry and VR, we are seeing an increase in the research about spatial audio in VR and virtual environments.

While the study of VR has been growing significantly in the last few years with the increased accessibility of VR technology, our literature review, and related meta-analysis, clearly highlights the lack of existing research that examines the importance of audio in VR, and how the control schemes used in VR can affect the importance of spatial audio and other design factors when it comes to virtual environments and experiences.

Our literature review and meta-analysis has also demonstrated that there is evidence that audio plays an important role in user immersion in virtual environments, as well as providing practical benefits to awareness, performance, and enjoyment. Whether it be in serious games, CVEs, decision support systems, or entertainment, audio quality and spatialisation greatly influence the user experience, and this can be made even more key as we move to more natural and intuitive control schemes. Despite this, most of the examples discussed here don't focus on audio, especially when compared to the visual elements of virtual spaces. While there is research within audio fields, there isn't enough being done in the connections between these various study areas, although they often go hand in hand and provide benefits to each other.

Our research aims to address this gap between the fields, taking the known advantages of audio and integrating and examining them more deeply in relation to virtual environments and VR. It especially looks at how the type of audio can influence design decisions for virtual spaces, and how the type of virtual experience being created can affect the importance of audio quality and design.

We also aim to fill the gap in research into how the type of control system (such as keyboard and mouse or HMD) used for virtual environments can impact user performance and enjoyment, especially in relation to how it can impact the importance of sound within that environment.

2.8 Conclusion

The goal of this literature review is to provide a view of current thinking within a range of areas related to spatial audio in VR, and to identify the requirement for and relevance of VR and spatial audio for decision making, awareness of surrounding environments, and the importance of the way a user interacts with the virtual environment. An additional aim is to also use readily available technologies and resources, such as game engines and spatial audio engines, to showcase the accessibility and availability of these resources. Where previously spatial audio required development of specific environments and engines, modern solutions have made it much easier for a user to implement a spatial audio engine within an existing game engine and create an experimental environment. Many of these technologies also integrate directly with VR equipment.

While our early work looked at marine simulation, geographical information systems, and collaborative virtual environments, this led to a more detailed focus on two specific areas: *the way a user interacts with a virtual environment, and how spatial audio can aid a user in decision making by providing information on their virtual surroundings without requiring their visual attention*. While our initial work looked at creating virtual recreations of the real world populated with data for collaborative decision making, it was found that the amount of information being visually displayed could be difficult for a user to understand. Auditory display was identified as an important aspect.

We have seen how VR and spatial audio are intrinsically tied to users' feelings of presence and awareness within virtual environments. We have also seen cases where the addition of audio can improve users' performance in a variety of tasks, and also allow them to be aware of changes in virtual surroundings or sonified information. Finally, we looked at how humans in nature react to and search for sound sources, and how those same methods can carry across to virtual environments with natural control schemes, creating a more intuitive and realistic feeling scenario.

3. Methodology and Experimental Design

3.1 Introduction

Within this chapter we will outline the research questions examined as part of this research, as well as provide details on the experiments created and performed. We will examine how the experiments were designed, what the goal of the experiment was, and how the user would perform each experiment. The quantitative and qualitative data gathered and why it is important is explained, as well as the specific methods used to gather the qualitative data through a questionnaire and a NASA-TLX survey. The technologies used for the creation of the experiment are detailed and justified. Finally, we discuss some of the limitations of the experiments and some of the research constraints that developed throughout the course of research, including the impact of COVID-19 on in-person experimentation.

3.2 Research Questions

As such, our research questions are seen in Table 3.1 in addition to the relevant experiments for each question. We also show a flow diagram of how our research questions led to each experiment in Fig 3.1.

Research Question	Relevant Experiment
"Is there a performance difference between fully spatialised audio and traditional panning audio, while searching for audio sources?"	*Experiment 1*
"Is there a performance difference between different control schemes while searching for audio?"	*Experiment 3*
"Does spatial audio improve a user's awareness of their surroundings in a virtual environment?"	*Experiment 1, Experiment 2, Experiment 3*
"Does the control scheme implemented for a virtual environment change how a user interacts with virtual sounds?"	*Experiment 3*
"Do different control schemes affect a user's enjoyment or perceived performance?"	*Experiment 3*

Table 3.1 Research questions and relevant experiments

3. Methodology and Experimental Design

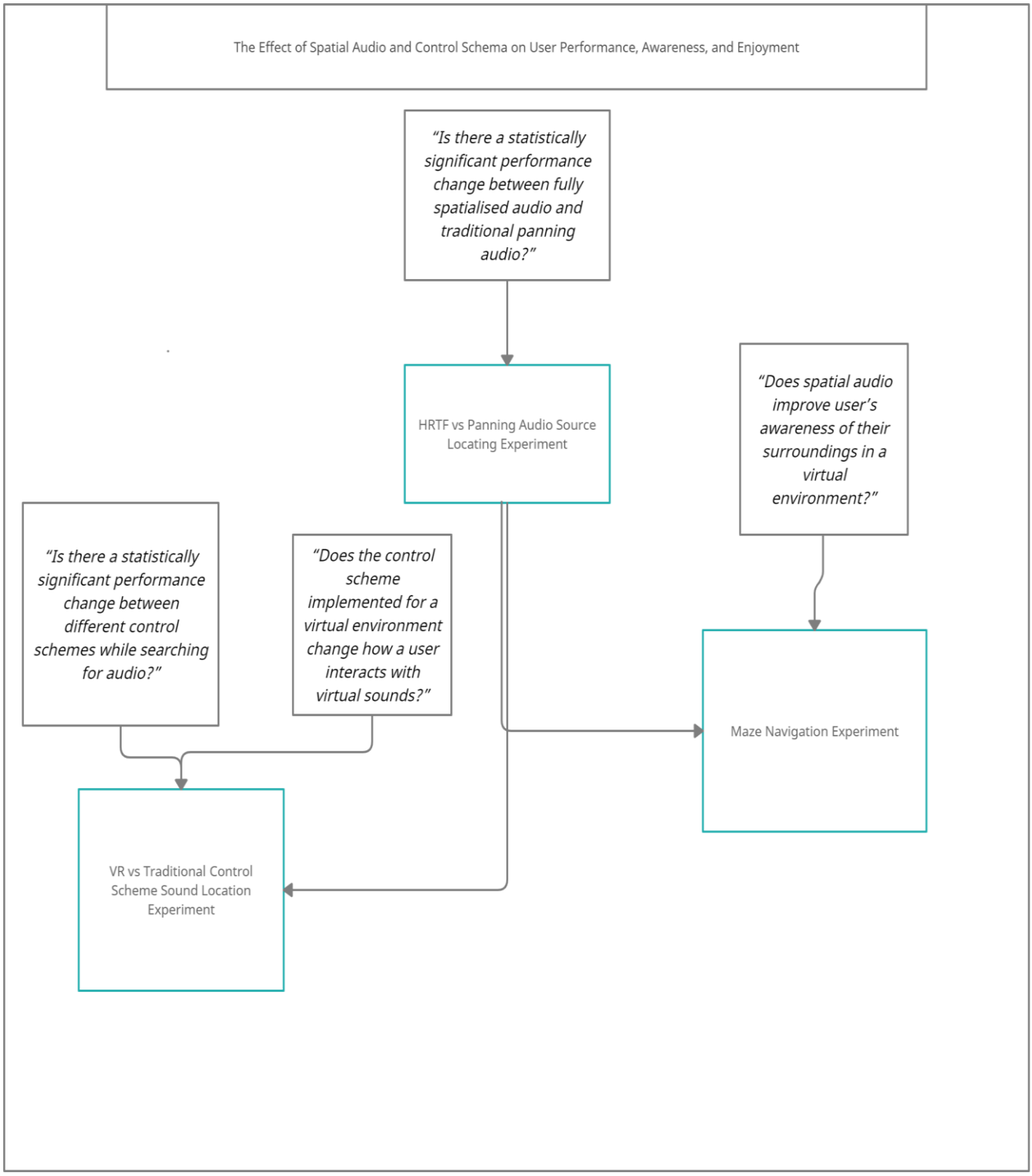


Fig 3.1 Research Questions and Experimental Flow Diagram

3.3 Summary of the experiments

In this section we will go through each of the experiments we created and performed to detail how they were created, why they were designed that way and the data gathered from each experiment. In Table 3.2 we have a summary of each experiment, its title, and its goal.

Experiment Number	**Experiment Title**	**Experiment Goal**
1	Locating sound sources between HRTF Spatial Audio and Panning Stereo Audio	To examine a difference in user performance while locating sound sources using either HRTF Spatial Audio or traditional stereo audio.
2	Navigating a virtual maze and avoiding obstacles using audio-visual and pure visual	To examine differences in performance and user movements while navigating a virtual maze and attempting to avoid moving obstacles with the use of audio or only visuals.
3	Performance at locating sound sources between Mouse + Keyboard and Head Mounted Virtual Reality	To examine differences in performance and search techniques while searching for sound sources using Mouse+Keyboard or Head Mounted Virtual Reality for user control.

Table 3.2 Experimental Summary

3.3.1 Experiment 1: Locating sound sources between HRTF Spatial Audio and Panning Stereo Audio

The first experiment was aimed at comparing HRTF Spatial Audio and Panning Stereo Audio in a virtual environment. The goal was to evaluate if there was a noticeable difference in user performance in locating an audio source within a virtual environment depending on if the audio was properly spatialised using HRTF or traditional panning stereo audio. It was most focused on research questions 1 and 3:

"Is there a performance difference between fully spatialised audio and traditional panning audio, while searching for audio sources?"

"Does spatial audio improve a user's awareness of their surroundings in a virtual environment?"

3.3.1.1 Development of Experimental Environment

Experiment 1 was performed with normal headphones on a computer with a mouse and keyboard. The user was in a 3D environment built in Unity, with only the ability to move their camera to look around them using their mouse. The user was surrounded by a series of grey cubes at pre-set points, totalling 26. In front of the user were 9 cubes, 3 situated above their eyeline in a row, three at their eyeline in a row, and 3 below their eyeline in a row. The same array was behind the user, and on the same plane as the user were a further 8 cubes, missing one at the user's set position. The environment was enclosed in a box, with each wall given a different colour so that users could orientate themselves and not become lost in the space. The visual fidelity of this environment was kept basic as it was not the focus of the given task and should not distract from the audio. The cubes were placed in multiple locations and elevations to create more possible

difficulty when sounds came from above and below the user. An example of the environment can be seen in Fig 3.2.

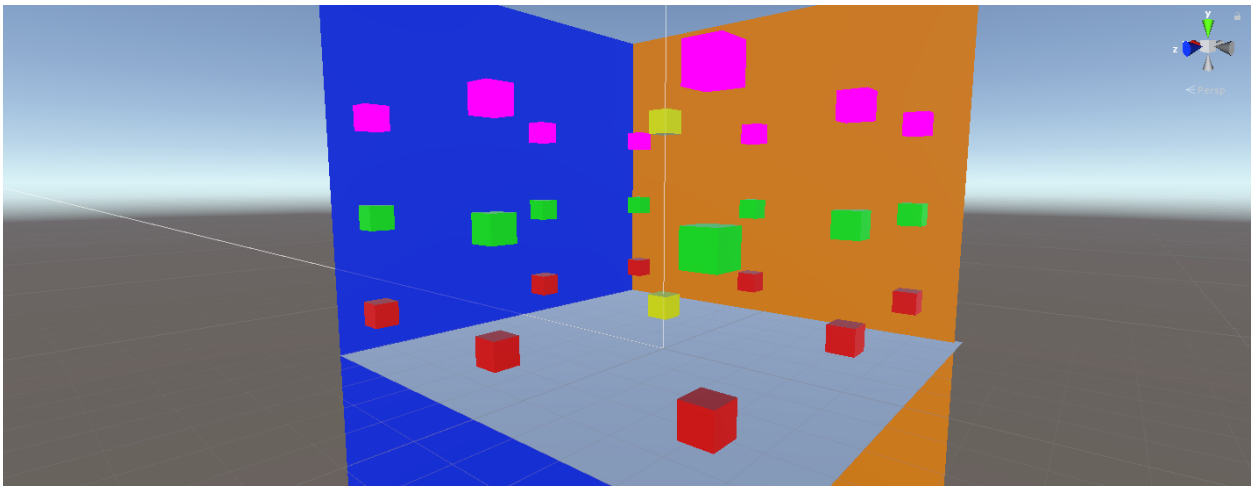


Fig 3.2 Experimental environment overview, with cubes coloured by group

The user Game Object[2] at the centre had a camera attached to it which was controlled using the mouse. There were also a variety of scripts which handled the raycasting[3] from clicking the mouse, as well as storing user data and triggering the sound activations. The walls and floor of the environment were simply coloured planes so that users wouldn't get lost in a blank background. Having coloured walls also allowed the user to turn around quickly while having an idea of how far they had rotated, as it gave them a point of reference.

The cubes are Unity cubes with colliders[4], allowing them to interact with the raycasting activated by the user. They have a script attached to them which handles whether or not it is the active cube, controlling its colour, if it has been clicked, the amount of guesses a user has made to find the cube, and controls starting and stopping the audio. The cube also has an audio source component which handles the sound [5]. There are

[2] A fundamental object in Unity which represents a character, scenery, object. It also stores functionality
[3] Raycasting is the process of shooting an invisible ray from a point, in a specified direction to detect whether any colliders lay in the path of the ray.
[4] Colliders are used to create a physical shape for a GameObject and are used for detecting collisions.
[5] An audio source component is a Unity component for objects which allow them to be a source of sound, causing the audio to play from a specific location, character or object in the scene

settings for sound volume, priority, looping, etc. It also handles settings for 3D sound when used with a spatialiser within Unity. The 3D sound settings include distances the object can be heard from, volume rolloff graphing, etc (Fig 3.3).

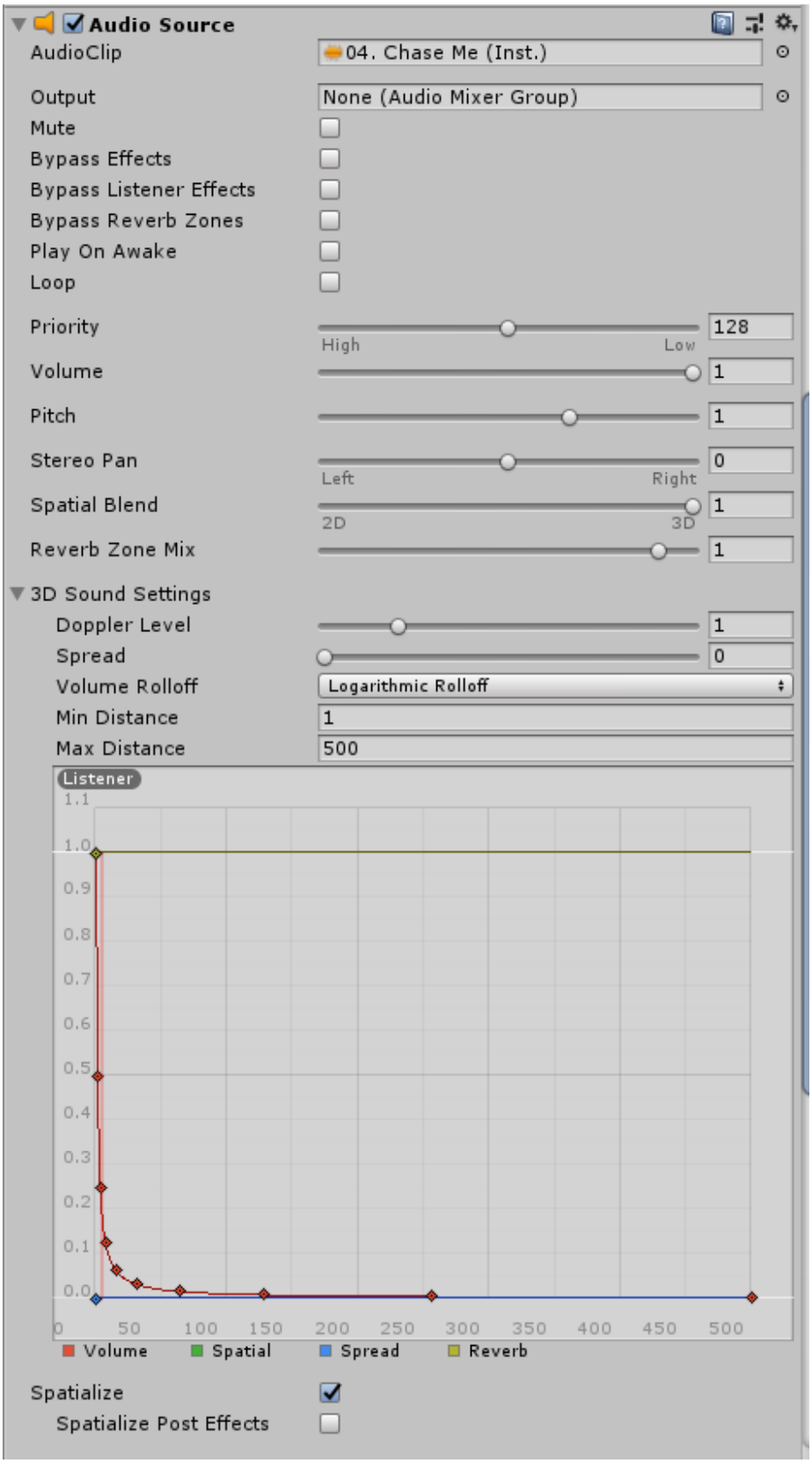


Fig 3.3 Unity Audio Source

3. Methodology and Experimental Design

Two versions of the experiment were created. One was built using the inbuilt Unity audio solution which used panning stereo audio for placing sound sources within the environment. This means that as a user moves either their character or camera in an environment, audio will seem to pan between either ear to let users know where the sound is coming from in their environment. This simple form of audio is the most common in games and generally not computationally expensive. It is optimised for a full stereo audio set, where multiple speakers can provide differing locations for the audio. As such it translates much more poorly to headphones, leading to issues discerning front from back, top from bottom, and general elevation.

The second version of the experiment is built using the Oculus HRTF audio solution built into Unity's editor. This converts Unity's sound source objects to use HRTF audio, adding some options for spatial audio changes. Compared to panning audio, HRTF attempts to replicate how sounds work in the real world. As a user moves their character or camera, the sound source applies a HRTF to the sound based on the sound source location and the user's location before sending the sound to the user. This changes the sound in such a way as to replicate how our brains receive the sound in real life. Ideally, this will make it simpler for the user to locate the source of the sound, as the location would sound similar to its location in real life. This should also reduce issues with handling elevation differences between the user and the sound source, as well as helping with the front-back issue.

However, this system has several limitations. HRTFs used by this algorithm are generic, meaning it will be different to a user's individual HRTF. This can be a small difference or a large difference depending on the user, which can cause issues. It also sounds different to how we are used to hearing sounds in virtual environments. People who are very used to panning audio may take time to adjust to a more realistic style of sound in a virtual space. There are also limitations to how we can mimic real world motions with a mouse and keyboard or a game controller. In the real world, we tilt and shift our head in a lot of angles while listening to sounds but doing so in a virtual environment with traditional control schemes is much more awkward and generally not

supported. It is also more computationally expensive than traditional panning audio techniques. It is possible that different headphones can create different situations, either through the quality/type of headphone, or potential damage. This is why it is usually better to perform this type of experiment in-person with identical equipment.

As the experiment was built in Unity as a game executable, it was possible to run the experiment on any type of machine, and remotely, which was attempted. However, there were many issues with remote testing, which are discussed later in this chapter.

The goal of the experiment was for users to correctly guess which cube was the source of the sound playing in the environment. The process would be repeated nine times before the experiment ended. The reason for nine iterations was to give the user a chance to understand the experiment and get data for multiple cubes, while not creating a lengthy experiment or allowing for too much training. The first three times the sound was played, the cube would light up purple, indicating to the user where the sound was coming from. This was to give the user a chance to accustom themselves to the controls, the environment, the sound, and the way the experiment would work. The remaining seven sounds came from unmarked cubes. The user task flow is shown in Fig 3.4.

3. Methodology and Experimental Design

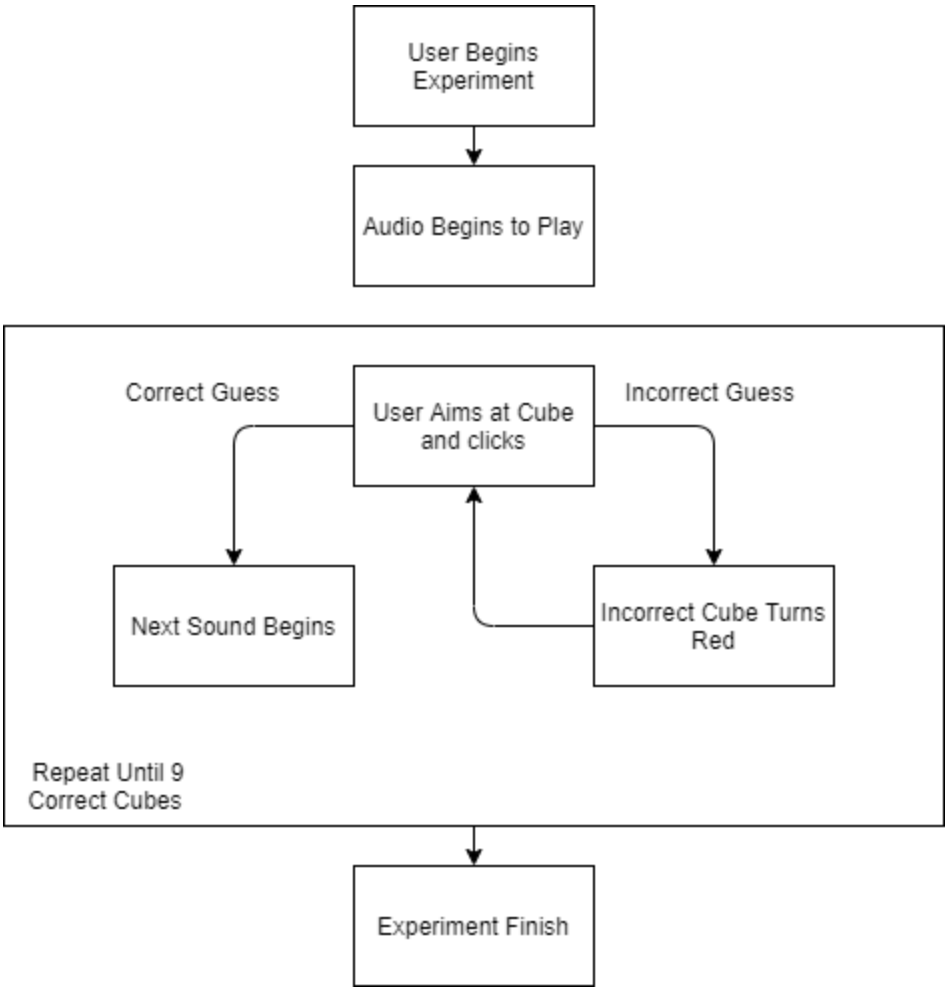


Fig 3.4 Experimental Flow

The user used the mouse to aim at the cube they thought was generating the sound, with a targeting reticle at the centre of the screen to make it clear where they were aiming. If the user clicked a cube that was incorrect, the cube would turn red, and they would be able to continue until they found the correct cube. Upon clicking the correct cube, the user would be given an indication of correct choice, all cubes would have their colour reset to grey, and the sound would start from the beginning at a new cube.

The cubes that generated sound were randomly selected from all available cubes. The sound used was a clip of instrumental music. Before the experiment, several

unrelated individuals were given possible noises in testing for the experiment, and feedback was that using the music compared to tones or similar noises was simpler and more pleasant for the user. It was also easier to tell when the sound reset. We have

After performing the experiment, the user would be asked to repeat the experiment a second time, with a different audio type. The type of sound spatialisation the user received first was randomised to avoid training affecting results too strongly. The goal was to view the difference in user performance between audio types.

Users were asked for their name, age, sex, email, location of experiment, and if they had any visual or auditory impairments or aids. As the user performed the experiment, the program logged this information, the type of audio used, and the time it took them to complete each sound location task. The program also calculated the average time it took the user to complete the experiment, removing the tutorial cubes at the beginning.

3.3.2 Experiment 2: Navigating a virtual maze using audio-visual and pure visual

The purpose of this experiment was to examine how a user can perform in an environment with and without audio. The audio lets the user know when they come close to a wall or to a moving object, such as the obstacles within the maze. The user is instructed to avoid touching the walls or the moving obstacles within the maze, while trying to navigate their way to the end point. The hypothesis is that *the presence of audio should allow the user to be more aware of their surroundings, especially outside of their vision, while a lack of audio leads to a user having to constantly look behind and to their sides, while making more mistakes as unexpected collisions occur*. While there is considerable research examining spatial audio for navigation and sound location, the focus in this experiment was to see awareness of and avoidance of moving, sonified objects, with aims to later perform the experiment comparing different control schemes

and the effect they have on the same type of experiment. This experiment primarily targets one of our research questions:

"Does spatial audio improve a user's awareness of their surroundings in a virtual environment?"

3.3.2.1 Developing the Experiment Environment

This experiment was designed in Unity similar to the other experiments, using the Steam Audio plugin for handling the spatialised audio attached to the walls and objects within the environment. The experiment is performed with a mouse and keyboard, a monitor, and headphones. All the equipment is the same as listed above for the second sound location experiment, minus the VR kit as this experiment did not have a VR element. The experiment took place in the same type of environment as the second sound location experiment with some of the previous cohort of subjects.

For this experiment, users were asked to navigate their way through two mazes, with a short tutorial space at the beginning to get accustomed to the controls. They were given two or three out of a selection of three mazes, with at least one maze having no audio components and at least one maze having audio components. Users were told of this in advance so that they wouldn't be confused at the lack of sound in only one maze. Each maze is different in layout and contains different obstacles. Users were told to avoid touching walls or obstacles as they attempted to navigate their way to the end of the maze. The general direction of their goal was given through a waypoint which appears on screen as a green arrow, showing through objects and walls. This was to give them an idea of their target direction regardless of if there were obstacles or walls in the way. Examples of the mazes can be seen in Fig 3.5, 3.6, and 3.7, with an example of some obstacle types seen in Fig 3.8.

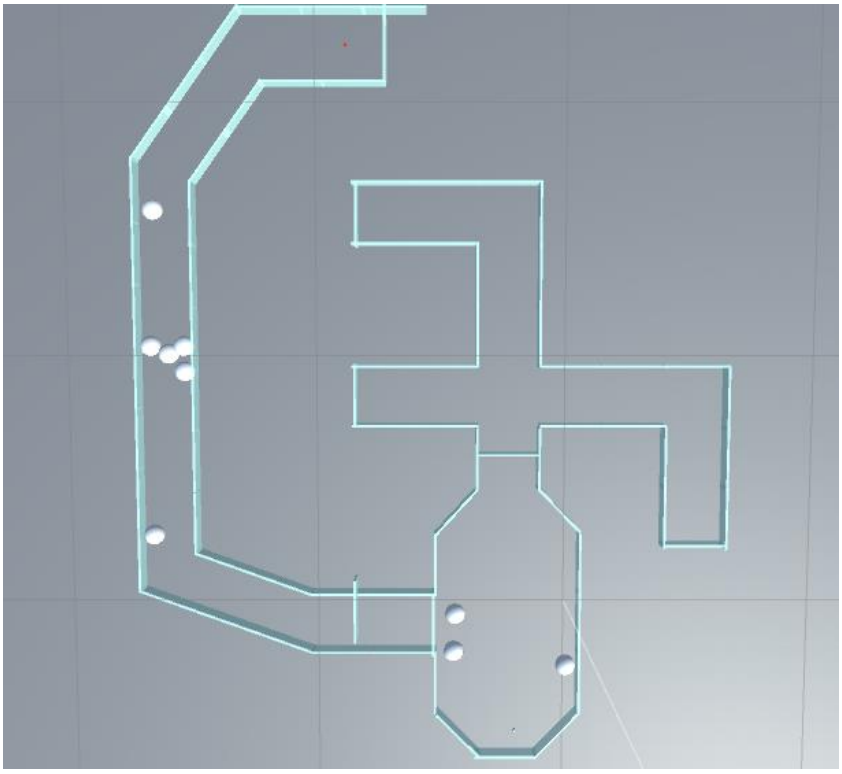


Fig 3.5 First Maze

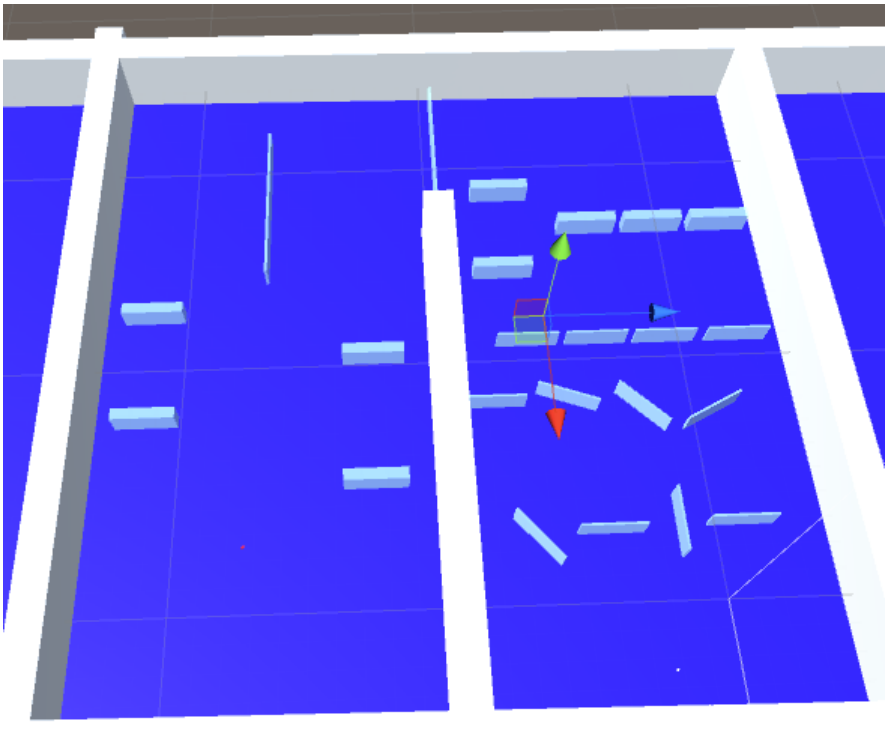


Fig 3.6 Second Maze

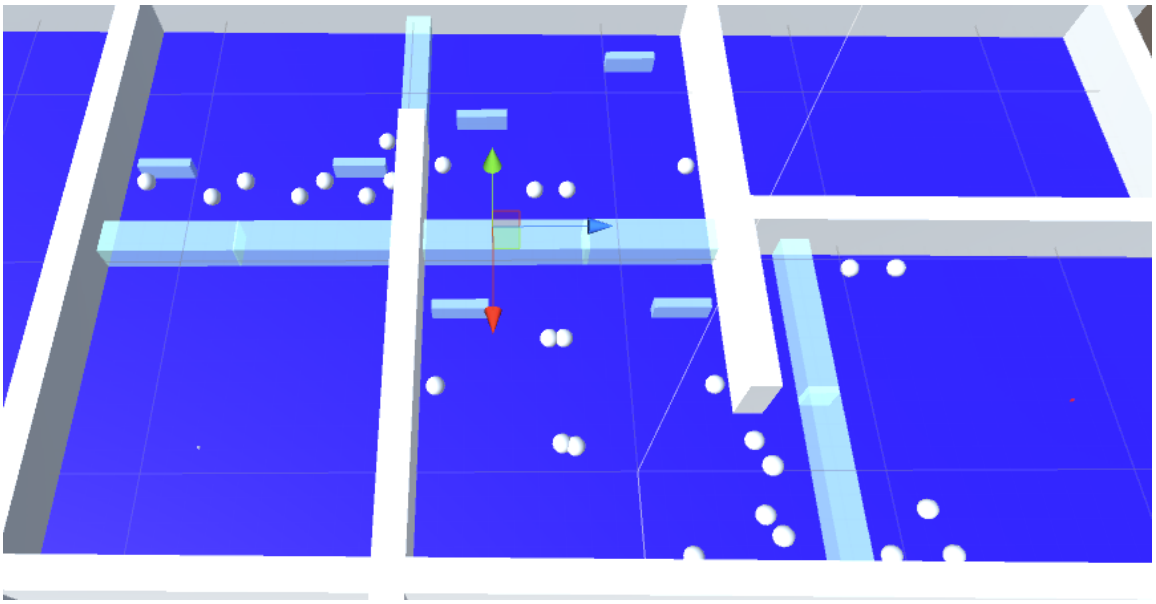


Fig 3.7 Third Maze

The maze was created using Unity models, as the visual fidelity of the environment isn't overly important to the task. The walls of the maze as well as the moving obstacles are Unity objects with sound source components attached to them, as well as Steam Audio components to control the sound spatialisation. To avoid sound clutter, the volume falloff distance is set so that the players hear the sound within a certain radius of an object or wall. The sound sounds like rushing water, which grows louder as it gets closer to the user and quieter as the user gets further away.

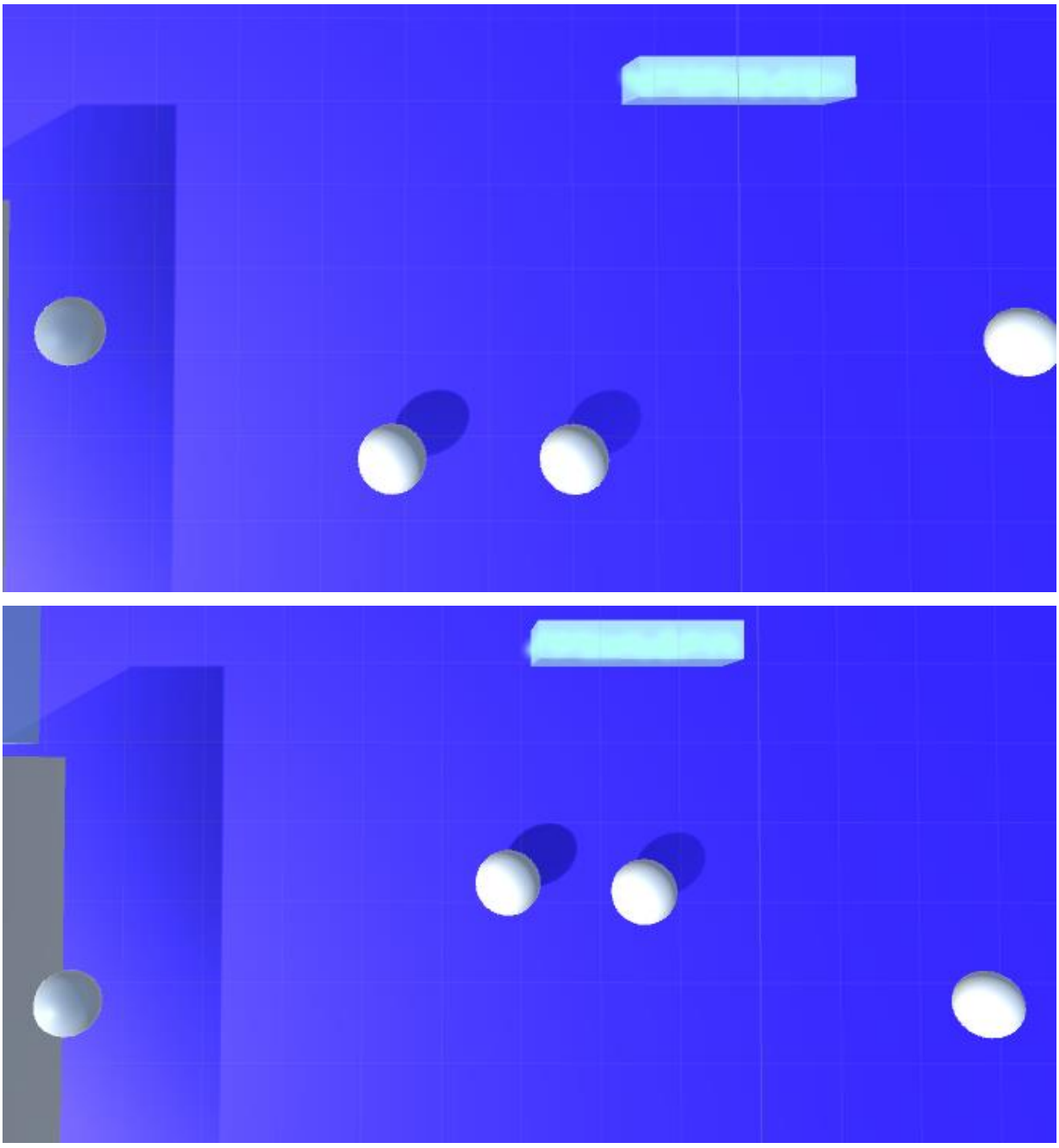


Fig 3.8 Movement of sphere obstacles over 2 seconds.

There are a variety of obstacle types of the course of the three mazes. There are boulder-like spheres that roll back and forth along a set path at a set speed, spinning walls that rotate slowly around a set point, blocks that move back and forth blocking

certain paths, and doorways which open and close. When the user has a collision with an object or wall, they hear a penalty noise, and further collisions are paused for 2 seconds to give the user a chance to adjust. The specific code that moves the objects can be seen in the appendix to aid in replication of these obstacle types, with an example of the Unity Editor options for the object in Fig 3.9. The objects were coded to be generic and reused throughout the experiment, with minor changes made based on speed and distance to move. The maze isn't particularly detailed, as the main difficulty is the objects moving in the way. The concept is that having spatial sound should let users remain aware of the moving parts around them, while a lack of audio will either force them to look around more often or lead to a higher number of mistakes. All objects follow a set path at a constant speed, so a user can predict their movement to a certain degree with patience. It also means that each run of a particular maze is the same over time, which makes it easier to replicate a user's travel through the maze using the tracked movement data discussed in the quantitative data section of this chapter.

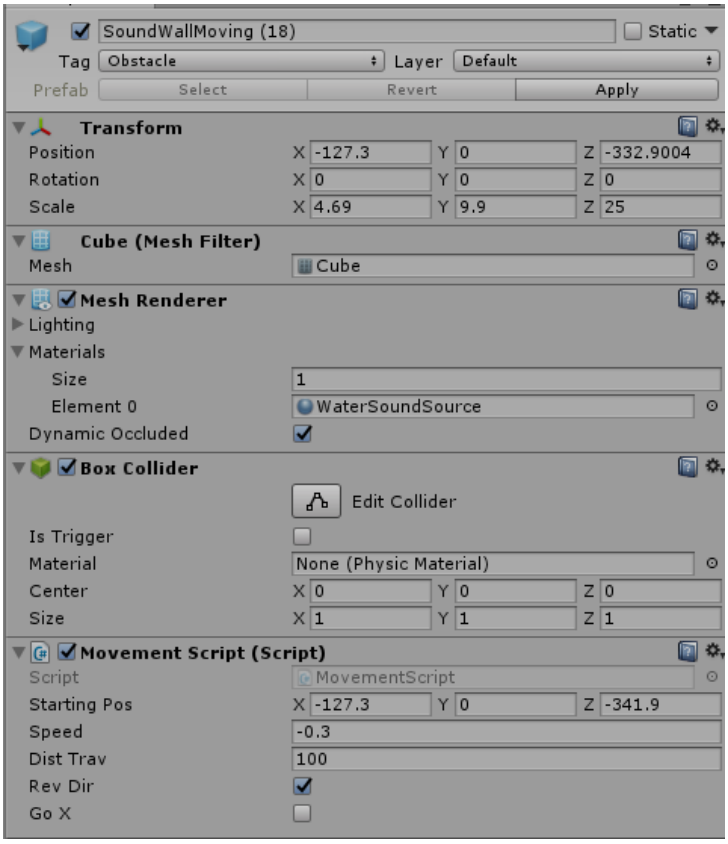


Fig 3.9 Sample of Moving Wall Components

3.3.2.2 Quantitative Data Gathered

For this experiment, the quantitative data that was gathered focused on the user's performance in navigating the maze. The first thing tracked was the *time it took for the user to navigate through a particular maze*. This is then supplemented by tracking collisions made with either the walls of the environment or moving obstacles within the environment. This information is used to determine a user's general performance within the environment, as well as their awareness of things outside of their eyesight. It also tracks what location they were at when they made a collision, allowing us to see where mistakes happen more frequently.

The more important factor tracked in this experiment is the *player's positional location as they performed the task*. This allows us to replicate the player's movements after the fact, to see the path they take, any objects they collide with, and how they choose to avoid objects in the environment. This information is stored as Vector3s[6], allowing for the positions and movements to be replicated in Unity using a driver which moves through the player's path. Compared to a video recording of the player's activities, this is a much more efficient use of storage space, and it also allows for other types of analysis. As well as simply looking at the path an individual player took, or where they made mistakes, or how they avoided the obstacles, we can create a heatmap[7] of the environment and examine patterns between multiple runs of the map. This allows us to pose such questions as: did players follow a similar path to each other, did the presence of audio make their decisions different, or was it more up to individual style and strategy? While the advantage of audio over no audio is more predictable, it is more interesting to look for how spatial audio emanating from objects outside of view could change user planning, strategy and decision making, as well as their awareness. Research has shown that humans use audio to maintain awareness of things outside of vision, and there is support

[6] A Vector3 is a data structure which can be used to store the x, y and z position of an object in an environment.
[7] A form of data visualisation using a variation between two colours to show activity at specific locations.

that the same concept can work for decision support systems or digital applications (Donmez et al, 2009; Brock et al, 2002).

3.3.3 Experiment 3: Locating sound sources between Mouse + Keyboard and Head Mounted Virtual Reality

This experiment was the primary experiment of the research performed, as it addresses the majority of our research questions and provides some of the most unique research, looking at the variety of ways control schemes can influence different factors. Its base is like the previous sound location experiment, but the goal of the experiment was altered, and the overall experience improved through improvements to the tracking of experimental data and a focus on in-person experimentation. The aim of this experiment was to evaluate the difference between movement control schemes when it comes to finding audio sources in a virtual environment. The focus was on a traditional mouse + keyboard setup with a computer monitor vs a VR test with a head-mounted display (HMD). This primarily targets three of our research questions:

"Does the control scheme implemented for a virtual environment change how a user interacts with virtual sounds?"

"Do different control schemes affect a user's enjoyment or perceived performance?"

"Is there a performance difference between different control schemes while searching for audio?"

The experiment looked at both qualitative and quantitative information from the users' performance as well as their personal opinions and used that information to better

analyse the differences and advantages between the different control and viewing schemes. We show the experimental flow and tracked data in Fig 3.10.

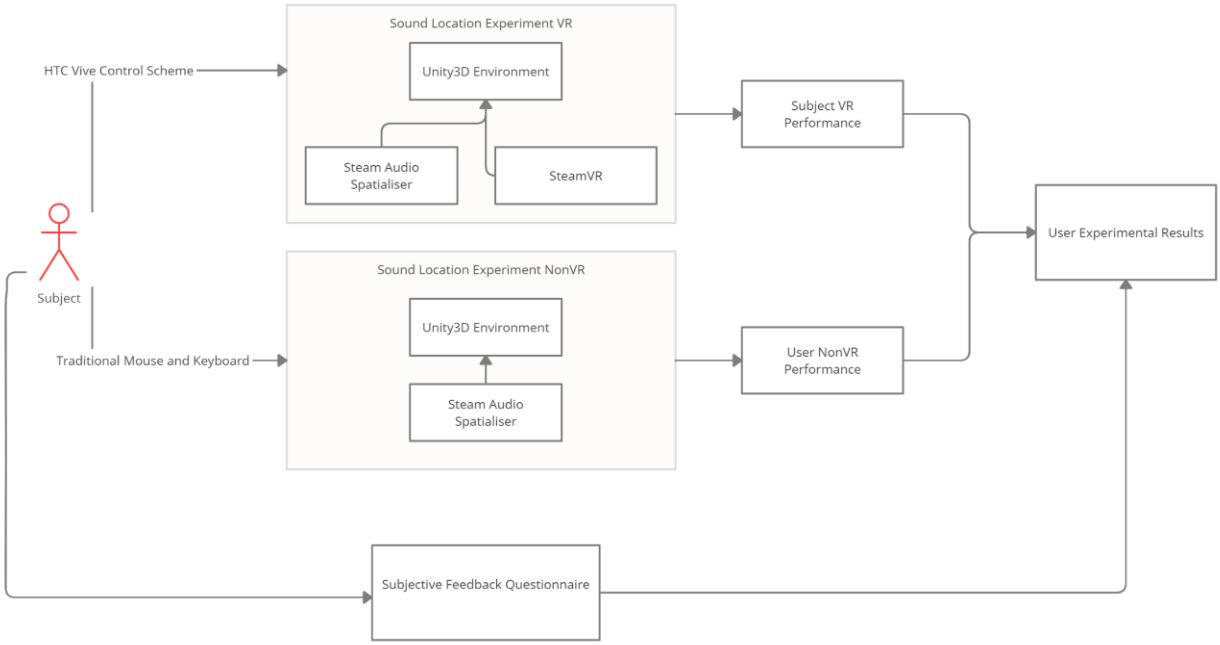


Fig 3.10 Experimental flow and results tracking

3.3.3.1 Development of Experimental Environment

The environment used for this experiment was based on the previous sound location experiment. The virtual environment itself is the same, with most changes being tied to the user input for VR controls, and the changes to the spatial audio framework used. Visually it remained similar to Fig 3.2.

The experiment was performed on an Asus ROG FX502VM-DM092T laptop, suitable for VR experiences and gaming. Having a powerful enough computer is necessary for any VR experiments and development, due to the higher load required, though for this experiment the environment is very simplistic and small. A Logitech G402 gaming mouse was used for mouse inputs, and a pair of Philips Bass+ SHL3075 Closed-

3. Methodology and Experimental Design

Back Headphones were used for audio in the experiment. A simple, inexpensive set of headphones were used for this experiment as they would allow us to examine from an accessible standpoint. As any changes to the sound are happening from a software standpoint, it is good to try and emulate an experience which could be accessible to more gaming inclined individuals. An HTC Vive VR kit was used for the VR control side of the experiment. The reasons for choosing this VR kit are listed in section 3.4, but in theory, most VR kits that provide headtracking should work for this type of experiment.

The test took place with a group of students in a private lab workspace, providing a quiet environment free from interruptions. Students were permitted to enjoy some VR experiences after the experiment concluded as compensation for their time. The students provided a variety of ages, backgrounds, and experience with VR or virtual environments/gaming. Each user performed one experiment, followed by a short break of between two and five minutes, before performing the second experiment. This break gave time to prepare the alternate experiment. It was randomised if they performed the VR experiment first or the traditional control method, to avoid training effects. Selected cubes were also randomised to avoid learning patterns.

The main changes with this experiment compared to the previous one are the implementation of Steam Audio for the spatialised audio component, the implementation of VR controls via the HTC Vive, and improvements to the quantitative data tracked by the software. Users were also given a questionnaire and survey after performing the test which aimed to gather qualitative feedback and opinions on the tasks and the differences between the control schemes.

The VR framework used within the Unity environment is the SteamVR Unity plugin (Valve, 2021). This plugin allows SteamVR to interface with the Unity editor and projects, which allows for simple use with multiple VR kits, including the HTC Vive, which was used for this experiment. While Unity has native support for Oculus, some other VR kits require a plugin such as SteamVR for development. Having support for SteamVR is very useful as many people who use VR kits for entertainment and gaming likely have SteamVR

installed, as it is used for playing VR games through Steam. Also, as SteamVR is made by Valve, the same company who made the SteamAudio plugin being used, it is likely to have less chance of integration problems.

SteamVR handles all the VR aspects of the project, though the main one for this experiment is the head tracking for camera control and movement. As the player moves and tilts their head around, the HTC Vive tracks their motions and feeds the information through SteamVR into the Unity environment, either the Editor or the running application. This allows the camera to move and shift, changing the player's view, and also changing the direction of the sound they are hearing.

This change of control scheme is one of the main focuses of the experiment, as it changes the user experience in several ways. First, the user has one less task to consider, which is mostly aimed at users with low experience levels in gaming and virtual environments. Inexperienced users will often be much slower and clumsier when it comes to navigating a virtual environment. While this experiment doesn’t involve character movement, which simplifies the experience, the act of using a mouse to rotate a camera can be something alien to a user. By using the VR HMD, they don’t have to think about that movement or control, they can simply move their head naturally and see the result happen immediately.

This natural feeling is also a core change caused by the control scheme change. Moving around physically is something that comes naturally to us, whereas moving and searching with a mouse and keyboard is a different experience. People learn habits through experience, but it is difficult to translate those between control schemes. The experience becomes more intuitive when using motion tracking, which can reduce the barrier of entry to experiences, and can lead to a change in how users try and do things within the environment (Seibert et al, 2018).

The other major change is that head tracking creates a much greater freedom of movement when it comes to head angles. Traditional keyboard and mouse movement

offers rotation on only 2 of the 3 axes, while you have freedom to angle yourself any way you can naturally with head tracking. This can vastly change the way users interact with the environment and simplifies control, allowing users to make movements without having to consider a keyboard input. To have a similar range of motion with keyboard and mouse requires extra controls which even experienced gamers are probably unaccustomed to. For this experiment, “tilt” controls were added to the keyboard, allowing users to shift their head left or right. Users were instructed of this additional change and had it demonstrated to them, but it added a level of difficulty and thought to the process. Comparatively, users using the HMD could just move as they would in real life while attempting to find a sound source. This combination of intuitive controls and wider range of freedom would hopefully lead to a more straightforward experience for the user that feels more natural.

The audio spatialisation framework was changed for this experiment as well. While the last sound location experiment used the Oculus HRTF spatialiser integrated in Unity, this experiment made use of the SteamAudio plugin developed by Valve, discussed previously. While the Oculus HRTF spatialiser provides spatial audio controls using Unity’s audio source objects, SteamAudio provides a much wider range of options and choices and seems to have a stronger background for use in real-time applications.

The framework is implemented by adding a Steam Audio Manager to the Unity scene and adding Steam Audio components to objects within the scene which will produce sound. This is in addition to the default sound source component built into Unity. The Steam Audio framework adds a spatialiser to the Unity Editor settings, overriding the Unity built in spatial audio settings (Fig 3.11). The settings for each individual sound object can then be tweaked using either of the audio components. For more complex environments and sounds, additional scripts are used for handling audio occlusion, pre-baking sound from sources and level geometry that don’t change, or various other more specific settings and changes. For this experiment, just the 3D sound settings and HRTF Interpolation are used.

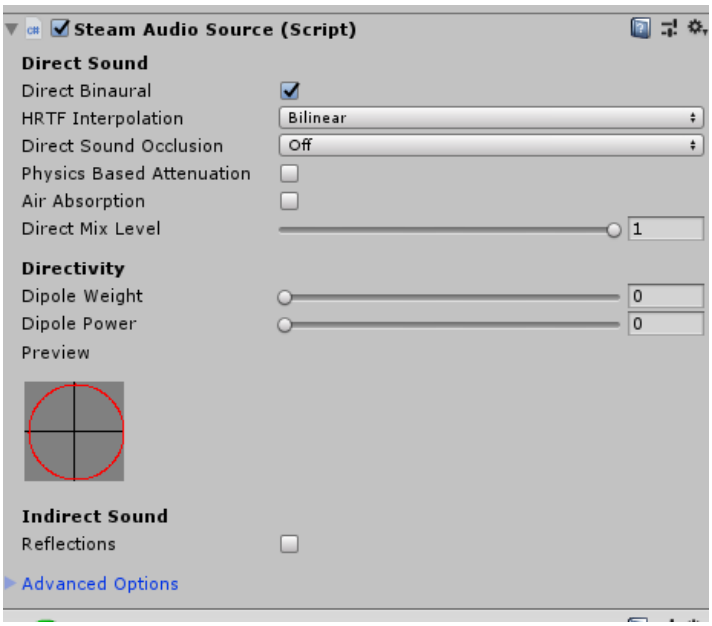


Fig 3.11 Steam Audio settings

3.3.3.2 Environment Limitation

One limitation of this experiment design is the relatively simple environment in which the experiment takes place. Compared to more complex real world or game environments where there are objects and geometry for the sound to be diffracted or occluded by, there are no obstacles for the sound to interact with. The realism of a sound and the effectiveness of its localising can be greatly affected by the environment it is within, with many sound spatialisers including real-time occlusion (Valve, 2016). An environment with very little for sound to interact with is something that is not common within games or the real world. However, there were reasons for the decision to focus on a simple environment.

The experiment was a static situation, where the only movement the user had was to rotate their head. As the focus of the experiment was on the camera/head movements and the sound location rather than any locomotion in the environment, it was simpler to have a static user position. This also reduces the learning barrier for users, as there is a minimal number of controls to learn, potentially levelling the playing field between highly

experienced gamers and lower experienced players. Because of this static environment for the experiment, any interaction between the audio and objects would also be static. However, a version of this experiment with moving sound sources would need to consider more complex environments, as they can have a greater impact on the perception of the sound in that situation.

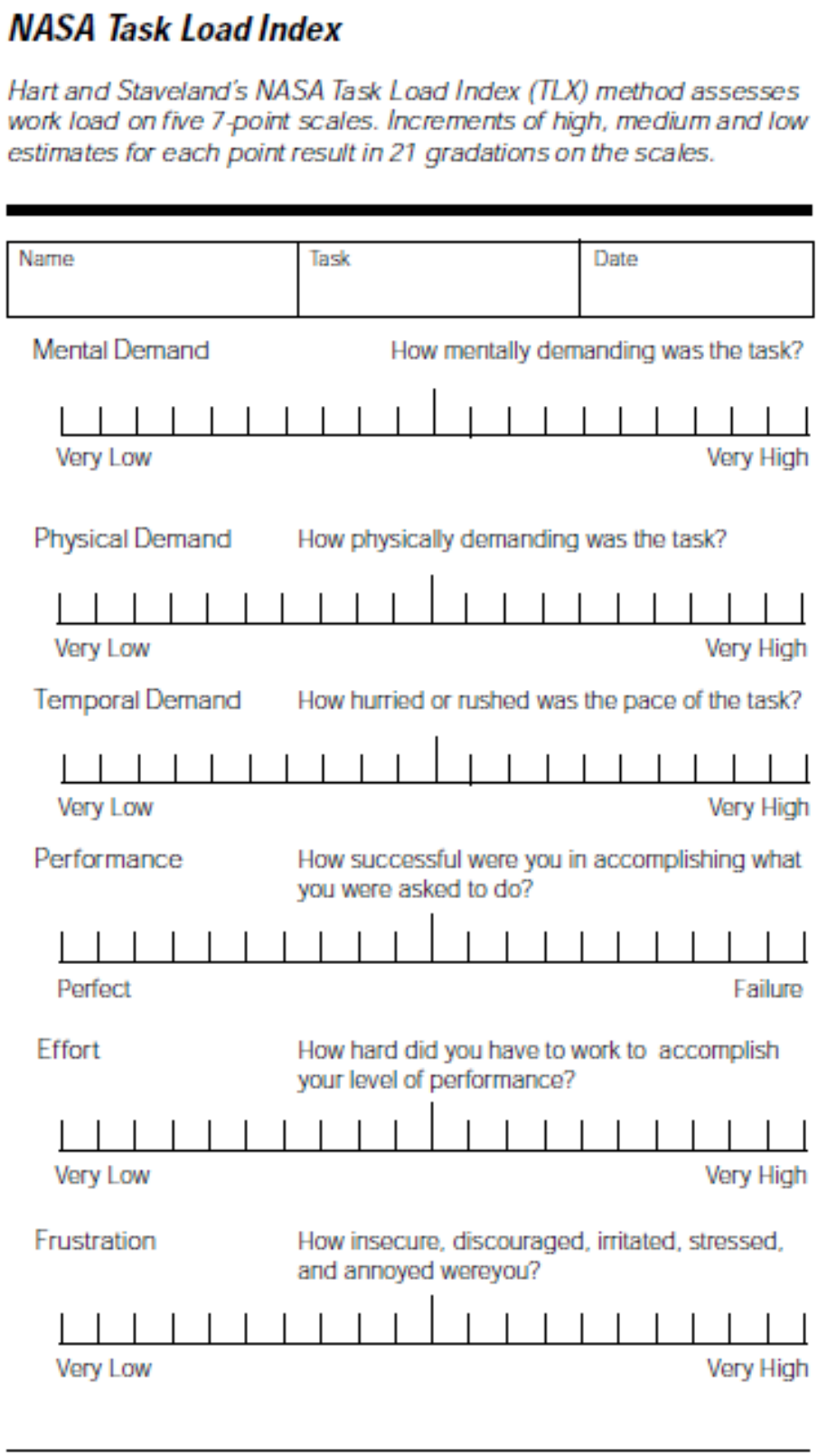
Also, with the growing popularity of the metaverse, we are going to see more VR and XR programs developed for low-experience users and with simple, focused functionality, such as townhalls and business meetings (Tu, 2021). Within these environments, controls are going to be more simplified and natural through the use of VR, and the environments are likely to be less complicated and more pre-planned, especially for use on more portable hardware. For these situations, the benefits and effects of VR and spatial audio in a simple environment are more beneficial.

Finally, some similar experiments examining spatial audio in VR have also made use of simple environments (Kuppanda et al, 2015; Mendonça et al, 2012; Moraes et al, 2020). Many of these experiments made use of simple environments as the focus was on the user being able to locate a sound source within view, at set distances, in a plain environment. Examining the impact of sound occlusion and refraction in complex environments is a different type of experiment, and one that is also important. Use of complex environments is a further step that could be considered when expanding upon the experiments performed as part of this work.

3.3.3.4 Qualitative Data Gathered

A variety of qualitative data was gathered as part of this experiment to supplement the quantitative performance data collected throughout the experiment. The two primary ways of gathering qualitative data was through a user questionnaire and interview presented after they performed the experiment fully, as well as a NASA-TLX survey they were asked to fill out (Fig 3.12). The NASA Task Load Index is a “widely used, subjective,

multidimensional assessment tool that rates perceived workload in order to assess a task, system, or team's effectiveness or other aspects of performance" (NASA, 2010). The questionnaire and interview were aimed at getting direct user feedback on the tasks completed, and their opinions on the differences between the two experimental control schemes. The NASA-TLX was aimed at getting their opinion on the overall difficulty of the tasks they were asked to complete.



NASA Task Load Index

Hart and Staveland's NASA Task Load Index (TLX) method assesses work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.

Name	Task	Date

Mental Demand — How mentally demanding was the task?

Very Low — Very High

Physical Demand — How physically demanding was the task?

Very Low — Very High

Temporal Demand — How hurried or rushed was the pace of the task?

Very Low — Very High

Performance — How successful were you in accomplishing what you were asked to do?

Perfect — Failure

Effort — How hard did you have to work to accomplish your level of performance?

Very Low — Very High

Frustration — How insecure, discouraged, irritated, stressed, and annoyed wereyou?

Very Low — Very High

Fig 3.12 NASA-TLX Survey, (NASA, 2010)

3. Methodology and Experimental Design

Where the first experiment asked for user information within the application itself, this experiment gathered that information through the questionnaire. Users were asked for their name, age, their experience with video games or virtual environments, and if they had any auditory or visual complications. Age is an interesting factor as there can be connections between age and hearing in reaction time, with older people having slower reaction times (Fozard et al, 1994; Der et al, 2006; Deary et al, 2005). Their experience with video games or virtual environments is a very important factor to consider for a few reasons.

Someone who is used to games and virtual environments is likely to be very comfortable with using computers, or comfortable with VR. Someone who isn't experienced with games is more likely to be slower using a mouse and keyboard for navigation in a virtual space, as it is something you learn through practice. This can potentially affect performance. As well as that, people who may not be used to virtual environments can become distracted, especially when it comes to Virtual Reality Novelty. When it comes to VR control schemes, especially more simplistic ones like this experiment, it is possible that the gap in skill can close, as moving with motion tracking is more natural than with a keyboard or controller.

However, there could be a benefit to a lack of experience. People who are unused to games may not have any preconceived strategies or concepts about virtual spaces. It was acknowledged earlier while looking at Counter Strike Global Offensive's implementation of HRTF that users often feel that more realistic audio sounds strange to them in a gaming environment. While switching to a new sound type after years of practice with a different type can cause an issue for a player, it is possible that people with only experience with simplified audio in virtual environments might feel a strange quality to the sound that can distract them. They also may be used to searching for a sound source in a traditional way using a mouse and keyboard, which might not be as effective, especially in a VR situation. Someone without experience may not stumble at this change in quality and may rely on their natural thoughts of how to find the source of a sound, which is not only more intuitive in VR, but can benefit greatly from properly spatialised audio.

Users were also asked a series of questions about their experience during the experiment. The questions are summarised in Table 3.3.

Question	***Rationale***
Was it harder to locate the sound sources in one of the experiments?	The first question is aimed at difficulty, if either of the experiments felt more difficult. Since the only difference in the experiment is the control scheme used, we are checking if the user felt like the control scheme caused any great difficulty change. Users could respond with VR, mouse and keyboard, or neither.
Was it more natural to find the sound sources in one of the experiments?	The second question asks about whether either method felt more natural. This is looking for whether users felt that either control scheme felt more intuitive or natural, to see if there was a difference in perceived intuitiveness to either control scheme. Users would again answer either VR, keyboard and mouse, or neither.
Was it difficult to tell where the audio was coming from in the environment?	The third question asks if it was difficult to tell where the sound came from in the environment. This is aimed at understanding perceived difficulty of the task for the user, regardless of control scheme. Users could write their own answer to this question.

Which control scheme did you prefer for this type of task?	The fourth question is about which control scheme they preferred. This is an important indicator, as people may have felt that one control scheme was more difficult, or less natural, and still prefer using it. As mentioned above, people will focus on what they enjoy more rather than what may be statistically better.
Any other feedback on the experiment?	The final question asks for any other feedback users had on the experiment, potentially providing information not gathered from the other questions or informing future experimental design.

Table 3.3 Feedback Questionnaire

The goal of these questions was to gauge the user's opinion on the experiment in contrast to the quantitative statistics of their performance. For something human focused, it is important to look at a user's feelings and perceptions rather than just performance (Wienrich et al, 2018). While an improvement in performance is an obvious indicator, it can be reduced or improved by a variety of factors, and when it comes to an experiment like this, practice and training could be necessary to notice a strong difference in performance over time. Often it can be just as valuable to have a user state that something "felt better", "felt more natural", "made sense" or "felt easier". Even if it is based on no performance difference, improving user experience is also a hugely positive thing, especially in serious games, entertainment, or VR as there is evidence that people will choose something that is less efficient if they enjoy it more (Martel et al., 2015; Ashdown et al., 2005).

In the NASA-TLX survey, users are asked to rate their feelings on a series of subjective scales describing the task they had been asked to perform. These scales measure Mental Demand, Physical Demand, Temporal Demand, Overall Performance, Effort, and Frustration Level. A traditional NASA-TLX involves pairwise comparisons of the individual subscales after rating, but this can be dropped, leading it to be called a "Raw TLX". The goal is to have the user rate the perceived demand on them from the task they completed based on the different scales. For this experiment, it was added to determine how difficult the experiment was, and in what particular ways.

3.3.3.4 Quantitative Data Gathered

A variety of quantitative data was gathered as part of this experiment to measure user performance in a variety of ways. All these factors are tracked by the experimental system throughout the user's task completion and saved as a CSV file on the computer's hard drive with the user's name that they were asked to input, as well as whether or not the experiment was performed in VR.

The first factor tracked is the *time taken to complete the experiment*. A timer starts when the user begins the experiment, and each time the user correctly identifies a sound source, the program tracks the difference in time between the current time and the last timestamp created, giving a value in milliseconds for each sound source task. This leads to a total of nine time readings, from three tutorial sources and six test sources. This time gives an indication of how long the task took in general, whether specific sources took a longer time to find, etc. This becomes more important when combined with some of the other data gathered.

The second factor being tracked is the *number of mistakes the user makes* during each sound source task. This again works by tracking the number of incorrect cubes clicked on before the user finds the correct cube for that particular sound, then resetting and repeating throughout the experiment. This is valuable as a form of user performance

alongside the time taken. If a user takes a long time to find a source, but makes few mistakes, they are likely being patient and searching closely. On the other hand, if a user finds a source quickly, but makes a comparatively high number of mistakes, they are likely finding the rough location of the sound and then brute forcing the solution.

The third factor being tracked is *which cube is making the sound* during each task of the experiment. This is important as each cube has different potential difficulty for the user, and by attaching mistakes and times to each cube, we can better discover which cubes caused the greatest difficulty for the user. There are 4 many categories the cubes can fall into:

1. *Higher Elevation to the user*
2. *Same Elevation as the user*
3. *Lower Elevation to the user*
4. *Directly Above/Below*

Different cubes can fall prey to different issues, such as the front-back problem and the difficulty inherent in pinpointing cubes' elevations. Therefore, being able to track what cubes are currently sounding and the mistakes and times attached to each of them can further support these potential difficulties.

The final factor being directly tracked is *the user's rotational value* throughout the experiment. Five times per second, the system takes a snapshot of the user's current rotational value and stores it. When the user finishes the experiment, this series of numbers is stored with the other factors tracked. The goal with this is to be able to replicate the user's movements as they perform the experiment to look for more detail, such as searching methods, areas that caused most issues, possible user distraction, or

differences in movement between experiment types. For example, we can better analyse when people use tilting controls in keyboard and mouse compared to the more natural head tracking tilt of VR, or we can see if people pan their view back and forth over the environment as they attempt to roughly locate the sound source. We can both replay the user's movements after the fact and look at individual sections of rotational data more closely, such as examining the user movements directly around a correct guess, or similar possibilities. Being able to analyse strategy, confusion, or frustration after the experiment has concluded can provide a much more comfortable period of analysis, as well as being significantly more cost effective compared to video recording each user's in-game performance.

For analysis of the quantitative data gathered, we used a two-tailed Welch's Two Sample T-Test as our primary analysis tool with a confidence interval of 95%. This test seems to be the best fit for our type of data gathered and has been used in some previous experiments similar to this (Kern et al, 2020; Hernandez, 2009). As well as this, we used a Shapiro-Wilk Normality test to check for normal distribution and an F-Test to test for equal variance in the populations.

3.4 Research Constraints

There were a variety of obstacles and issues encountered over this research, as is to be expected.

3.4.1 Experiment 1: HRTF vs Panning Audio Experiment

As this was the first experiment designed and performed, there were a number of issues related design, and due to the fact that the goal of the overall research was still at an early stage in the iteration process.

3. Methodology and Experimental Design

Firstly, in terms of data gathered, there were some oversights, such as that there was no tracking of user movement, which was implemented for later experiments. While not as key a difference between HRTF and Panning compared to different control schemes, it would have provided useful information about whether or not people searched differently depending on how the audio was presented to them. Due to HRTF providing a more realistic sound style, people could use a different search technique as they move their camera attempting to locate the sound. It also provides more detail on particularly long sound searches. A user who spends a large amount of time on a single sound source may have gotten frustrated while searching and started simply guessing all other squares. They also could have gotten distracted, which would have shown from the user not moving their camera at all for a period of time.

The most significant issue with this experiment was the attempt to perform it remotely. While the portability of something like Unity is convenient, performing something such as this remotely creates more issues than benefits. For the remote portion of the experiment, people who chose to take part were sent an executable for the two experiment types, an instruction sheet detailing what needed to be done and what equipment they might need. They would then perform the experiment, which would log the data and send it to a database set up online for the experiment, where it could be accessed by the creators. This would allow for more research subjects to be reached, especially as the experiment took place when there weren't many students around.

However, this led to issues when performing an experiment remotely, you can't guarantee the conditions of the experiment environment. It is possible the user will get interrupted part way through the experiment. It is possible that the experiment won't work correctly for them. They may not have the required equipment and not state that to the creator, or they may have unsuitable equipment, such as a trackpad instead of a mouse, or slightly damaged headphones. There is also no guarantee that the user will read the instructions clearly. With this experiment, due to the importance of sound, a loud environment would also be unsuitable. The biggest issue with this particular run of tests is that many users only performed one of the tests or performed the same one

twice. While it is possible to compare results from all HRTF and all Panning, it doesn't provide nearly as good information as comparing individual performance between both tests. Comparing the HRTF cohort and the Panning cohort can lead to a bigger effect from particularly strong or particularly weak individual performances. Ideally this experiment would have been repeated at a later stage after the other experiments were performed, but the impact of COVID-19 mentioned later this section and the timeframe of the research made this difficult.

This particular test of remote experimentation worked well as a learning experience. While some tests may be doable remotely, it simply doesn't suit certain experiments, and if it does, there needs to be a much stronger supervisory and one-to-one element between the subject and the organiser. Some similar experiments have also stated issues with remote experimentation (Andersen et al, 2021).

3.4.2 Experiment 3: VR vs Traditional Control Scheme Experiment

There are obstacles and issues that arise with this experiment. First, VR experiments are inherently more restrictive and complicated. Due to the expensive and bulky equipment, experimental times must be planned out in advance. It isn't possible to do multiple experiments simultaneously without having multiple pieces of expensive and heavy gear, such as headsets and computers powerful enough to handle VR. There is also a time cost with setup of equipment at the start, getting each person probably set up with the equipment so that it isn't blurry, too loose, too tight, and making sure that the equipment is correctly cleaned in between uses. There are already horror stories of events with VR where someone has had an eye infection and unknowingly passed it onto a large number of people due to improper cleaning of equipment. We had access to an HTC Vive and a laptop capable of handling VR, but compared to a traditional computer experiment, this is much more time intensive. This led to experiments taking more time to complete as it could only be performed one at a time, with time in between for cleaning

of equipment. This made the experiment more time intensive for the experiment organiser and reduced the total amount of experimentation that could be completed.

One issue occurred with the tracking of data from the experiment. The rotational data of the user which was gathered as a quaternion value[8], was trimmed as it was stored in the CSV due to the Unity ToString method. This issue wasn't noticed as it didn't affect the information being looked at in the Unity console window, and the output quaternions weren't noticed as rounded during initial testing. Because of this, the first batch of experiments have a lower granularity of rotational movement than was meant to be stored. This makes some analysis of small movements very difficult, but some analysis of greater search patterns, head tilts, and similar things can be performed. The issue was fixed for any later experimental groups.

3.4.3 COVID-19 Pandemic

A major issue for all in-person research over the last few years has been COVID-19 pandemic. Experiments like this can't be performed without people, or remotely. In theory people with their own VR kit could perform the experiment remotely but doing so has not only the issues mentioned in the previous experiment with remote testing, but the issues are made worse by the inability to properly test the experiment, as a difference in equipment could cause an issue in the experiment, or it might not work at all. Not only is there a risk with large groups or people sharing equipment (especially something face-mounted like an HMD) (Moore et al, 2021), but it was illegal to have any gatherings or meetings for the majority of this time. Between restrictions and the risks to the health and wellbeing of participants and experiment organisers, as well as the shutdown of most in-person learning, which was a primary source of participants, experiment results were limited to what was gathered before the beginning of the pandemic. Hopefully there will be more opportunity to repeat these experiments in the future, after this unprecedented situation is fully under control. There is also a possibility to consider how to better perform

[8] Quaternions in Unity are used to represent rotations and are based on complex numbers.

remote experimentation, especially as VR equipment becomes more commonplace in consumer homes.

Related to this, Ireland was at Level 5 Lockdown for the majority of time between March 2020 and November 2021, meaning it was impossible to arrange more test phases to get more data (Citizen's Information, 2020). As mentioned, this made gatherings of multiple individuals illegal for the majority of that time, which further made experimentation impossible. While there were some short periods of time throughout where restrictions were partially lifted, it was not enough time to organise a safe experiment environment for participants, especially if while aiming to reduce in-person interactions.

3.5 Technologies chosen

3.5.1 Unity for building environments/game controls

For the experiments developed as part of this research, the virtual environments were developed using the Unity Game Engine. While some of the specifics of this game engine have been discussed above, we will reiterate some key points in specific reference to the experiments created here. There are several well-known game engines that can fit a project such as this, such as Unreal Engine, Unity, or Godot. While each has its advantages and disadvantages, only one can be used in the end.

In terms of existing academic research, a variety of engines are used. Some older research used the Torque Game Engine (Moloney et al, 2004; Mateevitsi et al, 2008; Wyeld et al, 2007), as there weren't a lot of free 3D game engine options before more recent years. Around 2009/2010, Unity released a free version of their engine, and in 2015, Unreal also made their engine free to use. While there are caveats to their use, in terms of academic work, they can be used without cost. Between Unreal's history as a AAA game engine and Unity's high quality and early switch to a free model, these tools quickly became very common use in academic research for games and serious gaming

(Sharma et al, 2014; Klippel et al, 2020; Jäger et al, 2017). While some research continues to use lesser-known engines or self-developed systems, the majority has gravitated towards these two options, for simplicity, their integration with other systems, and the amount of learning resources and tutorials available.

The wealth of knowledge is a big reason why Unity was selected for this work. Because Unity has had a free version available for several years longer than Unreal, there are a lot of indie resources and free online tutorials developed over those years. The accessibility of this software to any user made it one of the most popular choices for indie and personal projects, which lead to more people making learning resources available for free. While not a perfect example, a Google search for Unity Tutorials returns about 125,000,000 results, while a search for Unreal Tutorials returns about 25,900,000. Obviously, quantity doesn't equal quality, but especially a few years ago, there just wasn't an equivalent amount of available learning materials.

In terms of VR support, Unity has great out-of-the-box integration tools for many popular VR technologies. Unity has worked directly with Oculus to provide integration plugins for Unity developers to develop VR environments and experiences quickly and easily (Wessels et al, 2020). They have also provided in-engine support for Microsoft Mixed Reality technologies, and support other VR technologies such as Google VR, HTC, and Valve's OpenVR through third party platforms and plugins (Unity, 2021). All of these support making working with VR much simpler and allow for focus on other areas of research.

Similarly, to its support for VR technologies, Unity has also supported growing improvements in spatial audio technologies, and many third party tools have been developed with Unity integration in mind. Unity previously provided panning audio and some simple spatialisation through its own toolset, before integrating Oculus HRTF and Microsoft HRTF directly into the engine. This shows the engine's awareness of the importance of spatial audio, and considering the companies it has integrated with, the inherent connection between VR experiences and spatial audio. As well as this, there are

third party spatial audio technologies that integrate directly with many game engines, including Unity. For the purpose of this work, Steam Audio was used, which will be discussed below.

3.5.2 HTC Vive for Virtual Reality headset

As the main experiment involved a VR element, a choice had to be made between all of the available options of VR kits. Each device has its own pros and cons, and we will detail some of the factors that led to the decision to use the HTC Vive for this experiment.

The first factor is if the kit can integrate with Unity, as Unity was already used for environmental development before the decision to look at VR. Luckily, most major VR devices have Unity integration as of more recent Unity versions. Oculus Rift has the most integrated support with Unity due to partnerships between the two companies. This is also why Unity has the Oculus Audio Spatialiser built into the editor. However, it is not difficult to get HTC Vive, Google VR, or other VR kits working with Unity through third party plugins, such as SteamVR. This plugin, discussed in more detail later, allows an HTC Vive to communicate with Unity and aids in development of VR experiences and games.

Obviously cost and availability is a major consideration when it comes to equipment usage. While there are some cheaper options for VR such as Google Cardboard at $15, many fall in the range of $500-$800, especially if you need motion tracking hand controllers, accurate head tracking, or other specifics (Google, 2021; Vive, 2021; Oculus, 2021). For this research we had access to a HTC Vive, which helped inform our choice, especially as the HTC Vive is one of the more expensive kits, providing a wealth of features, as well as hand motion controllers.

3.5.3 Steam Audio for audio spatialisation

As one of the primary focuses of this research thesis is spatial audio, having a robust spatial audio framework is of key concern. A few spatial audio options were examined, such as the possibility of creating a spatial audio framework from scratch, using the Microsoft or Oculus HRTF packages built into Unity, and finally, Steam Audio, which was ultimately chosen as the audio package for this project.

Steam Audio is a “full-featured audio solution”, created by Valve Software, creators of the Steam digital marketplace and many popular games such as Counter Strike, Portal, and Half Life (Valve, 2021). Valve as a company have been looking at spatial audio for a few years, choosing to develop a HRTF audio option for their game Counter Strike: Global Offensive several years after the release of the game, as looked at in Chapter 2. In 2017, they released Steam Audio, originally as a Unity plugin and a C API for other game engines, but it has quickly grown to support a variety of platforms (Windows, Mac, Linux, Android) as well as game engines and software packages (Unity, Unreal Engine, Fmod Studio). It was built with a focus on games and VR experiences, providing HRTF binaural audio rendering, and audio occlusion and reflection, all in real time. It has also been developed for good performance, due to its focus on games and VR experiences.

Compared to the Oculus and Microsoft HRTF packages integrated into the Unity Editor directly, Steam Audio provides a lot more control and flexibility. The integrated HRTF options provide some new settings to the existing Unity sound objects, but it doesn’t seem to provide deeper control, or have nearly as many functions as Steam Audio does. Since Steam Audio uses new audio objects, there is a lot more that can be seen up front in terms of options and changes.

Obviously, an ideal situation would involve making an audio framework from scratch, as that gives the creator the deepest level of editing and control. However, there are issues with this approach. It requires a much larger time and effort investment from the researcher, especially if they don’t come from an audio background. It can be harder

to find issues tied to the audio package, as it is quite difficult to replicate someone else's work when using custom frameworks. Similar to how using game engines makes certain fields of research or work more accessible to the largest group of people, using audio frameworks can allow more and more people to engage with the audio field and learn more about it. Hopefully, in a similar way to how certain game engines have become more popular in research, we can see more researchers working with publicly available tools, which will allow for great replicability of research and greater discussion between people.

3.6 Conclusion

Throughout this chapter we have discussed the different experiments that have been performed as part of this research project, how the experiments were created, how they addressed our research questions, and why they were designed in the way they were. We compared these experiments to similar experiments in existing literature as well as highlighting their unique coverage. We examined our choice to gather both qualitative and quantitative measurements within these experiments, discussing why it is important to measure not only the time taken within the experiments, but the number of mistakes made, where mistakes took place, and how the location or type of mistake made can lead to different conclusions. We also examined our analysis of user movements within the experiments and why they are important to monitor, such as how users can use different movement patterns depending on the presence of sound or potential experience, or how different search techniques can be used depending on the control scheme. This different approaches to movement highlight why measuring this information for analysis is important to our research. The measured factors and the ways of gathering qualitative data were discussed and compared to other experiments, including our use of both a user questionnaire and a NASA-TLX survey to gather both user opinions and perceived task load. We also covered the constraints on these research experiments, such as the impact of COVID-19 and problems faced with remote experimentation. Finally, we discussed the technologies and equipment used in this research, primarily the use of Unity and Steam Audio for the creation of the experimental environment and the use of the HTC Vive as the VR control method. We compared these choices to other potential game engines and audio spatialisers that currently exist.

4 Experimental Results

4.1 Introduction

In this section, we will be covering how our two primary experiments relate to specific research questions, examine the results gathered from each experiment, and draw conclusions from these results. Our two experiments being discussed are Experiment 2, *the Maze Navigation Experiment* aimed at measuring how sound can affect a user's awareness and ability to navigate in a virtual space, and Experiment 3, *the Sound Location Experiment* which looked at how varying control schemes can affect a user's ability to find sounds in a virtual space. We will discuss the subject cohort within each experiment, detailing the relevant information gathered about the subjects. Next we discuss Experiment 2, examining the quantitative data gathered on user performance throughout the maze, as well as movement patterns we saw develop among the users. We break this down into not only comparisons of audio and non-audio runs of the experiment, but also how experience seemed to have an impact on people's improvement in performance. Finally, we will discuss Experiment 3, where we examine the quantitative measurements of users' performance throughout the experiment, their head movements tracked throughout and the possible reasons for movement changes, and the qualitative data gathered from the questionnaire and NASA-TLX survey. We examine this qualitative data to look for evidence of user preference towards a particular control scheme, perceived difficulty differences between the two control schemes, and also the perceived workload difference between the two control schemes.

As outlined previously, the main research questions being asked in this thesis are summarised on Table 4.1.

Research Question	Relevant Experiment
"Is there a performance difference between fully spatialised audio and traditional panning audio, while searching for audio sources?"	*Experiment 1*
"Is there a performance difference between different control schemes while searching for audio?"	*Experiment 3*
"Does spatial audio improve a user's awareness of their surroundings in a virtual environment?"	*Experiment 1, Experiment 2, Experiment 3*
"Does the control scheme implemented for a virtual environment change how a user interacts with virtual sounds?"	*Experiment 3*
"Do different control schemes affect a user's enjoyment or perceived performance?"	*Experiment 3*

Table 4.1 Research Question Table

4.2 Subject Cohort

The subject cohort for the maze experiment was a group of 6 students, aged between 21 and 32 years old, with an average age of 26. None of the subjects reported any auditory or visual issues beyond need for glasses. The group contained 2 individuals identifying as female, and 4 identifying as male. With regards to experience with virtual environments and video games, 2 reported low experience, 1 medium experience, and 3 high level of experience. The experiments were performed over one day, with each person taking approximately 120 minutes to perform the experiment. The full details of the participants are detailed in Table 4.2.

Participant #	Age	Gender	Experience	A/V Issues
1	27	M	MODERATE	N/A
2	24	F	LOW	N/A
3	25	M	HIGH	N/A
4	24	F	LOW	N/A
5	24	M	HIGH	N/A
6	32	M	HIGH	N/A

Table 4.2 Maze Experiment Subject Details

The subject cohort for the sound location experiment was a group of 7 students, aged between 21 and 32 years old, with an average age of 25. None of the subjects reported any auditory or visual issues beyond the need for glasses. The group contained 4 individuals identifying as female, and 3 identifying as male. With regards to experience with virtual environments and video games, 2 reported low experience, 2 medium experience, and 3 high level of experience. The experiments were performed over one day, with each person taking approximately fifteen-twenty minutes to perform the experiment including preparation, breaks and the experiment itself. The full details of the participants are detailed in Table 4.3.

Participant #	Age	Gender	Experience	A/V Issues
1	27	M	MODERATE	N/A
2	24	F	LOW	N/A
3	25	M	HIGH	GLASSES
4	22	F	MODERATE	N/A
5	21	F	LOW	N/A
6	32	M	HIGH	N/A
7	24	F	HIGH	N/A

Table 4.3 Sound Location Experiment Subject Details

4 Experimental Results

Experience with gaming and virtual environments can make a significant difference in this experiment, as it involves moving and navigating through a virtual space using a mouse and keyboard. Someone with no experience doing this can struggle greatly with learning the controls, while someone with a lot of experience will find basic navigation very simple. They also may be more used to listening for audio cues in a virtual environment compared to beginner users and may be better able to react quickly to changing situations. We specifically asked for experience with games or virtual environments to include people who are used to VR but not games, or similar use of virtual environments. A visual of the experience breakdown is shown in Fig 4.1.

For these experiments, the main factors prepared for were any potential audio/visual issues which could affect the ability to perform the experiment and getting a variety of experience with games and virtual environments. With an audio focused experiment, any issues with deafness or reduced hearing ability could make it difficult or impossible for a user to perform the experiment and could lead to unexpected results. With regards to the user's gaming experience, getting a variety of gaming knowledge is important as experience levels could potentially have a positive or negative effect on the performance of the experiment (Kulshreshth et al, 2013; Martel et al, 2017). Being used to video games could make understanding the task more straightforward, having better performance with mouse and keyboard controls, and generally being more comfortable with the style of experiment. Alternatively, there can be learned behaviours which may influence their techniques and movements when using a VR setup. People with very little experience may struggle slightly more with mouse controls, but they also may more quickly adapt to a naturalistic VR control setup. They won’t have any learned behaviours from games and may rely on their natural intuition when it comes to moving their heads and searching for sounds. However, this area has yet to be studied as far as we have seen.

One drawback with this cohort is the size. It is a relatively small group, which can lead to issues with accuracy and the impact of outliers. Unfortunately, due to the VR nature of the experiments, there is limitation on where and how they can be performed,

and they must be performed individually with time in between users for proper hygiene control. Also, due to the COVID-19 pandemic, it wasn't possible to host in person experiments, especially with the usage of head mounted displays.

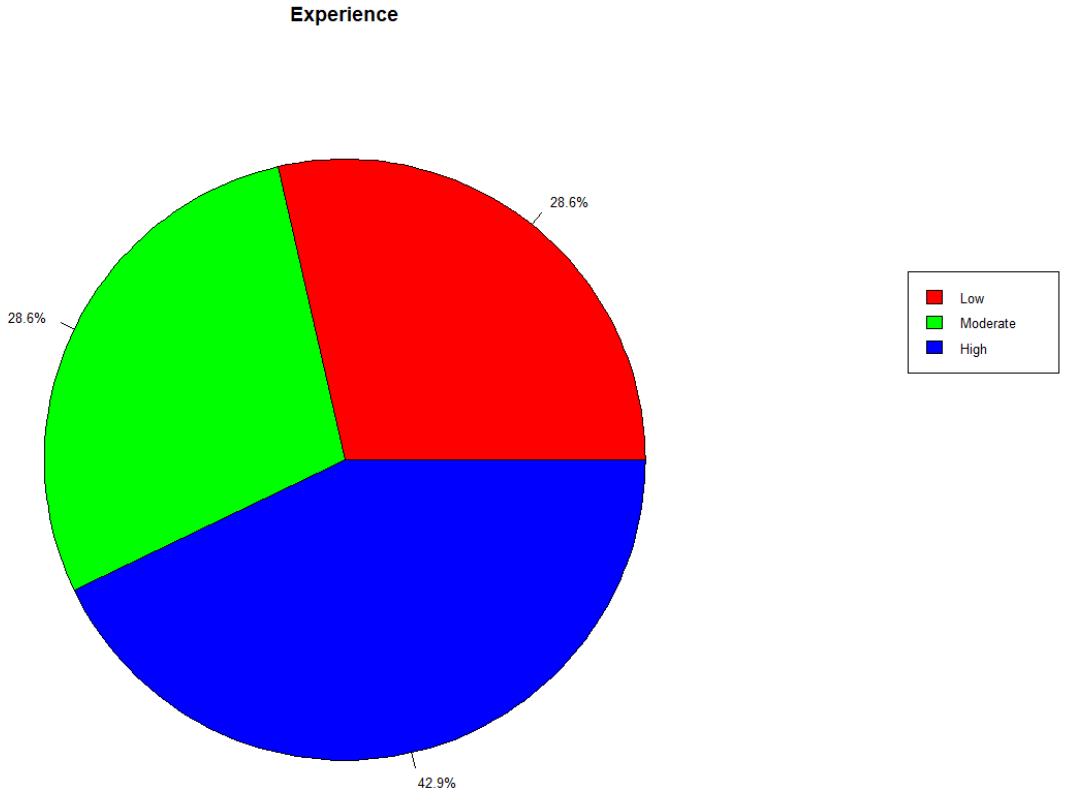


Fig 4.1 Participant Experience with virtual environments/games

4.3 Experiment 2: Maze Navigation Experiment

4.3.1 Overview

In Fig 4.2 we have an example of one of the mazes used.

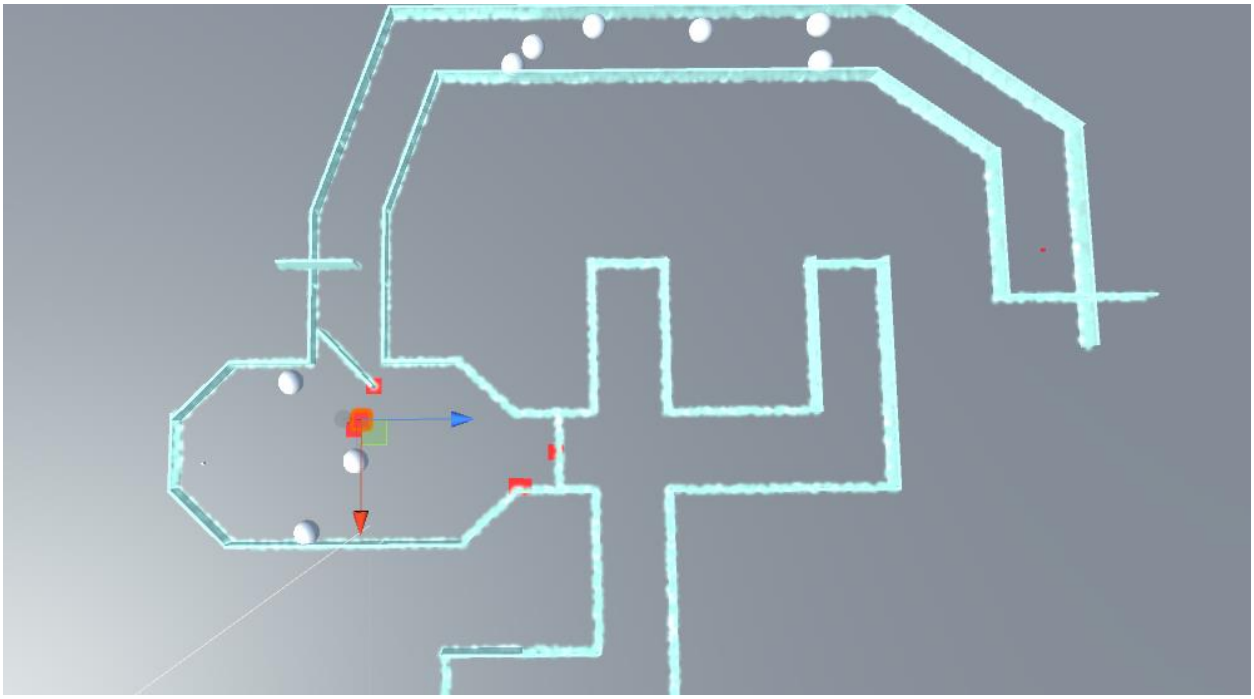


Fig 4.2 Maze Example

This experiment looked at whether or not spatial audio allow users to be aware of dynamic parts of their environment, and how much of a difference can this make to a user's pathfinding and ability to avoid obstacles? Compared to similar experiments which mostly focused on spatial audio for navigation of an environment or finding an object (Grohn et al, 2005; Walker et al, 2006), we wanted to examine the impact of audio in keeping users aware of dynamic, moving obstacles in their environment, helping to avoid these objects while moving through the maze. This experiment wanted to explore whether or not users would act differently with the presence of audio vs no audio. We analyse the

time it takes users to navigate a mixture of mazes, as well as the number of mistakes made, while having audio or not to warn them of moving obstacles in the environment. We also track the players' movements through the maze to look for patterns in navigation and where mistakes generally occurred. All users performed the first maze with audio as a practice/training section, half the users did the second maze with audio and the third maze without audio, and the remaining half did the second maze without audio and the third maze with audio.

4.3.2 Quantitative Results Discussion

In terms of user performance in the maze, two factors were measured. The factors tracked were the number of mistakes made (when they collided with objects in the environment or walls) and how long it took them to complete each maze. The performance breakdown by maze can be seen in Fig. 4.5 and Fig 4.6.

Overall performance was quite similar between both types of experiment, with a slightly better performance overall from the audio tests. The average mistakes made in audio mazes was 2.5 (median of 2), with the average mistakes in no audio being 3.83 (median 2.5) (Fig 4.7), and the average time in audio mazes was 99.18 seconds (median 102.19), with the average time in non-audio mazes being 103.63 seconds(102.99) (Fig 4.8). While this performance was slightly in favour of the audio iterations, it was not statistically significant based on a two-tailed Welch's T Test(p-value =0.832 for time taken and 0.483 for mistakes made) (Fig 4.3 and Fig 4.4). The maximum time was longer in non-audio mazes (155.82 vs 132.48) and the minimum was higher (57.22 vs 51.72). A breakdown of results can be seen in Table 4.4.

Experiment	Average	Median	StdDev	Max	Min
Audio Mistakes	2.5	2	2.95	7	0
Audio Time	99.18	102.19	34.94	132.48	51.72
No Audio Mistakes	3.83	2.5	3.37	9	1
No Audio Time	103.63	102.99	35.87	155.82	57.22

Table 4.4 Maze Experiment Results by Audio Presence

```
        Welch Two Sample t-test

data:  mazeaudiotime and noaudio_mazetime
t = -0.21769, df = 9.9931, p-value = 0.832
alternative hypothesis: true difference in means is not equal to
 0
95 percent confidence interval:
 -50.00102  41.10102
sample estimates:
mean of x mean of y
    99.18    103.63
```

Fig 4.3 Audio vs No Audio Time Analysis

```
> t.test(soundmazeaudiomistakes,noaudio_mazemistakes)

        Welch Two Sample t-test

data:  soundmazeaudiomistakes and noaudio_mazemistakes
t = -0.72908, df = 9.8265, p-value = 0.483
alternative hypothesis: true difference in means is not equal to
 0
95 percent confidence interval:
 -5.417889  2.751223
sample estimates:
mean of x mean of y
 2.500000  3.833333
```

Fig 4.4 Audio vs No Audio Mistake Analysis

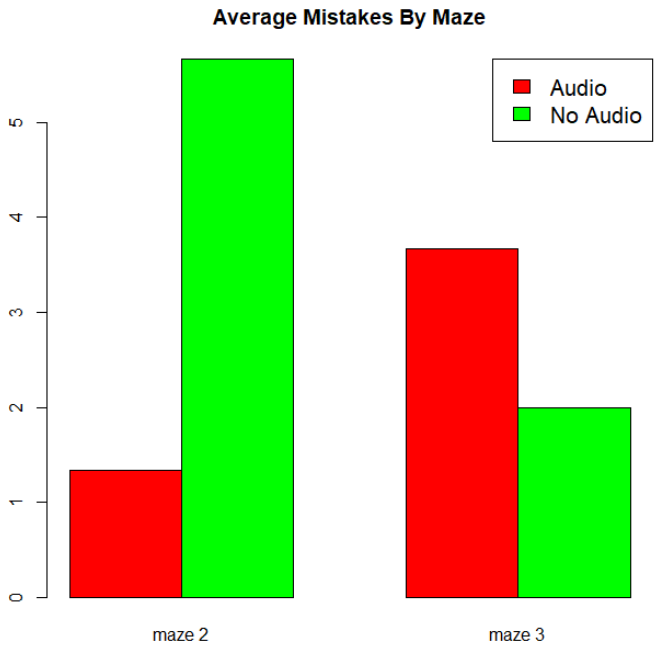


Fig 4.5 Average Mistakes Made by Maze

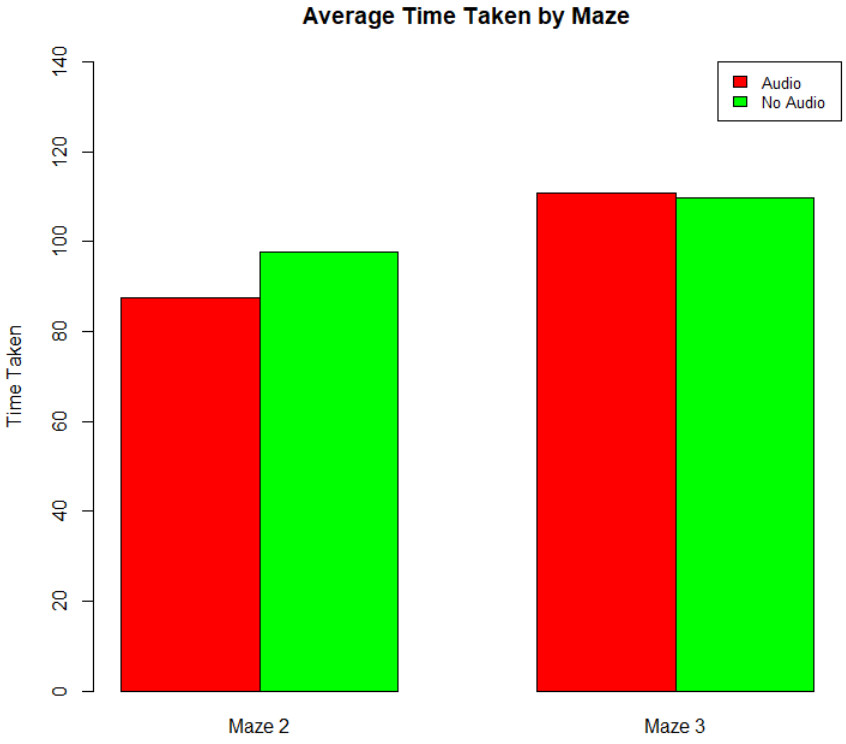


Fig 4.6 Average Time Taken by Maze

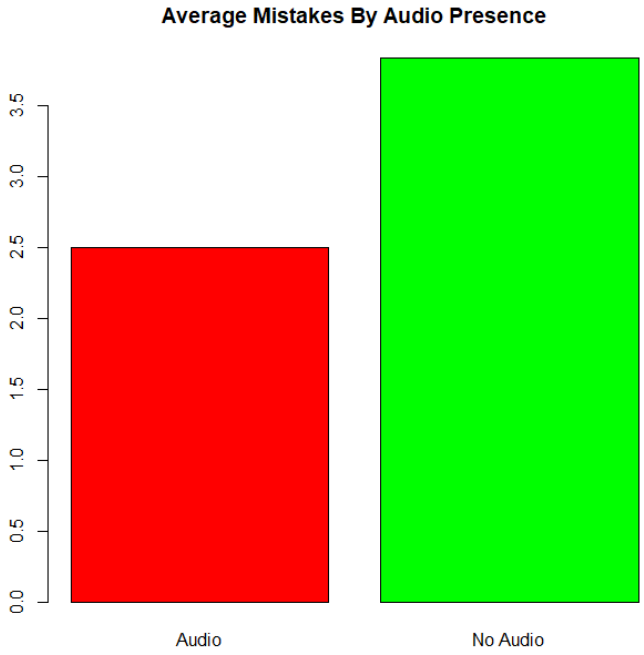


Fig 4.7 Average Mistakes By Audio Presence

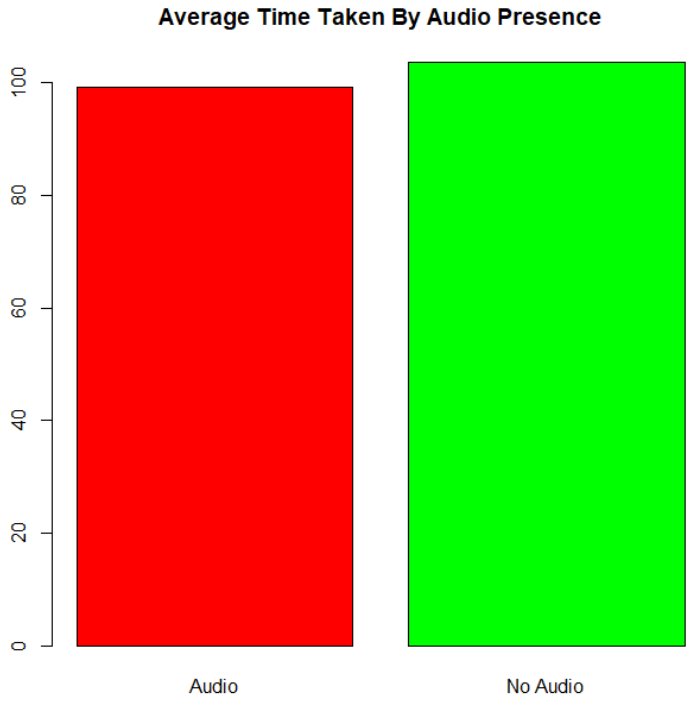


Fig 4.8 Average time By Audio Presence

What can be seen is that in terms of mistakes, every user made at least 1 mistake in all non-audio mazes, while some users did make 0 mistakes in some of the audio mazes. The range of mistakes for audio mazes was between 0 and 7, with non-audio mazes having between 1 and 9. All users who made 0 mistakes were highly experienced players, which can make them more comfortable with this sort of experiment, with two of those users also some of the fastest to complete the mazes. The third highly experienced player was the second slowest to complete the mazes. So, while experience in gaming is likely to give some advantages in terms of avoiding mistakes, it doesn't necessarily equate to speed through the maze. Having experience with gaming doesn't necessarily mean someone plays a lot of first-person games, which is what these controls most closely relate to. Those who reported having low or medium amounts of experience made significantly more mistakes, having an average of 5.667 mistakes compared to the highly experienced users having an average of 0.667 (Fig 4.10). This is a statistically significant difference after performing Welch's Two Sample T-Test(p-value=0.002358), which shows that for these sets of mazes, comfort and experience with virtual environments may have a greater effect than the presence of sound. Since the sound is mostly providing awareness of things out of view, those with less experience may be too slow to respond, or panic and not react appropriately. They are also evidently more prone to make mistakes with audio or no audio. A full breakdown of results can be seen in Table 4.5, with Fig 4.9.1 through 4.9.3 showing examples of the Shapiro - Wilk Normality Test, the F-Test and the Welch Two Sample T Tests performed.

Name	Age	Gender	Audio	Experience	TimeTaken	MistakesNum	MazeNum
1	27	M	False	moderate	106.64	7	2
1	24	M	True	moderate	125.66	7	3
2	25	F	True	low	78.72	4	2
2	24	F	False	low	99.34	3	3
3	24	M	False	high	57.22	1	2
3	32	M	True	high	75.06	0	3
4	27	F	False	low	128.92	9	2
4	24	F	True	low	131.44	4	3
5	25	M	True	high	51.72	0	2
5	24	M	False	high	73.84	1	3
6	24	M	True	high	132.48	0	2
6	32	M	False	high	155.82	2	3

Table 4.5 Individual Performance Table

```
> shapiro.test(lowexp)

        Shapiro-Wilk normality test

data:  lowexp
W = 0.90177, p-value = 0.3845

> shapiro.test(highexp)

        Shapiro-Wilk normality test

data:  highexp
W = 0.82162, p-value = 0.09114
```

Fig 4.9.1 Low-Medium Experience vs High Experience Shapiro Test

```
> var.test(highexp,lowexp, alternative="two.sided")

        F test to compare two variances

data:  highexp and lowexp
F = 0.12195, num df = 5, denom df = 5, p-value =
0.0375
alternative hypothesis: true ratio of variances is not equal to
1
95 percent confidence interval:
 0.01706475 0.87150998
sample estimates:
ratio of variances
         0.1219512
```

Fig 4.9.2 Low-Medium Experience vs High Experience F Test

```
> t.test(highexp,lowexp, alternative="two.sided")

        Welch Two Sample t-test

data:  highexp and lowexp
t = -4.9454, df = 6.2016, p-value = 0.002358
alternative hypothesis: true difference in means is not equal to
 0
95 percent confidence interval:
 -7.454582 -2.545418
sample estimates:
mean of x mean of y
0.6666667 5.6666667
```

Fig 4.9.3 Low-Medium Experience vs High Experience Mistakes

While there is a statistically significant difference between the mistakes made by highly experienced players in audio and no audio mazes (p-value=0.0286) (Fig 4.11), the sample number is small, making it difficult to make a strong claim. Compared to the previous experiment where a mixture of gaming experience was a bonus, this style of analysis may benefit more from having all users have a similar amount of experience, or simply larger groups of both low and high levels of experience.

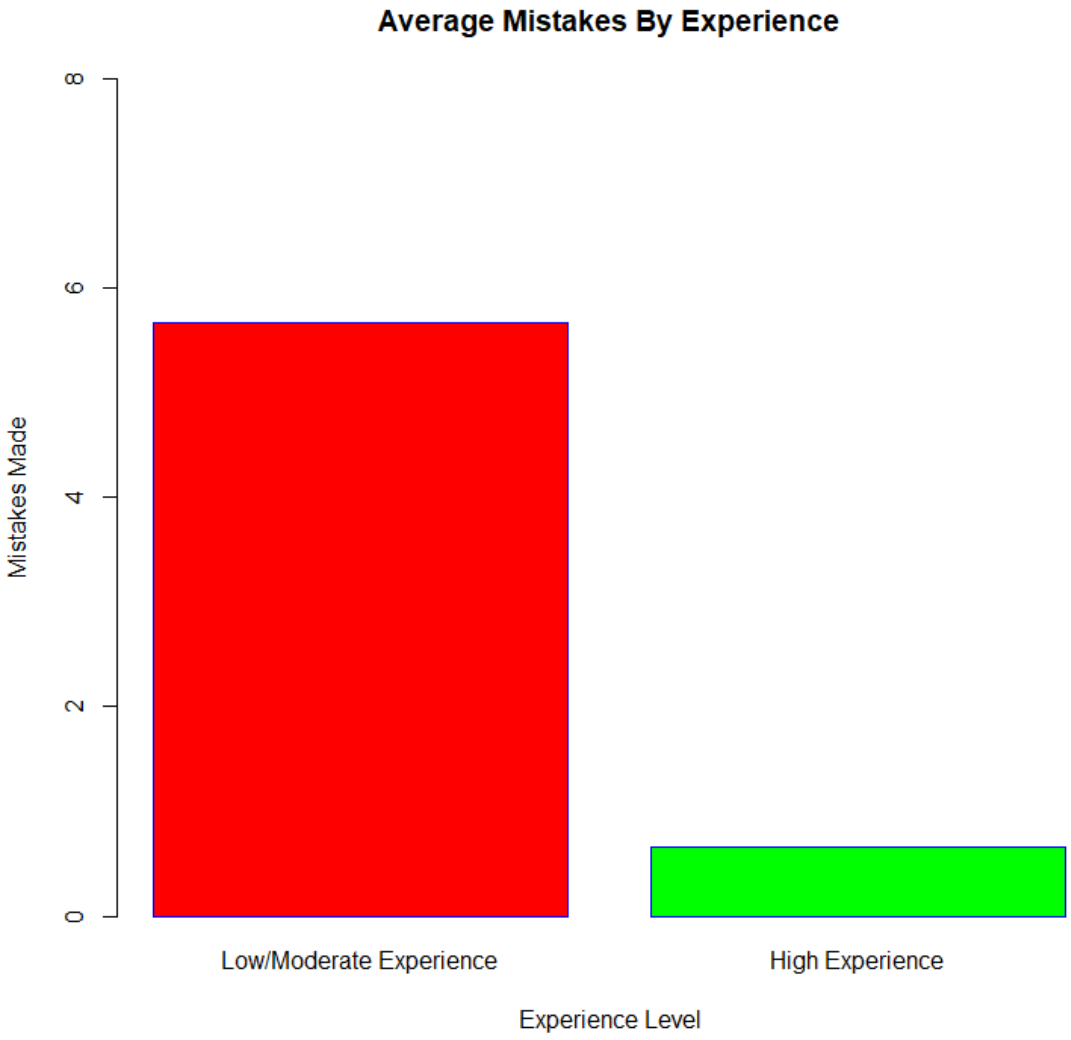


Fig 4.10 Average Mistakes by Experience

There also did not appear to be any statistically significant difference between male/female results. While there is evidence that males usually have faster reaction times (Jain et al, 2015), the longer time frame of finding the sound location in this experiment makes that less impactful. Also, as the range of ages is quite similar, the impact of age is less likely to have an impact (Deary et al, 2005). If there was a wider range of ages, this could cause more of an impact.

```
        Welch Two Sample t-test

data:  noaudio_mazemistakes_highexp and audio_mazemistakes_highe
xp
t = 4, df = 2, p-value = 0.0286
alternative hypothesis: true difference in means is greater than
 0
95 percent confidence interval:
 0.3600048       Inf
sample estimates:
mean of x mean of y
 1.333333  0.000000
```

Fig 4.11 High Experience Mistakes Comparison

4.3.2.2. User Movement Patterns

By tracking user movements through the mazes, we hoped to be able to get some insight into people's movement patterns and see if users acted differently based on the presence or lack of audio. It also lets us see where any mistakes occurred to see if certain obstacles were more difficult than others, or what situation caused the collision.

Generally, users followed similar paths to each other regardless of sound or no sound. Most users followed the fastest path through the maze direction wise, with little variation. We can see twists and turns where users moved to avoid obstacles which caused slight differences in their paths, but generally they followed similar lines. This does appear to be the optimal path, or close to it.

In Fig 4.12 and Fig 4.13, we have the paths taken through maze 3, with the first image showing each user's run as an individual colour, and the second screen splitting into groupings of sound and no sound, with sound in red and no sound in teal. We do see two of the teal users pull wider arcs at corners to avoid obstacles, where the reds tend to stay somewhat in the middle of the path as they travel through. By looking at this, it seems that there are some cases of users without audio trying to be safer in their movements, causing a change in movements compared to the audio group. This relates to our research question:

"Does spatial audio improve a user's awareness of their surroundings in a virtual environment?"

Users are taking different paths at certain sections depending on if they have audio or not. The non-audio users taking more careful paths can be representative of needing to look around more to keep track of nearby obstacles, where audio users can be riskier in their paths because they can hear any approaching objects.

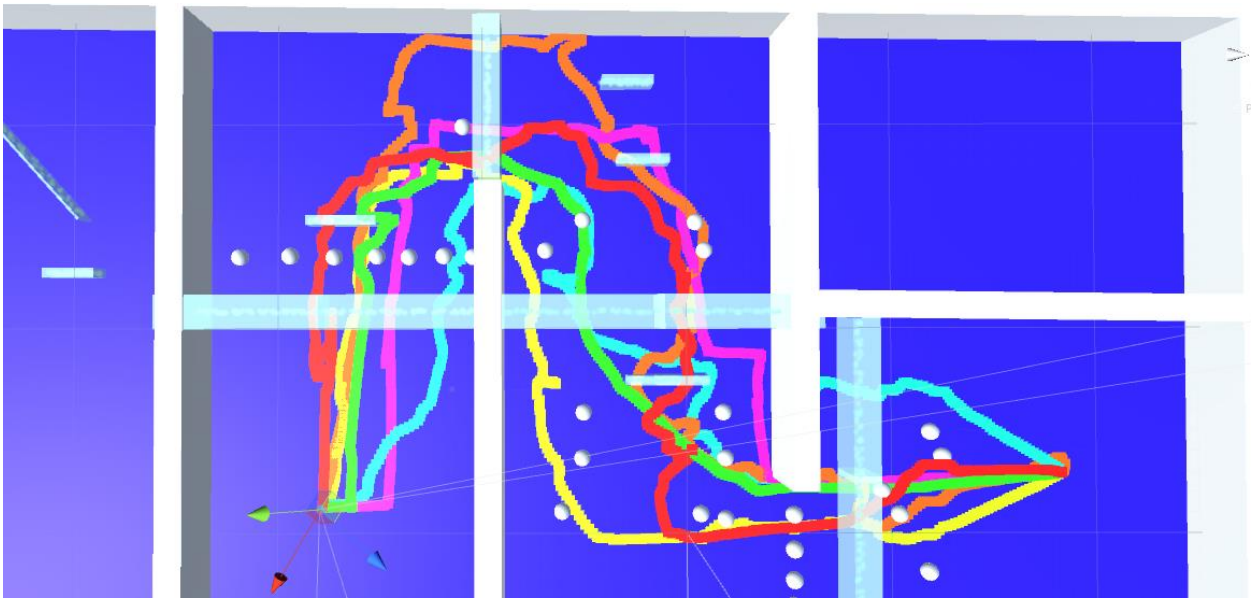


Fig 4.12 Maze 3 user paths

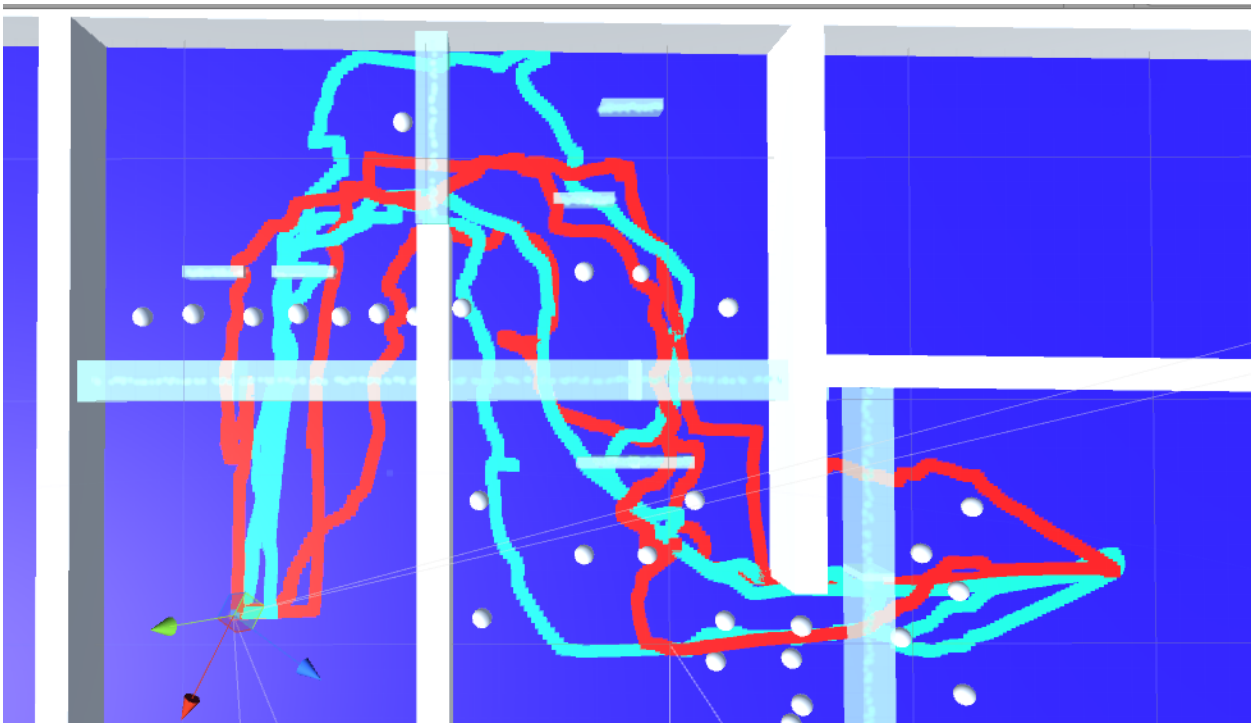


Fig 4.13 Maze 3 user paths by presence of sound

We can further examine this by splitting into the experience level groupings. Fig 4.14 is the same pathing, but with high experience users in red and low/mid experience users in teal. We can see that the two paths which pull off to the sides at corners are not only no sound runs, but also high experience runs.

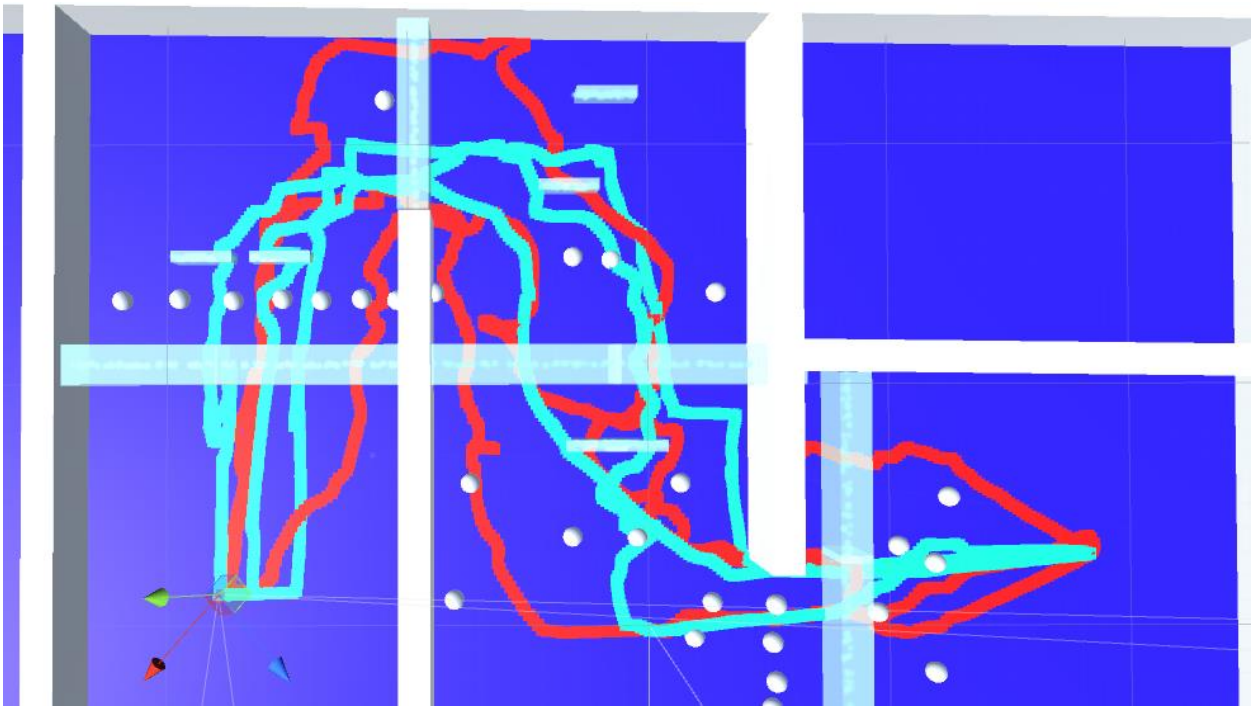


Fig 4.14 Maze 3 user paths by experience level

We also have the breakdown of where mistakes happened. We can see in Fig 4.15 where each mistake happened on maze 3, coloured by which user made the mistake. This can give us an idea of which user made the most mistakes in the maze, and if some of the mistakes were caused by the same obstacle. While there is a “cooldown” period after hitting an obstacle, it is possible for a user to not move fast enough or far enough, leading to an additional collision. This cooldown lasts for 2 seconds after a user has a collision and is in place to stop a user getting multiple errors from a single collision while they attempt to correct themselves or see where the obstacle is.

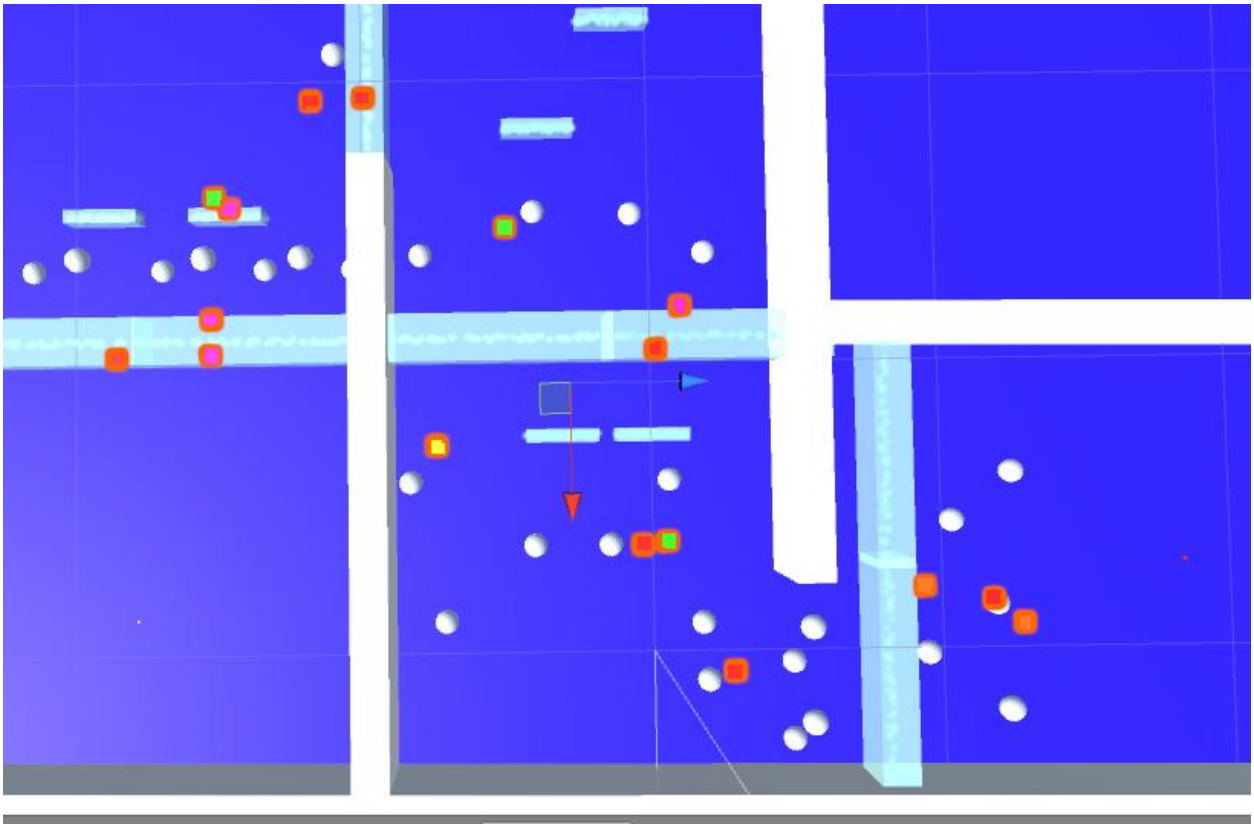


Fig 4.15 Maze 3 mistake locations by user

We can also look at a breakdown of mistakes by experience group. Fig 4.16 shows that for this maze, the vast majority of mistakes were made by the low experience group (14/17). We can also see that 2 of the 3 mistakes made by the high experience group were right at the end of the maze where the user may have been more focused on the goal in front of them, leading to more mistakes being made.

Looking at the mistakes by audio or no audio grouping can also give us some more information. Looking at Fig 4.17, we have the mistakes coloured by whether or not the experiment had audio (red) or no audio (teal). Adding this to our previous image, we know that all three of the high experience mistakes made were also no audio runs of the experiment. Having this visual element for analysis adds another level of understanding and makes things much clearer at a glance.

Further from that image, we can see that only one mistake made by the high experience group involved one of the moving wall obstacles, with the rest all being from the moving ball obstacles. By contrast, the low experience group had a lot of mistakes caused by wall obstacles, as well as by the ball obstacles.

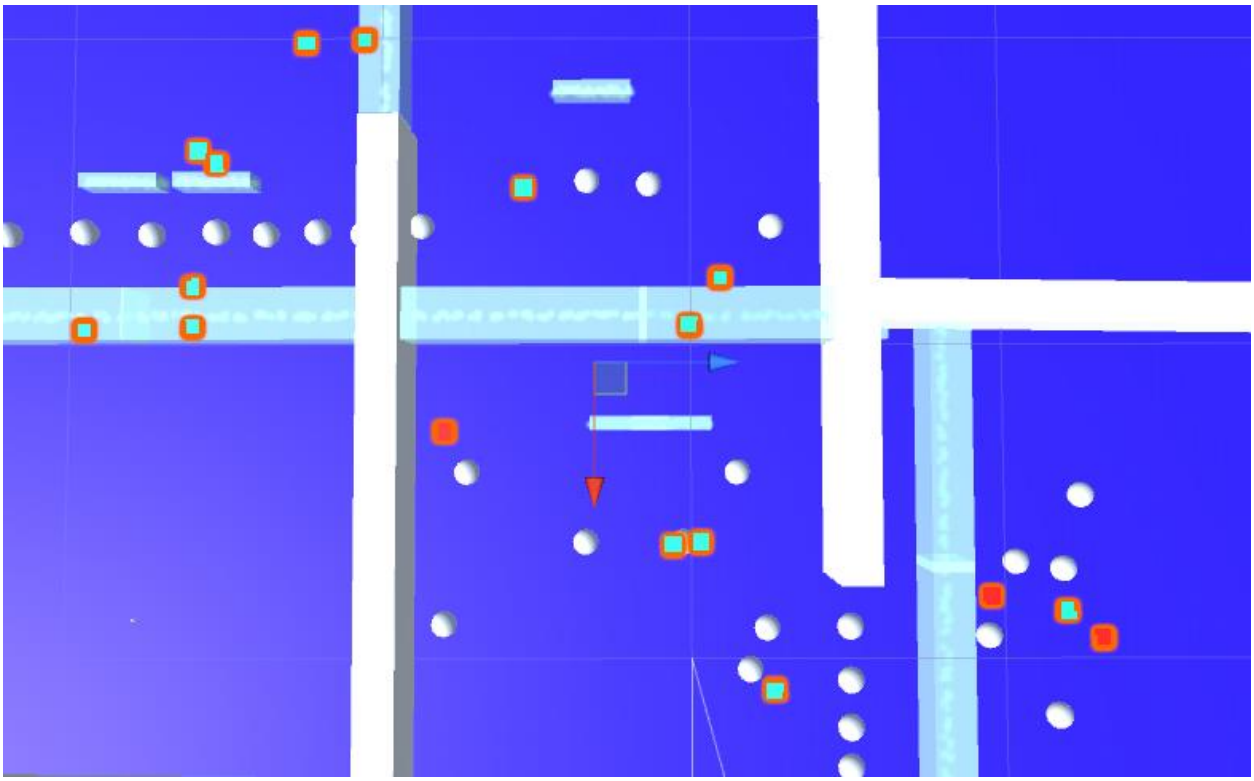


Fig 4.16 Maze 3 mistake locations by experience level

What we can gather from this experiment is that while general performance between audio and non-audio mazes was quite similar, experience with games and virtual environments could make it more impactful on performance to have sound. We also saw that the range of mistakes was much less with audio, allowing some users to make no mistakes throughout the maze. Finally, we saw that some different movement patterns developed at certain stages of the maze based on whether or not there was sound. This ties in with our research question, showing that spatial audio can have an impact on users' awareness of their surroundings in virtual environments.

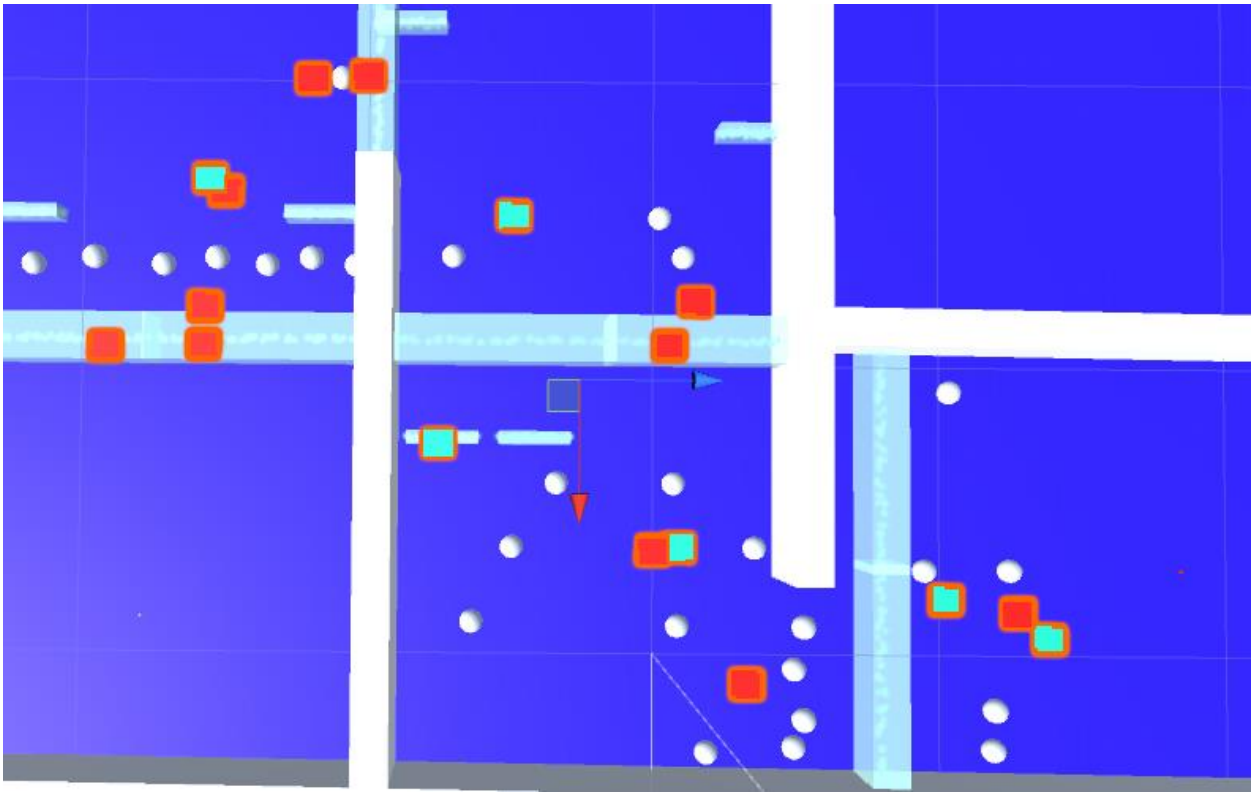


Fig 4.17 Maze 3 mistake locations by audio presence

4.4 Experiment 2: Varying Control Scheme Sound Location Experiment Results

4.4.1 Overview

This is the main experiment conducted as part of this research, relating most strongly to the majority of research questions. As mentioned before, this experiment focused on users locating spatialised sounds in a virtual environment, using two possible control schemes: (1) a mouse and keyboard, or (2) a VR headset.

4.4.2 Research Questions

The questions addressed by this experiment were:

"Is there a significant performance change between different control schemes while searching for audio?"

"Does the control scheme implemented for a virtual environment change how a user interacts with virtual sounds?"

"Do different control schemes affect a user's enjoyment or perceived performance?"

4 Experimental Results

With regards to the research questions, there were the following hypotheses:

(1) "There is a statistically significant performance change between different control schemes while searching for audio"

When it comes to working with sound and control schemes, a lack of training and experience can be an issue, with performance increase over a short time/short number of tests. It has been shown in real world experiences through popular video games that introducing HRTF and spatial audio often has a significant learning curve, as until users become used to the new sound and how to use it, they will often feel that it's strange or not useful (wickedplayer494, 2016; zipzapzooom, 2017; Campos et al, 2012). Over time though, users gain experience with the new sound system and quickly the benefits make it the norm. In Counter Strike Global Offensive, while HRTF audio was initially ignored by much of the user base upon its introduction, it has now become the default sound for many users, including some professionals who are notoriously slow to change (Night_Not_Day, 2020).

However, because of this need for training and experience, any significant change in performance may not be observable over the course of shorter experiments. Instead, most of the focus will be on how the user's interactions change between control schemes, as well as their subjective preference and opinion on performance.

(2) "Different control schemes change how a user interacts with virtual environments"

The hypothesis here is that by using a control scheme that is more natural to a user, they may use different techniques to locate the source of sounds within the environment or have different movement patterns while searching. Traditional mouse and keyboard or gamepad controls only really make use of two axes of movement. While it is possible to add control of all three axes with these setups, doing so is often quite awkward and generally unused in most games or virtual experiences. Game console controllers by default must use two joysticks for movement, and game engines for PC games which

provide generic control methods don't have functionality for tilting by default. However, with VR head tracking, it is possible for users to embrace all axes of movement for their head, allowing for more natural and intuitive movements and control. Doing so should allow users to use a greater degree of rotation and tilting while searching for sound, using techniques and strategies subconsciously used in the real world. This can potentially alleviate some issues with virtual audio spatialisation. Compared to emulating this tilting and rotation with traditional control schemes, head tracking doesn't add any additional workload to the user, as they don't need to actively think about pressing buttons to make use of these small tilts and movements. Instead, they can simply focus on listening and searching within the virtual environment.

(3) "Different control schemes are more or less enjoyable for users and can improve or disimprove perception of performance"

Regardless of actual performance variation between different control schemes, certain types of controls can be more enjoyable or fun for users to use (Martel et al., 2015; Ashdown et al., 2005). Whether it be due to uniqueness, intuitiveness, immersion, or a combination of factors, certain controls will appeal more and be a user's preference. More than just preference for enjoyability, some people will feel that a more enjoyable or natural system leads to better performance, even when there is no significant difference in their own performance. This preference can even cover a decreased performance. While performing tasks or activities where performance isn't the most important factor, people sometimes prefer to use less efficient or worse performance methods or equipment if they have more fun using those methods. Because of these factors, it is important to recognise a user's preferences and feedback as well as measurable task performance. It is not enough to simply say one system is better due to user performance when it may not be the most appealing system for the majority of users. There is also evidence that using a more enjoyable system can reduce user workload, both in physiological measurements (Chao et al, 2017) and in user perceived workload (Lum et al, 2018; Lackey et al, 2016). This can further inform user decisions on which type of system they prefer and would choose.

4.4.3 Quantitative Results Discussion

4.4.3.1 Grouping of Cubes

Within this experiment, cubes are broken into 4 distinct groups based on potential difficulty. The 4 groups and the cube numbers contained within are as follows:

Group 1 - Same Plane as User = 1, 6, 8, 10, 15, 17, 24, 26

Group 2 - Higher Plane than user = 2, 3, 4, 11, 12, 13, 21, 22

Group 3 - Lower Plane than user = 0, 5, 7, 9, 14, 16, 23, 25

Group 4 - Directly Above/Below = 18, 20

In Fig 4.18, the groups are coloured green (group 1), pink (group 2), red (group 3), and yellow (group 4).

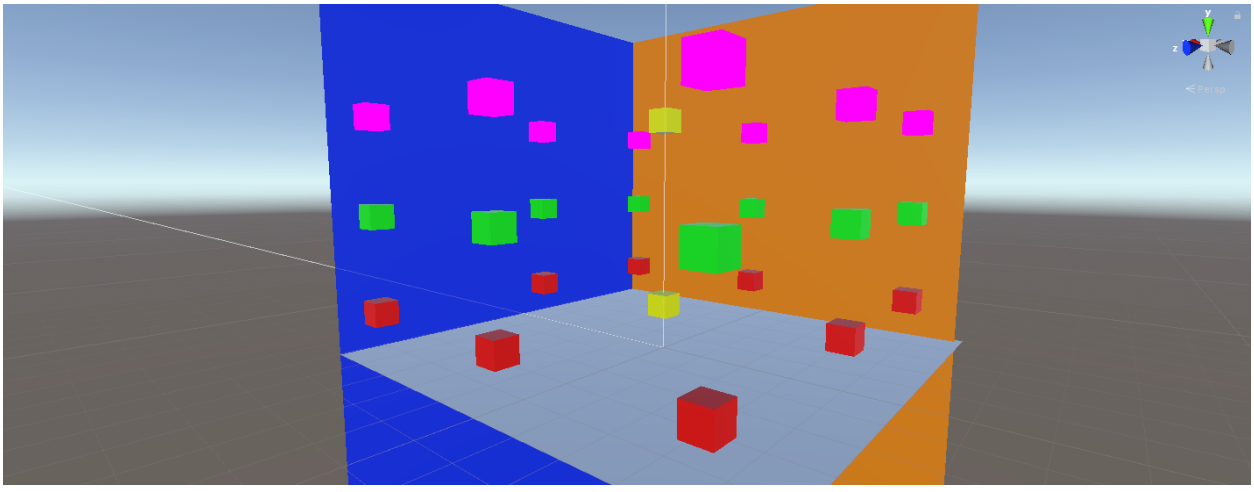


Fig 4.18 Cube locations by type grouping

4 Experimental Results

Group 1 cubes are the simplest cubes, as they are on the same elevation level as the user. To locate these, the user simply rotates around their ‘up’ (y) axis. While there can be issues with the front-back problem, these cubes have the fewest potential issues.

Group 2 and 3 cubes are similar. All cubes within these groups are either located at a higher elevation than the user (Group 2) or a lower elevation than the user (Group 3). This change in elevation combined with the front-back problem can cause additional issues with locating the spatialisation. Usually when we try to locate a sound at a different elevation to ourselves, we rely on either tilting our head at different angles or listening to how the sound is altered by the environment. Generalised HRTFs can lack clarity with regards to sound alteration, and traditional control schemes don’t provide a straightforward way to tilt the user’s “head” (on its z axis), making it more difficult to try and pinpoint these cubes.

Group 4 contains only 2 cubes: the one directly above the user and the one directly below the user. These cubes have the highest potential difficulty, as simply rotating in place won’t cause any change to the sound, and due to the only difference being elevation, there are similar issues to the back-front problem. Tilting of the head can shift the sound to the left or right, but this isn’t intuitive in traditional control schemes. However, if the user can figure out that the sound is either above or below them, there is much less error caused by guessing, as there are only 2 potential cubes.

In Fig 4.19 we show the proportion of each cube group appearing in each experiment.

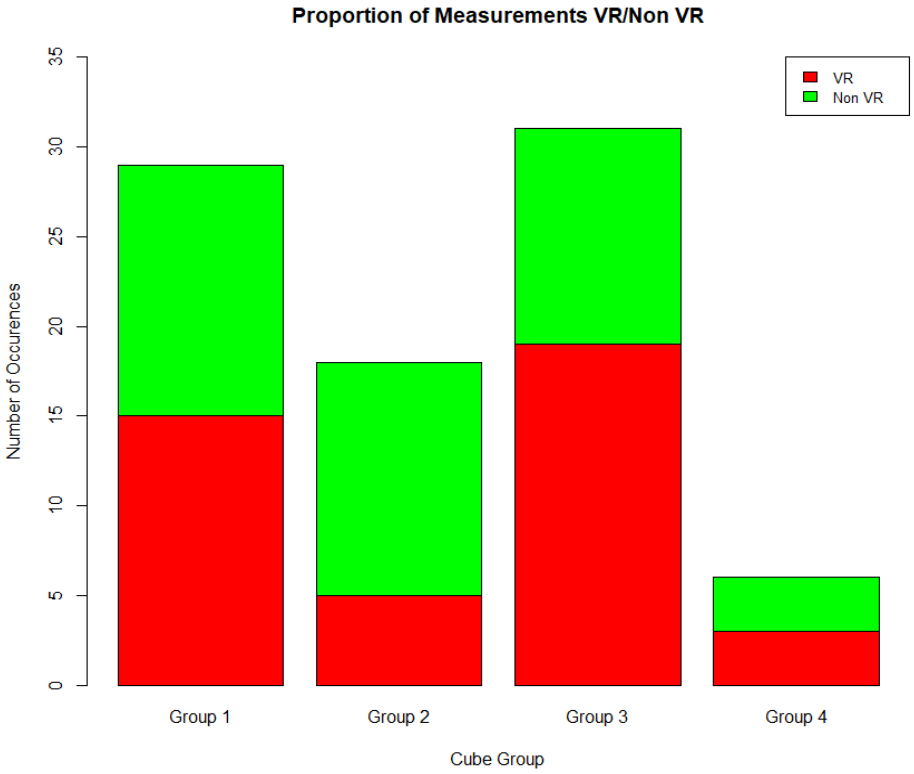


Fig 4.19 Proportion of Measurements VR/Non-VR

As the cubes selected by the experiment are random, there is a possibility that some individual cubes may appear more or less frequently overall. Generally, there was an even distribution of cube types between the experiments, though there was a heavier skew of group 2 cubes in Non VR experiments (13/18), with a higher amount of group 3 cubes in VR experiments (19/31), and very slightly more Group 1 cubes in VR experiments (15/29). This can all be seen in Fig 4.20. Due to the similarity in difficulty between Group 2 and 3, this slight skewing shouldn't cause an issue. Group 4 had an even split between both VR and Non-VR, with the smallest number of occurrences due to it having the least number of potential cubes. Ideally in future runs of the experiment, there would be a preference to have at least one occurrence of group 4 in each experiment, as that group has some interesting added challenges related to its placement.

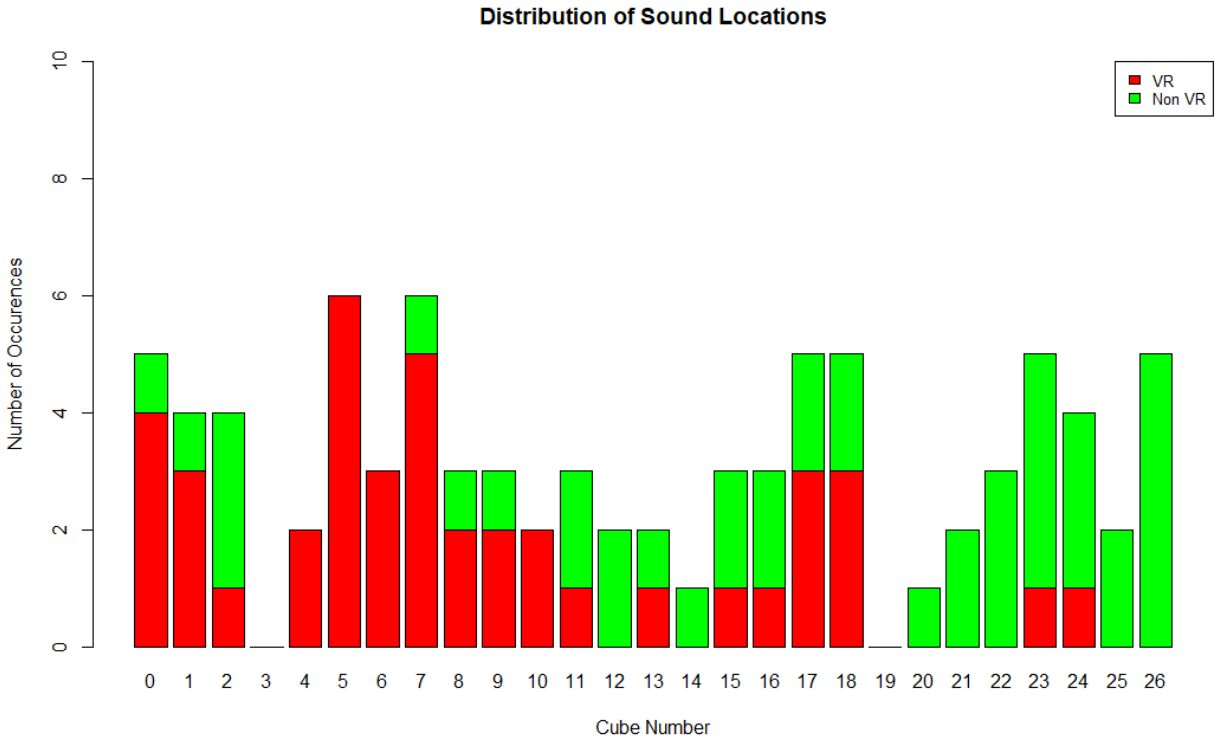


Fig 4.20 Distribution of Sound Locations

Here we have an overall breakdown of how many times each individual cube was made a sound source across the experiments, broken down into VR and No VR. There are a few instances of cubes that appear in only one experiment type or the other, but this is of less importance compared to what group they are part of.

4.4.3.2 Analysis of Cube Performance

Now we will examine how user performance varied for the grouping of cubes, looking at the differences in mistakes made and time taken between the VR and non-VR experiments.

In terms of overall difficulty of groups, while group 3 had the largest number of total mistakes (Fig 4.21), the highest average amount of mistakes was recorded by group 4 by far (Fig 4.22). Group 4 seemed to have a pattern of users having difficulty locating where the sound was coming from and simply clicking a large number of cubes until they eventually reached the correct one through the process of elimination. This is shown by two users having mistakes numbering 27 and 25 while the target cube was 18 in group 4. Both of these cases were in Non-VR groupings. The remaining instances of group 4 cubes involved much lower numbers of mistakes, with the other Non-VR resulting in 4 mistakes, and the three VR instances having 1, 1 and 0 mistakes. However, it is difficult to make strong claims about these results, as there isn't a huge amount of data to assess. Among the experiments, only 1 user had a group 4 cube appear in both their VR and non-VR experiments. They performed better in their VR experiment, but it is difficult to make a solid claim about this performance change due to the small sample size. Ideally, every user performing the experiment would have one instance of a group 4 cube in both of their experiments, so we could examine if the improvement in ability to locate the cube in VR continues to be the case.

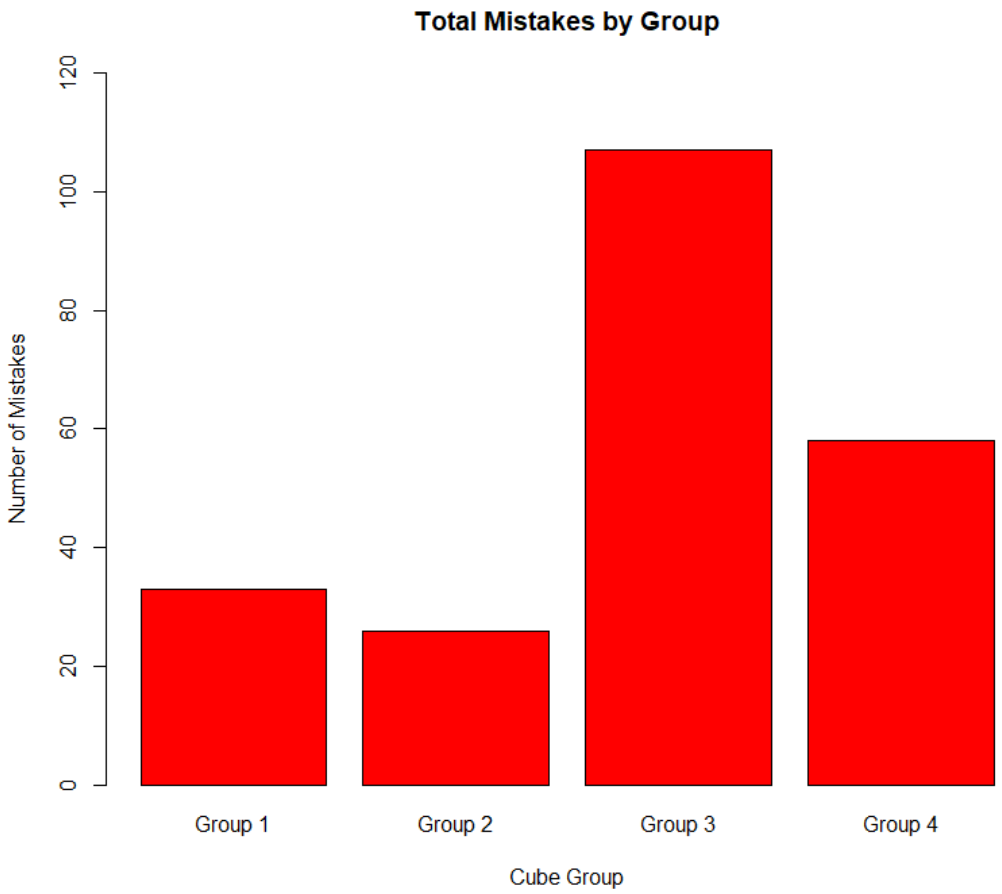


Fig 4.21 Total Mistakes By Group

Apart from group 4, performance in the other groups is quite close. Group 3 had more mistakes than group 2 or 1, which seems to be tied to 2 very high mistake values from 2 specific experiment results where users couldn't determine the location and started guessing all the cubes, similar to the issues in group 4. Issues with group 4 is expected, as it faces some very similar problems to the front-back problem, but with a verticality which compounds the issues. This is likely why similar experiments had limited verticality (Moraes et al, 2020, Kuppanda et al, 2015). We can see a breakdown of group performance in Table 4.6 and 4.7.

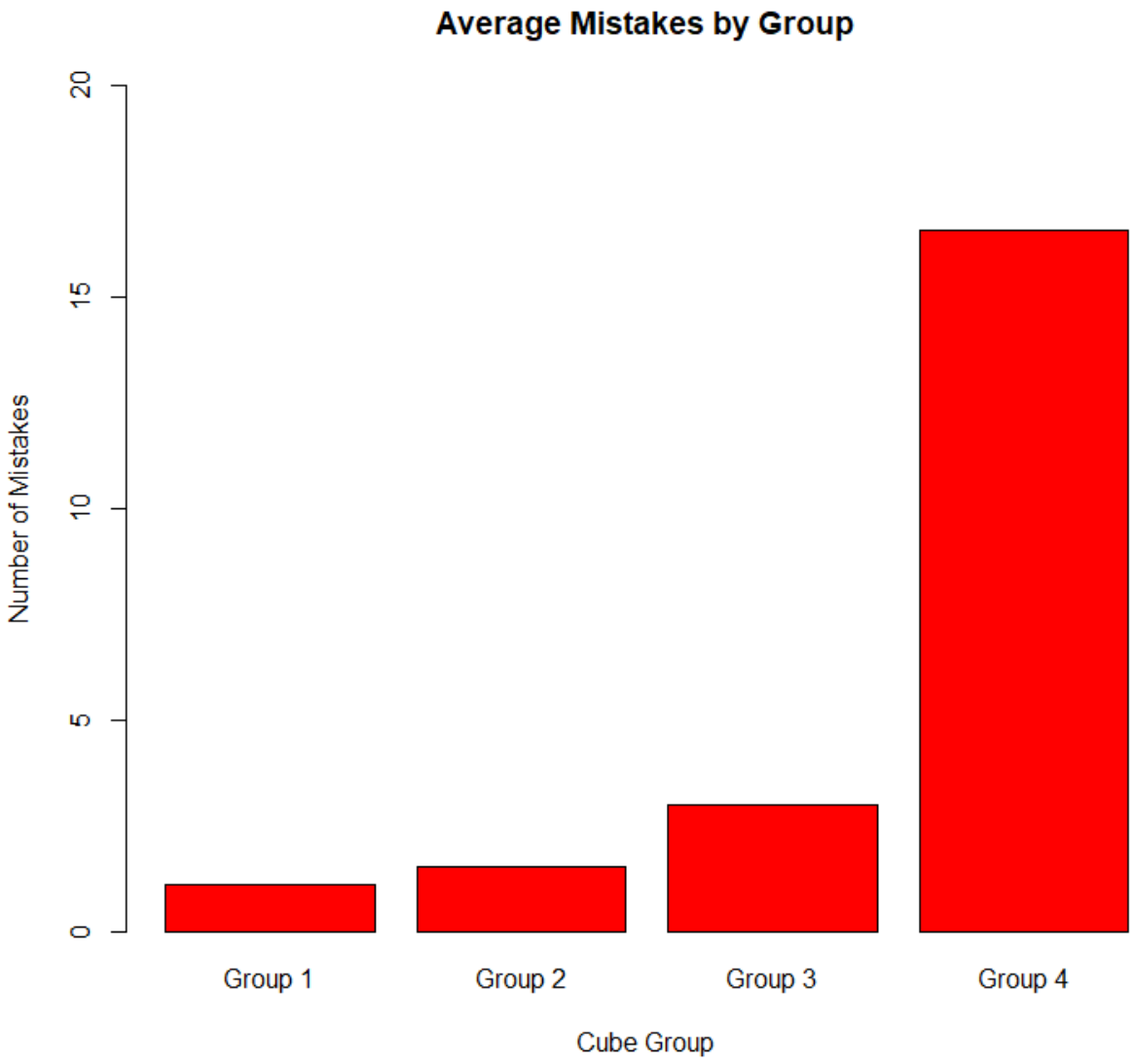


Fig 4.22 Average Mistakes By Group

One of the more interesting things is how close group 1 and 2 were in terms of average mistakes made. In theory, group 2 and 3 could cause more issues with determining location due to the change in elevation, but users seemed to perform similarly from these experiments.

Group	Average	Median	StdDev	Max	Min
1	1.098	2	5.334	7	0
2	1.536	1	1.432	7	0
3	2.983	1	1.854	25	0
4	16.6	2.5	12.738	27	0

Table 4.6 Group Mistakes Results

Group	Average	Median	StdDev	Max	Min
1	11.922	16.266	32.333	48.639	2.147
2	9.951	8.426	11.658	23.333	2.321
3	24.321	7.767	6.448	169.402	1.547
4	65.448	25.601	51.494	126.025	3.479

Table 4.7 Group Time Results

Now moving on from looking at the groups as a whole, we are looking at performance differences between VR and Non-VR within these groups, looking at the number of mistakes made (Fig 4.23). Looking at the averages, user performance on Group 1 and group 3 performs quite similarly, with user VR performing very slightly worse in terms of mistakes made. For group 2, VR did have more mistakes on average (2.375 vs 0.952). Group 2 had the smallest amount of VR results, having only 5 out of 18 results, so it is possible that the size of the sample could influence the average results.

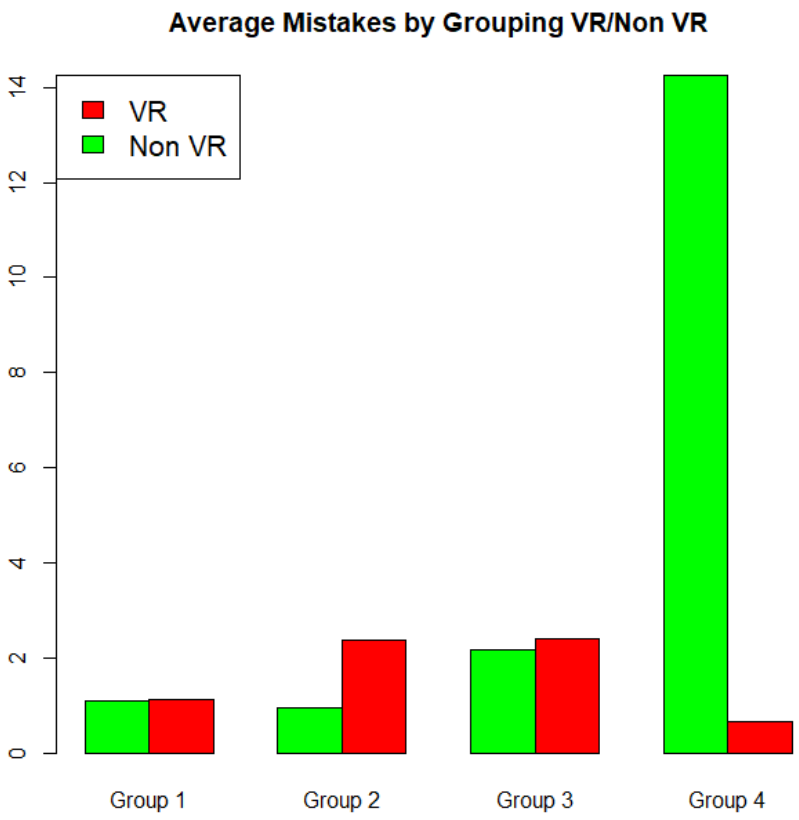


Fig 4.23 Average Mistakes by Grouping VR vs Non-VR

When we look at group 4, we see a huge difference in performance. The VR experiments performed quite well with the group 4 cubes, but the non-VR experiments performed quite poorly in terms of mistakes made. Group 4 is quite a small sample size, though they have equal appearance rate between VR and non-VR experiments. In terms of comparable performance, it will probably be better to look at some of the individual results and see if the spike in mistakes with group 4 is more linked to the user's overall performance, or if that specific style of cube caused significant issues unrelated to the user's overall performance.

Breaking down the mistakes by individual cube can provide some more information (Fig 4.24), though due to the different appearance rates between VR and non-VR, it may not always be an equal comparison. There are individual cubes that were only selected in VR or in non-VR. However, this is why we break them into the groupings of cubes which are of a similar style and difficulty for overall comparison.

While most cubes had similar amounts of mistakes, the main outliers for the VR experiment were cubes 2 and 5. Cube 2 only appeared once in VR, leading to an average of 5 mistakes. Cube 5 had several appearances in the VR experiment with quite similar performance but had one incredibly poor instance with 25 mistakes.

Within the non-VR results, there was generally similar performance with the two outliers being cubes 18 and 20, both part of Group 4. For cube 20, it only showed up once throughout the experiment, having a poor performance of 27 mistakes. Cube 18 showed up in both VR and non-VR experiments, with the VR results performing quite well and the non VR results being quite a bit poorer, with one quite bad result of 25 mistakes.

Generally, across both experiments the average mistakes made were quite similar, as we saw from the cube groupings. The largest difference is with the performance of grouping 4, cubes 18 and 20, which performed very poorly in the non-VR performances. We can see a full breakdown of group performance in VR and Non VR in Tables 4.8 and 4.9.

Group	Average	Median	StdDev	Average	MedianTime	StdDev
1	1.119	2	6.683	13.494	10.979	40.757
2	2.375	1	1.033	11.613	7.271	12.017
3	2.389	1	3.209	20.403	9.662	9.745
4	0.667	1	0.577	7.127	6.453	4.027

Table 4.8 Group Results for VR

Group	Average	Median	StdDev	Average	Median	StdDev
1	1.094	2	1.779	12.586	24.359	12.743
2	0.952	1	1.781	8.247	9.807	11.664
3	2.179	1	0.816	20.718	7.213	4.942
4	14.25	25	12.741	65.865	93.465	43.566

Table 4.9 Group Results for Non VR

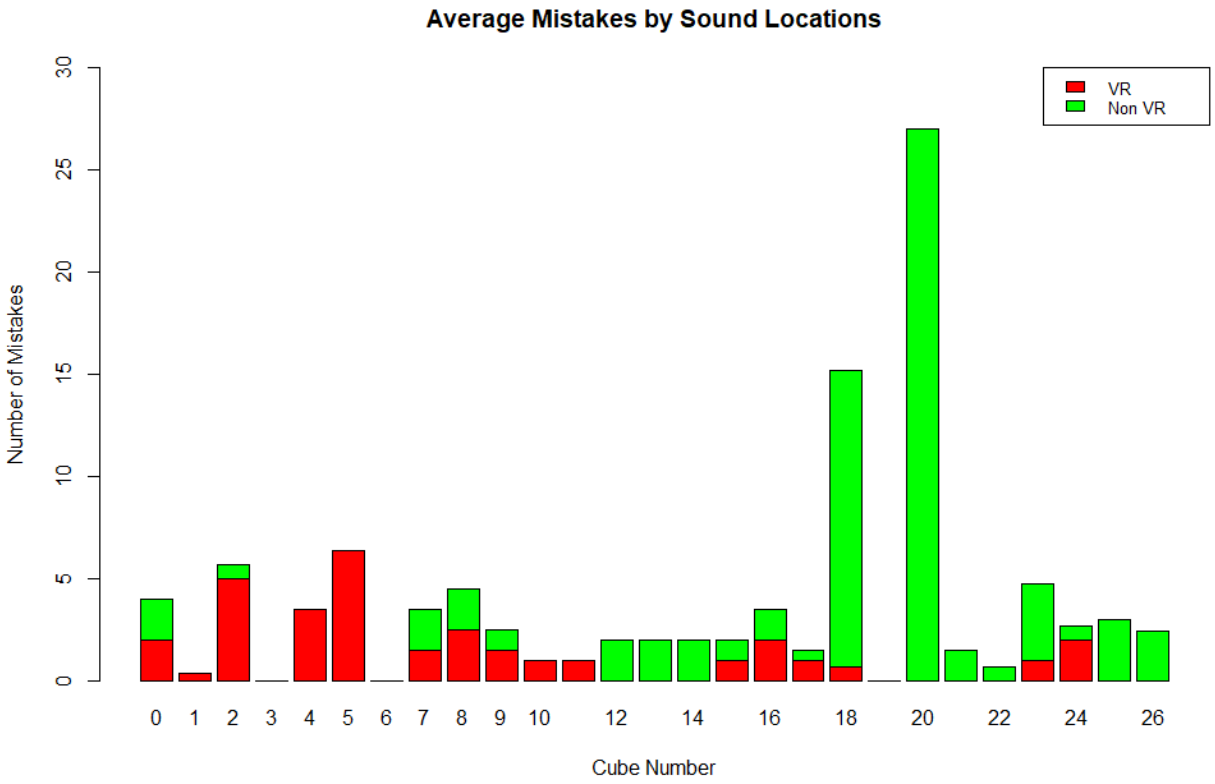


Fig 4.24 Average Mistakes by Sound Location

Looking at the times taken during the experiments (Fig 4.25) when using our cube groupings, we can see that the time taken on average was of a similar performance between VR and non-VR for Groups 1, 2, and 3. The greatest difference is once again in Group 4, where the non VR average time taken is over ten times the average VR time. Fig 4.26 shows the average time taken for each cube in VR and Non VR.

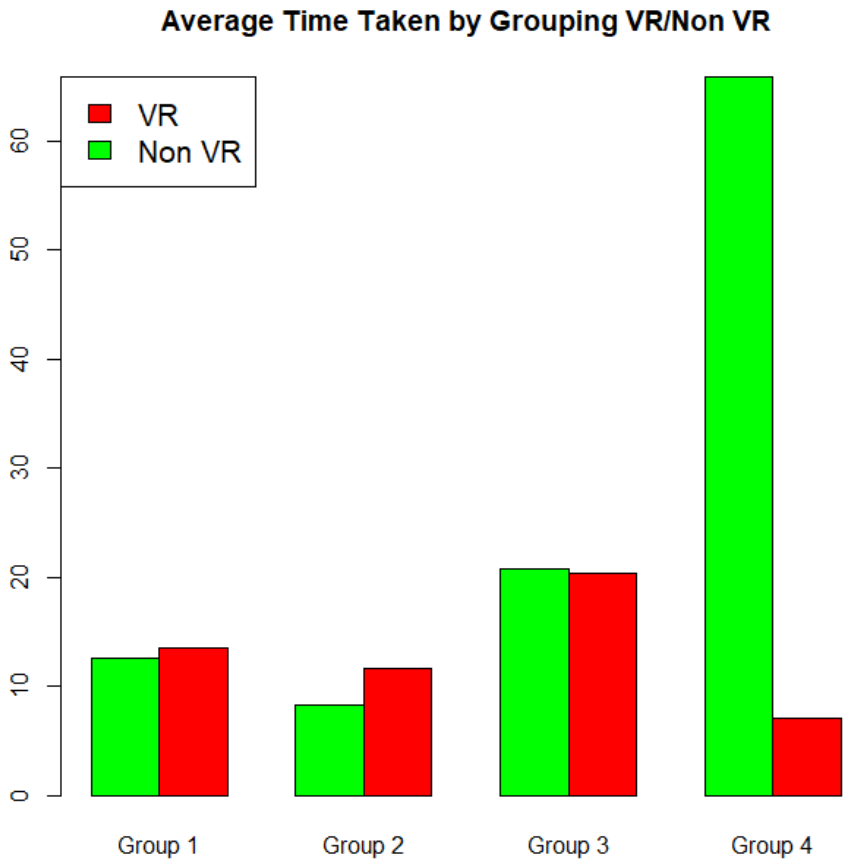


Fig 4.25 Average Time Taken by Grouping VR vs Non-VR

The improved performance of VR in group 4 is expected. Group 4 cubes behave similarly to the front-back problem, which we know to cause sound location issues in traditional audio. However, rotating the head to try and locate the elevation difference of these cubes is unnatural using traditional control schemes as it requires unfamiliar keyboard controls. However, in VR, the natural head motions we use to locate similar sounds are intuitive to use. This aids with locating that sound type. This fits with existing literature about issues with the front-back problem (Letowski, 2016; Link and Lehnhardt, 1966; Van Soest, 1929; Majdak et al., 2010; Kato et al, 2003).

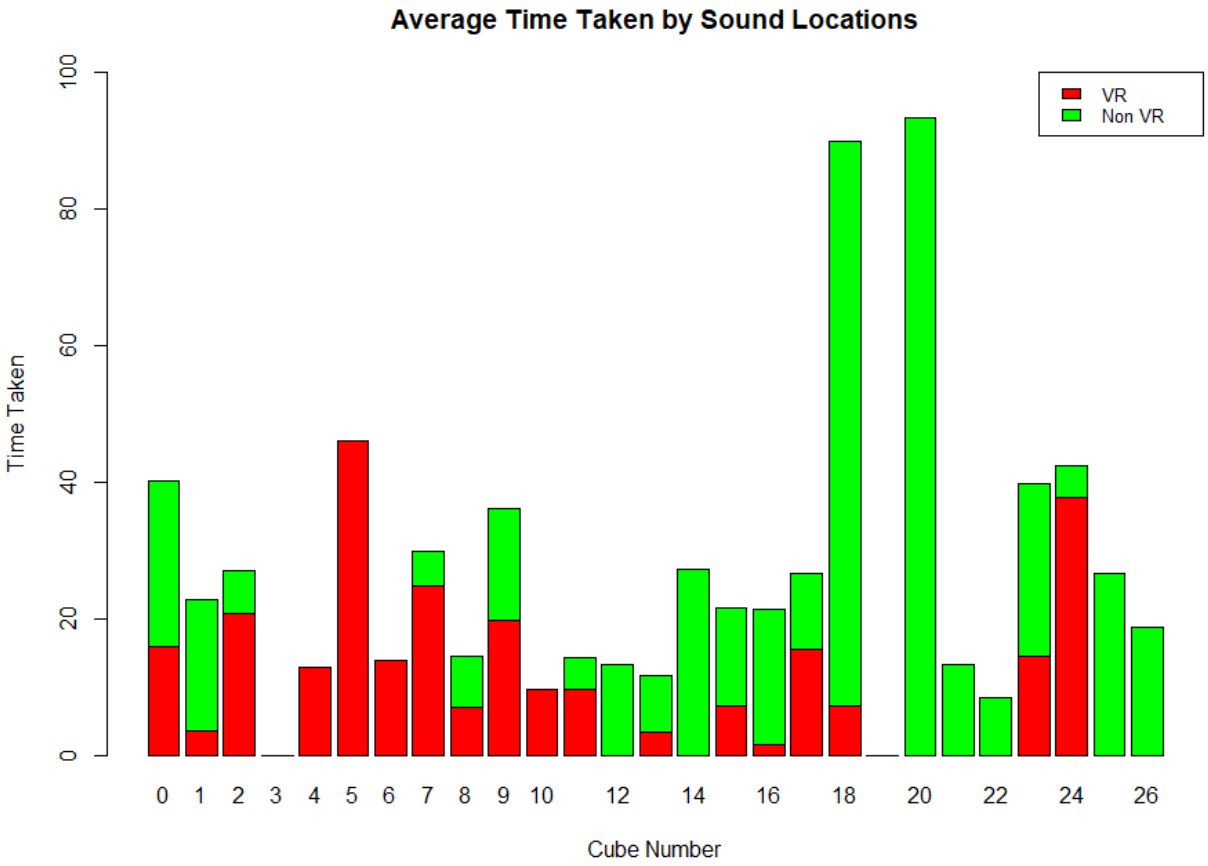


Fig 4.26 Average Time Taken by Sound Locations

4.4.3.3 User Experience Differences

A User's individual performance between VR and non-VR was usually quite similar, with occasional spikes of a single difficult cube. Generally, performance differences were not statistically significant in regard to mistakes made, though when it came to time taken, there were some users who had a statistically significant performance increase from non VR to VR (p-value = 0.0446 for one such case) (Fig 4.27), with some other users having results just above the confidence threshold of 0.05 (p-value = 0.053755306 or 0.057952323). Some of these users performed the Non-VR experiment first, and others performed the VR version first, so this isn't a result of training and getting used to the experiment type.

```
        Welch Two Sample t-test

data:  loc_experiment11_nonvr and loc_experiment11_vr
t = 2.6469, df = 5.1082, p-value = 0.04464
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  0.6200987 34.8319376
sample estimates:
mean of x mean of y
 22.14602   4.42000
```

Fig 4.27 Individual Time Taken Comparison T-Test

The three users who had these performance increases are also the 3 users who reported having a high level of experience in video games and virtual environments. It is possible that this level of experience can make it easier to adapt to the experiment, or that they felt more comfortable in the environment and were better able to focus on the experiment itself.

Looking closer at the differences between highly experienced and low/moderate experienced users, there is a statistically significant difference in time taken (p-value = 0.06773) and mistakes made (p-value = 0.01325) between these two groups (Fig 4.27 through Fig 4.30). Highly experienced users performed better, both taking less time than the other users, and making less mistakes. The high experience group also had a lower maximum number of mistakes on a single cube of 6, compared to the other group having a max of 27. The more experienced group seemed to get less frustrated while searching out the cube or were better able to make out where the sound was coming from.

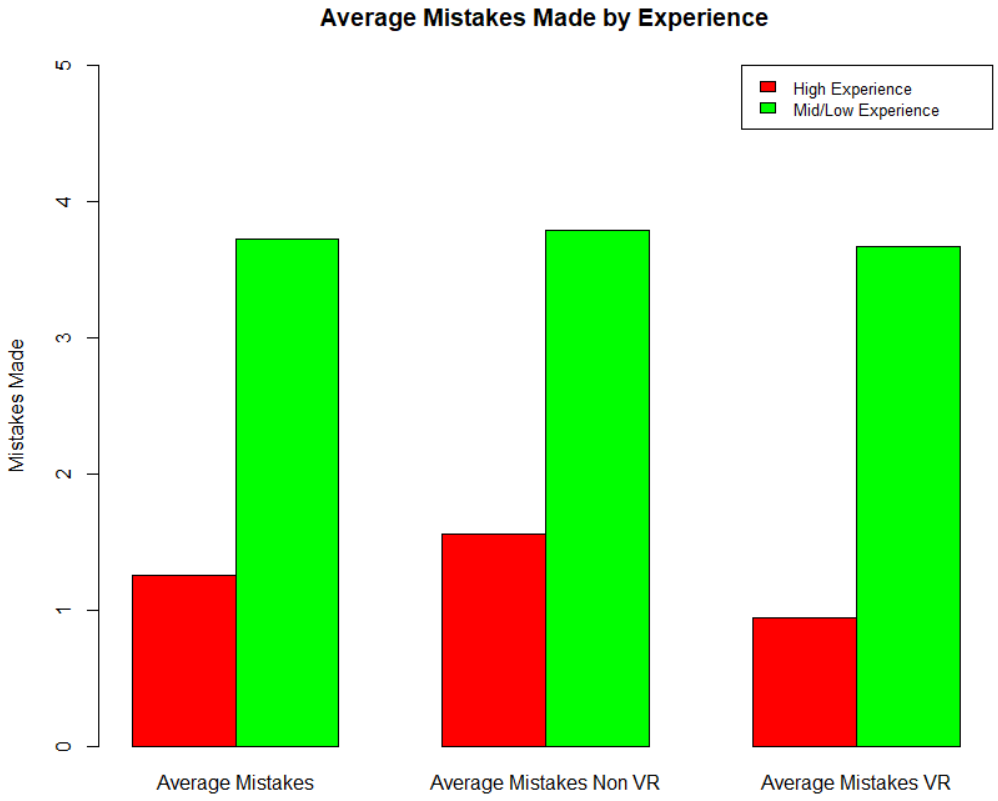


Fig 4.27 Average Mistakes by Experience Level

```
        Welch Two Sample t-test

data:  loc_time_highexp and loc_time_lowexp
t = -1.8583, df = 64.038, p-value = 0.06773
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -19.9375178   0.7210348
sample estimates:
mean of x mean of y
 13.58649  23.19474
```

Fig 4.28 Comparison of time taken between experience level groups

```
        Welch Two Sample t-test

data:  loc_mistakes_highexp and loc_mistakes_lowexp
t = -2.5653, df = 51.705, p-value = 0.01325
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -4.4187202 -0.5396131
sample estimates:
mean of x mean of y
 1.250000  3.729167
```

Fig 4.29 Comparison of mistakes made between experience level groups

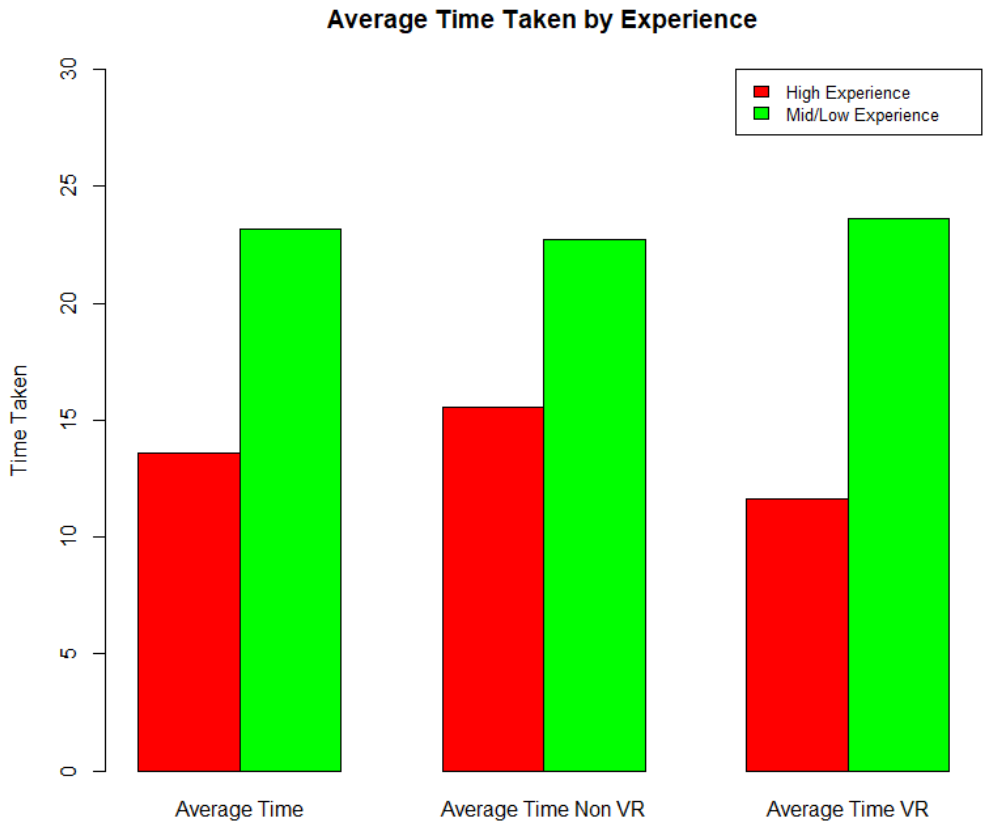


Fig 4.30 Average Time Taken by Experience Level

Examining the groups separately, there is not a statistically significant performance difference between VR and non-VR in neither the highly experienced group nor the low/moderately experienced group. While the higher group is closer to being statistically

significant (p-value = 0.103299694 for mistakes and 0.1310559 for times), it does not quite reach a level that would ensure improvement.

4.4.3.4 User Movements and Search Patterns

The way users move and attempt to locate sounds in the environment can change depending on the control scheme they are using to navigate the virtual environment. To try and better analyse these differences which can be hard to find from quantitative data, we tracked the user's rotational movements throughout the experiment so that we can analyse the user's procedure. We use a driver program built into our Unity environment to read in the user's rotational data and replicate the movements tracked during the course of the experiment (Fig 4.31 and Fig 4.32).

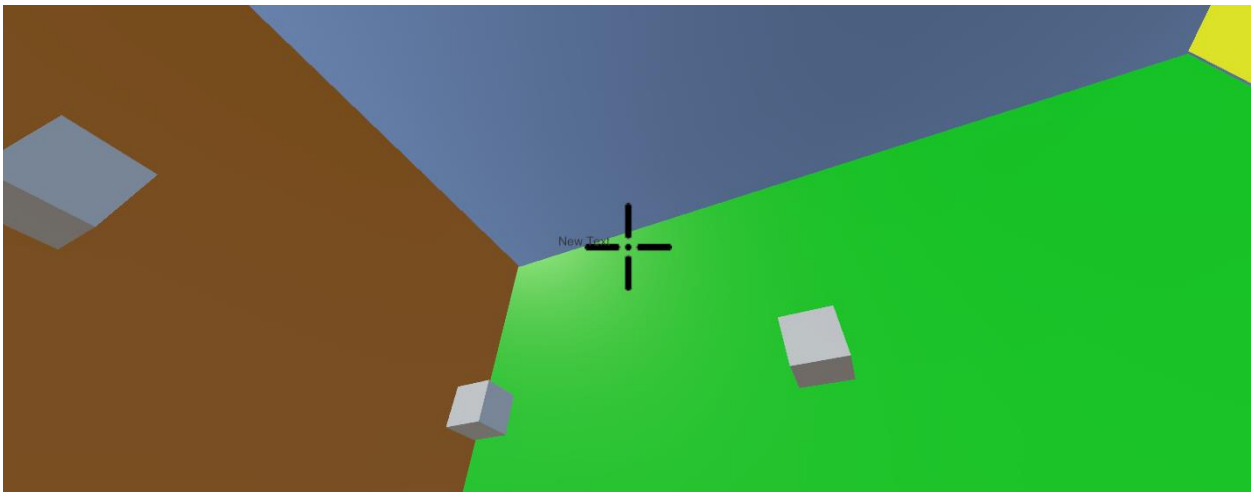


Fig 4.31 User viewpoint in VR

Generally, when a user first attempts to find a sound source, they start with broad sweeping movements while listening for changes in the audio. They use this to get a rough location of where the sound is based on which ear they are hearing the sound from as they rotate. This is the case in both VR and non-VR iterations of the experiment.

Once the user narrows down a rough location, they make smaller sweeping motions to confirm which specific column they need to look at. This leaves them with 3 cubes they feel are most likely to be the sound source.

The vertical element is what causes an issue, especially for users using the traditional mouse and keyboard control scheme. While the instructions given before the experiment stated that there were keyboard controls to tilt the camera angle to the side, no users made use of this functionality. Such a thing is rare in most games or virtual environments, and when games do have a tilting function, it is usually for viewing changes to look around corners. We don't often think actively about using these angle changes to try and determine audio elevation differences. For traditional control scheme users, they would try and make an educated guess based on the spatialised audio they heard while moving. This puts a lot of importance on the accuracy of the audio spatialiser.

However, when it came to the VR control scheme, users instinctively made use of head tilts to help them find the audio source (Fig 4.31 and Fig 4.32). While it wasn't usually a conscious thing being done, some users spent more time tilting their head while performing their searching sweeps, which can be seen from the tilted angle of the camera. This was especially seen in the highly experienced users who had to locate the sound source when it was a group 4 type.

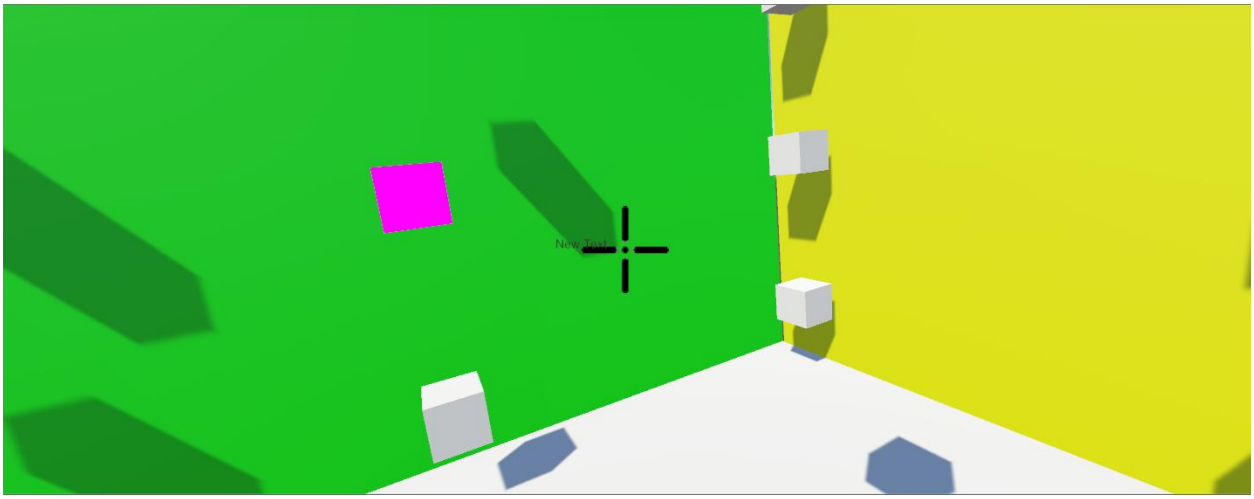


Fig 4.32 User viewpoint in VR

Doing this provides the user with more information as there will be changes to the audio heard from the ear pointed towards the sound source. This can let the user zero in on which part of the column the sound is coming from. Users can do this subconsciously, as doing so is similar to how humans would locate the sound in the real world. However, more exaggerated movements can provide more information, and doing so is less natural. This can be an issue, as when using generalised HRTF for the audio spatialisation, our natural subconscious movements may not be as accurate, due to the difference in the sound compared to what we would expect.

4.4.4 Qualitative Results Discussion

4.4.4.1 User Feedback Questionnaires

As part of the experiment, users were given a questionnaire asking a series of questions about their experience with the experiment, the results of which will be covered here. Each participants response is shown in Table 4.10.

Participant #	Experience	Q.1.	Q. 2.	Q. 3.	Q. 4.
1	MODERATE	Mouse + Keyboard	VR	Not too difficult	VR
2	LOW	VR	VR	A little difficult	VR
3	HIGH	No	VR	Not too difficult	VR (felt more natural)
4	MODERATE	VR	Mouse and Keyboard	Sometimes	VR
5	LOW	Mouse and Keyboard	VR	Not too difficult	VR
6	HIGH	No	VR	Sometimes, but easier over time	VR
7	HIGH	Mouse and Keyboard	VR	Not too difficult	VR

Table 4.10 User Questionnaire Responses

Q.1. Was it harder to locate the sound sources in one of the experiments?

In terms of reported difficulty, 2 users reported no difference between the two control schemes, 2 reported that the VR control scheme was more difficult, and 3 reported that the mouse and keyboard control scheme was more difficult. The two users who reported no difference between both control schemes were people who had a high level of experience with video games and virtual environments, which may have contributed to them being more comfortable with this type of experiment. Of the two users who reported having more difficulty with the VR experiment, one commented afterwards that they may have been more distracted by the VR control scheme, which could have affected their performance and perceived difficulty.

Overall, in terms of reported difficulty, there was a relatively even split between the three options, with a slight edge towards keyboard and mouse being more difficult.

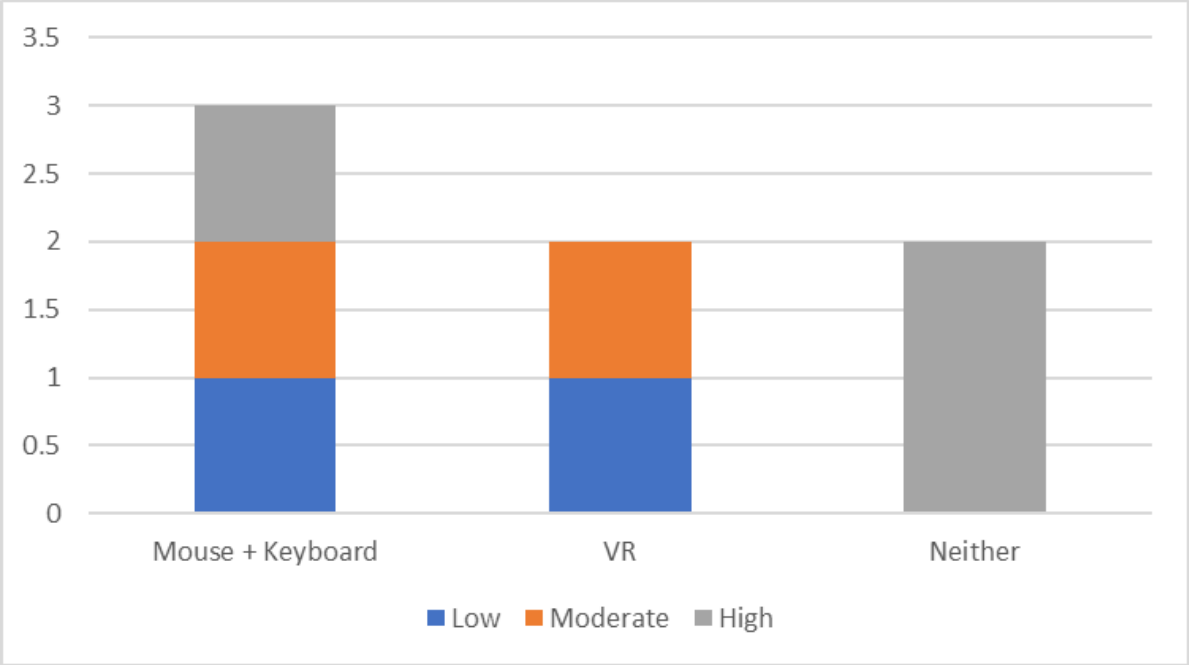


Fig 4.33 Was either method more difficult

Q. 2. Was it more natural to find the sound sources in one of the experiments?

Regarding which control scheme was more natural to use, almost all users (85.7%) reported that the VR control scheme felt more natural. Only one user reported that using a keyboard and mouse was more natural. This user was of moderate experience with virtual environments and games and was also one of the users who reported higher difficulty with VR. From examining their performance in the experiment, this user performed significantly worse in their VR experiment, while performing quite well in their mouse and keyboard experiment. It is possible that having some experience with traditional gaming can make the task of switching to a new control scheme more difficult.

Having a control scheme feel more natural is something that can be of greater importance to users with little experience with traditional methods of control in virtual environments. It can be quite common for people who are new to mouse-controlled cameras to have issues adapting to the style of movement, especially when it comes to rotating the camera while moving, or fine control of camera aiming. Being able to remove that learning roadblock by controlling the camera in a way that comes natural to a user

increases accessibility and user enjoyment, as seen by the overwhelming preference towards the VR control setup.

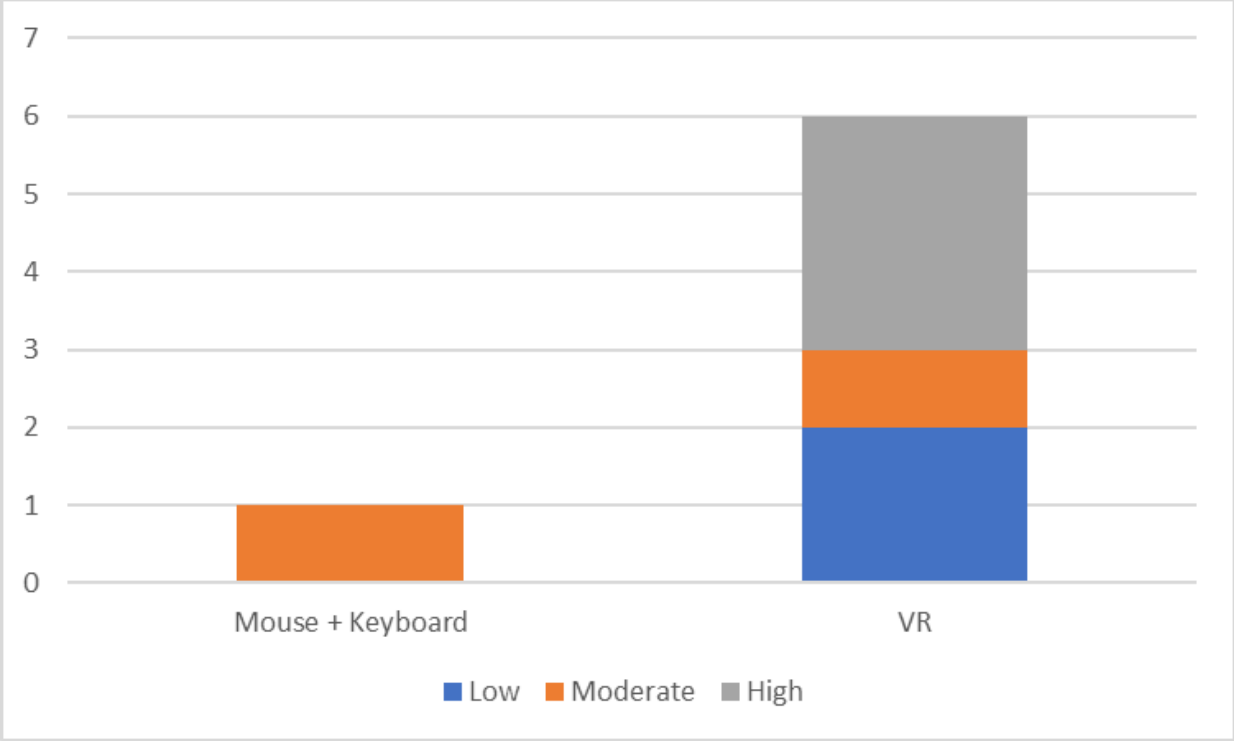


Fig 4.34 Was either method more natural

Q. 3. Was it difficult to tell where the audio was coming from in the environment?

For general difficulty of the task, most of the users stated that locating where the sound was coming from wasn't too difficult, 2 users stated there was some difficulty, and 1 user stated there was a higher level of difficulty. One of the users who stated some difficulty said that as the experiment continued, it became easier to figure out where the audio was coming from, and the difficulty reduced. The one user who had a larger amount of difficulty was someone with very low experience with games and self-reported that they may have been distracted by the VR aspect. Generally, the users with more experience with gaming adapted to the task more quickly and found it less difficult. The task becoming easier over time also fits with what we saw about the impact of training (Mendonça et al, 2012) and reduced cognitive load as we become used to the task (Jordi et al, 2009).

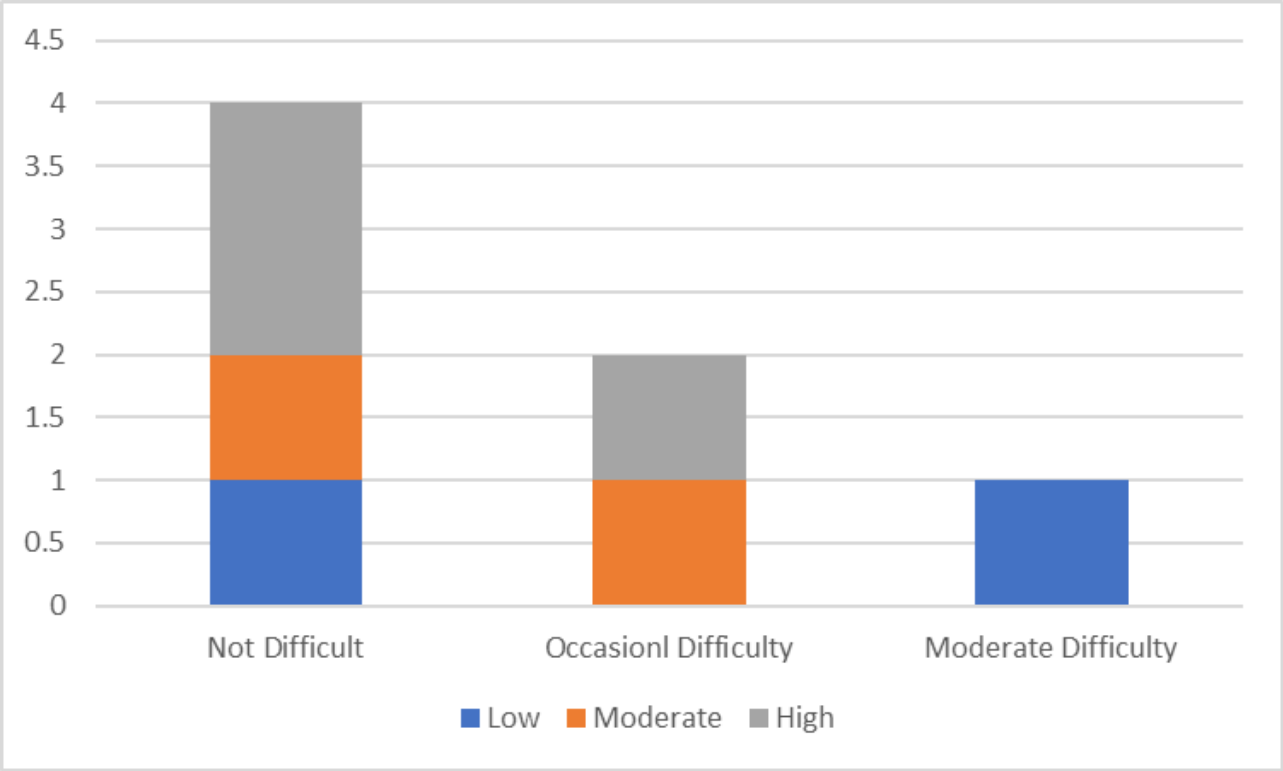


Fig 4.35 Was it difficult to tell where the sounds were coming from feedback

Q. 4. Which interaction system did you prefer for this type of task?

When it came to a preference of control scheme used, every user preferred the usage of VR compared to mouse and keyboard. Regardless of performance or perceived difficulty, users reported that they would strongly prefer using VR compared to a more traditional control scheme, because of intuitiveness, enjoyment, or novelty. This fits with other research that shows that users will gravitate towards things that they enjoy more or find interesting rather than focus on performance of a task (Kulshreshth et al., 2013; Martel et al., 2015; Ashdown et al., 2005).

Q. 5. Any other feedback on the experiment?

One user commented that they may have been distracted by the novelty of VR, as it wasn't something they had used prior to the experiment. This is a known factor, especially as VR is only becoming more widespread and available in recent years. This can be countered to an extent by allowing users to partake in a VR "experience" before experimenting, which allows them to become used to the equipment and be less distracted by the new technology. However, it can be difficult to fully remove that sense of wonder from completely new users.

Two users reported that the greatest difficulty was locating the box at higher or lower elevations compared to themselves, while locating the particular column was more straightforward. These were both users who were highly experienced and performed quite well at the experiment, scoring higher in the VR version of the experiment. This fits with known issues of spatial audio, where it can be more difficult to account for higher or lower elevations of sound sources. In theory, this can become easier when users have full head rotation as in real life, as they can tilt their heads to try and determine changes in the sound. However, while this is something we do subconsciously in the real world, it may take more of a conscious effort virtually, and is quite difficult to implement in a keyboard and mouse control scheme.

4.4.4.2 NASA-TLX Results

To broaden the results from the user questionnaire, we also asked each participant to fill out a NASA-TLX survey after each experiment part (after performing a VR experiment or Non-VR experiment). We hoped to use this in combination with the direct questionnaires to see patterns in user's perception of performance, difficulty, and frustration. The main categories we are examining are Performance, Effort, and Frustration. For Performance, we measure if user's feel they performed better in one type of experiment or the other, for Effort we want to see if the user felt that one type of task was more difficult than the other, and for Frustration we want to see if there was a noticeable difference in frustration levels between both types. A lower rating in

Performance indicates a perceived performance closer to perfect, a lower rating in Effort means less effort was required to complete the task, and a lower rating in Frustration means the user felt a lower amount of stress and frustration while performing the task. The full breakdown of user responses is shown in Table 4.11.

With regards to Performance, people generally felt that they performed better in the VR experiments (average score of 6.71, median of 5) compared to the Non-VR experiments (average score of 7.57, median score of 6). The users who scored the highest difference between the two Performance ratings were users who had a single very poor performance on an individual cube, leading to the experiment containing that cube being marked as a worse performance. It is possible this standout moment influenced their overall feelings on their performance. Also, two of the users who had the largest gaps in their ratings were also users who rated VR as more difficult in the questionnaire and responded as moderate and low experience levels in regard to video games.

When looking at Effort, we see again that people felt that VR took less effort (average of 7.86, median of 5) compared to Non-VR (average of 8.57, median of 6). Generally, the low and moderate users ranked the Effort as higher (11.75 average for Non VR, 10.75 average for VR) compared to the high experience group (5 average for Non VR, 3.67 average for VR). Three participants felt the level of effort was the same between both tests, three felt that VR was less effort, and only one felt that VR was more effort compared to Non-VR.

Finally, users felt less frustrated in VR (average of 7, median of 5) compared to Non-VR (average of 7.43, median of 6). The low experience users reported the largest amounts of frustration, with the moderate and high-level experience groups having generally lower frustration scores and slightly lower scores for frustration. The biggest gap was with one user (score of 4 for Non-VR and 10 for VR) who also felt that VR was more difficult than mouse and keyboard. This user performed poorly with one cube in VR,

which could be tied to their level of frustration with one control scheme compared to the other.

Participant	VR	Mental	Physical	Temporal	Performance	Effort	Frustration
1	No	14	4	4	6	17	6
1	Yes	10	4	4	3	14	5
2	No	11	9	9	8	13	12
2	Yes	11	8	9	11	13	12
3	No	5	1	4	3	3	4
3	Yes	4	1	4	2	3	3
4	No	6	4	8	10	3	4
4	Yes	6	5	8	14	4	10
5	No	14	5	10	6	14	15
5	Yes	13	5	10	3	12	11
6	No	4	2	2	14	4	6
6	Yes	4	3	2	9	4	4
7	No	4	3	3	6	6	5
7	Yes	4	3	3	5	5	4

Table 4.11 User NASA-TLX Responses

These results show that the control scheme used to perform a task or interact with an environment does affect a user's enjoyment and perceived performance, answering our research question. It also shows that different control schemes can impact a user's frustration at a task, and even reduce perceived workload while performing a task. This is important to recognise as it should be considered while designing virtual environments and experiences for users, regardless of user experience or the particular task at hand.

4.5 Conclusion

Over the course of this chapter, we have discussed the results from the experiments performed and the implications these results may have while referring to the appropriate research questions. With the maze navigation experiment we saw that while performance between audio and non-audio runs were similar, higher experience users were more likely to have an increased performance in audio compared to non-audio. We also saw that user movement through the maze did change between audio and non-audio mazes, highlighting that the presence of audio can change the way a user acts in an environment and their decision making. While discussing the results of the sound location experiment, we saw that while overall performance between VR and traditional control schemes was similar, certain sound locations were much easier for users to locate in VR compared to traditional control schemes. We also saw that experienced users were more likely to perform significantly better in VR compared to traditional control schemes. From looking at user head movements in the experiment, we saw that VR control schemes led users to use more natural search techniques compared to traditional control schemes, altering the way they interacted with the environment. By examining the qualitative measurements of the questionnaire, we saw that not only did the majority of users find VR control schemes more natural, but all users preferred using VR control schemes regardless of performance differences or perceived difficulty. Finally, through the NASA-TLX survey, we saw that people felt that VR took less effort, was less frustrating, and had better performance compared to traditional control schemes.

5 Conclusions and Further Research

5.1 Summary of Contributions

We will now summarise the findings from our experiments, both in terms of how they relate to each research question, and more general findings that were not originally searched for.

Regarding our literature review and meta-analysis of the relevant fields of research, we highlighted the impact that audio and particularly spatial audio can have on user immersion, user experience, and even user performance. We also showed that there is a particular research gap regarding spatial audio and the use of audio in serious games and virtual reality research. While examining review papers covering those fields, we saw that there was often no discussion or focus on the presence of audio in the reviewed papers, and quite often few to none of the reviewed papers made mention of audio themselves. All of this is despite the existing literature showing the power of spatial audio through performance improvement and improved user experience and presence.

5.2 Research Questions

In Table 5.1, we reiterate our research questions and their relevant experiments.

Research Question	Relevant Experiment
"Is there a performance difference between fully spatialised audio and traditional panning audio, while searching for audio sources?"	*Experiment 1*
"Is there a performance difference between different control schemes while searching for audio?"	*Experiment 3*
"Does spatial audio improve a user's awareness of their surroundings in a virtual environment?"	*Experiment 1, Experiment 2, Experiment 3*
"Does the control scheme implemented for a virtual environment change how a user interacts with virtual sounds?"	*Experiment 3*
"Do different control schemes affect a user's enjoyment or perceived performance?"	*Experiment 3*

Table 5.1 Research Questions

5 Conclusions and Further Research

(1) *"Is there a performance difference between fully spatialised audio and traditional panning audio, while searching for audio sources?"*

With Experiment 1 we aimed to examine differences in performance between spatial audio and traditional stereo audio in a sound source location test. As outlined previously, there were many issues with this first version of the experiment, primarily with the decision to attempt to perform the experiment remotely, which led to limited usable results.

From the results gathered, there was not a significant difference in performance between the two, which does fit with existing research (Mendonça et al, 2012) and general user feelings from our research into HRTF implementation in Counter Strike. To really benefit from the advantages of spatial audio, a certain degree of training is required, something that is supported by Mendonça's research and by anecdotal evidence provided by users on game related forums. Until a user becomes used to the differences in spatial audio compared to stereo audio, performance will remain mostly unchanged. Ideally, we would perform this experiment multiple times over a period of time with a set group of participants to examine any performance difference after training. This examination of the effect of training was not originally in the scope of our experimentation but is an interesting area of research to pursue in the future.

(2) *"Is there a performance difference between different control schemes while searching for audio?"*

From Experiment 3, we saw that generally there wasn't a statistically significant performance change between VR and mouse and keyboard control schemes while attempting to locate the sound sources. Overall performance in terms of mistakes made and time taken was quite similar between the two control schemes, though VR did perform quite significantly better when it came to locating Group 4 cubes (those directly above and below the user). However, it is difficult to make a strong claim on this as the amount

of data on those cubes is small, so replication of the experiment with emphasis on those locations is necessary.

In terms of individual performance, some users had a statistically significant improvement in time taken to find the cubes when using VR compared to when they used mouse and keyboard. These users all self-reported as highly experienced with games and virtual environments, so this performance increase could be tied to familiarity with game-like environments, making it easier to adapt to the environment or more comfortable and able to focus on the given task.

There were significant performance differences between users of high experience compared to those of low/moderate experience. It seems that regardless of control scheme, a user's experience with virtual environments and video games makes a large difference in their performance in this type of task.

In comparison to other research, performance changes when moving between different control schemes often depend on the type of task being performed. As seen by Kulshreshth et al (2013) when comparing head tracking and non-head tracking in different types of games, some games would have a performance increase when using head tracking, such as FPS or racing games, while flight games didn't have a statistically significant improvement. It is possible that similar to spatial audio vs panning audio, larger improvements will be seen with larger training periods. Also seen by Kulshreshth and in other research (Andersen et al, 2021), the experience level of the user can often impact their ability to improve at the task. Both authors saw a difference in performance between both groups based on experience. Andersen found that low-experience users benefited more from individualised HRTFs, while experienced players could perform better even with generalised HRTFs.

What we can gather from these results is that in the short term, there doesn't appear to be a statistically significant change in performance at simple sound location when using head mounted VR compared to traditional mouse and keyboard setups.

However, there also isn't a decrease in performance, so there isn't necessarily a drawback in shifting to a different control scheme for certain tasks. We also see that highly experienced individuals are more like to have a significant performance increase when switching control schemes and over the course of the experiment, which is important knowledge when designing similar experiments, finding subject cohorts, and designing VR experiences. As VR is growing more popular with the greater public, it is important to be aware of differences between experience groups, and also on how the difficulty or complexity of a task can impact user performance in different control schemes.

(3) "Does spatial audio improve a user's awareness of their surroundings in a virtual environment?"

In Experiment 2, the maze navigation experiment, overall performance across the whole cohort was similar across both audio and non-audio maze runs. However, when it came to non-audio mazes, all users made at least 1 mistake, whereas in audio mazes, some users achieved a perfect score. There was also a significant difference in performance based on user experience, with the highly experienced group making less mistakes and taking less time to complete the mazes.

When looking at user performance within these groups, the highly experienced group did have a significant performance difference between audio and non-audio mazes, performing better when audio was present. However, the sample size of this group is small, so further replication of this could strengthen the claim. This does fit with similar findings in Experiment 3, where highly experienced users are better able to benefit from differences and advantages between experiments.

Research which has examined audio location in navigation experiments (Grohn et al, 2005; Walker et al, 2006) has shown that the presence of audio is preferrable compared to a visual only environment, even if the difference in short term performance is minor. This is similar to the results we saw here. The greater performance increase

among highly experienced users is also similar to some work such as that done by Kulshreshth (2013).

(4) "Does the control scheme implemented for a virtual environment change how a user interacts with virtual sounds?"

By examining the player movement data gathered from Experiment 3, we can see how the player attempted to search for the source of the sound around them. While we don't have a level of detail to see very minute rotations, we can see the broader movements made by the user. Generally, between both control types, users would follow a similar pattern of panning their head left and right, homing in on the general direction the sound was coming from. As they got closer to the perceived source, these movements would become smaller and more controlled. This approximate location of the source was mostly the same between VR and mouse and keyboard. However, the difference usually came from determining the elevation of the sound.

When using a mouse and keyboard, after finding the column they believed the sound to originate from, users would try to determine the elevation by turning left and right to try and determine if the sound was higher or lower. Usually, it would come down to an educated guess based on what sounded "right". For the VR control scheme, users would tilt their heads at different angles, seeing if there was a difference in the sound. This appeared to mostly be an unconscious process, something that was simply a natural part of trying to locate the sound. While this wasn't always used and didn't always lead to a correct target, it was noticeable.

In the mouse and keyboard control scheme, there were keyboard shortcuts that would allow the user to tilt the camera in a similar way to head tracking, which was made clear in the documentation given to the user beforehand. However, none of the users made use of this. Some users when asked said they simply didn't remember or think of using it. In this case, the more naturalistic control scheme did lead to a change in how the

user attempted to search. With training, it is possible this change could lead to a performance increase.

The importance of these results is that as use of VR and head tracking controls becomes more common, more design considerations need to be made as to how they differ from previous virtual environments. By showing that people act differently and more similarly to real life when using these types of controls, it can emphasise the need to include realistic audio and physics to these types of environments and plan how inexperienced users may act differently compared to those highly experienced in virtual environments.

(5) *"Do different control schemes affect a user's enjoyment or perceived performance?"*

From Experiment 3, when comparing the usage of a traditional mouse and keyboard control scheme and a head mounted VR control scheme, we saw from user feedback that users overwhelmingly preferred the usage of VR (100%), regardless of any effect on their performance. Many users also felt that the mouse and keyboard control scheme was the more difficult of the two (~43%), with the remainder split between no difference (~28%) and VR being the more difficult control scheme (~28%). Finally, almost all users felt that VR was the more natural feeling control scheme (~86%).

The preference towards using VR is in line with some of the existing research we've examined in our literature review, as it is still a novel and entertaining technology for many people (Seibert et al, 2018; Andersen et al, 2021; Schuemie et al., 2001; Martel et al., 2015; Ashdown et al., 2005). Regardless of perceived difficulty, feelings of natural usage, and actual performance, all users said they would prefer to use VR in this type of environment in the future. In fact, some users stated the natural feeling of the control scheme is what made them prefer it for this type of task.

We also saw from the NASA-TLX results that there was some reduction in perceived workload when using the VR control scheme compared to the mouse and keyboard controls. People generally felt less of a mental demand with VR, and also seemed to be happier with their own personal performance.

The importance of these results is to highlight that the way in which we interact with a virtual environment has an impact on how we perceive the task. Even if there isn't a significant change in how well a user performs a task, they may feel that they did better or had an easier time depending on the control scheme they used. This tied in with the way people changed their movements while searching, and the general feeling of VR being more "natural", show that considerations should be made based not only on the quantifiable differences between control types, but on user's personal feelings. This is also similar to a lot of research discussed previously which examined the qualitative differences with using VR or head tracking controls compared to traditional methods (Andersen et al, 2021; Schuemie et al., 2001; Martel et al., 2015; Ashdown et al., 2005). Users overwhelmingly prefer to use VR or head tracking systems compared to traditional controls and tend to enjoy this type of control regardless of performance changes positive or negative.

The overall implication of the results of this thesis is to try and highlight how the type of control scheme used to interact with a virtual environment should be considered when designing virtual experiences. Beyond just the obvious change of controls, the way in which users consciously and unconsciously move and act within the environment is changed by the type of control scheme. This impact must be considered, especially as the metaverse continues to grow. We also aimed to show how sound can play an important part in virtual spaces and how users behave in those spaces. Finally, we tried to show how regardless of actual performance or experience, users prefer to use VR in virtual spaces, feel that VR is a more natural way to interact with a virtual space and often perceive these more positive control schemes as easier or leading to better performance. By leveraging this preference and perceived impact, we can improve future virtual spaces and aim to make virtual experiences that will be more appealing to wider audiences.

5.3 Future Research

5.3.1 Greater Examination of Effect of Elevation

Similar experiments have mostly used set elevations at the same level as the user when analysing sound location in virtual environments. We saw from our experiment that elevation can create greater difficulty in determining the source of a sound, and this is especially true when the sound is directly above or below the user. From our results we saw that VR did perform noticeably better when it came to locating a sound source directly above or below the user, but the amount of data we had on this was limited. In further replications of this type of experiment, it would be beneficial to guarantee a placement of a sound source in a position directly above or below the user in each experiment, increasing the number of tests for that position and hopefully determining if VR does in fact continue to provide an improvement to performance.

5.3.2 Movement Differences Between Experience Levels

It was seen in our research that the level of experience people had with video games caused noticeable differences in performance, and there were some differences in movements of those groups within the experiments themselves. It could be an interesting area of research to examine more closely how head movements might differ between high and low experience groups of subjects. There is a clear difference in player movement of a character, as those with more experience are more used to moving a character in a virtual space. When it comes to potential differences in head movements, especially in terms of minor movements or while searching for sound sources, there is a benefit to expanding the research to see if there is still a performance gap between the two groups, or if the more natural and intuitive movements could be found to be mostly similar.

5.3.3 Replication through Further Studies

Ideally the specific types of experiments covered in this thesis would be replicated with larger subject cohorts so that the results discussed throughout could be verified. Since these experiments were performed with smaller cohorts, being able to show that the results maintain true with more data would strengthen research claims.

Replication of these experiments should be relatively straightforward, either through usage of the executables created in this thesis, or creating similar work based on what is covered both in the discussions and design chapters, as well as the code in the appendices. The primary reason for not having done this already is due to a number of factors, including (1) the focus of the research on the design of an experimental framework to conduct the evaluations, which can be used for further studies; (2) the focus to complement the study with the gathering of qualitative data during the experiment, and (3) the on-going COVID-19 pandemic restricting gatherings of people and adding health risks to the usage of head-mounted display units, which impacted the recruitment of additional subjects Future work will expand the subject cohorts as it becomes more possible to organise in-person experimentation.

5.3.4 Impact of Training

We've seen that user experience can create differences in performance and comfort in virtual environments, and there is evidence that training can cause improvement in user performance, especially when it comes to spatial audio (Mendonça et al, 2012). Ideally in any replication of the experiments in this thesis or in similar ones, the experiments would be performed two or more times with the same subject cohort, to study the impact of training with the system. From Mendonça's research and from feedback of gamers playing Overwatch and Counter Strike, we know that there is an adjustment period with spatial audio, where it can cause no benefit or potentially cause issues until the user has become more accustomed to the sound changes. Being able to

further verify this theory, as well as seeing if the effect is the same when comparing VR to traditional control schemes, could benefit spatial audio experiments in virtual environments going further.

5.3.5 Impact on Higher Complexity Tasks

With regards to the experiments performed in this thesis, the tasks given to the subjects were quite simplified. The audio location task had clear indicators of where the sound could be coming from, and the user had no capacity for locomotion, being in a locked position and only being able to rotate their “head”. It is possible that spatial audio could have a greater impact in an environment where users are fully locomotive, able to move freely in an environment. Having unmarked sound sources or moving sound sources could also have an impact on user performance. There is room for further experimentation with environments and tasks that more closely resemble either games or training environments seen in serious games, and to examine the impacts of control schema and spatial audio in this more complex situation.

As part of this, one future experiment is to perform the maze navigation experiment with audio included but compare VR and traditional mouse and keyboard control schemes. We have seen the general preference towards VR control schemes and the more intuitive feel to it, so maze navigation while avoiding sonified obstacles is a potentially useful next step of complexity.

5.4 Research Challenges Due to COVID-19

During the course of the research covered in this thesis, the global COVID-19 pandemic began. The maze navigation and varying control scheme experiments were performed at the end of 2019, and the original plan was to repeat the experiments in early 2020 to increase the amount of data gathered and strengthen the findings from early

analysis of the first set of experiments. Unfortunately, before that could be done, COVID grew to be quite a serious situation, and it was simply not possible to perform the experiments. After the documented issues experienced trying to perform early experimentation remotely, and especially with the usage of VR equipment, it wasn't feasible for the experiments to be performed outside of an in-person setting. As well as this, there are the other impacts of COVID and lockdown on mental health.

Ireland in particular had a very strong response to COVID, being in a high state of lockdown for almost all of the time between March 2020 and September 2021. With the restrictions in place, it wasn't possible to perform in person experimentation, and would have been potentially dangerous with the usage of head-mounted equipment in particular. This is a major factor in the limited sample size of the experiments performed.

During the time where it wasn't possible to perform further experiments, the research work was adapted to focus on further expansion of current research and results, as well as planning for future work. During this time the meta-analysis of review papers in related fields was performed, allowing us to further highlight the lack of focus on audio in fields such as serious games, VR, and virtual environments. While the research gap was clear from papers examined throughout the thesis, being able to take the time to study these reviews further highlights not only the lack of overall research, but that researchers themselves aren't necessarily aware of the importance of audio in these fields in technologies.

We were also able to more closely examine the impact of user experience on experiments involving virtual environments and spatial audio. Similar research usually made some mention of user experience but didn't delve too deeply into the effect it could have on user performance or strategies used. Originally there wasn't much focus on how experience could make people able to improve more quickly, or change how people would move in the experiments, but the extra time allowed us to analyse the differences more closely in these user experience groups.

References and Bibliography

Anderson, E. F., McLoughlin, L., Liarokapis, F., Peters, C., Petridis, P., & de Freitas, S. (2010). Developing serious games for cultural heritage: A state-of-the-art Review. Virtual Reality, 14(4), 255–275. https://doi.org/10.1007/s10055-010-0177-3

Ashdown, M., Oka, K., & Sato, Y. (2005). Combining head tracking and mouse input for a gui on multiple monitors. Conference on Human Factors in Computing Systems - Proceedings, 1188–1191. https://doi.org/10.1145/1056808.1056873

Ballora, M., Giacobe, N. A., & Hall, D. L. (2011). Songs of cyberspace: an update on sonifications of network traffic to support situational awareness. Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications 2011, 8064(May), 80640P. https://doi.org/10.1117/12.883443

Baños, R. M., Botella, C., Rubió, I., Quero, S., García-Palacios, A., & Alcañiz, M. (2008). Presence and emotions in virtual environments: The influence of stereoscopy. Cyberpsychology and Behavior, 11(1), 1–8. https://doi.org/10.1089/cpb.2007.9936

Bathke, D. J., Haigh, T., Bernadt, T., Wall, N., Hill, H., & Carson, A. (2019). Using Serious Games to Facilitate Collaborative Water Management Planning Under Climate Extremes. Journal of Contemporary Water Research & Education, 167(1), 50–67. https://doi.org/10.1111/j.1936-704x.2019.03311.x

Begault, D. R., & Pittman, M. T. (1996). Three-Dimensional Audio Versus Head-Down Traffic Alert and Collision Avoidance System Displays. International Journal of Aviation Psychology, 6(1), 79–93. https://doi.org/10.1207/s15327108ijap0601_5

Beig, M., Kapralos, B., Collins, K., & Mirza-Babaei, P. (2019). An Introduction to Spatial Sound Rendering in Virtual Environments and Games. The Computer Games Journal, 8(3–4), 199–214. https://doi.org/10.1007/s40869-019-00086-0

References and Bibliography

Bell, J. T., Fogler, H. S., & Arbor, A. (1995). The Investigation and Application of Virtual Reality as an Educational Tool. Proceedings of the American Society for Engineering Education, (2513), 1–11. Retrieved from https://www.researchgate.net/profile/Hs_Fogler/publication/247920944_The_Investigation_and_Application_of_Virtual_Reality_as_an_Educational_Tool/links/55f721fb08ae07629dbfcfee.pdf

Bellotti, F., Berta, R., De Gloria, A., & Margarone, M. (2002). Using 3D sound to improve the effectiveness of the advanced driver assistance systems. Personal and Ubiquitous Computing, 6(3), 155–163. https://doi.org/10.1007/s007790200016

BENFORD, S., BOWERS, J., FAHLÉN, L. E., & GREENHALGH, C. (1994). Managing Mutual Awareness in Collaborative Virtual Environments, (August 1994), 223–236. https://doi.org/10.1142/9789814350938_0018

Bente, G., & Breuer, J. (2010). Why so serious? On the Relation of Serious Games and Learning. Eludamos - Journal for Computer Game Culture, 4(1), 7–24. Retrieved from http://www.eludamos.org/index.php/eludamos/article/viewArticle/vol4no1-2

Bishop, I. D., & Stock, C. (2010). Using collaborative virtual environments to plan wind energy installations. Renewable Energy, 35(10), 2348–2355. https://doi.org/10.1016/j.renene.2010.04.003

Bonneel, N., Suied, C., Viaud-Delmon, I., & Drettakis, G. (2010). Bimodal perception of audio-visual material properties for virtual environments. ACM Transactions on Applied Perception, 7(1), 1–16. https://doi.org/10.1145/1658349.1658350

Brock, D., Stroup, J. L., & Ballas, J. A. (2002). Effects of 3D Auditory Display on Dual Task Performance in a Simulated Multiscreen Watchstation Environment. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 46(17), 1570–1573. https://doi.org/10.1177/154193120204601709

Broderick, J., Duggan, J., & Redfern, S. (2017). Using Auditory Display Techniques to Enhance Decision Making and Perceive Changing Environmental Data Within a 3D Virtual Game Environment, (Icad), 147–151. https://doi.org/10.21785/icad2017.012

References and Bibliography

Broderick, J., Duggan, J., & Redfern, S. (2018). The Importance of Spatial Audio in Modern Games and Virtual Environments. 2018 IEEE Games, Entertainment, Media Conference, GEM 2018, 507–510. https://doi.org/10.1109/GEM.2018.8516445

Broderick, J., Duggan, J., & Redfern, S. (2016). Using game engines for marine visualisation and collaboration. 2016 International Conference on Image, Vision and Computing, ICIVC 2016, 96–101. https://doi.org/10.1109/ICIVC.2016.7571280

Bronkhorst, A. W., Veltman, J. A., & Van Breda, L. (1996). Application of a three-dimensional auditory display in a flight task. Human Factors, 38(1), 23–33. https://doi.org/10.1518/001872096778940859

Calderón, A., & Ruiz, M. (2015). A systematic literature review on serious games evaluation: An application to software project management. Computers and Education, 87, 396–422. https://doi.org/10.1016/j.compedu.2015.07.011

Canossa, A., Drachen, A., & Sørensen, J. R. M. (2011). Arrrgghh!!! - Blending quantitative and qualitative methods to detect player frustration. Proceedings of the 6th International Conference on the Foundations of Digital Games, FDG 2011, 61–68. https://doi.org/10.1145/2159365.2159374

Cao, S., Nandakumar, K., Babu, R., & Thompson, B. (2020). Game play in virtual reality driving simulation involving head-mounted display and comparison to desktop display. Virtual Reality, 24(3), 503–513. https://doi.org/10.1007/s10055-019-00412-x

Chang, E., Kim, H. T., & Yoo, B. (2020). Virtual Reality Sickness: A Review of Causes and Measurements. International Journal of Human-Computer Interaction, 36(17), 1658–1682. https://doi.org/10.1080/10447318.2020.1778351

Chao, C. J., Wu, S. Y., Yau, Y. J., Feng, W. Y., & Tseng, F. Y. (2017). Effects of three-dimensional virtual reality and traditional training methods on mental workload and training performance. Human Factors and Ergonomics In Manufacturing, 27(4), 187–196. https://doi.org/10.1002/hfm.20702

References and Bibliography

Checa, D., & Bustillo, A. (2020). A review of immersive virtual reality serious games to enhance learning and training. Multimedia Tools and Applications, 79(9–10), 5501–5527. https://doi.org/10.1007/s11042-019-08348-9

Chen, G., Li, B., Tian, F., Ji, P., & Li, W. (2012). Design and implementation of a 3D ocean virtual reality and visualization engine. Journal of Ocean University of China, 11(4), 481–487. https://doi.org/10.1007/s11802-012-2112-6

Cheng, C. I., & Wakefield, G. H. (2001). Introduction to head-related transfer functions (HRTFs): Representations of HRTFs in time, frequency, and space. AES: Journal of the Audio Engineering Society, 49(4), 231–249.

Chittaro, L., & Buttussi, F. (2020). Learning safety through public serious games: A study of “prepare for impact” on a very large, international sample of players. IEEE Transactions on Visualization and Computer Graphics, 1–12. https://doi.org/10.1109/TVCG.2020.3022340

Citizen’s Information (2020, Mar, 27). New measures to combat COVID-19 from midnight on 27 March 2020. https://whatsnew.citizensinformation.ie/2020/03/27/new-measures-to-combat-covid-19-from-midnight-on-27-march-2020/

Coburn, J. Q., Freeman, I., & Salmon, J. L. (2017). A Review of the Capabilities of Current Low-Cost Virtual Reality Technology and Its Potential to Enhance the Design Process. Journal of Computing and Information Science in Engineering, 17(3). https://doi.org/10.1115/1.4036921

Collins, K., & Kapralos, B. (2019). Pseudo-haptics: leveraging cross-modal perception in virtual environments. Senses and Society, 14(3), 313–329. https://doi.org/10.1080/17458927.2019.1619318

Conference, I., & Display, A. (2015). REALTIME SONIFICATION AND VISUALISATION OF NETWORK METADATA (The NetSon Project) David Worrall Fraunhofer Institute for Integrated Circuits IIS, 337–339.

References and Bibliography

Connolly, T. M., Boyle, E. A., MacArthur, E., Hainey, T., & Boyle, J. M. (2012). A systematic literature review of empirical evidence on computer games and serious games. Computers and Education, 59(2), 661–686. https://doi.org/10.1016/j.compedu.2012.03.004

Dargar, S., Kennedy, R., Lai, W., Arikatla, V., & De, S. (2015). Towards immersive virtual reality (iVR): a route to surgical expertise. Journal of Computational Surgery, 2(1). https://doi.org/10.1186/s40244-015-0015-8

Deary, I. J., & Der, G. (2005). Reaction time, age, and cognitive ability: Longitudinal findings from age 16 to 63 years in representative population samples. Aging, Neuropsychology, and Cognition, 12(2), 187–215. https://doi.org/10.1080/13825580590969235

Debashi, M., & Vickers, P. (2018). Sonification of network traffic flow for monitoring and situational awareness. PLoS ONE (Vol. 13). https://doi.org/10.1371/journal.pone.0195948

Debashi, M., & Vickers, P. (2018). Sonification of Network Traffic for Detecting and Learning about Botnet Behavior. IEEE Access, 6, 33826–33839. https://doi.org/10.1109/ACCESS.2018.2847349

Der, G., & Deary, I. J. (2006). Age and sex differences in reaction time in adulthood: Results from the United Kingdom health and lifestyle survey. Psychology and Aging, 21(1), 62–73. https://doi.org/10.1037/0882-7974.21.1.62

Dernoncourt, F. (2012). Replacing the computer mouse. MIT CSAIL Student Workshop, 1–9. Retrieved from http://francky.me

Diemer, J., Alpers, G. W., Peperkorn, H. M., Shiban, Y., & Mühlberger, A. (2015). The impact of perception and presence on emotional reactions: A review of research in virtual reality. Frontiers in Psychology, 6(JAN), 1–9. https://doi.org/10.3389/fpsyg.2015.00026

Donmez, B., Cummings, M. L., & Graham, H. D. (2009). Auditory decision aiding in supervisory control of multiple unmanned aerial vehicles. Human Factors, 51(5), 718–729. https://doi.org/10.1177/0018720809347106

References and Bibliography

Ebrahimi, E., Babu, S. V., Pagano, C. C., & Jörg, S. (2016). An empirical evaluation of visuo-haptic feedback on physical reaching behaviors during 3D interaction in real and immersive virtual environments. ACM Transactions on Applied Perception, 13(4), 1–21. https://doi.org/10.1145/2947617

Feng, Z., González, V. A., Amor, R., Lovreglio, R., & Cabrera-Guerrero, G. (2018). Immersive virtual reality serious games for evacuation training and research: A systematic literature review. Computers and Education, 127, 252–266. https://doi.org/10.1016/j.compedu.2018.09.002

FISCHER H. (1970). Modelling systems of operations research, (Modellsysteme der Operationsforschung), 91–94.

Fleming, T. M., Cheek, C., Merry, S. N., Thabrew, H., Bridgman, H., Stasiak, K., … Hetrick, S. (2014). Serious games for the treatment or prevention of depression: A systematic review. Revista de Psicopatologia y Psicologia Clinica, 19(3), 227–242. https://doi.org/10.5944/rppc.vol.19.num.3.2014.13904

Flood, S., Cradock-Henry, N. A., Blackett, P., & Edwards, P. (2018). Adaptive and interactive climate futures: Systematic review of “serious games” for engagement and decision-making. Environmental Research Letters. https://doi.org/10.1088/1748-9326/aac1c6

Fozard, J. L., Vercruyssen, M., Reynolds, S. L., Hancock, P. A., & Quilter, R. E. (1994). Age differences and changes in reaction time: The Baltimore longitudinal study of aging. Journals of Gerontology, 49(4). https://doi.org/10.1093/geronj/49.4.P179

Freeman, G., & Maloney, D. (2021). Body, Avatar, and Me: The Presentation and Perception of Self in Social Virtual Reality. Proceedings of the ACM on Human-Computer Interaction, 4(CSCW3), 239:1–239:27. Retrieved from https://doi.org/10.1145/3432938

Freina, L., & Ott, M. (2015). A literature review on immersive virtual reality in education: State of the art and perspectives. Proceedings of ELearning and Software for Education (ELSE)(Bucharest, Romania, April 23--24, 2015), (April), 8.

References and Bibliography

Gaver, W. W. (1986). Auditory Icons: Using Sound in Computer Interfaces. Human-Computer Interaction, 2(2), 167–177. https://doi.org/10.1207/s15327051hci0202_3

Godin, K. (2019, July, 11). Spatial sound in Unity. https://docs.microsoft.com/en-us/windows/mixed-reality/develop/unity/spatial-sound-in-unity

Google. (2021). Google Cardboard. https://arvr.google.com/cardboard/

Gorbanev, I., Agudelo-Londoño, S., González, R. A., Cortes, A., Pomares, A., Delgadillo, V., … Muñoz, Ó. (2018). A systematic review of serious games in medical education: quality of evidence and pedagogical strategy. Medical Education Online, 23(1). https://doi.org/10.1080/10872981.2018.1438718

Gore, A. (1998). The Digital Earth. Australian Surveyor, 43(2), 89–91. https://doi.org/10.1080/00050348.1998.10558728

Greenwood, P., Sago, J., Richmond, S., & Chau, V. (2009). Using game engine technology to create real-time interactive environments to assist in planning and visual assessment for infrastructure. 18th World IMACS Congress and MODSIM 2009 - International Congress on Modelling and Simulation: Interfacing Modelling and Simulation with Mathematical and Computational Sciences, Proceedings, (July), 2229–2235. Retrieved from http://www.mssanz.org.au/modsim09/F8/greenwood.pdf

Gröhn, M., Lokki, T., & Takala, T. (2005). Comparison of Auditory, Visual, and Audiovisual Navigation in a 3D Space. ACM Transactions on Applied Perception, 2(4), 564–570. https://doi.org/10.1145/1101530.1101558

Hall, W., & Keynes, M. (1908). AudioGPS : spatial audio in a minimal attention interface. Personal and Ubiquitous Computing, 6(4), 253–259.

Hamidi, F., & Kapralos, B. (2009). A Review of Spatial Sound for Virtual Environments and Games with Graphics Processing Units. The Open Virtual Reality Journal, 1(1), 8–17. https://doi.org/10.2174/1875323x00901010008

References and Bibliography

Held, R. M., & Durlach, I. (1983). Telepresence, 109–112.

Heller, L. M., Schenker, A., Grover, P., Gardner, M., & Liu, F. (2017). Evaluating Two Ways to Train Sensitivity to Echoes to Improve Echolocation (pp. 159–166). https://doi.org/10.21785/icad2017.053

Henriquez, P., Alonso, J. B., Ferrer, M. A., & Travieso, C. M. (2014). Review of automatic fault diagnosis systems using audio and vibration signals. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 44(5), 642–652. https://doi.org/10.1109/TSMCC.2013.2257752

Hermann, T., Hunt, A., & Neuhoff, J. (2011). The Sonification Handbook. The Sonification Handbook.

Hernandez, J. (2009). Auditory Display in Virtual Environments to Assist Navigational and Interaction, (September).

Hookham, G., & Nesbitt, K. (2019). A Systematic Review of the Definition and Measurement of Engagement in Serious Games. In ACM International Conference Proceeding Series. https://doi.org/10.1145/3290688.3290747

Huang, H., Solah, M., Li, D., & Yu, L. F. (2019). Audible panorama: Automatic spatial audio generation for panorama imagery. Conference on Human Factors in Computing Systems - Proceedings, (Chi), 1–11. https://doi.org/10.1145/3290605.3300851

Indraprastha, A., & Shinozaki, M. (2009). The Investigation on Using Unity3D Game Engine in Urban Design Study. ITB Journal of Information and Communication Technology, 3(1), 1–18. https://doi.org/10.5614/itbj.ict.2009.3.1.1

Innes, J. E., & Booher, D. E. (2000). Public Participation in Planning : New Strategies for the 21st Century, 1–39.

Ivanov, S. V., Kosukhin, S. S., Kaluzhnaya, A. V., & Boukhanovsky, A. V. (2012). Simulation-based collaborative decision support for surge floods prevention in St. Petersburg. Journal of Computational Science, 3(6), 450–455. https://doi.org/10.1016/j.jocs.2012.08.005

References and Bibliography

Jäger, A., & Hadjakos, A. (2017). Navigation in an Audio-only First Person Adventure Game (pp. 152–158). https://doi.org/10.21785/icad2017.033

Jain, A., Bansal, R., Kumar, A., & Singh, K. (2015). A comparative study of visual and auditory reaction times on the basis of gender and physical activity levels of medical first year students. International Journal of Applied and Basic Medical Research, 5(2), 124. https://doi.org/10.4103/2229-516x.157168

Jensen, M., & Melzack, N. R. (2004). SpiderWorld and SnowWorld, (July).

Kapralos, B., F. Moussa, K. Collins, and A. Dubrowski. 2017. “Levels of Fidelity and Multimodal Interactions.” In Techniques to Improve the Effectiveness of Serious Games, edited by P. Wouters and H. van Oostendorp, 79–101. New York: Springer.

Kapralos, B., Moussa, F., Collins, K., & Dubrowski, A. (2017). Instructional Techniques to Facilitate Learning and Motivation of Serious Games. Instructional Techniques to Facilitate Learning and Motivation of Serious Games, 79–101. https://doi.org/10.1007/978-3-319-39298-1

Kato, M., Uematsu, H., Kashino, M., & Hirahara, T. (2003). The effect of head motion on the accuracy of sound localization. Acoustical Science and Technology, 24(5), 315–317. https://doi.org/10.1250/ast.24.315

Kern, A. C., & Ellermeier, W. (2020). Audio in VR: Effects of a Soundscape and Movement-Triggered Step Sounds on Presence. Frontiers in Robotics and AI, 7(February), 1–13. https://doi.org/10.3389/frobt.2020.00020

Khalid, S., Ullah, S., Ali, N., Alam, A., Rabbi, I., Rehman, I. U., & Azhar, M. (2019). Navigation AIDS in collaborative virtual environments: Comparison of 3DML, audio, textual, arrows-casting. IEEE Access, 7, 152979–152989. https://doi.org/10.1109/ACCESS.2019.2948285

References and Bibliography

Klippel, A., Zhao, J., Oprean, D., Wallgrün, J. O., Stubbs, C., La Femina, P., & Jackson, K. L. (2020). The value of being there: toward a science of immersive virtual field trips. Virtual Reality, 24(4), 753–770. https://doi.org/10.1007/s10055-019-00418-5

Knight, J. F., Carley, S., Tregunna, B., Jarvis, S., Smithies, R., de Freitas, S., … Mackway-Jones, K. (2010). Serious gaming technology in major incident triage training: A pragmatic controlled trial. Resuscitation, 81(9), 1175–1179. https://doi.org/10.1016/j.resuscitation.2010.03.042

Kulshreshth, A., & La Viola, J. J. (2013). Evaluating performance benefits of head tracking in modern video games. SUI 2013 - Proceedings of the ACM Symposium on Spatial User Interaction, 53–60. https://doi.org/10.1145/2491367.2491376

Kuppanda, T., Degara, N., Worrall, D., Thoshkahna, B., & Meinard, M. (2015). VIRTUAL REALITY PLATFORM FOR SONIFICATION EVALUATION Fraunhofer Institute for Integrated Circuits IIS , Erlangen , Germany International Audio Laboratories , Erlangen , Germany, 117–124. Retrieved from http://opensoundcontrol.org/

Laamarti, F., Eid, M., & El Saddik, A. (2014). IJFS 2020 - An overview of serious games (needed to review Sen X). International Journal of Computer Games Technology, 2014.

Lackey, S. J., Salcedo, J. N., Szalma, J. L., & Hancock, P. A. (2016). The stress and workload of virtual reality training: the effects of presence, immersion and flow. Ergonomics, 59(8), 1060–1072. https://doi.org/10.1080/00140139.2015.1122234

Lawlor, S., Neumann, T. (2016, March). Overwatch - The Elusive Goal: Play by Sound. GDC. https://www.gdcvault.com/play/1023317/Overwatch-The-Elusive-Goal-Play

Lewis, M., & Jacobson, J. (2002). Games engines in scientific research. Communications of the ACM, 45(1), 27–31. https://doi.org/10.1145/502269.502288

Li, W., Chen, G., Kong, Q., Wang, Z., & Qian, C. (2011). A VR-Ocean system for interactive geospatial analysis and 4D visualization of the marine environment around Antarctica. Computers and Geosciences, 37(11), 1743–1751. https://doi.org/10.1016/j.cageo.2011.04.009

References and Bibliography

Lin, H., Chen, M., & Lu, G. (2013). Virtual Geographic Environment: A Workspace for Computer-Aided Geographic Experiments. Annals of the Association of American Geographers, 103(3), 465–482. https://doi.org/10.1080/00045608.2012.689234

Lin, H., Chen, M., Lu, G., Zhu, Q., Gong, J., You, X., … Hu, M. (2013). Virtual Geographic Environments (VGEs): A New Generation of Geographic Analysis Tool. Earth-Science Reviews, 126, 74–84. https://doi.org/10.1016/j.earscirev.2013.08.001

Liu, S. H., Yu, N. H., Chan, L., Peng, Y. H., Sun, W. Z., & Chen, M. Y. (2019). PhantomLegs: Reducing virtual reality sickness using head-worn haptic devices. 26th IEEE Conference on Virtual Reality and 3D User Interfaces, VR 2019 - Proceedings, (March), 817–826. https://doi.org/10.1109/VR.2019.8798158

Liu, X., Zhang, X., Chi, T., Qu, H., & Jiang, Y. (2009). Study on China Digital Ocean Prototype System. 2009 WRI World Congress on Software Engineering, WCSE 2009, 1, 466–469. https://doi.org/10.1109/WCSE.2009.337

Lohse, K. R., Hilderman, C. G. E., Cheung, K. L., Tatla, S., & Van Der Loos, H. F. M. (2014). Virtual reality therapy for adults post-stroke: A systematic review and meta-analysis exploring virtual environments and commercial games in therapy. PLoS ONE, 9(3). https://doi.org/10.1371/journal.pone.0093318

Lum, H. C., Greatbatch, R., Waldfogle, G., & Benedict, J. (2018). How immersion, presence, emotion, & workload differ in virtual reality and traditional game mediums. Proceedings of the Human Factors and Ergonomics Society, 3, 1474–1478. https://doi.org/10.1177/1541931218621334

Malimbo. (2017, November, 1). Feedback regarding HRTF and Sound Occlusion 31/10/17. https://www.reddit.com/r/GlobalOffensive/comments/7a3ec7/feedback_regarding_hrtf_and_sound_occlusion_311017/

Maloney, D., & Freeman, G. (2020). Falling Asleep Together: What Makes Activities in Social Virtual Reality Meaningful to Users. CHI PLAY 2020 - Proceedings of the Annual Symposium on Computer-Human Interaction in Play, 510–521. https://doi.org/10.1145/3410404.3414266

References and Bibliography

Maloney, D., Freeman, G., & Wohn, D. Y. (2020). “Talking without a Voice”: Understanding Non-Verbal Communication in Social Virtual Reality. Proceedings of the ACM on Human-Computer Interaction, 4(CSCW2). https://doi.org/10.1145/3415246

Martel, E., & Muldner, K. (2017). Controlling VR games: control schemes and the player experience. Entertainment Computing, 21, 19–31. https://doi.org/10.1016/j.entcom.2017.04.004

Mateevitsi, V., Sfakianos, M., Lepouras, G., & Vassilakis, C. (2008). A game-engine based virtual museum authoring and presentation system. Proceedings - 3rd International Conference on Digital Interactive Media in Entertainment and Arts, DIMEA 2008, 451–457. https://doi.org/10.1145/1413634.1413714

Mayer, I., Zhou, Q., Keijser, X., & Abspoel, L. (2014). Gaming the Future of the Ocean : The Marine Spatial Planning Challenge 2050 Policy Making in the Wake of Complexity, 150–162.

McCarthy, L., & Olsen, K. N. (2017). A “looming bias” in spatial hearing? Effects of acoustic intensity and spectrum on categorical sound source localization. Attention, Perception, and Psychophysics, 79(1), 352–362. https://doi.org/10.3758/s13414-016-1201-9

Meghanathan, R. N., Ruediger-Flore, P., Hekele, F., Spilski, J., Ebert, A., & Lachmann, T. (2021). Spatial sound in a 3D virtual environment: All bark and no bite? Big Data and Cognitive Computing, 5(4), 1–16. https://doi.org/10.3390/bdcc5040079

Mendonça, C., Campos, G., Dias, P., Vieira, J., Ferreira, J. P., & Santos, J. A. (2012). On the improvement of localization accuracy with non-individualized HRTF-based sounds. AES: Journal of the Audio Engineering Society, 60(10), 821–830.

Merlo, A., Dalcò, L., & Fantini, F. (2012). Game engine for Cultural Heritage New Opportunities in the Relation Between Simplified Models and Database. 18th International Conference on Virtual Systems and Multimedia (VSMM), 623–628.

Microsoft. (2015, December, 1). Spatial Audio. https://www.microsoft.com/en-us/research/project/spatial-audio/

References and Bibliography

Microsoft. (2019, July, 11). Audio in mixed reality. https://docs.microsoft.com/en-us/windows/mixed-reality/design/spatial-sound

Microsoft. (2020). Mixed Reality Toolkit. https://github.com/Microsoft/MixedRealityToolkit-Unity

Moghadam, K., Banigan, C., & Ragan, E. D. (2020). Scene Transitions and Teleportation in Virtual Reality and the Implications for Spatial Awareness and Sickness. IEEE Transactions on Visualization and Computer Graphics, 26(6), 2273–2287. https://doi.org/10.1109/TVCG.2018.2884468

Moghimi, M., Stone, R., Rotshtein, P., & Cooke, N. (2016). The Sense of embodiment in Virtual Reality. Presence: Teleoperators & Virtual Environments, 25(2), 81–107. https://doi.org/10.1162/PRES

Moldoveanu, A. D. B., Stanica, I., Dascalu, M. I., Bodea, C. N., Flamaropol, D., Moldoveanu, F., … Unnpórsson, R. (2017). Virtual environments for training visually impaired for a sensory substitution device. 2017 Zooming Innovation in Consumer

Mordor Intelligence. (2021). GAMING MARKET - GROWTH, TRENDS, COVID-19 IMPACT, AND FORECASTS (2021 - 2026). https://www.mordorintelligence.com/industry-reports/global-gaming-market

Moustakas, N., Floros, A., Rovithis, E., & Vogklis, K. (2019). Augmented audio-only games: A new generation of immersive acoustic environments through advanced mixing. Proceedings of the audio engineering society 146th international convention, Dublin, Ireland.

Electronics International Conference: Galvanize Your Creativity, ZINC 2017, 26–29. https://doi.org/10.1109/ZINC.2017.7968654

Moll, J., Huang, Y., & Sallnäs, E. L. (2010). Audio makes a difference in haptic collaborative virtual environments. Interacting with Computers, 22(6), 544–555. https://doi.org/10.1016/j.intcom.2010.06.001

References and Bibliography

Moore, N., Dempsey, K., Hockey, P., Jain, S., Poronnik, P., Shaban, R. Z., & Ahmadpour, N. (2021). Innovation During a Pandemic: Developing a Guideline for Infection Prevention and Control to Support Education Through Virtual Reality. Frontiers in Digital Health, 3(July), 1–5. https://doi.org/10.3389/fdgth.2021.628452

Morrison-Smith, S., Hofmann, M., Li, Y., & Ruiz, J. (2016). Using audio cues to support motion gesture interaction on mobile devices. ACM Transactions on Applied Perception, 13(3). https://doi.org/10.1145/2897516

NASA. (2010). NATA-TLX. NASA. https://humansystems.arc.nasa.gov/groups/tlx/

Nehme, C. E., & Cummings, M. L. (2006). Audio Decision Support for Supervisory Control of Unmanned Vehicles: Literature Review. MIT Report for Charles River Analytics, 1–22.

Nelson, W. T., Hettinger, L. J., Cunningham, J. A., Brickman, B. J., Haas, M. W., & McKinley, R. L. (1998). Effects of localized auditory information on visual target detection performance using a helmet-mounted display. Human Factors, 40(3), 452–460. https://doi.org/10.1518/001872098779591304

Niederriter, B., Rong, A., Aqlan, F., & Yang, H. (2020). Sensor-Based Virtual Reality for Clinical Decision Support in the Assessment of Mental Disorders. IEEE Conference on Computatonal Intelligence and Games, CIG, 2020-August, 666–669. https://doi.org/10.1109/CoG47356.2020.9231896

Night_Not_Day. (2020, December, 2). Night's Ultimate Audio Guide for CS:GO [2020 Edition]. https://www.reddit.com/r/GlobalOffensive/comments/k57r6a/nights_ultimate_audio_guide_for_csgo_2020_edition/

Norman, G., Dore, K., & Grierson, L. (2012). The minimal relationship between simulation fidelity and transfer of learning. Medical Education, 46(7), 636–647. https://doi.org/10.1111/j.1365-2923.2012.04243.x

The NPD Group/Games Market Dynamics: US

References and Bibliography

Oakley, I., McGee, M. R., Brewster, S., & Gray, P. (2000). Putting the feel in “look and feel.” Conference on Human Factors in Computing Systems - Proceedings, 2(1), 415–422. https://doi.org/10.1145/332040.332467

Oculus. (2021). Quest 2. https://www.oculus.com/quest-2/

Oculus. (2021). OVRAudio Developer Reference Manual. https://developer.oculus.com/reference/audio/v32/

Pang, A., & Wittenbrink, C. (1997). Collaborative 3D visualization with CSpray. IEEE Computer Graphics and Applications, 17(2), 32–41. https://doi.org/10.1109/38.574676

Perrott, D. R., Cisneros, J., Mckinley, R. L., & D’Angelo, W. R. (1996). Aurally aided visual search under virtual and free-field listening conditions. Human Factors, 38(4), 702–715. https://doi.org/10.1518/001872096778827260

Persson, M., & Wide, P. (2007). Sensor Source Intelligence Cell, 1–5.

Petrock, V. (2021, April). US Virtual and Augmented Reality Users 2021. https://www.emarketer.com/content/us-virtual-augmented-reality-users-2021

Pranckevičius, A. (2018, March). Releasing the Unity C# source code. https://blog.unity.com/technology/releasing-the-unity-c-source-code

Prange, A., Chikobava, M., Poller, P., Barz, M., & Sonntag, D. (2017). A multimodal dialogue system for medical decision support in virtual reality. SIGDIAL 2017 - 18th Annual Meeting of the Special Interest Group on Discourse and Dialogue, Proceedings of the Conference, (August), 23–26. https://doi.org/10.18653/v1/w17-5504

Pryor, T. A. (1994). Development of Decision Support Systems. MIS Quarterly, 4(4), 61–73. https://doi.org/10.1007/978-1-4612-2698-7_4

References and Bibliography

Radecki, A., Bujacz, M., Skulimowski, P., & Strumiłło, P. (2020). Interactive sonification of images in serious games as an education aid for visually impaired children. British Journal of Educational Technology, 51(2), 473–497. https://doi.org/10.1111/bjet.12852

Rajguru, C., Obrist, M., & Memoli, G. (2020). Spatial Soundscapes and Virtual Worlds: Challenges and Opportunities. Frontiers in Psychology, 11(November). https://doi.org/10.3389/fpsyg.2020.569056

Ramsey, K. (2009). GIS, modeling, and politics: On the tensions of collaborative decision support. Journal of Environmental Management, 90(6), 1972–1980. https://doi.org/10.1016/j.jenvman.2007.08.029

Ravyse, W. S., Seugnet Blignaut, A., Leendertz, V., & Woolner, A. (2017). Success factors for serious games to enhance learning: a systematic review. Virtual Reality, 21(1), 31–58. https://doi.org/10.1007/s10055-016-0298-4

Rebenitsch, L., & Owen, C. (2016). Review on cybersickness in applications and visual displays. Virtual Reality, 20(2), 101–125. https://doi.org/10.1007/s10055-016-0285-9

Ricciardi, F., & De Paolis, L. T. (2014). A Comprehensive Review of Serious Games in Health Professions. International Journal of Computer Games Technology, 2014. https://doi.org/10.1155/2014/787968

Rizzo, A., Parsons, T. D., Lange, B., Kenny, P., Buckwalter, J. G., Rothbaum, B., … Reger, G. (2011). Virtual reality goes to war: A brief review of the future of military behavioral healthcare. Journal of Clinical Psychology in Medical Settings, 18(2), 176–187. https://doi.org/10.1007/s10880-011-9247-2

Robinson, G. M., Hardman, M., & Matley, R. J. (2021). Using games in geographical and planning-related teaching: Serious games, edutainment, board games and role-play. Social Sciences & Humanities Open, 4(1), 100208. https://doi.org/10.1016/j.ssaho.2021.100208

References and Bibliography

Rojas, D., Kapralos, B., Hogue, A., Collins, K., Nacke, L., Cristancho, S., … Dubrowski, A. (2013). The effect of sound on visual fidelity perception in stereoscopic 3-D. IEEE Transactions on Cybernetics, 43(6), 1572–1583. https://doi.org/10.1109/TCYB.2013.2269712

Rovithis, E., Floros, A., Moustakas, N., Vogklis, K., & Kotsira, L. (2019). Bridging audio and augmented reality towards a new generation of serious audio-only games. Electronic Journal of E-Learning, 17(2), 144–156. https://doi.org/10.34190/JEL.17.2.07

Rubio-Tamayo, J. L., Barrio, M. G., & García, F. G. (2017). Immersive environments and virtual reality: Systematic review and advances in communication, interaction and simulation. Multimodal Technologies and Interaction, 1(4), 1–20. https://doi.org/10.3390/mti1040021

Rumiński, D. (2015). An experimental study of spatial sound usefulness in searching and navigating through AR environments. Virtual Reality, 19(3–4), 223–233. https://doi.org/10.1007/s10055-015-0274-4

Rumsey, F. (2012). Spatial Audio. Spatial Audio, 1–240. https://doi.org/10.4324/9780080498195

Russell, M. E. B., Hoffman, B., Stromberg, S., & Carlson, C. R. (2014). Use of Controlled Diaphragmatic Breathing for the Management of Motion Sickness in a Virtual Reality Environment. Applied Psychophysiology Biofeedback, 39(3–4), 269–277. https://doi.org/10.1007/s10484-014-9265-6

Sahai, A., Wefers, F., Pick, S., Stumpf, E., Vorländer, M., & Kuhlen, T. (2016). Interactive simulation of aircraft noise in aural and visual virtual environments. Applied Acoustics, 101, 24–38. https://doi.org/10.1016/j.apacoust.2015.08.002

Salinas, E.-L. (2005). Effects of communication mode on social presence, virtual presence, and performance in collaborative virtual environments. Presence: Teleoperators and Virtual Environments, 14(4), 434–449.

Scheucher, B., Bailey, P. H., Gütl, C., & Harward, J. V. (2009). Collaborative Virtual 3D Environment for Internet-Accessible Physics Experiments. International Journal of Online and Biomedical Engineering (IJOE), 5(5), 65. https://doi.org/10.3991/ijoe.v5s1.1014

Schl, M., Modelling, E., Abstract, G., Gis-based, T., Tugai, T., & Simulation, I. (2001). Application of a GIS-based Simulation Tool to Analyze and Communicate Uncertainties in Future Water Availability in the Amudarya River Delta. Engineering.

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

Schuemie, M. J., Van der Straaten, P., Krijn, M., & Van der Mast, C. A. P. G. (2001). Research on presence in virtual reality: A survey. Cyberpsychology and Behavior, 4(2), 183–201. https://doi.org/10.1089/109493101300117884

Serafin, S., Geronazzo, M., Erkut, C., Nilsson, N. C., & Nordahl, R. (2018). Sonic Interactions in Virtual Reality: State of the Art, Current Challenges, and Future Directions. IEEE Computer Graphics and Applications, 38(2), 31–43. https://doi.org/10.1109/MCG.2018.193142628

Seymour, N. E. (2008). VR to OR: A review of the evidence that virtual reality simulation improves operating room performance. World Journal of Surgery, 32(2), 182–188. https://doi.org/10.1007/s00268-007-9307-9

Sharma, S., Jerripothula, S., Mackey, S., & Soumare, O. (2014). Immersive virtual reality environment of a subway evacuation on a cloud for disaster preparedness and response training. IEEE SSCI 2014 - 2014 IEEE Symposium Series on Computational Intelligence - CIHLI 2014: 2014 IEEE Symposium on Computational Intelligence for Human-Like Intelligence, Proceedings. https://doi.org/10.1109/CIHLI.2014.7013380

Sheridan, T. B. (n.d.). Musing On Telepresence, 120–126.

Shu, Y., Huang, Y. Z., Chang, S. H., & Chen, M. Y. (2019). Do virtual reality head-mounted displays make a difference? A comparison of presence and self-efficacy between head-mounted

displays and desktop computer-facilitated virtual environments. Virtual Reality, 23(4), 437–446. https://doi.org/10.1007/s10055-018-0376-x

Simmons, A., Hatfield, G., Snowdon, P., & Siewert, C. (2016). The Oxford handbook of the philosophy of perception. Choice Reviews Online, 53(06), 53-2597-53–2597. https://doi.org/10.5860/choice.194275

Snowdon, D., Churchill, E. F., & Munro, A. J. (2001). Collaborative Virtual Environments: Digital Spaces and Places for CSCW: An Introduction. https://doi.org/10.1007/978-1-4471-0685-2_1

Softline. (1982, January). Softline. https://archive.org/details/Softline_1982_01

Steamcharts. (2020). Counter-Strike: Global Offensive. http://steamcharts.com/app/730

Tanaka, N., & Takagi, H. (2004). Virtual reality environment design of managing both presence and virtual reality sickness. Journal of Physiological Anthropology and Applied Human Science, 23(6), 313–317. https://doi.org/10.2114/jpa.23.313

Unity. (2017, June). See how Cinemachine can revolutionize your cameras. https://www.youtube.com/watch?v=r1SkOoJJRAA

Unity. (2021). FAQ: Licensing & activation. https://unity3d.com/unity/faq/2491

Unity. (2021). Audio Spatialiser SDK. https://docs.unity3d.com/Manual/AudioSpatializerSDK.html

Unity. (2021). Getting started with VR development in Unity. https://docs.unity3d.com/Manual/VROverview.html

Unreal (2015). If You Love Something, Set It Free. https://www.unrealengine.com/en-US/blog/ue4-is-free

References and Bibliography

Valladares-Rodríguez, S., Pérez-Rodríguez, R., Anido-Rifón, L., & Fernández-Iglesias, M. (2016). Trends on the application of serious games to neuropsychological evaluation: A scoping review. Journal of Biomedical Informatics, 64, 296–319. https://doi.org/10.1016/j.jbi.2016.10.019

Valve. (2016, December, 7). Holiday Season never sounded so good. https://blog.counter-strike.net/index.php/2016/12/17260/

Valve. (2021). A BENCHMARK IN IMMERSIVE AUDIO SOLUTIONS FOR GAMES AND VR. https://valvesoftware.github.io/steam-audio/

Vinnikov, M., Allison, R. S., & Fernandes, S. (2017). Gaze-contingent auditory displays for improved spatial attention in virtual reality. ACM Transactions on Computer-Human Interaction, 24(3). https://doi.org/10.1145/3067822

Vive. (2021). Vive Pro Full Kit. https://www.vive.com/eu/product/vive-pro-full-kit/

Walker, B. N., & Lindsay, J. (2006). Navigation performance with a virtual auditory display: Effects of beacon sound, capture radius, and practice. Human Factors, 48(2), 265–278. https://doi.org/10.1518/001872006777724507

Wang, R., DeMaria, S., Goldberg, A., & Katz, D. (2016). A systematic review of serious games in training: Health care professionals. Simulation in Healthcare, 11(1), 41–51. https://doi.org/10.1097/SIH.0000000000000118

Wang, S., Mao, Z., Zeng, C., Gong, H., Li, S., & Chen, B. (2010). A new method of virtual reality based on unity3D. 2010 18th International Conference on Geoinformatics, Geoinformatics 2010, 1–5. https://doi.org/10.1109/GEOINFORMATICS.2010.5567608

Wenzel, E. M., Stone, P. K., Fisher, S. S., & Foster, S. H. (1990). A system for three-dimensional acoustic “visualization” in a virtual environment workstation (pp. 329–337). https://doi.org/10.1109/visual.1990.146398

References and Bibliography

Wessels, A., Clarke, C. (2020). Unity Technologies and Oculus Expand Strategic Partnership. https://unity.com/our-company/newsroom/unity-technologies-and-oculus-expand-strategic-partnership

Westerberg, A., & Schoenau-Fog, H. (2015). Categorizing video game audio - An exploration of auditory-psychological effects. In ACADEMICMINDTREK 2015 - Proceedings of the 19th International Academic Mindtrek Conference (pp. 47–54). https://doi.org/10.1145/2818187.2818295

wickedplayer494. (2016, December, 6). Counter-Strike: Global Offensive update for 12/7/16 (12/8/16 UTC, 1.35.6.0). https://www.reddit.com/r/GlobalOffensive/comments/5h3yld/counterstrike_global_offensive_update_for_12716/

William R. Sherman and Alan B. Craig. 2002. Understanding Virtual Reality: Interface, Application, and Design. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

Wu, R., & Yu, G. (2016). IMPROVEMENTS IN HRTF DATASET OF 3D GAME AUDIO APPLICATION Acoustic Lab , School of Physics and Optoelectronics , South China University of Technology , 185–190.

Wu, W., Morina, R., Schenker, A., Gotsis, A., Chivukula, H., Gardner, M., … Heller, L. M. (2017). EchoExplorer(TM): A Game App for Understanding Echolocation and Learning to Navigate Using Echo Cues, (Icad), 237–240. https://doi.org/10.21785/icad2017.040

Wyeld, T. G., Carroll, J., Ledwich, B., Leavy, B., Gibbons, C., & Hills, J. (2007). The ethics of Indigenous storytelling: Using the Torque Game Engine to support Australian aboriginal cultural heritage. 3rd Digital Games Research Association International Conference: “Situated Play”, DiGRA 2007, 261–268.

Yamin, M. M., Katt, B., & Nowostawski, M. (2021). Serious games as a tool to model attack and defense scenarios for cyber-security exercises. Computers & Security, 110, 102450. https://doi.org/10.1016/j.cose.2021.102450

References and Bibliography

Yung, R., & Khoo-Lattimore, C. (2019). New realities: a systematic literature review on virtual reality and augmented reality in tourism research. Current Issues in Tourism, 22(17), 2056–2081. https://doi.org/10.1080/13683500.2017.1417359

Zhang, X., Luo, J., Dong, W., HU, X., & Gao, L. (2013). Exploring and visualizing three-dimensional ocean data in the China Digital Ocean Prototype System. Journal of Coastal Research, 65(65), 1081–1085. https://doi.org/10.2112/si65-183.1

zipzapzooom. (2017, January, 18). Teams at majors using HRTF?. https://www.reddit.com/r/GlobalOffensive/comments/5ooyn0/teams_at_major_using_hrtf/