



---

# When Worlds Collide

The Effects of Augmented Virtuality on Presence, Workload, and Input Performance

---

*Author:*  
Jacob H CLARKSON

*Supervisor:*  
Prof. Edwin H BLAKE

*A thesis submitted in fulfillment of the requirements  
for the degree of Master of Science*

*in the*

Department of Computer Science

December 8, 2017

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF.

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.



University Of Cape Town

## *Abstract*

Faculty of Science  
Department of Computer Science

Master of Science

### **When Worlds Collide**

by Jacob H CLARKSON

Head-Mounted Displays (HMDs) offer, more than any easily accessible technology that has come before, the sensation of presence – that feeling that you are “really there” in a virtual world. However, HMDs cut the wearer off from the real world, making even trivial interactions, such as having a drink or typing, difficult and frustrating. In the home context where these devices are most likely to be used, such interactions are commonplace, and in order to execute them, users have to remove the HMD (“peep”), breaking their sense of presence. How, then, can real-world interactions during HMD usage be facilitated such that presence is damaged as little as possible?

Previous work indicates that Augmented Virtuality (AV), a technique that allows the wearer of an HMD to see through it when they need to, is a promising answer to this question. However, direct comparisons between AV and VR that thoroughly account for presence and workload are lacking. To corroborate previous findings, and to address some of the gaps in the current literature, we conducted a quantitative user experiment to compare our own implementation of AV to VR in terms of presence, workload, and typing performance. The experiment followed a between-groups design with participants selected via pseudo-random convenience sampling of university students.

To simulate the context of home usage – an extended immersive session that must occasionally be interrupted – we designed a mixed reality game that periodically required the player to interact with real-world objects before they could proceed. Participants in the experimental group played the game using our AV system to assist them in completing the required real-world tasks. Participants in the control group used pure VR to play the game and had to peep. This allowed us to directly compare AV to VR in terms of the levels of presence and workload experienced. These data were gathered using post-hoc self-report questionnaires. To measure and compare typing performance under various conditions, we created desktop, VR, and AV versions of a typing test that participants had to complete.

We found that typing performance in AV was significantly better than in VR, but did not reach the levels achieved in baseline desktop conditions. While there was not a significant difference in the overall level of workload associated with using AV compared to VR, participants in the AV condition were able to interact successfully with the real world without having to remove the HMD, and reported being significantly less frustrated than those in the VR condition. Finally, AV users reported significantly higher levels of presence than those who used VR.



## *Acknowledgements*

I would first like to thank my supervisor, Professor Edwin Blake, for inspiring me to begin this research and for providing me with the resources and guidance that made it possible. Thank you for the time and energy you have put into this work, but also for the welcome digressions from it, which were opportunities to learn about Fourier transforms, yoga, and many things in-between. I would also like to thank Bryan Davies, my friend and colleague, for his moral support and for always being interested. Thanks too to the members of VExInG for testing prototypes and providing valuable feedback at various stages of this project. Finally, Louise: thank you for everything.



# Contents

<b>Abstract</b>	<b>iii</b>
<b>Acknowledgements</b>	<b>v</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Augmented Virtuality . . . . .	2
1.2 Research Aims . . . . .	3
1.3 Summary and Outline . . . . .	5
<b>2 Background</b>	<b>7</b>
2.1 Virtual Reality . . . . .	7
2.2 Presence - Definitions and Theory . . . . .	11
2.2.1 What We Mean by Presence . . . . .	11
Is technology involved? . . . . .	11
What is the phenomenon a property of? . . . . .	12
What is the source of the stimuli? . . . . .	12
How is technology perceived? . . . . .	12
What aspect of the phenomenon is of interest? . . . . .	12
Virtual Spatial Telepresence . . . . .	12
2.2.2 Theories of Presence Formation . . . . .	13
Formation of the Spatial Situation Model . . . . .	14
From SSM to presence . . . . .	15
Summary of presence formation . . . . .	16
2.2.3 Factors Influencing Presence . . . . .	16
2.2.4 The Importance of Presence . . . . .	17
2.3 A Problem with HMDs . . . . .	18
2.3.1 Input Performance . . . . .	18
2.3.2 Real-World Interaction . . . . .	19
2.3.3 Diminished Sense of Presence . . . . .	20
2.4 Augmented Virtuality . . . . .	21
2.4.1 Design Approaches . . . . .	21
Blending . . . . .	22
Transitions . . . . .	22
2.4.2 Implementation Approaches . . . . .	23
2.4.3 Measurement Approaches . . . . .	24
Presence . . . . .	24
Workload . . . . .	24
Input performance . . . . .	25
2.5 Previous Findings Regarding User Experience in AV . . . . .	27
2.6 Conclusion . . . . .	28

<b>3</b>	<b>Methods</b>	<b>31</b>
3.1	Apparatus . . . . .	32
3.1.1	The Virtual Environment . . . . .	32
3.1.2	The Game . . . . .	33
	Tutorial, controls, and user interface . . . . .	35
	Beginning the game . . . . .	37
	Puzzle one - temple of the horse . . . . .	37
	Puzzle two - temple of the lion . . . . .	39
	Puzzle three - temple of the dragon . . . . .	40
3.1.3	AV System . . . . .	40
	Hardware . . . . .	40
	AV software . . . . .	43
3.1.4	Input Capture . . . . .	45
	Baseline typing test . . . . .	47
	VR typing test . . . . .	47
	AV typing test . . . . .	48
3.2	Experimental Design . . . . .	48
3.3	Measures . . . . .	49
3.3.1	Presence . . . . .	49
3.3.2	Workload . . . . .	49
3.3.3	Input Performance . . . . .	50
3.4	Participants . . . . .	50
3.5	Procedure . . . . .	51
3.5.1	Pre-experimental Phase . . . . .	51
3.5.2	Experimental Phase . . . . .	52
	Game play and subjective measures . . . . .	52
	Typing test . . . . .	52
3.5.3	Post-experimental Phase . . . . .	53
3.5.4	Processing Phase . . . . .	53
3.6	Summary . . . . .	53
<b>4</b>	<b>Results</b>	<b>55</b>
4.1	Presence . . . . .	56
4.2	Workload . . . . .	58
4.3	Typing . . . . .	62
4.3.1	Baseline . . . . .	63
4.3.2	Within Subjects . . . . .	63
4.3.3	Between Groups . . . . .	65
4.4	Summary . . . . .	68
<b>5</b>	<b>Discussion</b>	<b>69</b>
5.1	Presence . . . . .	69
5.1.1	General Presence and Spatial Presence . . . . .	70
5.1.2	Involvement . . . . .	71
5.1.3	Realism . . . . .	71
5.2	Workload . . . . .	71
5.2.1	Mental and Physical Demand . . . . .	72
5.2.2	Temporal Demand and Effort . . . . .	72
5.2.3	Performance . . . . .	73
5.2.4	Frustration . . . . .	73
5.3	Typing . . . . .	73

5.3.1	Baseline Typing Performance . . . . .	74
5.3.2	Within-subjects Typing Performance . . . . .	74
5.3.3	Between Groups Typing Performance . . . . .	75
5.4	Summary and Limitations . . . . .	76
5.4.1	Presence . . . . .	76
5.4.2	Workload . . . . .	76
5.4.3	Typing . . . . .	77
<b>6</b>	<b>Conclusions</b>	<b>79</b>
6.1	Presence . . . . .	80
6.1.1	General and Spatial Presence . . . . .	80
6.1.2	Involvement . . . . .	82
6.1.3	Summary . . . . .	83
6.2	Workload . . . . .	84
6.2.1	Summary . . . . .	84
6.3	Typing Performance . . . . .	85
6.4	Future Work . . . . .	86
6.5	Conclusion . . . . .	87
<b>A</b>	<b>Assets Used</b>	<b>89</b>
<b>B</b>	<b>List of Hardware and Specifications</b>	<b>93</b>
B.1	Hardware Specifications . . . . .	93
B.2	Software Specifications . . . . .	93
<b>C</b>	<b>Igroup Presence Questionnaire - Participant Version</b>	<b>95</b>
<b>D</b>	<b>NASA Task Load Index</b>	<b>99</b>
<b>E</b>	<b>Ethics Clearance</b>	<b>103</b>
<b>F</b>	<b>Informed Consent Document</b>	<b>111</b>
<b>G</b>	<b>Tutorial Script</b>	<b>113</b>
<b>H</b>	<b>Statistical Analysis of Touch Typist Performance</b>	<b>115</b>
<b>I</b>	<b>IGroup Presence Questionnaire Items</b>	<b>117</b>
<b>J</b>	<b>Minimum String Distance Statistic Algorithm</b>	<b>119</b>
	<b>Bibliography</b>	<b>121</b>



# List of Figures

2.1	Perspective Painting . . . . .	7
2.2	Panoramic Painting . . . . .	8
2.3	Early Stereoscopes . . . . .	8
2.4	The Sensorama . . . . .	9
2.5	Head-Mounted Displays . . . . .	10
2.6	The Rabbit-Duck Illusion . . . . .	13
2.7	The Reality/Virtuality Continuum . . . . .	21
2.8	Graphical Representation of Different Types of AV . . . . .	22
3.1	Overview of Virtual Environment . . . . .	34
3.2	Tutorial Level . . . . .	36
3.3	Reticule . . . . .	36
3.4	Message Window . . . . .	37
3.5	Temple of the Horse . . . . .	38
3.6	First Puzzle . . . . .	38
3.7	First Clue . . . . .	39
3.8	Temple of the Lion . . . . .	39
3.9	Second Puzzle . . . . .	40
3.10	Second Clue . . . . .	40
3.11	Temple of the Dragon . . . . .	41
3.12	Third Puzzle . . . . .	41
3.13	Third Clue . . . . .	41
3.14	Hardware for Hand Tracking and Video Pass-through . . . . .	42
3.15	AV System Hardware Setup - Front View . . . . .	43
3.16	AV System Hardware Setup - Side View . . . . .	44
3.17	XBox One Controller . . . . .	44
3.18	Stages of AV Implementation . . . . .	44
3.19	AV System in Use . . . . .	46
3.20	Baseline Typing Test . . . . .	47
3.21	VR Typing Test . . . . .	47
3.22	AV Typing Test . . . . .	48
3.23	Clue Locations . . . . .	52
4.1	Outliers in Presence Data . . . . .	57
4.2	Average Presence Scores . . . . .	58
4.3	Mean Workload Scores . . . . .	59
4.4	Box Plot of Workload Factors . . . . .	60
4.5	Mean Baseline Typing Performance . . . . .	64
4.6	Treatment Typing Performance . . . . .	67
6.1	Comparison Between VR Peeping and AV . . . . .	81
J.1	Minimum String Distance Statistic Algorithm . . . . .	119



# List of Tables

4.1	Interpretation of Effect Sizes . . . . .	56
4.2	Descriptive Statistics for IPQ . . . . .	57
4.3	Mann-Whitney U Test for Non-Normally Distributed Presence Data . .	59
4.4	T-Tests for Normally Distributed Presence Data . . . . .	59
4.5	Descriptive Statistics for Responses to NASA-TLX . . . . .	61
4.6	T-Tests for Normally Distributed Workload Data . . . . .	61
4.7	Mann-Whitney U Tests for Non-Normally Distributed Workload Data	62
4.8	Descriptive Statistics for Baseline Typing Data . . . . .	63
4.9	T-Tests for Baseline Typing Data . . . . .	63
4.10	Paired-samples t-tests for Within-Subjects Typing Data (Control Group)	65
4.11	T-Tests for Within-Subjects Typing Data (Experimental Group) . . . . .	65
4.12	Descriptive Statistics for Treatment Typing Data . . . . .	66
4.13	Mann-Whitney U Tests for Non-Normally Distributed Typing Data . .	66
4.14	T-Tests for Normally Distributed Typing Data . . . . .	66
5.1	Comparison of presence factor scores to IPQ database . . . . .	69
A.1	Visual Assets Used . . . . .	90
A.2	Audio Assets Used . . . . .	91
H.1	Descriptive Statistics for Touch Typists . . . . .	115
H.2	T-tests Comparing Touch Typists . . . . .	115
I.1	Items of the IGroup Presence Questionnaire Divided by Factor . . . . .	117



# List of Abbreviations

<b>AR</b>	<b>Augmented Reality</b>
<b>AV</b>	<b>Augmented Virtuality</b>
<b>E</b>	<b>Effort</b>
<b>ERF</b>	<b>Egocentric Reference Frame</b>
<b>F</b>	<b>Frustration</b>
<b>GTS</b>	<b>Gross Typing Speed</b>
<b>HMD</b>	<b>Head Mounted Display</b>
<b>INV</b>	<b>Involvement</b>
<b>IPQ</b>	<b>IGroup Presence Questionnaire</b>
<b>KSPC</b>	<b>Key Strokes Per Character</b>
<b>MD</b>	<b>Mental Demand</b>
<b>MR</b>	<b>Mixed Reality</b>
<b>MSD</b>	<b>Minumum String Distance</b>
<b>NASA-TLX</b>	<b>NASA Task Load Index</b>
<b>NTS</b>	<b>Net Typing Speed</b>
<b>P</b>	<b>Performance</b>
<b>PD</b>	<b>Physical Demand</b>
<b>PERF</b>	<b>Primary Egocentric Reference Frame</b>
<b>PRES</b>	<b>General Presence</b>
<b>REAL</b>	<b>Realism</b>
<b>SP</b>	<b>Spatial Presence</b>
<b>SSM</b>	<b>Spatial Situation Model</b>
<b>TD</b>	<b>Temporal Demand</b>
<b>TPH</b>	<b>Theory of Perceptual Hypotheses</b>
<b>TTFKP</b>	<b>Time To First Key Press</b>
<b>UI</b>	<b>User Interface</b>
<b>VE</b>	<b>Virtual Environment</b>
<b>VOIP</b>	<b>Voice Over IP</b>
<b>VR</b>	<b>Virtual Reality</b>



*To my parents,  
for everything they have done for me.*



## Chapter 1

# Introduction

In 2016, consumer-level Virtual Reality (VR) devices arrived in the form of Head-Mounted Displays (HMDs) such as the Oculus Rift and HTC Vive. VR has seen a massive resurgence in recent years, and by 2020 the industry is expected to be worth approximately \$30 billion (Digicapital, 2016). Although projections like this may seem overly optimistic, technology has come a long way since the 90s' VR bubble burst. The problems previously faced by consumer-grade VR experiences (such as low resolution, small field of view, uncomfortable weight, motion sickness-inducing latency, and poor graphical quality (Hutchison, 2007)) have largely been addressed (McGill et al., 2015).

Given the expected ubiquity of HMDs, it is important to consider how and where they will be used, and who will be using them. Since personal entertainment (video games in particular) will be a primary function of consumer-level VR, it is of particular importance to consider the context of home usage, where these media are most likely to be consumed (McGill et al., 2015). Despite the fact that video games are meant to place players in a virtual world to the deliberate exclusion of the real world, a considerable number of real-world interactions still take place during play (Foehr, 2006). Examples of these interactions include the use of peripherals such as a keyboard, or taking a drink. A typical scenario is therefore that a user engages in an extended immersive session that is occasionally interrupted when they want or need to do something in the real world. The complete visual exclusion of the real world by HMDs, while making them more immersive, makes these natural and frequently executed real-world tasks far more difficult.

While these difficulties alone are worthy of note, the way that users are currently trying to mitigate them creates a more interesting problem. To be able to interact with the real world effectively, users are resorting to "peeping" (temporarily lifting the HMD off to see reality). This defeats the very purpose of the HMD, which is to provide a sense of presence – that "sense of being there" in a virtual world. Presence is a vital aspect of the experience of playing video games, which, as we mentioned, is one of the most common use-cases for consumer HMDs. The main reason that we play games is because we enjoy playing them; enjoyment has been conclusively shown to determine intention to play games and use other entertainment media (Wu, Wang, and Tsai, 2010). Not only has presence been shown to enhance the enjoyment of games (Tamborini and Skalski, 2006), but also to be a component of that enjoyment in the first place (Weibel and Wissmath, 2011). Tamborini and Skalski (2006) predicted that with ever improving technology, games would become "the *ultimate* presence-inducing medium". The immersive VR games that are now accessible thanks to HMDs demonstrate the accuracy of that prediction. Presence is central to games and games research, and games, as a means of inducing presence, are also central to presence research.

Presence is not only important for games, however, having also been found to improve the usefulness of serious VR applications such as training simulations and remote surgery (Regenbrecht, Schubert, and Friedmann, 1998), and to be a vital aspect of the effectiveness of virtual environments in general (Witmer and Singer, 1998; Slater and Wilbur, 1997). Presence is the primary goal of all VR (Steuer, 1992). It is therefore important for the developers of VR systems and software to maximise the amount of presence that users can experience, and to make sure that this presence can be maintained. Issues that lead to breaks in presence, or otherwise diminish it, such as common real-world interactions during HMD usage, should therefore be handled in such a way that their negative effects on presence are mitigated, allowing VR applications to reach their maximum potential.

In short, wearing an HMD makes it difficult to provide input, or to interact with real world objects other than input devices, and can therefore lead to diminished user experience, particularly with respect to presence. The question, then, is how these real-world interactions can be facilitated such that presence is diminished as little as possible. Our research sought to answer this question by testing the effects of a Mixed Reality technique known as Augmented Virtuality on presence, real-world interaction, and input performance.

## 1.1 Augmented Virtuality

Augmented Virtuality (AV) is a subset of Mixed Reality techniques that blend the real world and a virtual world. Whereas the more widely known Augmented Reality (AR) focuses on the addition of virtual artefacts to the otherwise mostly real world, AV is concerned with adding a subset of the real world to an otherwise virtual one. AV, in effect, allows a user to “see through” their HMD into the real world, allowing one to interact with the real world without having to remove the HMD.

Some previous work has sought to solve the problem of visual cut-off with respect to specific issues. Some researchers, for example, have focused on facilitating text-entry while wearing an HMD (Walker et al., 2017; Walker, 2017; Lee and Kim, 2017; Lin et al., 2017). Others have come up with solutions that make it easy to use a mobile phone while remaining immersed in VR (Desai, Pena-Castillo, and Meruvia-Pastor, 2017), and some have attempted to break down the social barriers between the HMD user and the non-immersed other (Chan and Minamizawa, 2017). However, a great strength of AV is that it can be a general solution to all of the problems caused by visual cut-off. If implemented well enough, AV would allow people to do all of the things they might usually want to, such as type, check their phone, take a note, have a drink, or interact with another person, without removing the HMD. Limiting the view of reality to only a subset necessary to facilitate such tasks means that users could execute them while still receiving stimuli from the virtual world, allowing them to interact in the real world without completely leaving the virtual one. An existing theory of presence formation allows us to explain how this might lead to higher levels of presence.

Wirth et al. (2007) propose that presence formation is a two-step process (see Section 2.2.2). During the first step, VR users form a mental representation of the virtual world based on spatial cues. During the second step, this mental representation becomes the user’s primary point of reference for where they really are, through unconscious hypothesis testing. Not only does this theory explain how presence may form and how it may be broken, but it can also inform our understanding of the effects of AV on presence, which is our principle interest. Because the AV user is

receiving fewer stimuli from the real world (that conflict with their hypothesis that the virtual world is where they really are) than if they removed the HMD, it is possible that AV allows real-world interactions to be executed without presence being broken at all. Alternatively, because AV can allow the user to continue receiving stimuli from the virtual world, even while interacting with the real world, the user's mental representation of the virtual world can be maintained. Even if the user is not actually present in the virtual world during real-world interaction, AV may therefore facilitate a recovery of presence, because they do not have to begin the presence formation process from the beginning when they return to the virtual world. This would lead to a higher overall level of presence during an extended session that is occasionally interrupted.

Previous work on AV has explored its ability to facilitate real-world interactions and mitigate the damage to presence that they can cause (McGill et al., 2015; Budhiraja et al., 2015). It has been found that AV implementations that keep the amount of reality shown to a minimum are preferred by users to "peeping" or blindly interacting without removing the HMD, reduce the workload of interacting with the real world and providing input, help with text-entry, and preserve the sense of presence that users have in the virtual world.

We believe that AV is the most promising solution to the problems caused by visual cut-off. However, research in this area is in its infancy. Direct comparisons between AV and VR in terms of presence and workload are lacking or absent, with previous work having focused primarily on comparing various AV designs to each other. There has also been no attempt to explain *why* AV may preserve presence during real-world interactions. Furthermore, the effects of AV implementations that show as little of reality as possible on input performance have not been explored. Finally, though the importance of the home or office context has been emphasised, previous work has relied on impractical implementations.

We therefore decided to implement and test our own version of an AV system to address these deficiencies.

## 1.2 Research Aims

Our overall research goal was to compare AV to pure VR. In order to do this, various aspects of user experience and performance under AV and VR conditions would need to be measured.

The importance of presence for VR technologies is widely acknowledged. How presence is affected by any manipulations to VR is therefore an important consideration for those manipulations. Specifically, we were interested in the overall level of presence that users experience during an extended period in VR.

Perhaps the most obvious issue caused by visual cut-off is the hindrance of otherwise trivial real-world interactions. A successful AV system should facilitate such interactions, or in other words, reduce the subjective workload that users experience while executing them.

Finally, not only does visual cut-off make it difficult to interact with objects that are not related to the VR experience, but with objects that are related, such as input devices. Of such devices, the standard keyboard is surely the most ubiquitous, as is its primary function – text-entry. The use of a keyboard is also strongly affected by visual cut-off (more so than simpler input devices such as controllers), and text-entry performance can be measured thoroughly.

The aim of this research was therefore to develop our own implementation of an AV system, and test its effects on presence and workload when real-world interactions were required, including its impact on typing performance, compared to VR baselines. This led to the development of the following research questions and corresponding hypotheses:

1. Is the overall level of presence experienced over an extended session in an immersive virtual setting higher in Augmented Virtuality than in Virtual Reality when the session has to be interrupted by interactions with the real world?

$H_{A0}$  There is no difference between presence scores in Augmented Virtuality and in Virtual Reality.

$H_{A1}$  Presence scores are higher in Augmented Virtuality than in Virtual Reality.

$H_{A2}$  Presence scores are lower in Augmented Virtuality than in Virtual Reality.

2. Is it easier to interact with real-world objects in Augmented Virtuality than in Virtual Reality when users have to remove the HMD?

$H_{B0}$  Workload for performing real-world interactions is the same in Augmented Virtuality as it is in Virtual Reality.

$H_{B1}$  Workload for performing real-world interactions is lower in Augmented Virtuality than it is in Virtual Reality.

$H_{B2}$  Workload for performing real-world interactions is higher in Augmented Virtuality than it is in Virtual Reality.

3. Is typing performance better in Augmented Virtuality than in Virtual Reality?

$H_{C0}$  There is no difference in typing performance between Augmented Virtuality and a Virtual Reality baseline.

$H_{C1}$  Typing performance is better in Augmented Virtuality than in a Virtual Reality baseline.

$H_{C2}$  Typing performance is worse in Augmented Virtuality than in a Virtual Reality baseline.

In order to answer these research questions, we conducted a user experiment that required the development of several pieces of apparatus. In order to measure the effects of AV on presence, we needed participants to be able to feel present in a virtual environment. To this end, we created a virtual world capable of eliciting a sense of presence. To give participants a reason to explore this virtual world, and a reason to interact with various real-world objects, we developed a Mixed Reality game that would take place in this virtual world and periodically require players to execute an action in the real world. This strategy also allowed us to closely simulate the way HMDs are used in the home context, with users taking part in an extended play session that is interrupted by occasional real-world interactions. We designed an implementation of AV that participants could use to interact with the real world while wearing an HMD. Our implementation did not require any modifications to the environment in which the system would be used. Finally, several pieces of software were created in order to measure typing performance under VR and AV conditions.

## 1.3 Summary and Outline

For the VR user, it is inevitable that the real world must impinge on the virtual one. Developers of VR systems and software, if the industry is to affect as many users as market projections suggest, should care about reducing the negative effects on user experience that these impingements can have. Actions as simple as having a sip of tea, and as ubiquitous as typing on a keyboard, are currently not well supported in VR. To effectively execute these actions, users have to remove the HMD, which defeats the primary purpose of the device: presence. AV is the most promising solution to these problems, allowing users to interact with the real world while preserving their sense of presence in the virtual world. We implemented our own version of an AV system and compared it to VR by evaluating its effectiveness in terms of presence, workload, and typing performance, aiming to support previous positive results and explore gaps in the current literature. The remainder of this dissertation presents the investigation of our research questions.

Chapter 2 starts off with a brief history of VR and establishes our teleological stance towards it. We go on to provide a detailed explication of presence and its theoretical underpinnings, before discussing the problem of visual cut-off caused by HMDs. Previous work regarding the design and implementation of AV, as well as experimental approaches and results in the area, are presented. Finally, pertinent measurement strategies are discussed.

Chapter 3 reintroduces our research objectives before going in to detail about the methods we used to achieve them. Our experimental apparatus – the virtual environment, game, AV system, and typing software we created – is thoroughly explained, as is the design of the experiment itself. The chapter concludes with information about our chosen measures and materials, participants, and experimental procedure.

In Chapter 4, the results of our experiment and all subsequent statistical analyses are presented. The chapter is divided into sections – presence, workload, and typing – that correspond with our research questions.

A discussion of these results can be found in Chapter 5. Once again, each section in this chapter corresponds to one of our research questions, and provides an interpretation of our results in relation to previous work. The overall results from each measure, as well as the contributions of individual sub-factors of each measure, are discussed in detail.

Finally, our research questions are answered in Chapter 6, where our conclusions and possible avenues for future work are presented.



## Chapter 2

# Background

In this chapter, we introduce Virtual Reality (VR) and some of the problems faced by its most recent and popular implementation: the consumer Head-Mounted Display (HMD). We are primarily concerned with the user experience that VR offers, and provide an explication of presence, a core component of this experience. We go on to discuss usability challenges faced by the HMD user and how these can negatively affect presence. Subsequently, we introduce Augmented Virtuality (AV) as a potential solution to these problems, critically examining several previous works in the area with respect to design, implementation, and experimentation. Finally, we discuss the measurement of presence, input performance, and workload – metrics necessary for the evaluation of a Mixed Reality system.

### 2.1 Virtual Reality

The term *virtual reality* only rose to popularity in the late 80s, but many of the qualities we associate with VR today were attracting attention long before this. In some ways, the lineage of VR can be traced all the way back to the fifteenth century, when the rules of linear perspective were first introduced to painting – adding, through spatial cues, the illusion of depth to a two-dimensional image (Figure 2.1).



FIGURE 2.1: Masolino da Panicale's *St. Peter Healing a Cripple and the Raising of Tabitha* (1423). One of the earliest examples of linear perspective painting, using monocular spatial cues to simulate a sense of depth. White lines show the vanishing point. From Tyler and Kubovy (2004).

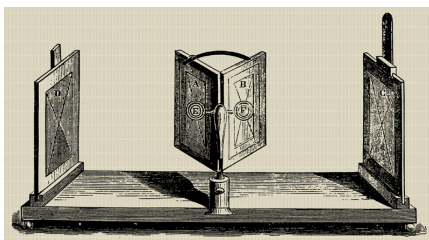
The painters of the Renaissance wanted you to feel as though you were actually looking at a scene, rather than just a painting of it. The eighteenth and nineteenth centuries saw several additional developments, including the exploration of increased field of view by way of dramatic 360-degree murals hung on the inside of a cylindrical surface and designed to be viewed from within. This idea is exemplified by the work of Franz Roubaud (Figure 2.2). Roubaud wanted you to feel like you

were *in* the scenes his paintings depicted, or as John Ruskin put it, to “put the spectator, so far as art can do, in the scene represented” (Clifton, Scatone, and Fetvaci, 2009).

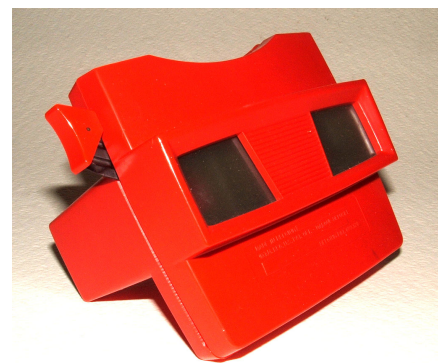


FIGURE 2.2: Franz Roubaud’s *Battle of Borodino* (1812) on display in a 360-degree viewing room built for the purpose. (*Battle of Borodino Panorama*).

Around the same time, Sir Charles Wheatstone created the first stereoscope (Figure 2.3a), demonstrating the importance of binocular depth perception and showing how it could be achieved by showing slightly offset views of the same image to each eye. The nineteenth century also saw the birth of photography, and by 1939, the stereoscope, detailed colour photographs, and lenses to unite the offset images were combined in the View-Master (Figure 2.3b). With its photorealism, the View-Master took the ideals of the Renaissance painters further, and allowed would-be travellers to transport themselves to the depicted locations.



(a) The original Wheatstone stereoscope. Mirrors in the centre allowed each eye to be shown a slightly offset version of the same image (on the far left and right), simulating stereoscopic 3D. (*Wheatstone Mirror Stereoscope*).



(b) The View-Master. Advances in colour film technology allowed a series of small images to be viewed in 3D using a Viewmaster. This is a later model (1962), but the design is nearly unchanged from the original. (*View-Master Model G*).

FIGURE 2.3: Early Stereoscopes.

In the meantime, motion pictures had taken off. Now not only could we trick the human brain into perceiving depth, but motion too. In the nineteen-fifties, Morton Heilig thought that to further draw people into the cinematic experience, more sensory modalities needed to be accommodated. And so the Sensorama was born, combining stereoscopic cinema, stereo sound, proprioceptive feedback, and even smells (Figure 2.4). Applying what had already been learnt about static images to movies, and stimulating more senses, Heilig hoped to put viewers *into* the films he created for the Sensorama. By the end of the nineteen-fifties, we had almost all of the ingredients we needed for VR, save one: the *virtual*. Ivan Sutherland would soon provide it.



FIGURE 2.4: The Sensorama. As the text on the device indicates, the user was treated to a 3D movie, stereo sound, vibrations, aromas, and wind. (*The Sensorama Machine*).

The Sword of Damocles (Sutherland, 1968) was, by all accounts, the first VR HMD (Figure 2.5a). For the first time, images of wholly computer-generated artefacts could be presented in stereoscopic 3D, and respond appropriately to movements of the user's head. This introduction of kinetic depth effects and motion parallax served to further convince the brain that it was perceiving in three dimensions. HMD technology was in its infancy, and while the rudimentary wire-frame rooms and objects that could be displayed fell far short of Sutherland's hypothetical "Ultimate Display" (Sutherland, 1965), it was nonetheless a triumph, with Sutherland optimistically concluding the accompanying publication by saying that users "uniformly remark on the realism of the resulting images". The rapid advancement of computing power in the following decades brought about numerous advancements to display and tracking technology, resulting in many attempts to put VR in the mainstream: VPL's EyePhone and Dataglove, various VR arcade systems, SEGA's VR glasses, and Nintendo's Virtual Boy. However, the experience that these devices could offer did not justify their cost; VR was not yet ready for commercial success. Continual technological advancement has changed that, however, with 2016 seeing the release of several consumer-oriented HMDs.



(a) Sutherland's Sword of Damocles was the first VR HMD. (Sutherland, 1968).



(b) The Oculus Rift: currently at the forefront of consumer VR technology. (Oculus Rift).

FIGURE 2.5: Head-Mounted Displays: 1968 - 2016.

Modern HMDs (Figure 2.5b) differ from 2D monitor-based desktop setups in several ways, with motion tracking and stereoscopic displays among the most obvious. It is therefore tempting to try to define VR based on technologies, such as these, that separate it from more traditional systems. In fact, many early definitions of VR were based purely on the technology that was involved (Steuer, 1992). However, the human observer has always been the common element. For hundreds of years, the creators of media have been exploring ways to heighten the way in which consumers experience it. To this end, we have discovered (in very roughly chronological order) the importance of monocular spatial cues, large fields of view, photorealistic images, binocular spatial cues, high framerates, multimodal stimuli, proprioceptive feedback, kinetic spatial cues, and various interaction methods. These are *immersive* technologies – objectively measurable system-centric capabilities that influence the extent to which a constructed environment can be delivered (Slater, 2003; Slater, 1999; Slater, Usoh, and Steed, 1994). The current generation of consumer HMDs are the culmination of all of these properties and technologies<sup>1</sup>. Although our discovery and use of these elements is vastly separated by both time and technology, they all serve a common purpose: to draw the consumer into the medium. What linear perspective did for painting, the View-Master did for photography, and the Sensorama did for film, HMDs attempt to do for virtual environments. The user is made to feel that they are actually *in* a virtual environment, rather than just looking at it.

What really separates VR from the traditional desktop setup, then, is not technological, but experiential (Slater and Wilbur, 1997). Implementations may differ, but the primary goal of all VR is to create in the user a heightened sense of *presence* (Steuer, 1992).

<sup>1</sup>With the perhaps merciful exception of olfactory stimulation.

## 2.2 Presence - Definitions and Theory

Many media create their own worlds – mediated environments – for the consumer to enjoy, explore, or work in. It is quite possible for the reader, player, watcher, or worker, to feel as though they are actually in the mediated environment, rather than their physical one. Loosely speaking, this subjective experience, “a sense of being there” in that mediated environment, is presence (Slater and Wilbur, 1997; Steuer, 1992; Witmer and Singer, 1998; IJsselsteijn et al., 2000; Weibel and Wissmath, 2011).

Presence is important for games, and the effectiveness of virtual environments has often been linked to the amount of presence experienced by users (Witmer and Singer, 1998; Regenbrecht, Schubert, and Friedmann, 1998). Existing media effects, such as the enjoyment of video games, are known to be intensified by presence (Tamborini and Skalski, 2006). To put it another way, the sense of presence in a virtual world is a core component of the pleasure of playing games (Weibel and Wissmath, 2011).

One of the primary experiential differences between VR and traditional computer setups is the level of presence that is possible. Presence is a main goal of VR, and given its importance in games and the fact that gamers are the primary target market for HMDs, it is important to take presence into account when designing HMD-related systems and software.

### 2.2.1 What We Mean by Presence

In 1980, Minsky (1980) introduced the idea of telepresence. His focus was on teleoperation – remote control of distant equipment using instruments with high quality sensory feedback. Teleoperation would allow people to work at remote locations as if they were really there. The biggest challenge, he thought, would be achieving “that sense of being there”. In the early 90s, this became known as presence, and it has been researched extensively since then, resulting in dozens of competing, overlapping, and even contradictory definitions and underlying theoretical frameworks being generated (Lombard and Jones, 2015). This lack of clarity has been noted by many authors (Slater, 2003; Lombard and Jones, 2015; Schubert, 2009; Wirth et al., 2007; Cummings and Bailenson, 2016; Van Baren and IJsselsteijn, 2004), leading to pleas for standardisation of terminology (Slater, 2003; Lombard and Jones, 2015) and, in some cases, abandonment of the term presence altogether (Slater, 2009). Lombard and Jones (2015) propose a definitional framework which aims to clarify and standardise discussions and definitions of presence. In accordance with this framework, we define presence as *Virtual Spatial Telepresence*.

We arrive at this definition through the answers to several key questions.

#### Is technology involved?

The term telepresence has largely been conflated with presence since the 90s. However, it is useful to maintain the distinction for definitional purposes. In so doing, we can distinguish between scenarios that involve technology (telepresence), and those that do not (presence) (Lombard and Jones, 2015).

Instances of presence that do not include technology include *corporeal copresence* (Zhao, 2003), which is focused on properties of communication where people are physically located in the same place, or broad conceptualisations that extend to the

sensation of presence we can experience regarding the real external world (Heeter, 1992)<sup>2</sup>.

Within our context, technology is most certainly involved, so what we are interested in is a form of *telepresence*.

### **What is the phenomenon a property of?**

Although there are some cases where presence or telepresence can be thought of as an objective quality of a person or technology<sup>3</sup>, it is far more often thought of as a subjective experience of an individual (Lombard and Jones, 2015). It is this subjective experience of stimuli in which we are interested.

### **What is the source of the stimuli?**

Stimuli being experienced can be either internal or external in origin (Lombard and Jones, 2015). Internal stimuli include phenomena such as hallucinations (Heeter, 1992). External stimuli are those that come from the physical world around us, including, as in our context, technological sources.

### **How is technology perceived?**

It is widely acknowledged that telepresence is a multidimensional construct, and although the prevailing theories have their differences, many have one idea in common: that telepresence is the perception of non-mediation (Lombard and Ditton, 1997; Schuemie et al., 2001; ISPR, 2017). That is to say, a person experiencing telepresence no longer perceives the medium that is conveying stimuli.

In our case the stimuli are being provided for us via technology (the HMD). *Virtual Telepresence* (Lombard and Jones, 2015) is defined as this phenomenon of inaccurately perceiving the technology responsible for these stimuli as not being there, or not being responsible.

### **What aspect of the phenomenon is of interest?**

A review of the literature (Lombard and Ditton, 1997) revealed six broad conceptualisations of telepresence. Of particular import is the notion of telepresence as transportation – the feeling that “you are really there”. Within the context of VR, this notion of transportation is what the term presence most commonly refers to (Schuemie et al., 2001), and is more accurately defined as *Virtual Spatial Telepresence* (Lombard and Jones, 2015).

### **Virtual Spatial Telepresence**

To reiterate, we define presence as *Virtual Spatial Telepresence* for the purposes of this research. It is the subjective experience of “being there” in a computer-generated virtual environment. This virtual environment consists of external stimuli presented via technology, and the person experiencing presence has an inaccurate perception of the involvement of this technology.

<sup>2</sup>Even without the help of technology, it is entirely possible for one to *not* feel present in their own reality (Biocca, 2003). One need only count the pairs of glazed over eyes in a boring lecture to confirm this.

<sup>3</sup>In literature regarding teleconferencing, for example, telepresence is the technology that facilitates it, not the user’s experience of using it (Lichtman, 2006).

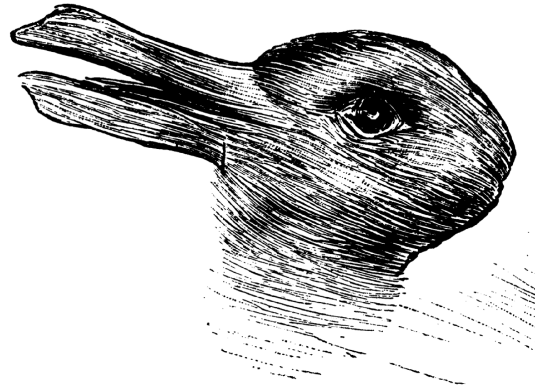


FIGURE 2.6: The Rabbit-Duck Illusion. This illusion illustrates the notion of the gestalt-shift. Depending on which aspects of the image one attends to, either a rabbit or a duck is perceived. (*Rabbit-Duck Illusion*).

### 2.2.2 Theories of Presence Formation

We now have a clear definition of presence. To fully understand how it forms, and therefore why certain factors facilitate or hinder its formation, it is important to understand the theoretical underpinnings of presence. As we mentioned in Section 2.2.1, dozens of competing presence theories have been developed over the years, each with their own strengths and weaknesses (see Nunez (2007) for a thorough review). The Process Model put forward by Wirth et al. (2007) (also known as the MEC model and Two-Level model) is based on a definition of presence that agrees with our own, unifies many of the concepts put forward in the extensive presence literature (Hartmann et al., 2015), and connects presence to well established cognitive theories. Wirth's model is particularly well defined, and there is a significant amount of evidence suggesting its validity despite its relative youth (Hartmann et al., 2015; Nunez, 2007).

Building upon previous work on the mechanisms of human perception, Slater (2002) and Slater and Steed (2000) suggest the notion that presence arises from the organisation of sensory data into an "environmental gestalt", or perceptual hypothesis about an environment. A key element of perceptual theory is that perception arises from continual selection between hypotheses that are based on the stream of sensory data we are receiving (Gregory, 2015). This process is exemplified by the famous rabbit-duck illusion, shown in Figure 2.6. The image presents two competing perceptual hypotheses: this is a duck, or this is a rabbit. Depending on which you select, your eye-movements and the features to which you attend change, therefore changing what you are "seeing". In a virtual environment (VE), one is also presented (in all but the most immersive example imaginable) with competing perceptual hypotheses, this time about what environment you are in: the real, or the virtual (Slater, 2002). From this viewpoint, presence occurs when the hypothesis that the VE is one's actual environment is consistently selected over the hypothesis that the real world is one's actual environment, and any competing signals from the real world, such as the weight of an HMD and its cables, are ignored. A break in presence occurs when a signal from outside the VE is, for whatever reason, salient enough that one switches their attention to it. This is similar to the switch between seeing the duck and the rabbit in Figure 2.6.

Wirth's Process Model (Wirth et al., 2007) significantly expands on these ideas. Broadly speaking, the Process Model consists of two stages. First, a mental model (Spatial Situation Model) of a virtual environment is constructed by the person confronted with it. Once this has happened, through a series of cognitive "tests", this mental model of a space becomes the person's primary frame of reference for where they are. When this occurs, the person is experiencing presence. We will now look at these two stages in more detail.

### Formation of the Spatial Situation Model

A Spatial Situation Model (SSM) is a mental representation of an environment. In the context of presence, we are interested in the representation of a virtual environment. This kind of spatial cognition, the development of an internal model of a virtual space, is a key component of presence (Regenbrecht, Schubert, and Friedmann, 1998; Prothero et al., 1995; Slater and Usoh, 1993a; Slater and Usoh, 1993b; Wirth et al., 2007). In order for the SSM of a VE to form, the user must be paying attention to the mediated stimuli. This attention can either be involuntary or voluntary (controlled) (Wirth et al., 2007). The role of attention in presence formation is acknowledged in many other models (e.g. (Slater, Usoh, and Steed, 1994; Steuer, 1992; Regenbrecht, Schubert, and Friedmann, 1998)).

Many aspects of digital media, mainly to do with their form, are able to manipulate people to subconsciously attend to certain stimuli. A basic example of this is the involuntary orientation of the eyes or head towards a moving element (Posner, 1980). There are a great deal of media factors capable of wresting our attention which we will not cover here, but as Wirth et al. (2007) summarise: "a continuous stream of highly detailed information should sustain users' involuntary attention allocation more effectively than an interrupted and/or less-detailed stream of input".

Not all attention is involuntary, however. Even in the absence of conspicuous formal stimuli, the user may purposely direct their attention to something because they find the content interesting. If the content of media lies within the user's domain-specific interest, they will be more motivated to pay (controlled) attention to it (Sacau et al., 2005; Wirth et al., 2007). Personal factors therefore also play a role.

Controlled and involuntary attention are not mutually exclusive, and for most media both will be involved. It is presumed that the more immersive the medium in question, the less controlled attention is required for the formation of presence (Wirth et al., 2007). This assumption is corroborated by Nunez (2007), who mentions that due to the highly immersive nature of VR, which provides a continuous stream of rich stimuli and limits distractors, that automatic attention would be very active. Personal factors are therefore less of a contributing factor for presence in VR than they are in low-immersion or non-immersive media.

Once a user is paying attention to the medium, they notice spatial cues. Spatial cues are essential for the formation of an SSM (Wirth et al., 2007), as they allow us to identify what we are looking at as a space. These cues are most often visual and include occlusion, convergence, motion parallax, and stereopsis, among others (Nunez, 2007). The more spatial cues offered by a medium, the better the SSM we are able to form (Wirth et al., 2007; Gysbers et al., 2004). However, a vivid SSM alone is not enough to ensure presence (Nunez, 2007; Gysbers et al., 2004).

### From SSM to presence

An *egocentric reference frame* (ERF) is a mental model of a space that is organised from a first-person perspective (Mou and McNamara, 2002). SSMs of mediated environments can act as ERFs. In such cases, there will be competition for which ERF, the mediated or the real-world ERF, is the primary one. A person's *primary egocentric reference frame* (PERF) (Wirth et al., 2007) is that with which they associate their spatial perception and possible actions. At a given point, if a user's PERF is the mediated environment, they are experiencing presence (Wirth et al., 2007). Without egocentricity, an SSM cannot act as an ERF, and therefore cannot ever challenge the real-world PERF. Several other models also acknowledge the importance of this egocentricity (Regenbrecht, Schubert, and Friedmann, 1998; Slater, 2002).

The Theory of Perceptual Hypotheses (TPH) (Bruner and Postman, 1949) offers an explanation of how a mediated ERF may become one's PERF.

Perception, according to the TPH, is divided into three stages (Wirth et al., 2007):

1. Provision of expectation hypothesis
2. Input of information about the object of perception
3. Confirmation or disproof of the hypothesis

During the first stage, prior experience informs the perceiver what sorts of stimuli they should be attentive to. They form a hypothesis of expectation about the scenario that confronts them based on things that they have perceived previously. ("This space seems to be a room, based on other rooms that I have experienced.").

In the second stage, information about the target of perception is absorbed. The expectation hypothesis in the first stage determines which aspects of the target of perception are attended to the most. ("This space is enclosed by four walls, a floor, and a ceiling.").

Based on the information gathered about the target of perception during the second stage, the expectation hypothesis is either accepted or repudiated. If accepted, the process of testing the expectation hypothesis comes to an end – we have successfully perceived ("Yes. This is a room."). If not, the perception process starts again at the first stage.

The perceptual hypothesis that we are interested in is the *medium-as-PERF-hypothesis* (Wirth et al., 2007). This hypothesis arises when a media artefact provides a sufficiently strong SSM and inherent ERF, challenging the user's currently active perceptual hypothesis that the real world is their PERF. If the hypothesis that the mediated world is the PERF is confirmed repeatedly through the stimuli it conveys, the phenomenon of presence occurs. However, if information that contradicts this hypothesis is received, the hypothesis is disproved, and presence is broken. Let us use a simple example to consider how this hypothesis may be tested.

Assume that a person wearing an HMD has already formed a strong SSM of a virtual space. This SSM is an ERF – the SSM has been formed from the first-person perspective afforded by the HMD. This mediated ERF is now competing with the real-world ERF, and the medium-as-PERF-hypothesis arises. If this hypothesis is true, then the person expects that their spatial perception and actions will be associated with the virtual world. A simple turn of the head is enough to subconsciously test if this is the case. The action is carried out in the virtual world thanks to the HMD tracking, and their visual perception of the virtual space is updated accordingly via the screens in the HMD. The medium-as-PERF-hypothesis passes the test, and the person is now present; the virtual world is now their PERF. In order to stay

present, action-perception expectations such as this must continually be met (Wirth et al., 2007; Slater, 2002).

The stronger the SSM and ERF, the harder it is to break presence (the more evidence is required to disprove the medium-as-PERF-hypothesis) (Wirth et al., 2007), but a failure by the system to meet the expectation tied to an action (such as turning your head), or a salient enough stimulus from the real world (such as the harsh tug of a cable attached to the HMD) could be enough for the real world to become the PERF once again, and presence to be broken, since the association of perception and/or action with the virtual world is broken (Wirth et al., 2007).

### Summary of presence formation

In summary, the Process Model of presence formation is divided into two main parts: the formation of the SSM and ERF, and a hypothesis testing stage. Evidence suggests that the formation of the SSM and ERF are indeed separate from the adoption of the ERF as PERF through perceptual evidence (Nunez, 2007). In the first stage, user attention is focused on spatial cues organised from an egocentric perspective to create a mental model of the virtual space. The resulting ERF challenges the current PERF, and after confirming that the ERF offered by the medium conforms to the user's expectations, it becomes the new PERF and the user is now present. The better the initial SSM, the more likely presence is to occur, and the more conflicting stimuli will be needed to break it.

Wirth's (Wirth et al., 2007) theory allows us to explain how presence forms and how it may be broken, and explains the results of previous work with respect to the factors that have been found to influence presence.

### 2.2.3 Factors Influencing Presence

How immersive the system delivering a VE is, and therefore the extent to which it is able to facilitate a sense of presence (Slater and Wilbur, 1997; Wirth et al., 2007), is affected by several factors.

From a hardware perspective, the following dimensions are important for immersion (Slater and Wilbur, 1997). The higher a system scores in each of these dimensions, the more immersive it is, and the greater its affordance of presence:

- *Inclusiveness* - extent to which the real world is shut out.
- *Vividness* - resolution and graphical fidelity.
- *Surrounding* - field of view.
- *Proprioceptive Matching* - match between optical flow and proprioception.
- *Extensiveness* - range of sensory modalities that are accommodated.

The importance of *inclusiveness* is twofold. The more inclusive a system is, the fewer stimuli from the real world can be perceived. During the first stage of presence formation, this means that there is less competition for the attention required to form a strong SSM. Once presence has been established, conflicting stimuli that may negate the medium-as-PERF-hypothesis are less likely to reach the user, so presence is less likely to be broken. Stimuli that suggest the existence of the device conveying the stimuli should also be minimised (the device is part of the real world), so systems with faster update rates, less aliasing, and lower weight would be considered more *inclusive* than systems that fail in these areas and draw attention to themselves.

A continuous and rich stream of sensory stimuli are required both for the formation of a strong SSM, and to continually reconfirm the medium-as-PERF-hypothesis. *Vividness* and *surrounding* are therefore both important during both stages of presence formation. Stereoscopic 3D could be considered to fall under the umbrella of *vividness* and, along with motion tracking of the user's head, supports spatial cues such as stereoscopic depth and motion parallax, both of which aid in the creation of a strong SSM, and both of which have been found to improve presence (Hendrix and Barfield, 1996a).

*Proprioceptive matching* is the match between what the eyes see and what the body feels. The head tracking offered by modern HMDs is an excellent example – when the user turns their head, their view of the virtual environment is updated exactly as they would expect it to be. Head tracking has been found to positively effect presence to a large extent (Hendrix and Barfield, 1996a). Greater levels of *proprioceptive matching*, such as allowing users to walk as a means of input rather than use a controller, for example, have been found to increase presence further (Slater, Usoh, and Steed, 1995). A component of the second phase of the formation of presence is the association of a user's possible actions with the virtual world (Wirth et al., 2007). A high level of *proprioceptive matching* achieves exactly this, by giving users the expected visual (or any other modalities that are supported by the system) feedback from the VE based on their movement.

The more *extensive* a system, the more of a user's sensory modalities it is able to stimulate, and to greater effect. For any stimuli that can be used to convey spatial cues, *extensiveness* is therefore important for the construction of an SSM, and the stronger the SSM, the more likely that presence will be achieved (Wirth et al., 2007). For example, stereoscopic visual displays lead to higher levels of presence than monoscopic displays (Hendrix and Barfield, 1996a), and spatialized audio leads to higher levels of presence than non-spatialized audio (Hendrix and Barfield, 1996b). *Extensiveness* can also play a role in the second phase of presence formation. A system that uses haptic feedback, for example, which stimulates the sense of touch, can provide better feedback about the association of user action with the virtual world than a system that does not.

Other factors which can influence presence that are not hardware-centric include *egocentric representation* in the VE, and *plot*. Egocentric self-representation in the VE is important for a sense of presence (Slater and Wilbur, 1997; Regenbrecht, Schubert, and Friedmann, 1998). It allows all of the spatial information conveyed by the system to be organised according to a first-person perspective, which is essential for the formation of an SSM that can act as an ERF (Wirth et al., 2007). *Plot*, depending on personal factors such as domain-specific interest, can offer a sense of narrative immersion and can further prompt users to pay controlled attention to the VE.

Together, these factors contribute to how immersive a system is. A meta-analysis has shown that these immersive properties do indeed affect presence, and are closely aligned with those put forward in the Process Model (Cummings and Bailenson, 2016). Considering the hardware factors mentioned here, it should come as no surprise that HMDs, which exploit all of them, offer a significantly more immersive experience than alternatives. As Wirth puts it, this explains the “demonstrated power of such media to evoke and sustain spatial presence”.

### 2.2.4 The Importance of Presence

Presence is thought to have various effects: to cause a subjective sensation of place, to affect task performance, to elicit responses to stimuli as if they were real, and to

affect the level of simulator sickness experienced in VR (Schuemie et al., 2001). There is, however, contention surrounding these effects. Some have found that there is a small positive correlation between task performance and presence (Kim and Biocca, 1997), whereas others claim there is no correlation at all (Welch, 1999), and others still suggest there may be a negative correlation (Ellis, 1996). Several studies have found a significant correlation between presence and emotional responses such as fear (Regenbrecht, Schubert, and Friedmann, 1998; Slater, Pertaub, and Steed, 1999), but the effect sizes are mostly small and the causal nature of the relationship has yet to be established (Schuemie et al., 2001). Finally, simulator sickness in VR has also been found to be both positively (Witmer and Singer, 1998) and negatively (Slater, Steed, and Usoh, 1995) correlated with presence.

These uncertainties lead Schuemie et al. (2001) to question “the usefulness of presence” and its study, with the notable exception that if we accept the definition of presence as a subjective sensation, it becomes a goal in itself for entertainment media such as video games. This is precisely in accord with our definition of presence, and the context of use with which this research is concerned.

## 2.3 A Problem with HMDs

Of the various ways in which VR can be implemented, the HMD is the most practical for consumers. HMDs are (relatively) small, light, and inexpensive, while offering high fidelity visuals and audio.

However, the visual cut-off they cause is a double-edged sword. On the one hand, this cut-off makes the system more *inclusive*, stopping stimuli from the real world from challenging the medium-as-PERF-hypothesis, increasing the likelihood that it will be repeatedly confirmed and presence will occur or continue to occur (Wirth et al., 2007). On the other hand, not only are users cut off from the tools and peripherals they may need for whatever activity they are engaged in (such as input devices) (Billinghurst and Kato, 1999), but also from all other real-world objects they may wish to interact with (McGill et al., 2015; Budhiraja et al., 2015).

Visual cut-off leads to a diminished capacity to provide input to the system and interact with real-world objects, and the strategies that users are currently employing to mitigate these deficiencies can negatively affect presence.

### 2.3.1 Input Performance

Certain tasks, such as text-entry, are very difficult to accomplish without being able to see what one is doing (McGill et al., 2015). Although a subset of keyboard commands, or the mapping of a controller, can be memorised and used for activities such as gameplay (Billinghurst and Kato, 1999), these are inadequate for tasks that require a greater bandwidth of input. Certain situations may also call for users to switch between multiple input methods. Current HMD users rate the inability to provide input into VR as a greater impediment to widespread adoption than the hitherto rather common HMD-induced nausea (McGill et al., 2015). Because of its ubiquity and ease of measurement, we will focus on text-entry via a standard keyboard as an example of a commonly required input task that is difficult to execute while visually cut off.

In the home context, HMDs are most likely to be used to play video games. Text-entry in games is more common than one may think, and especially in online multiplayer games, is remarkably prevalent. Almost every online multiplayer game has

a text chat box in a corner of the heads-up display (Peña and Hancock, 2006). An analysis of communication in the game *World of Warcraft* (Blizzard, 2004) showed that nearly half a million text messages were sent in the 11775 recorded sessions, an average of about 40 text messages per session (where the average session lasted 57 minutes) (Suznjevic, Dobrijevic, and Matijasevic, 2009). This amounts to an average of more than one text message being sent every two minutes per player. In fact, the dominant form of communication in games of several genres is text (Herring et al., 2009). The effectiveness of voice-based communication for cooperative endeavours and task coordination (Jensen et al., 2000) lends itself to use in those situations, but text-based chat is used for socio-emotional communication to a large extent (Peña and Hancock, 2006; Herring et al., 2009). Text-based communication is seemingly entrenched in the social and emotional aspects of multiplayer gaming. This is perhaps no more evident than in the recent case where text-based communication was not included in the popular *Metal Gear Solid V:MGS Online* (Konami, 2015), leading to outrage among the franchise's fan base. To quote one Mr. GGscrub, "Wtf like, it was an essential part of the charm of MGO2! How the hell is it missing?" (gamefaqs, 2015). The sentiment is echoed by CanOpener74 ("really not a good change, this... sadness... pervades.") and many others (gamefaqs, 2015; Steam, 2016).

Although there are several communication and text-entry methods alternative to typing, they are not ideal in all situations. Games like Counter-Strike (CS) (Valve, 2016) provide players with a small set of predefined audio clips (predominantly strategic in nature) which can be used whenever required. Of course, this limits the scope of communication which is possible. Research has shown that text-based chat is used in CS when such predefined utterances are not adequate, or the content of the communication is not strategic in nature (Manninen, 2001).

Voice Over IP (VOIP) is a commonly used communication method among gamers. However, for various reasons, gamers do not always favour it. Despite its potential for improving the performance of a team, an uncoordinated or overly large team all talking at once and trying to shout over each other can render VOIP far less useful, leading some players to prefer its absence (Manninen, 2001). This seems to be particularly true in cases where multiplayer games often occur between people that do not know each other, with "voice for friends, text for randoms" being a common preference. Even when VOIP is preferred, it cannot always be used. Having to speak out loud can be a privacy concern or disruptive to those around you. Voice-to-Text solutions suffer the same issues.

Several recent works have acknowledged that there is currently no method of text-entry alternative to the standard keyboard that is suitable for VR (Walker et al., 2017; Lin et al., 2017; Lee and Kim, 2017; McGill et al., 2015).

We have discussed how, even in the context of games, people will likely need to enter text via a standard keyboard. This need will undoubtedly be larger in other contexts of HMD usage. While there are many alternatives to the ubiquitous keyboard, none of them is good enough to completely replace it, which explains recent research efforts to facilitate its use in VR.

### 2.3.2 Real-World Interaction

The objects that people wish to interact with while they are playing games are not limited to peripherals used for playing the games. Intuitively, this seems obvious. Studies have shown that 67% of the time spent playing computer games is also spent doing something else (Foehr, 2006). Forty percent of the actions constituting that "something else" are real-world interactions not directly linked to the game being

played, such as eating or drinking, reading, homework, and using a mobile phone (Foehr, 2006).

Several researchers have tackled the issue of real-world interactions during HMD usage. Desai, Pena-Castillo, and Meruvia-Pastor (2017), for example, created a system that allows you to interact with a smartphone while wearing an HMD. Others have focused on the problem of interacting with other people (Chan and Minamizawa, 2017). Others still have attempted to solve the problem more generally (McGill et al., 2015; Budhiraja et al., 2015). It is clear that visual cut-off causes problems with real-world interactions and currently, HMD users rate their ability to interact with objects as extremely ineffective (McGill et al., 2015).

### 2.3.3 Diminished Sense of Presence

HMDs have existed since the late 60s, but have only recently started to see widespread everyday use thanks to affordable consumer-oriented technology. Although this first consumer-level VR hardware has only recently begun shipping, earlier development versions of the hardware (such as the Oculus Rift DK1 and DK2 (Oculus, 2016)) have been in use by early adopters for some time. This has allowed for the usage of these modern devices to be studied, and to find out how people are attempting to overcome the visual cut-off created by HMDs. The solutions that users have come up with, “peeping” and “groping”, are not particularly good.

Peeping involves temporarily lifting the device off one’s eyes in order to gain awareness of, or interact with, the real world. HMD users agree that having to resort to peeping is frustrating (McGill et al., 2015). Moreover, peeping totally shatters the sense of presence. When the user lifts the headset off their head, the *inclusiveness* of the system is attenuated, and all of the stimuli necessary to support the active medium-as-PERF-hypothesis are replaced with stimuli from the real world. All of the user’s sensory modalities will now confirm that the real world is, in fact, their PERF.

Groping is an attempt to interact with the real world without removing the HMD. User’s blindly try to navigate the real world using their memory and sense of touch. In this case, although the audio-visual stimuli necessary to maintain presence are still there, the medium-as-PERF-hypothesis faces a significant challenge. The tactile cues that the user receives from the real world conflict with the stimuli that he/she is receiving from the system, creating the contentious perceptual hypothesis that the real world is their PERF. Furthermore, the stimuli that the user is still receiving from the system are likely to be completely ignored, since the user’s controlled attention is being aimed squarely at the real world where their objective lies, overriding all but the most salient provocations of involuntary attention that the system can conjure. Presence is sure to be broken.

It is important to note that the visual cut-off created by wearing an HMD does not necessarily interfere with the usability of the HMD itself <sup>4</sup>, but rather with the usability of peripherals and other objects. The ways in which current HMD users have to work around visual cut-off to engage in common activities is frustrating, reducing the quality of the experience as a whole, and interferes with the very goal of VR – heightened presence.

If VR is to be as ubiquitous as traditional desktop computing is today, the usability issues created by HMDs need to be better addressed. Mixed Reality may offer a

---

<sup>4</sup>Quite the opposite, in fact, considering Slater’s definition of immersion and the importance of *inclusiveness*.

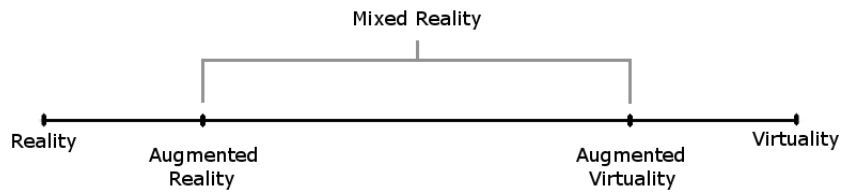


FIGURE 2.7: The Reality/Virtuality Continuum. After Milgram and Kishino (1994).

solution to these problems, mitigating the negative impacts of visual cut-off while preserving presence as much as possible.

## 2.4 Augmented Virtuality

Mixed Reality (MR) is a subset of VR-related technologies which aim to merge elements of real and virtual worlds (Milgram and Kishino, 1994). Augmented Reality (AR), the addition of virtual objects into the display of an otherwise real world, is probably the best known incarnation of MR, and several AR applications, such as *Pokemon GO* (Niantic, 2016), have achieved mainstream commercial success.

There are several different ways that MR can be achieved (Milgram et al., 1995). For consumer HMDs, the most common implementation of MR is video “see-through” (Edwards, Rolland, and Keller, 1993). This involves the mounting of one or more digital video cameras to the exterior of an HMD, and passing the images obtained from these cameras directly to the displays within the HMD. In effect, this allows the user to see through the HMD into the real world.

MR implementations can be thought of as occupying a point somewhere along a “reality/virtuality continuum” (Milgram and Kishino, 1994; Milgram and Colquhoun, 1999) as illustrated in Figure 2.7. AR is closer to the Reality side of the spectrum, where the display of an entirely real environment has virtual artefacts added to it (predominantly real with few virtual elements).

In Augmented Virtuality (AV), on the other hand, certain real objects are made visible in an otherwise entirely virtual environment (predominantly virtual with few real elements). AV falls close to the Virtual side of the Mixed Reality spectrum. Users see a world predominantly constructed from digital artefacts which is augmented with key elements of reality made visible over the virtual world. This solves the problem of visual cut-off as, quite simply, an AV user is no longer visually cut off from real world objects with which they wish to interact. There are several possible AV designs and implementations, differing according to how much of the real world is shown, and when or why the real world is shown at a given juncture. We will now elucidate these design approaches, and consider their value in terms of solving the issues caused by visual cut-off during HMD usage.

### 2.4.1 Design Approaches

The design approaches of AV systems differ mainly according to two factors. The first is the amount of reality that is shown. The second is when this reality is shown, or at what point the system adds reality to the virtual world.

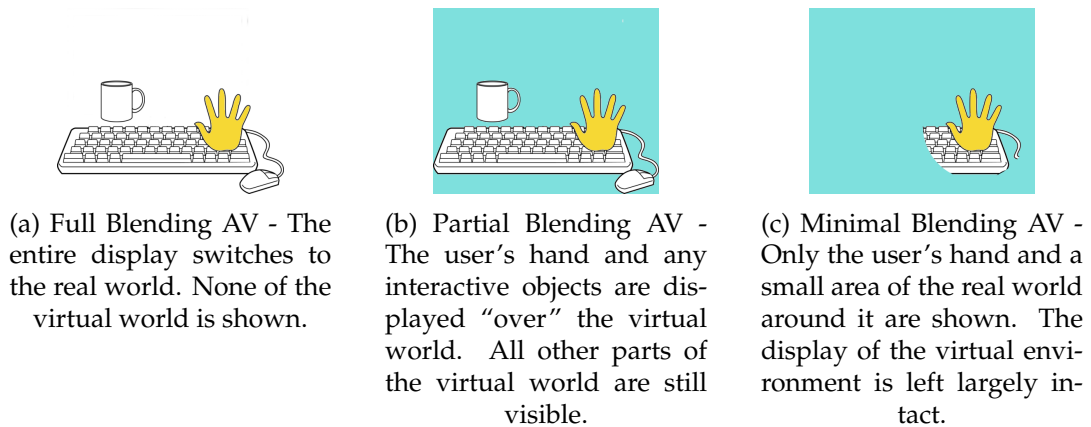


FIGURE 2.8: Graphical Representation of Different Types of AV - Adapted from McGill et al. (2015). The black and white parts of the images represent the real world in front of the user. The yellow hand is the hand of the user, held out in front of them. The blue areas represent areas of the visual field that are being occluded by the display of a virtual environment.

## Blending

The first way in which AV design may differ is the *amount* of the real world that is shown. This can vary from completely replacing the view of the virtual world with the view of the real world, to overlaying only small amounts of reality over the virtual world. McGill et al. (2015) refer to the amount of reality shown as the level of *blending* (how much of the real is "blended" into the virtual). The level of blending chosen determines where on the Reality/Virtuality continuum a particular design of AV will fall.

*Full blending* (McGill et al., 2015) (Figure 2.8a) completely replaces the view of the virtual world with that of the real world, and therefore falls closer to the real side of the spectrum. On the other side of the AV sub-spectrum is *minimal blending* (McGill et al., 2015) (Figure 2.8c), where only a very small amount of the real world is ever shown (such as just the area around the user's hands), leaving the display of the virtual world largely intact. *Partial blending* (McGill et al., 2015) (Figure 2.8b) falls between these two extremes, allowing key objects to be viewed over the virtual world, irrespective of the position of the user's hands. Another method, *inset blending* (Metzger, 1993), involves overlaying a small windowed view (or inset) of the real world in a fixed location of the display of the virtual world.

Experiments conducted by McGill et al. (2015) and Budhiraja et al. (2015) show that partial and minimal blending are the preferable designs. Restricting the view of reality to only that which is needed to execute real-world interactions means that fewer stimuli from the real world reach the user, and stimuli from the virtual world can still reach the user, thereby reducing the challenge to presence.

## Transitions

The second way in which AV designs may differ is *when* reality is shown (regardless of how much will be shown). This transition can be either explicitly requested by the user of the system (*user-initiated* (McGill et al., 2015)), or the system can attempt to predict when the user requires the transition (*inferred* (McGill et al., 2015)).

An example of a user-initiated transition would be the use of a button press or voice command, or some other gesture unrelated to the reason that someone needs to view reality, to bring about the view of reality. On the other hand, an inferred transition would be the automatic switching between VR and AV based on a user action which may be the first step in executing a desired task – reaching out for a real-world object, for example.

McGill et al. (2015) compared the level of workload associated with user-initiated and inferred transitions, and showed that user-initiated transitions are preferable in this respect.

### 2.4.2 Implementation Approaches

In general, video see-through (somehow allowing the user to see through the HMD into the real world) is the first step of the blending process. Both McGill et al. (2015) and Budhiraja et al. (2015) achieved this video see-through by allowing the HMD to display the live feed from a colour webcam. In order to maintain the relationship between proprioception and optical flow, the webcam is attached to the front of the HMD as close to the position of the eyes as is possible. McGill et al. (2015) used a single colour webcam, whereas Budhiraja et al. (2015) used two, which allowed a stereoscopic 3D view of the real world. Other implementations that have less general applications have used different blending implementations. Desai, Pena-Castillo, and Meruvia-Pastor (2017), for example, used an application running on a motion-tracked smartphone to superimpose the smartphone's display over the virtual environment. Additionally, the Microsoft Kinect has been used to allow the intrusion of proximate persons into the virtual world of the HMD user (McGill et al., 2015). However, the use of cameras mounted on an HMD is the most common way that video see-through is implemented. The major strength of this method is that it is a general solution to the problem of visual cut-off. If implemented correctly, it could be used for interacting with any object at all, as opposed to the more specific solutions which only facilitate interaction with a particular object (e.g smartphone (Desai, Pena-Castillo, and Meruvia-Pastor, 2017).

Once a view of reality is available, implementations differ in how the subset of reality to be shown is selected. This can be achieved using chroma-keying (McGill et al., 2015), colour segmentation (Budhiraja et al., 2015), or image recognition (Desai, Pena-Castillo, and Meruvia-Pastor, 2017; McGill et al., 2015). The benefit of using chroma-keying is that any real-world objects can be segmented out of the view of reality without any additional setup, making it a good general solution to the problem of visual cut-off. The drawback of this method is that it requires the area in which the HMD is used to be converted into a "green screen". Colour segmentation will rely on objects of interest being particular colours, or being marked in some way. Image recognition is the least general approach, as it will only segment specific objects, such as the smartphone used in research conducted by Desai, Pena-Castillo, and Meruvia-Pastor (2017). An implementation as generally useful as chroma-keying, but without the impractical setup requirements would be the best approach.

In order for inferred transitions into AV to be possible, the tracking of hands, other people, or objects of interest is necessary. This has been achieved in several ways, including image recognition (Desai, Pena-Castillo, and Meruvia-Pastor, 2017; McGill et al., 2015) and colour segmentation (Budhiraja et al., 2015). Image recognition is the preferable approach for hand recognition, as it can be achieved without the need to wear markers of some sort, and does not depend on environmental factors such as lighting conditions.

### 2.4.3 Measurement Approaches

The assessment of an AV implementations requires the measurement of several aspects of user interaction and experience. Previous work has focused on the levels of presence and workload experienced by users, as well as the speed and accuracy of providing input via a standard keyboard.

#### Presence

Presence can be measured either objectively or subjectively (Van Baren and IJsselsteijn, 2004). As virtual environments become more realistic, so too do our physiological responses to these digital stimuli – our bodies react to virtual entities or scenarios the same way they would to their real counterparts (Van Baren and IJsselsteijn, 2004). The more real the environment, the more evident our physiological responses will be, and their measurement can therefore yield insight into the level of presence that a subject is experiencing. However, establishing validity for objective measures is a challenge, as physiological responses may be triggered by stimuli other than those intended (Van Baren and IJsselsteijn, 2004). Furthermore, the equipment necessary for many objective measures of presence is expensive, and in some cases may be obtrusive enough to interfere with the experience of experimental participants (Van Baren and IJsselsteijn, 2004).

It has been argued that since presence is primarily a subjective experience, self-report offers its most direct measurement (IJsselsteijn et al., 2000; Sheridan, 1992). Questionnaires are the most common and frequently used form of subjective presence measure: they are cost-effective, facilitate administration and analysis, do not interfere with user experience, and have been found to demonstrate validity and sensitivity (Van Baren and IJsselsteijn, 2004).

The Igroup Presence Questionnaire (IPQ) (Schubert, Friedmann, and Regenbrecht, 2001) was constructed by combining items from several notable pre-existing presence questionnaires (such as the SUS (Slater, Usoh, and Steed, 1994) and PQ (Witmer and Singer, 1998)) with several new items. The IPQ has been found to be reliable and valid (Schubert, Friedmann, and Regenbrecht, 2001) and to demonstrate sensitivity, differentiating between multiple levels of presence in several studies (Van Baren and IJsselsteijn, 2004). Schuemie et al. (2001) recommend the use of either the IPQ or the ITC-SOPI (Lessiter et al., 2001) due to their demonstrated reliability and validity. The IPQ has been used extensively in research regarding virtual environments and games, including research into AV such as that conducted by McGill et al. (2015). Its use therefore facilitates comparison to previous work.

Finally, the IPQ is based on the same definition of presence presented in Section 2.2.1 and theoretical underpinnings that are very similar to those discussed in Section 2.2.2, with particular focus on the idea that presence arises from the construction of a mental model of a VE organised from an egocentric perspective (Schubert, Friedmann, and Regenbrecht, 2001). For these reasons, the IPQ is the most appropriate choice of presence measure for this research.

#### Workload

Broadly speaking, there are three ways in which workload can be measured: psychophysiological measures, performance measures, and subjective measures (such as self-report rating scales) (Cain, 2007; Miller, 2001).

Psychophysiological measures focus on the measurement of physiological responses (such as heart rate, eye movement, and respiration) to the psychological

states induced by an activity (Cain, 2007). This approach to measuring workload is problematic for several pragmatic reasons: the required equipment is expensive and often invasive, individual differences require baseline states to be captured for each individual under assessment<sup>5</sup>, and it is difficult to know which physiological responses are the correct ones to measure for a given activity (Cain, 2007). On a deeper level, it seems that the link between physiological responses and workload is poorly understood (Kramer, 1991; Eggemeier and Wilson, 1991). Several studies have shown that psychophysiological measures of workload do not change as workload does, or are too sensitive to extraneous effects to be useful (Corwin et al., 1989). Psychophysiological measures of workload will no doubt improve as the theory and technology behind them do, but for now their use is not generally recommended (Cain, 2007), and is not practical for the purposes of this research.

Intuitively, the easier a task is, the less workload there is in its execution, and the better a person will perform at the task. Due to this, and the fact that one of the primary motivators behind assessing workload in a given situation is to learn how to maximise the productivity of people working in it, performance metrics are one of the ways that workload is commonly measured. Performance measures are highly dependent on the task at hand but speed, accuracy, error rates, etc. are common examples of the metrics used (Cain, 2007). The main criticism of this method is that it doesn't actually measure workload at all (Hart and Wickens, 1990). For example, two different people may perform the same task equally well, even though one finds it very difficult and the other very easy (De Waard, 1996). For the purposes of this research, the subjective experience of workload that performance measures miss out on is more important than, for example, how quickly a real-world interaction takes place. This is because each user is different, and we feel that in the home context, performance is less of an issue than how users feel about that performance.

Subjective workload is a psychological construct, and it is therefore appropriate that it be measured in a subjective manner (Cain, 2007). The NASA Task Load Index (NASA-TLX) (Hart and Staveland, 1988) and the Subjective Workload Assessment Technique (SWAT) (Reid and Nygren, 1988) are self-report measures that have both seen extensive use. Thousands of studies have made use of the NASA-TLX, approximately forty-two percent of which aimed to evaluate some sort of audio-visual display or input device (Hart, 2006). In conclusion to a comprehensive review, Cain (2007) suggests that the NASA-TLX be used over other measures, noting that SWAT is a viable but laborious alternative.

Due to its many benefits and its suitability in our context, and to facilitate comparison to previous work, the NASA-TLX was chosen as the workload measure for this research.

### **Input performance**

Most previous work in AV has focused on the measurement of text-entry via a standard keyboard as an indication of the system's ability to facilitate input. Text-entry is measured in terms of speed and accuracy.

The method of text-entry analysis proposed by MacKenzie and Soukoreff (2002) is a comprehensive and widely used measure of text-entry accuracy. Their technique is based on the Levenshtein Minimum String Distance (MSD) statistic (Levenshtein, 1966), used in conjunction with the Key Strokes Per Character (KSPC) statistic.

Most text-entry research is based on user experiments that require participants to transcribe a piece of text. Perfect typing would result in the text entered by the

---

<sup>5</sup>The capturing of such baselines is a delicate and error-prone process in and of itself.

user (the transcribed text) being exactly the same as the text that was prescribed to them (the presented text). MSD allows us to judge the accuracy of a user's text-entry by comparing the presented text to the transcribed text. Specifically, MSD calculates the number of primitive edits that are required to get from one string to another. For example, if a participant is presented with 'duck', and they transcribe 'dcuk', then the MSD is equal to two: in order to get from the transcribed text to the prescribed text, two edits are required: the 'c' must be replaced with a 'u', and the 'u' must be replaced with a 'c'. The algorithm to determine the MSD is presented in Appendix J.

In order to derive an error rate from the MSD, the following formula is used (Soukoreff and MacKenzie, 2003):

$$ErrorRate = \frac{MSD(P,T)}{\overline{S}_A} \times 100\%$$

where  $P$  and  $T$  are the presented and transcribed strings respectively, and  $\overline{S}_A$  is the mean length of the elements in the set of alignment strings (ASCII representations of the differences between  $P$  and  $T$ ). This formula is a revised version of an earlier formula (Soukoreff and MacKenzie, 2001) that was shown to be inaccurate in certain circumstances (MacKenzie and Soukoreff, 2002).

The major advantage of using the MSD error rate is that participants can enter text naturally. They can merely be asked to enter the prescribed text quickly and accurately, without the need to artificially constrain them to not making any mistakes at all. However, since the MSD only compares the presented and transcribed text, it does not measure errors that were made and then corrected while the text was being transcribed: a participant who committed many errors and corrected them will produce an identical string to one who entered the text perfectly. In order to account for these "corrected errors", another statistic, Key Strokes Per Character (KSPC) (Soukoreff and MacKenzie, 2003) can be used:

$$KSPC = \frac{|InputStream|}{|TranscribedText|}$$

where the input stream is the exact sequence of keys pressed (including keys such as backspace and delete) that are used to create the transcribed text. A KSPC value of 1 would indicate that the transcribed text was entered without any corrections<sup>6</sup>. The more errors that are made and corrected, the higher the KSPC value would be.

However, KSPC is unable to differentiate between a scenario where few errors were made, but many keystrokes were used to fix them, and one where many errors were made, but they were fixed in few keystrokes. Furthermore, KSPC is dependent on input method (for example, the ITU E 1.161 telephone keypad requires more than one key stroke per character). Finally, in the current formulations, it is difficult to combine MSD and KSPC.

In order to meaningfully combine both of these statistics, we must first divide the characters in the input stream into distinct classes (Soukoreff and MacKenzie, 2003): C (correct), IF (incorrect fixed), INF (incorrect not fixed), and F (fixes). C characters are those in the transcribed text that are correct, IF characters are those in the input stream that are not in the transcribed text (excluding editing keys such as backspace or delete), INF characters are identified with the MSD algorithm presented in Figure J.1, and F characters are editing keys in the input stream. The MSD error rate and

<sup>6</sup>This is the case provided that the input device being used has a one to one mapping of keys to characters, such as a standard keyboard.

KSPC can then be re-written as follows:

$$MSDErrorRate = \frac{INF}{C + INF} \times 100\%$$

$$KSPC = \frac{C + INF + IF + F}{C + INF}$$

Finally, the total error rate can be calculated:

$$TotalErrorRate = \frac{INF + IF}{C + INF + IF}$$

Measuring the speed of text-entry is trivial. Speed is simply the number of words typed per minute (WPM), where a word is usually defined as five consecutive characters. This measure is known as Gross Typing Speed (GTS). A more useful measure is Net Typing Speed (NTS), which combines GTS and accuracy to provide a realistic measure of typing speed that takes uncorrected errors (and the time it would take to fix them) into account. NTS is calculated as follows:

$$NTS = GTS - \frac{INF}{t}$$

where GTS is gross typing speed, INF is the number of uncorrected errors, and t is time in minutes. Using these measures of speed and accuracy, text-entry can be thoroughly measured.

## 2.5 Previous Findings Regarding User Experience in AV

Previous work in the area has yielded several useful and promising results regarding AV. However, there are areas that have yet to be fully explored.

McGill et al. (2015) conducted a series of user experiments comparing typing performance, workload, and presence in different AV and VR conditions. Their work focused mainly on comparing different AV implementations to each other. They reported that users experienced a significantly greater amount of presence in minimal and partial blending conditions, compared to full blending or pure VR, when users had to interact with the real world. However, only a single item of the IPQ (from the General Presence factor, see Appendix I) was used to compare the level of presence experienced in VR directly to that experienced in the blending conditions. The impact of AV on Spatial Presence, Involvement, and Realism (measured by the other 13 items of the IPQ) compared to VR was not discussed. While it was shown that partial and minimal blending are the superior AV designs, it was not shown in sufficient detail how much better they were than VR, and what components of presence were most affected by the AV conditions. McGill et al. (2015) also found that typing performance was significantly better in partial blending conditions than in pure VR. However, typing performance in minimal blending AV was not assessed. Finally, it was found that inferred transitions between VR and AV lower the workload experienced by users compared to user-initiated transitions, but the workload experienced by VR users compared to inferred transition AV users was not discussed. A more thorough direct comparison between AV and VR, in terms of workload and presence, is required. So too is an assessment of typing performance in minimal blending AV. Furthermore, the way in which AV was implemented by McGill et al. was fairly

impractical, as blending was controlled via chroma-keying, which requires significant alterations to the HMD user's environment. For the home or office context, a more practical implementation would be desirable.

In a similar experiment, Budhiraja et al. (2015) found that their version of inferred minimal blending (called OHC) was better than alternatives. OHC scored better in terms of presence and overall satisfaction than pure VR or full blending conditions during real-world interactions. Additionally, Budhiraja et al. (2015) found that inset blending AV was inferior to OHC. While OHC users experienced a higher degree of presence than users in other conditions, no information was given about the components of presence and how they were affected individually. In fact, it is hard to tell exactly how presence was measured at all<sup>7</sup>. OHC relied on colour segmentation to control the level of blending. While this method does not require alterations to the user's environment, it is a slightly less general solution to the problem of visual cut-off than that proposed by McGill et al. (2015), as only certain objects were segmented from the rest of the environment. For example, users were easily able to locate and grasp a bright orange cup (the only object used in these experiments) but the use of a keyboard or other objects of interest would not necessarily have been facilitated.

Both McGill et al. (2015) and Budhiraja et al. (2015) measured presence by administering a self-report questionnaire after participants had spent a period of time in an immersive virtual setting. During this period, participants occasionally had to interact with objects in the real world<sup>8</sup>. What was therefore being measured was an "overall" level of presence for the entire period. Both of these works suggest that AV can preserve presence in such situations. The Process Model of presence formation can explain these results in one of two ways: either AV allows real-world interactions to be executed without presence being broken at all, or AV facilitates the return to presence after real-world interactions are complete<sup>9</sup>.

While promising in their own ways, other implementations such as those that let you use only your phone (Desai, Pena-Castillo, and Meruvia-Pastor, 2017) or only your keyboard (Lin et al., 2017) in VR are not general enough to be considered solutions to the issue of visual cut-off. An AV system that can enable users to perform any real-world task they might usually want to, such as the one developed by McGill et al. (2015), is far more desirable than these single use-case implementations.

## 2.6 Conclusion

Previous work has demonstrated that AV applications can preserve a sense of presence while facilitating interactions with the real world, improving the experience of using HMDs in home or office settings. Only a subset of reality need be incorporated into the virtual world to facilitate interaction, and incorporating too much of the real world is detrimental to presence. Furthermore, users prefer inferred transitions to user-initiated ones, and inferred transitions lower the workload of users. In concert, these findings suggest that a form of *inferred minimal blending* is the most promising AV candidate for our purposes: facilitating real-world interactions while preserving presence.

<sup>7</sup>Budhiraja et al. (2015) use the terms presence and immersion interchangeably, and mention that their questionnaire was "inspired by the core modules" of another work (Schild, LaViola, and Masuch, 2012) that used multiple metrics.

<sup>8</sup>We believe that this is a good strategy, as it closely resembles what would happen in the home/office context.

<sup>9</sup>See Section 6.1.1 for a discussion of these ideas.

The findings regarding presence presented by McGill et al. (2015) and Budhiraja et al. (2015) make sense according to the theory of presence formation presented in section 2.2.2. The low levels of presence experienced in the pure VR control conditions are explained by the fact that users had to deliberately break immersion to execute the required tasks. In the case of full blending, although still being mediated by the HMD, all stimuli presented are of the real world rather than the virtual environment, causing the medium-as-PERF-hypothesis to be disproven, and presence to be lost. The TPH also explains some of the failings of inset AV. Although the stimuli presented by the inset do not provide as much of a challenge to the medium-as-PERF-hypothesis (as the large majority of visual cues are still of the virtual environment), its small size, and the size discrepancy between what the inset shows and the real world, requires more controlled attention from the user, shifting their focus away from the stimuli required to maintain presence. Inferred minimal blending AV only provides stimuli pertaining to the real world when required, and then only just enough to perform tasks in the real world. The challenge to the medium-as-PERF-hypothesis is small enough, and the SSM facilitated by the immersiveness of the HMD is strong enough, that presence can be largely maintained.

Consumer HMDs are finally making VR accessible to the masses. However, there is a tension between the way these devices provide the immersion necessary for a heightened sense of presence and the way that people interact with computers in the home context. Previous work indicates that a form of inferred minimal blending AV has the potential to mitigate usability issues induced by visual cut-off while preserving the sense of presence intended by HMD usage. Previous work has not fully explored input performance in Minimal Blending AV, particularly with a standard keyboard. Previous work has also not provided a thorough comparison of presence and workload in AV compared to VR. Previous implementations have either not been good general solutions to visual cut-off, or have been impractical due to modifications required to the user's environment. We therefore propose an implementation of minimal blending AV that does not require any modifications to the user's environment or peripherals, and a user experiment that will allow us to evaluate it in terms of presence, workload, and input performance.



## Chapter 3

# Methods

The aim of this research was to investigate the effects of Augmented Virtuality (AV) on presence, workload, and input performance compared to pure Virtual Reality (VR). Our research questions and corresponding hypotheses were as follows:

1. Is the overall level of presence experienced over an extended session in an immersive virtual setting higher in Augmented Virtuality than in Virtual Reality when the session has to be interrupted by interactions with the real world?

$H_{A0}$  There is no different between presence scores in Augmented Virtuality and in Virtual Reality.

$H_{A1}$  Presence scores are higher in Augmented Virtuality than in Virtual Reality.

$H_{A2}$  Presence scores are lower in Augmented Virtuality than in Virtual Reality.

2. Is it easier to interact with real-world objects in Augmented Virtuality than in Virtual Reality when users have to remove the HMD?

$H_{B0}$  Workload for performing real-world interactions is the same in Augmented Virtuality as it is in Virtual Reality.

$H_{B1}$  Workload for performing real-world interactions is lower in Augmented Virtuality than it is in Virtual Reality.

$H_{B2}$  Workload for performing real-world interactions is higher in Augmented Virtuality than it is in Virtual Reality.

3. Is typing performance better in Augmented Virtuality than in Virtual Reality?

$H_{C0}$  There is no difference in typing performance between Augmented Virtuality and a Virtual Reality baseline.

$H_{C1}$  Typing performance is better in Augmented Virtuality than in a Virtual Reality baseline.

$H_{C2}$  Typing performance is worse in Augmented Virtuality than in a Virtual Reality baseline.

In order to answer these questions, we developed an AV system. Presence and workload, two of the measures chosen to evaluate the system, are subjective; we were interested in people's experience of using this system. As such, a series of user experiments was the most suitable method of evaluation. Our first research question was concerned with the effect of AV on a user's level of presence, and in order to measure this, we had to create a virtual environment in which a sense of presence could be induced and maintained. We also needed a reason for users within this virtual environment to interact with the real world (in order to answer our second

research question). Given our focus on home usage, and that the primary target market for HMDs is computer gaming, we decided that developing an immersive game within our virtual environment was a natural way to prompt users to interact with the real world while exploring the virtual environment. Furthermore, this would create an extended immersive experience with occasional disruptions, which very closely simulates the context in which HMDs are used. Our final research question, regarding input performance, necessitated the development of several pieces of software that could be used to record and measure users' typing proficiency under various conditions.

In this chapter, the details of the apparatus, methods, and measures used to answer our research questions are presented.

## 3.1 Apparatus

The apparatus required to answer our research questions consisted of four major components:

1. The virtual environment
2. The game
3. The AV system
4. Software for capture and analysis of keyboard input

We will now describe these components<sup>123</sup>.

### 3.1.1 The Virtual Environment

The virtual environment we created was a small island, shown in Figure 3.1. Only free assets, primarily from the Unity Asset Store<sup>4</sup> and Freesound.org<sup>5</sup>, were used to make it<sup>6</sup>. In Section 2.2.3, we discussed some of the factors that are known to contribute to presence. To make a sense of presence in our virtual world possible, we attempted to create a relatively realistic and good looking environment that leveraged these factors. In this section we provide a few examples of how we did this.

All possible player actions were accompanied by appropriate audio-visual feedback in the VE to emphasise the fact that their actions had effects in the VE. For example, when players moved they heard footsteps, when they pressed buttons the buttons lit up and made a sound, and doors swung open and creaked when pushed. This was to make players feel like they were acting in the virtual world rather than in the real world. This association of actions with the VE is a hallmark of presence.

If an action in a VE has unexpected or unrealistic consequences that do not conform to the SSM of the VE that the player has formed, it might be interpreted by the player as evidence against the medium-as-PERF-hypothesis, preventing presence

---

<sup>1</sup>The game and VR typing tests can be found here: <https://github.com/jacobHclarkson/AV-System>,

<sup>2</sup>The desktop typing test can be found here: <https://github.com/jacobHclarkson/Post-Experimental-Typing-Test>.

<sup>3</sup>The typing analysis software can be found here: <https://github.com/jacobHclarkson/Text-Entry-Analsysis>.

<sup>4</sup><https://www.assetstore.unity3d.com>

<sup>5</sup><https://freesound.org>

<sup>6</sup>For a full list of assets used, please see Appendix A.

from occurring. We therefore tried to have interactive objects behave in a physically realistic manner wherever possible. For example, objects that could be picked up would, when dropped, fall to the ground and collide with other objects and make a noise as they did so. For the same reasons, we also strived for realism and consistency with respect to non-interactive elements: game geometry cast realistic shadows, vegetation swayed in the breeze, environmental sound effects changed as players moved between the different areas of the island (from the beach where waves could be heard to a wooded area where wind rustled the leaves, for example), footsteps sounded different depending on the surface underfoot, *et cetera*. Additionally, the 3D models that we chose were relatively high quality and consistent in terms of art style (which we aimed to keep realistic rather than minimalist or cartoonish).

The Oculus Rift delivered this VE at a high resolution and low update rate, and afforded low latency head tracking, giving important proprioceptive matching by directly connecting the physical actions of the user to natural/expected outcomes in the VE. The HMD also allowed the environment to be explored from a first-person (egocentric) perspective, ensuring that the SSM of the VE created by the player could serve as an ERF – a precondition for presence. The HMD also provided a very large field of view (FOV) compared to a normal desktop display (approximately 80 degrees, which covers most of the human binocular field of view of 114 degrees).

In addition to monoscopic depth cues such as occlusion and relative size, the head tracking of the HMD allows motion parallax to play a bigger role than traditional desktop setups in conveying spatial cues. The HMD, which relays a distinct image to each eye, also provides stereopsis. These spatial cues were emphasised by the use of rendering effects such as ambient occlusion and environmental fog for longer distances. Using dynamic positional audio for all audio sources in the environment, we were also able to convey spatial cues (distance and direction) in the auditory channel. For example, the volume of torches crackling became louder the closer they were, and the sounds of the beach where the players started faded as they moved away.

Finally, the HMD completely excluded external visual stimuli by covering the user's eyes, and partially excluded external auditory stimuli with in built headphones. Together, the VE and HMD were used to deliver a highly immersive experience that was capable of eliciting a sense of presence.

### 3.1.2 The Game

The island served as the setting for a simple puzzle game that would give players a reason to explore and to interact with real-world objects. Three rooms were placed on the island, each containing a puzzle that would require some interaction with real-world objects to solve. To complete the game, players had to find the rooms and successfully complete the puzzles in them. Our game gave participants a reason to use our AV system to accomplish tasks outside the virtual environment as they might do in a normal home or office context, thereby allowing us to address our second research question. This extended immersive experience punctuated by brief real-world interactions closely models the way that HMDs are used in these contexts. Creating a mixed reality game was also a step towards future work suggested by Budhiraja et al. (2015), who mention that AV allows for the possibility of games that allow physical interactions with the real environment to somehow affect the virtual one.



FIGURE 3.1: Top down view of the virtual environment in which the puzzle game took place. 'S' indicates the start location of the player. 'A', 'B', and 'C' indicate the locations of the rooms containing puzzles that required real-world interaction in order to be solved.

We felt that the puzzle game genre was better suited to a mixed reality implementation in our context than any alternatives, and did not have a significant skill demand that may frustrate less experienced players or sap attention required to come to terms with the novelty of AV.

Each of the puzzles required players to locate and interact with objects on the desk in front of them to retrieve clues necessary to solve them. These real world interactions were designed to require several steps:

1. Orientation: Players had to figure out which direction they were facing in relation to the desk and turn towards it (players played the game in a swivel chair to help them turn in the game world and were not always facing the desk during gameplay). Players then had to put down the controller they were using to play the game.
2. Identification and Location: Players had to find the correct object on the desk, as specified by the in-game objective.
3. Manipulation: Players had to reach for the object and manipulate it in some way so as to expose the clue.
4. Examination: Finally, players had to examine the clue closely enough to correctly interpret it before relocating and picking up the controller once again.

Some of these steps needed to be repeated for each clue, depending on the player. The particular real-world interactions required were chosen to ensure that the system was properly tested and used often enough to garner useful data. The decision to incorporate the real-world interactions into the game was made as we felt that this would more accurately simulate the context of home use – an extended immersive experience punctuated by brief real-world interactions, such as gamers occasionally wanting to take a drink.

Before playing the game, participants were introduced to the controls and user interface through a brief tutorial.

### **Tutorial, controls, and user interface**

To ensure that participants all had the same level of understanding of the user interface (UI), basic mechanics, and the notion that real-world interaction would be required and how to use the AV system, a short tutorial was created. For simple games where mechanics can be discovered through experimentation, a tutorial is not necessary. However, in cases where a game or its mechanics are complex, a tutorial can improve player engagement (Andersen et al., 2012). We felt that the notion of a mixed reality game was uncommon enough to most, and that the unprompted discovery of the correct usage of our AV system in a short time frame was unlikely enough, that a tutorial was necessary. Improper or absent explanation of the functioning of our game and system may have lead to diminished user experience, potentially skewing our results.

The tutorial level was a simple room in which players learned how to move in the game world, interact with objects, and understand the UI whereby they would receive instructions in the game. This tutorial level is shown in Figure 3.2<sup>7</sup>.

Turning was controlled via the positional tracking of the HMD. To turn, players turned their head or body left or right (swivelling in their chair if necessary).

---

<sup>7</sup>Different versions of this tutorial level were created for the control and experimental conditions. See Section 3.5.1.

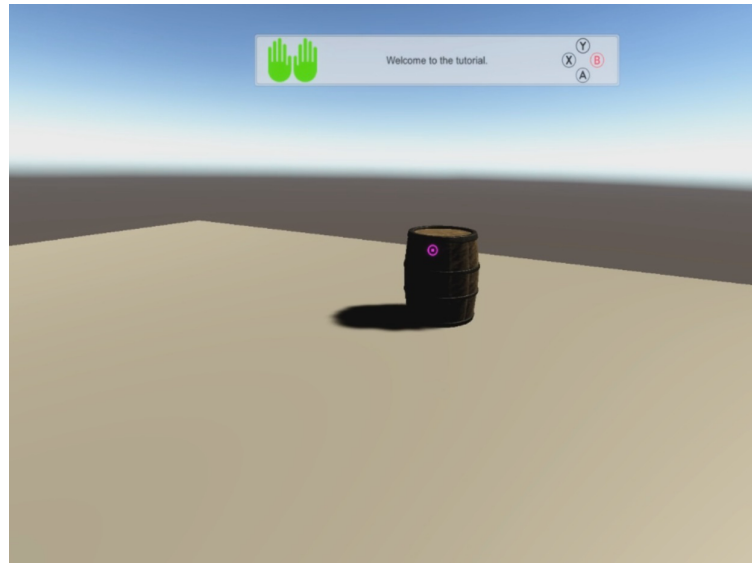


FIGURE 3.2: Tutorial Level. This simple tutorial level introduced players to the controls, user interface, and interaction mechanics of the game.

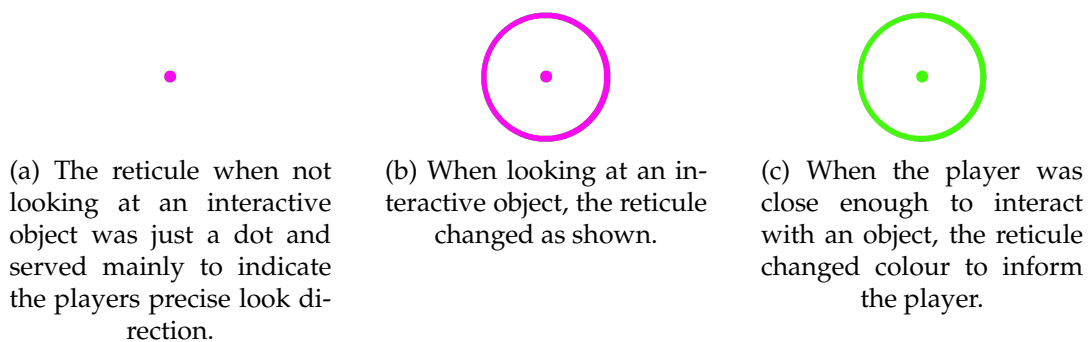


FIGURE 3.3: Reticule. This reticule was displayed in the center of the players visual field and provided information about look direction and objects that were looked at.

Movement was controlled using the left analogue stick of the controller (see Section 3.1.3).

An interactive object (a barrel) was placed in the tutorial level so that players could learn how to interact with objects (see Figure 3.2). The reticule in the centre of the screen gave players visual feedback about whether the object they were looking at was interactive or not, and whether they were close enough to interact with it if it was. Figure 3.3 shows and explains the reticule. When close enough to interact with an object, players could press the 'A' button on the controller to do so. The effect of the interaction depended on the object – some objects, like the barrel, could be picked up and dropped, other objects could be activated/deactivated, and others could be collected.

Players received their objectives in the game through a message window (Figure 3.4A and 3.4B). Whether or not this window was showing could be toggled by pressing the 'B' button on the controller. The presence or absence of an icon (the green hands, shown in Figure 3.4B) to the left of the message itself informed players if the instruction pertained to the game world or to the real world. If the icon was not shown, the instruction was referring to something the player had to do in the



FIGURE 3.4: Message window. All in-game instructions were given to players using text messages. Instructions pertaining to the game world appeared as in A above. Instructions pertaining to the real world were accompanied by the green hands icon as in B above.

game world (e.g. go to a particular location). If the icon was shown, it meant that the instruction was referring to something the player had to do in the real world (e.g. interact with a particular object on the desk in front of them).

For participants in the control condition, the tutorial explained that when they were instructed to interact with an object in the real world, they would have to lift the HMD in order to do so (the AV system was disabled). For participants in the experimental condition, the tutorial explained how to use the AV system to interact with real-world objects and gave them time to practise using it.

Once players had completed the tutorial, they had all of the information they needed to be able to play the game.

### Beginning the game

<sup>8</sup>The game started with the player on the edge of the island. The surrounding water and cliffs meant that they would be sure to find the path that lead to the various game objectives. The starting location and path are visible in Figure 3.1.

The game objectives were placed far apart and the player's move speed was fairly slow. This ensured that players would take time to get between objectives during which they were focused solely on the virtual world and could develop a sense of presence, and would not feel that they were being constantly interrupted by the real world. The slow movement speed was also a precaution taken to reduce the possibility of motion sickness occurring.

As the game began, players were instructed to "Find The Temple of the Horse" (Figure 3.5) via the message window presented in Section 3.4. To let players know that they had found the correct objective, and to act as an attractor, the temple had a large statue of a horse outside it that was easily visible from the path that players would have to take.

### Puzzle one - temple of the horse

<sup>9</sup>After finding the temple of the horse and getting close enough to it, players were instructed to enter it. Within the temple was an altar with four interactive flames and a locked door (Figure 3.6). Interacting with these flames changed their colour. To unlock the door, players had to change the colour of the flames to the correct combination.

<sup>8</sup>Full video play-through: <https://www.youtube.com/watch?v=Xatc1fsQQmU>.

<sup>9</sup>The first puzzle starts here: <https://youtu.be/Xatc1fsQQmU?t=1m59s>.



FIGURE 3.5: The Temple of the Horse. This temple contained the first puzzle in the game and was located at label 'A' in Figure 3.1.



FIGURE 3.6: The first puzzle that required the player to look for a clue in the real world. Each flame cycled through three colours when the player interacted with it. To unlock the door and progress, the player had to set the colours of the flames to the correct combination.



FIGURE 3.7: Clue for the first puzzle, indicating the required combination of flame colours. This clue was placed under a cup on the desk in front of the player.



FIGURE 3.8: Temple of the Lion. This temple contained the second puzzle in the game and was located at label 'B' in Figure 3.1.

Upon finding and entering this temple, players received a message telling them to “Look under the cup” on their desk, where a piece of paper with the correct colour combination for the flames had been placed (Figure 3.7). Players had to locate the cup on the desk in front of them and read the clue to know which combination was correct.

After solving the puzzle, the door unlocked and swung open, revealing a key which players were instructed to collect. Players were then instructed to “Find The Temple of the Lion”. As players exited the temple of the horse, a large lion statue was visible in the distance (located at 'B' in Figure 3.1), that would attract them to the next objective.

### Puzzle two - temple of the lion

<sup>10</sup>Once players had located the temple of the lion (Figure 3.8), they had to enter it using the key retrieved from the previous puzzle. Within this temple were a locked door and six buttons labelled 1-6 in Roman numerals (Figure 3.9). To unlock the door, players had to press these six buttons in the correct order.

When players entered this temple, they were instructed to “Read the message in the bottle”. Players then had to locate a small glass bottle on the desk in front of them and extract a piece of paper indicating the correct sequence of button presses (Figure 3.10).

After solving the puzzle, the door unlocked and the player was instructed to collect a small golden statue of a dragon before being told to “Find the Temple of the Dragon”. Upon exiting the temple of the lion, a large statue of a dragon at 'C' in Figure 3.1 was visible to the player, once again informing them of the direction they must go.

<sup>10</sup>The second puzzle starts here: <https://youtu.be/Xatc1fsQQmU?t=4m14s>.



FIGURE 3.9: The second puzzle. Players had to press the six labelled buttons in the correct order to progress.



FIGURE 3.10: Clue for the second puzzle, indicating the required order of button presses. This clue was placed inside a small glass bottle on the desk in front of the player.

### Puzzle three - temple of the dragon

<sup>11</sup>The temple of the dragon (Figure 3.11) contained the final puzzle of the game. The temple contained six tables, each with a unique pattern of candles placed on it, and an altar with two small statues: a lion and a horse. When players entered the temple, the small dragon statue they had previously collected was placed on the altar alongside the others (Figure 3.12). To solve the puzzle, these statues had to be picked up and placed on the correct tables.

Players were instructed to “Look inside the envelope” on their desk which contained three cards indicating which statue should be placed on which table (Figure 3.13). After locating and opening the envelope, players looked at the cards and placed the statues. When they had successfully done this, the game came to an end.

#### 3.1.3 AV System

Due to the promising results obtained in previous work with respect to preserving presence and facilitating real-world interaction (See Section 2.6), the AV system implemented was a form of inferred minimal blending AV. The system was implemented in the Unity3D game engine, and comprised several pieces of hardware.

#### Hardware

The Head-Mounted Display used was the Oculus Rift CV1. It was important to use a highly immersive display, as immersion is a prerequisite for the sensation of presence. The Oculus Rift CV1, compared to earlier versions used in previous work

<sup>11</sup>The third puzzle starts here: <https://youtu.be/Xatc1fsQQmU?t=7m12s>.



FIGURE 3.11: The Temple of the Dragon. This temple contained the third and final puzzle in the game and was located at label 'C' in Figure 3.1.

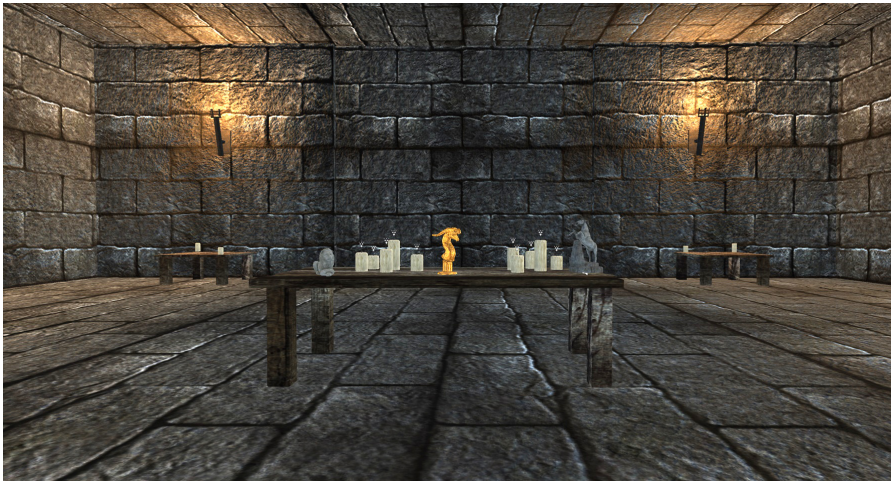


FIGURE 3.12: The final puzzle. The three small statues pictured had to be placed on the correct tables around the sides of the room. Number and placement of candles differentiated the tables.

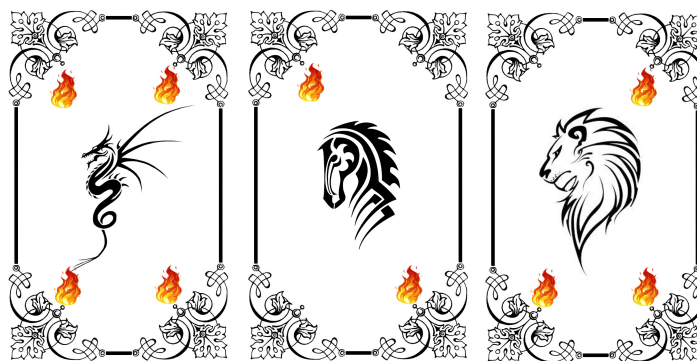


FIGURE 3.13: Clues for the third puzzle, indicating the table on which each statue should be placed. These three cards were placed in a sealed envelope on the desk in front of the player.

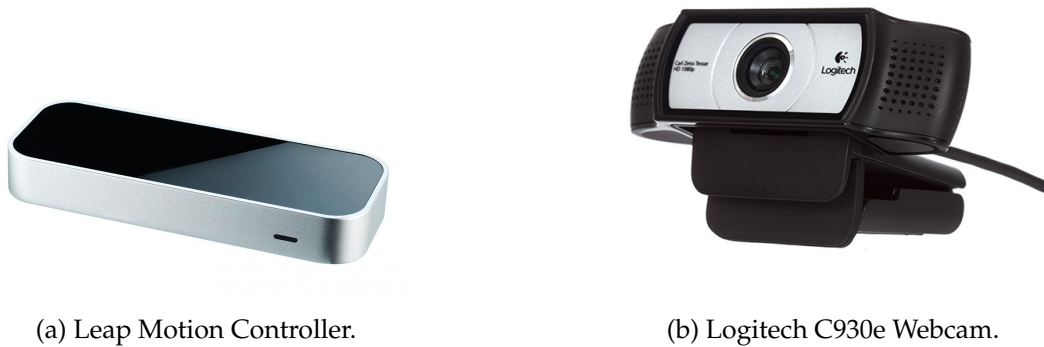


FIGURE 3.14: Hardware for Hand Tracking and Video Pass-through. The Leap Motion controller (A) is used for detection and tracking of a user's hands. The webcam (B) is used to pass video footage of the real world into the virtual world.

(McGill et al., 2015; Budhiraja et al., 2015), scores higher in terms of both *vividness* and *surrounding* (see Section 2.2.3), and is, at the time of writing, at the forefront of immersive display technology. The HTC Vive, a competitor to the Oculus Rift, would have been a viable alternative. The Rift was ultimately chosen due to better availability and the ease of development it offered when this research commenced.

In order for the hands of the user to be tracked (a requirement for both minimal blending and inferred transitions between VR and AV), a Leap Motion controller<sup>12</sup> was used (Figure 3.14a). This small device uses twin infrared cameras to provide low-latency tracking of a user's hands in three dimensions. The use of this device for hand tracking meant that no modifications to a potential user (e.g. wearing tracking markers) or their environment (e.g. green-screening) needed to be made. In addition, the device is well-suited to use with an HMD, as its small size allows it to be attached to the front of the HMD.

Video see-through for the AV system was achieved using a Logitech C930e, a high resolution colour webcam (Figure 3.14b). Initially, we had used the twin cameras of the Leap Motion controller for video see-through as well as hand tracking. While this strategy had the major benefit of allowing the view of the real world to have stereoscopic depth, which made its use for gross motor tasks feel slightly more natural, the images provided were black and white and very low resolution, severely limiting the usefulness of the AV system. For example, the screen of a mobile phone would appear black, and keys on a keyboard could not be discerned. The high-res colour webcam provided a better looking and more useful view into the real world.

Figures 3.15 and 3.16 show the components assembled for AV use. The webcam and Leap Motion are attached to the front of the HMD as close to the eye line as possible to maintain consistency between visual and proprioceptive cues, while leaving as many of the HMDs tracking points uncovered as possible. As Figure 3.16 shows, however, there is some distance between the eye position, hand tracker, and webcam. The actual position of the user's hands is therefore slightly offset from their apparent position. We correct for this as much as possible in software, but a slight discrepancy between optical flow and proprioception remains. The size of the tracking frustum of the Leap combined with the relatively small vertical FOV offered by

<sup>12</sup><https://www.leapmotion.com/>



FIGURE 3.15: AV System Hardware Setup: Leap Motion controller and webcam attached to the front of an HMD (Oculus Rift CV1).

the HMD meant that the user's hands had to be near the centre of the user's natural FOV to be tracked and seen.

Other hardware included a desktop computer that met the system specifications required to power the HMD and other peripherals<sup>13</sup>, as well as a standard Xbox One controller (Figure 3.17) to provide input in the game and tutorial.

### AV software

The Unity3D game engine<sup>14</sup> was used to implement the AV system. This engine was chosen because of the included developer tools necessary for using both the Oculus Rift and the Leap Motion controller. Furthermore, this same engine could also be used to create the virtual environment and implement the game in which the AV system would be used.

Using the Leap Motion controller, Unity3D allows the positions of a user's hands in the real world to be converted into coordinates in the virtual environment which, relative to the virtual viewpoint, closely match the real world. When a user's hands enter the tracking frustum of the Leap Motion controller, they are detected, and we gain access to their coordinates in the virtual environment. We will refer to these coordinates as *hand-L* and *hand-R* for the left and right hands respectively.

Unity3D allows the live feed of a webcam to be applied to any object as a texture. This texture, which we will refer to as the *webcamtexture*, is updated at the frame rate supported by the particular webcam used (30fps in our case). Due to a limitation of Unity, the *webcamtexture* had to be rendered at a lower resolution than the maximum offered by the webcam we used. Using the full resolution (1080p) resulted in update rates to the webcam texture of less than 1 frame per second, which is obviously totally unusable for AV, which requires real-time update rates. Lowering the requested resolution solves this issue at the cost of a minor reduction in visual fidelity, particularly for fine details, of the world as seen by the webcam.

The combined use of the *webcamtexture* and hand coordinates allowed inferred minimal blending AV to be achieved. Our implementation worked as follows:

<sup>13</sup>Please see Appendix B for the full system specifications.

<sup>14</sup><https://unity3d.com>

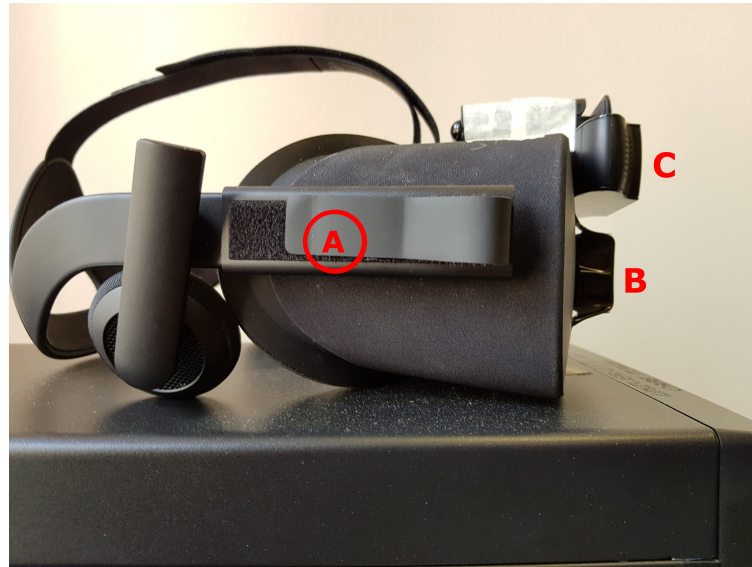


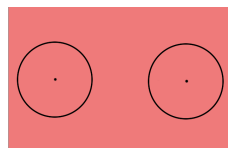
FIGURE 3.16: AV System Hardware Setup. Side view. 'A' marks the approximate position of the eye while the system is in use. The Leap Motion Controller is marked 'B'. The webcam is marked 'C'.



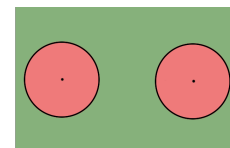
FIGURE 3.17: Xbox One Controller. This controller was used to play the game. The left analogue stick was used for movement. The 'A' button was used for interaction. The 'B' button was used to toggle the display of the current objective in the UI.



(a) Hands are detected. A plane taking up the entire FOV is created, its texture supplied by the live feed of the webcam, and it is rendered.



(b) Transparent disks are created, centred at the game world coordinates corresponding to the hand positions. As they are transparent and there is no other game geometry, their texture information is taken from the webcam texture behind them, and they are rendered.



(c) All other game geometry is rendered, occluding the webcam texture. The depth mask shader applied to the disks prevents interposing geometry from being drawn, so the view through the disks onto the webcam texture is not occluded.

FIGURE 3.18: The three stages of our AV implementation. Red indicates areas of the display showing the real world via the webcam. Green indicates game geometry. Dots indicate hand position.

1. When a hand, or hands, are detected, the webcamtexture is applied to a plane that takes up the entire field of view. This plane is textured and rendered before any other game geometry. If there were no game geometry, one would just be viewing the world as seen by the webcam. (Figure 3.18a).
2. Two small disks are created, centred at *hand-L* and *hand-R*, between the view-point and the webcamtexture. These disks, or *viewports* use a depth mask shader which has two effects. The first effect is to make the viewports transparent, meaning they take their texture information from whatever is behind them (at this point, the webcamtexture). The second effect of the shader is to prevent any game geometry later in the render queue from being drawn behind the viewports. Effectively, this means that the “view” through these transparent windows onto the webcamtexture can never be blocked by interposing game geometry. (Figure 3.18b).
3. Finally, all other game geometry is rendered. This geometry occludes the webcamtexture itself, so that it can no longer be seen, but the shader on the viewports mentioned in step two prevents geometry between them and the webcamtexture from being drawn, turning them into “windows” into the real world. (Figure 3.18c).

Figure 3.19 shows the system in use. The AV system gave players enough real-world stimuli to perform interactions in the real world, while still allowing stimuli from the VE to reach the player.

### 3.1.4 Input Capture

Typing on a standard keyboard is a “rich form of interaction that requires a high-bandwidth feedback loop” (McGill et al., 2015). This means that it is a task that is relatively difficult to execute, compared to something like the use of a controller, when one or more sensory modalities are removed. Because of this difficulty, its ubiquity as a form of input, and the ease with which it can be precisely measured, typing was chosen as the task that we used to measure input performance. To evaluate typing performance, several versions of a typing test were created. Each of these programs captured user input during use, which was later analysed using a separate program. While we would have liked to incorporate typing directly into the game to simulate a scenario such as in-game chat, there was no practical way to do this while still gathering enough typing data to facilitate its accurate measurement. The software created to measure typing performance was therefore kept separate from the game.

Each test began with the user being prompted to press space to start. After a five second countdown a window appeared, displaying text (the prescribed text) that the user had to type. The prescribed text was randomly generated from a list of commonly used English words<sup>15</sup> whenever the test was run. The user’s input (transcribed text) appeared in an input field below the prescribed text. When the user reached the end of each line of prescribed text, they had to hit enter in order for the next line of prescribed text to be displayed and the input field to be cleared. This was repeated for one minute in each version of the test, after which the gathered input was saved in a text file and the application automatically quit.

Each of the typing tests recorded the following information:

---

<sup>15</sup><https://gist.github.com/deekayen/4148741>



FIGURE 3.19: AV System in use. This is a screenshot from the game described in Section 3.1.2. The player has received a message in the message window telling them to 'Look under the cup'. The green icon to the left of this message indicated to the player that the instruction pertained to the real world. They have reached forward and their hand has been detected by our AV system, allowing the real world to be shown via the webcam in an area around the player's hand position. Using this window into the real world, they have managed to locate the cup.

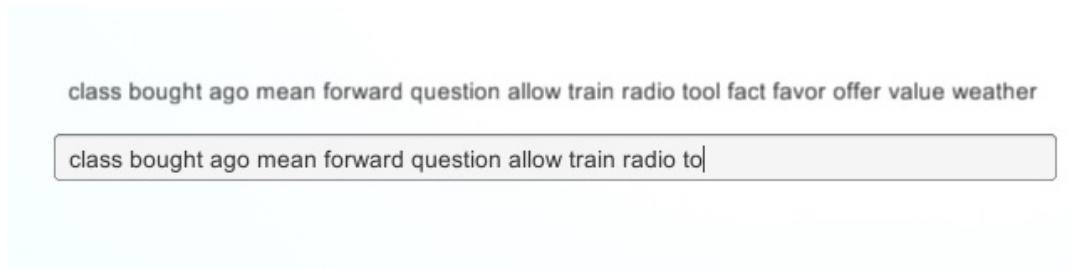


FIGURE 3.20: Baseline typing test. Prescribed text (to be typed by user) is shown at the top; transcribed text is shown in the input field below.

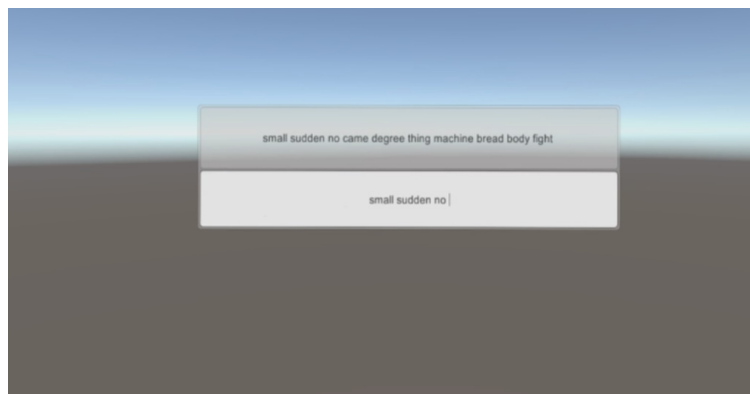


FIGURE 3.21: VR typing test. Horizon is shown in the VR version to provide visual feedback for head movement. The box containing the prescribed and transcribed text is always in the direction in which the user is looking.

- Prescribed Text - The randomly generated string of words that the user had to type.
- Raw Input - String of all input given by the user, including corrective keystrokes such as backspace.
- Transcribed Text - Text submitted by the user.
- Time to First Key Press (TTFKP) - Time taken to start typing.

### Baseline typing test

The baseline typing test was a traditional two-dimensional desktop application as shown in Figure 3.20. Results from this test were to serve as a point of comparison to the other tests for each participant.

### VR typing test

This version of the typing test (Figure 3.21) simulated typing while using an HMD (visually cut off from reality).

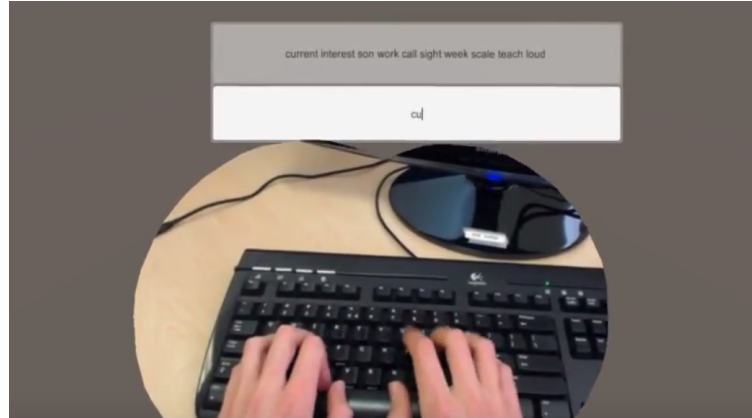


FIGURE 3.22: AV typing test. Same as the VR typing test in Figure 3.21, but with the use of the AV system enabled, allowing the user to see their hands and the keyboard while typing.

### AV typing test

This version of the typing test was the same as the VR typing test, except that the AV system was enabled, allowing users to see their hands and the keyboard during the test. This is shown in Figure 3.22.

## 3.2 Experimental Design

In order to investigate our research questions and confirm or deny our hypotheses, a quantitative user experiment was designed. The experiment followed a between-group design, with each participant being subject to one condition, experimental or control.

This design was chosen to avoid contamination due to carryover effects. The between-group design allowed for the required data to be collected without the need for a lengthy washout period (which would have been required to avoid fatigue and learning effects in a within-subjects alternative) and potential participant attrition between treatments. A between-group design was therefore more practical for our purposes.

Although between-group designs can often require more participants than within-subjects designs to reveal significant effects, a pilot study and subsequent power analysis suggested that the number of participants we would require was reasonable (see Section 3.4). The sensitivity of between-group design to expectancy effects and extraneous environmental variables was mitigated as far as possible. To avoid assignment bias, participants were randomly assigned to either the control or experimental condition.

The design was single-blind (via omission): participants were not given any information regarding measures or hypotheses, and were not aware of the differences between conditions. While preferable, a double-blind design was not possible as there was no way to conceal the condition from the experimenter due to the hardware requirements of the experiment. Therefore, experimenter bias was mitigated by limiting contact with participants before and during the experiment, with pertinent information disseminated to each participant via identical documents and scripted dialogue. To reduce the risk of any extraneous environmental factors, all experiments took place under laboratory conditions: the same venue was used for all

experiments; each participant used exactly the same hardware; and care was taken to ensure that there were no external distractions. The experimenter remained present for the duration of each experiment as a precaution<sup>16</sup>, but did not interfere in any way. Notes were taken in instances where anything unusual occurred that may have affected the gathered data.

The two conditions were VR (control condition), and AV (experimental condition). In both conditions, participants used a Head-Mounted Display (HMD) while playing the puzzle game described in Section 3.1.1. In the control version of the game, all aspects of the AV system were disabled, and in the control version of the tutorial, participants were told that instructions pertaining to the real world would require them to remove the HMD. In the experimental version of the game, the AV system was enabled, and in the corresponding tutorial, players were taught how to use the AV system.

The independent variable (IV) was whether or not the AV system was active. The dependent variables (DVs) were presence, workload, and typing performance.

### 3.3 Measures

Presence and workload were measured using subjective self-report questionnaires. Input performance was measured objectively.

#### 3.3.1 Presence

In order to measure presence, the Igroup Presence Questionnaire (IPQ) (Section 2.4.3) was used. A pen and paper version of this questionnaire was filled out by participants, and is included in Appendix C.

Our definition of presence (Section 2.2.1) identifies it as a subjective experience. For this reason, and to avoid current issues regarding its objective measurement (Section 2.4.3), the subjective measurement of presence, through self-report questionnaires, was most suitable.

The IPQ has been found to be both reliable and valid (Schubert, Friedmann, and Regenbrecht, 2001) and to demonstrate sensitivity (Van Baren and IJsselsteijn, 2004). The IPQ consists of three subscales (Spatial Presence, Involvement, and Experienced Realism) as well as one general item, the “sense of being there”. Each subscale is scored individually by calculating the average score of each item associated with it, except for the single general item, which is scored directly.

#### 3.3.2 Workload

The NASA-TLX questionnaire (see Section 2.4.3) was used to measure workload. This can be found in Appendix D.

Within the context of games, a player’s subjective conception of workload is more important than their actual performance. As such, and because of the problems with psychophysiological workload measurement (Section 2.4.3), a subjective measure of workload was suitable for our purposes.

The NASA-TLX has been found to be reliable, valid, and sensitive (Hart and Staveland, 1988), and measures workload according to six subscales: Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration.

---

<sup>16</sup>In case the participant required aid due to VR-induced discomfort.

Participants assign a weight for each subscale to indicate its relative perceived importance, as well as an absolute score to indicate its perceived magnitude. A total workload score can be calculated by taking the average weighted score for all subscales (see Hart and Staveland (1986)), or the weighted scores themselves can be used to examine each factor individually (Hart, 2006).

### 3.3.3 Input Performance

Text files recorded during the typing tests were processed using a Java application<sup>17</sup>. The application used the recorded prescribed, transcribed, and raw input strings to find the number of correct keystrokes, corrected errors, and uncorrected errors (see Section 2.4.3 for an explanation of how these are identified). These tallies were used to calculate the overall accuracy as a percentage using the method proposed by Soukoreff and MacKenzie (2003). The total number of characters in the transcribed text string, along with the number of uncorrected errors committed, were used to calculate the net typing speed in words per minute (WPM) using the following formula:

$$GTS = W/(t)$$

$$NetTypingSpeed = GTS - (INF/(t))$$

Where *GTS* is Gross Typing Speed, *W* is the number of words (where a word is defined as a sequence of 5 consecutive characters), *t* is time in seconds (60 in our case), and *INF* is the number of uncorrected errors.

## 3.4 Participants

Pseudo-random convenience sampling was used to recruit participants. Computer Science undergraduate and fourth year students were the easiest to access and were canvassed face to face. Ethics clearance and permission to access students for experimental purposes were obtained from the University of Cape Town Faculty of Science prior to recruitment (Appendix E).

Stereoscopic 3D in games has been found to increase immersion and presence (Schild, LaViola, and Masuch, 2012). For this reason, people with visual impairments precluding them from perceiving stereoscopic 3D were excluded from the study, as the level of presence they could have experienced may have been affected. Touch typists were not excluded from the study, as time to first key press, one of the metrics involved in text-entry analysis, was likely to be affected by the absence or presence of the AV system, no matter the proficiency of the typist. Whether or not candidates were able to touch type was recorded nevertheless, in case their inclusion affected the gathered data. We anticipated that few, if any, of our participants would have had prior experience with VR or mixed reality. We therefore restricted our participants to people who had at least some experience with computers and computer games, as it was possible that people with no experience would be overwhelmed by the plethora of new technology and experiences in the experiment, which may have negatively impacted their experience and skewed the data. Furthermore, although HMDs are immersive enough that controlled attention contributes less to the formation of presence than in non-immersive media, selecting participants who

<sup>17</sup>This application can be found here: <https://github.com/jacobHclarkson>.

were gamers meant that games were likely to be part of their domain-specific interests, therefore increasing the chance that they would pay controlled attention and become present (see Section 2.2.2 for more detail).

A pilot test was conducted before the experiment commenced. This test was conducted to ensure that the system was functioning as intended and to ascertain an estimate of the final sample size that would be required to reveal significant effects. Four participants took part in the pilot test, with two participants in each condition. Power analysis (with a confidence of 0.8 and a significance criterion of .05) indicated that a minimum sample size of 18 would be required to produce statistically significant results.

A total of 24 participants were recruited from the University of Cape Town student population. Most of the participants were Computer Science students between the ages of 19 and 25. Four of the participants were female, and the remaining 20 were male. Seven participants were touch typists, and all participants had prior gaming experience.

Participants, once recruited, were randomly allocated to either the control or experimental group in order to mitigate assignment bias.

## 3.5 Procedure

Each participant took part in one experimental session where they were exposed to one condition (AV or VR). The procedure for each experimental session was divided into four phases: pre-experiment, experiment, post-experiment, and processing.

### 3.5.1 Pre-experimental Phase

Before each participant arrived, the real-world clues for the puzzle game were placed on the desk the participant would be using. They were placed in the same locations for each experiment. The arrangement of clues is shown in Figure 3.23.

As soon as they arrived, the participant was given a printed informed consent document (Appendix F). This document informed participants that they would be playing a mixed reality game while wearing an HMD, filling out some questionnaires, and that they would be remunerated for their efforts. They were also warned about the possibility of experiencing nausea while wearing the HMD, and told that they could withdraw from the experiment at any time for any reason.

Once the participant had read through this document, they were required to sign to acknowledge their informed consent before continuing. If they were unwilling to consent to further participation in the experiment, they were allowed to leave. Otherwise, the experimenter assisted the participant in putting on the HMD correctly and navigated the participant to the lens spacing calibration instructions in the Oculus menu<sup>18</sup>. Participants then followed these instructions and were asked to notify the experimenter once they had done so.

The experimenter then launched the tutorial level to allow the participant to become acquainted with the interface and controls that would be required in the game. Participants were verbally guided through this tutorial according to a script (Appendix G). There were two versions of the tutorial level and script, one for the control condition and one for the experimental condition. Instructions pertaining to the usage of the AV system were excluded from the script used in the control condition.

---

<sup>18</sup>This was done to ensure that the image was as sharp as possible for each participant.

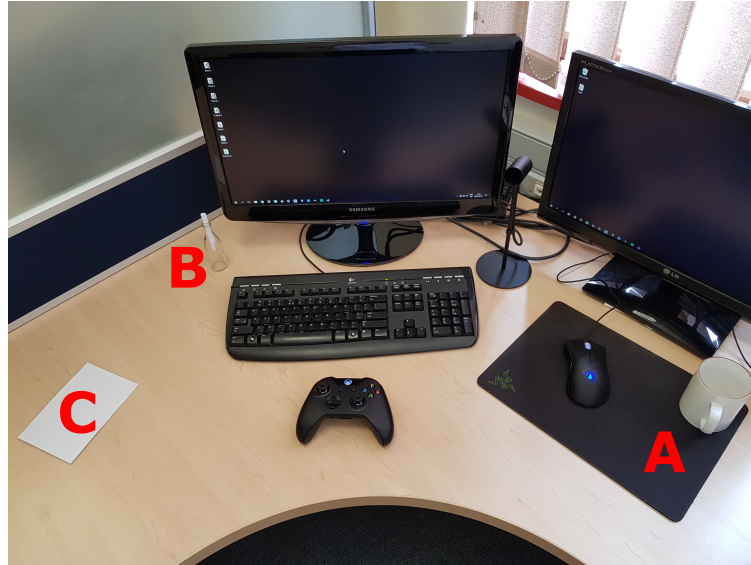


FIGURE 3.23: Clue locations. Before each experiment, the clues were concealed and arranged on the desk at which the experiment took place. The first clue was placed under the cup at 'A'. The second clue was rolled up and placed inside the small glass bottle at 'B'. The final clues were placed in sealed envelope at 'C'. Only a single monitor was used in the experiment, and only for the desktop version of the typing test.

Once participants had completed the tutorial, the experimenter started the game that would be played during the experimental phase.

### 3.5.2 Experimental Phase

During the experimental phase, the participants played our immersive 3D game while wearing an HMD. During the course of this game, they were required to perform several tasks, including interacting with real-world objects. See Section 3.1.1 for details about the game.

#### Game play and subjective measures

The participant played through the puzzle game. Immediately after completing the game, the IPQ was administered to measure the sense of presence that the participant had felt in the virtual environment. Subsequently, the NASA-TLX was administered to ascertain the workload experienced by the participant during the game.

#### Typing test

After the questionnaires were filled in, the participant completed two typing tests. The first test was a baseline using a conventional desktop computer monitor (Section 3.1.4) and was administered in both the control and experimental conditions. The second typing test (treatment) was done while wearing the HMD. The VR version of the treatment typing test (Section 3.1.4) was given to those in the control condition, and the AV version (Section 3.1.4) to those in the experimental condition.

We did not feel that it was necessary to interleave the baseline and treatment typing tests within the groups, as a single one-minute typing test was very unlikely

to offer enough of an opportunity for either learning or fatigue effects to occur. All participants therefore completed the baseline test before the treatment test. Scores on the baseline test served as a point of comparison to the treatment test scores, but also ensured that the control and experimental groups had, on average, similar levels of proficiency to begin with. Without the baseline, it would not have been possible to tell if one group happened by chance to contain much better typists, therefore skewing the results.

### 3.5.3 Post-experimental Phase

Finally, the participant's typing speed in non-VR conditions was captured using the baseline typing test. The e-mail address and student number of the participant were then recorded<sup>19</sup>, and they were remunerated for their participation. Finally, the participant asked any questions they had about the experiment and were allowed to leave.

### 3.5.4 Processing Phase

Once the participant had left, data gathered during the session via the apparatus were backed up, and their responses to questionnaires were filed.

After all experiments were complete, scores for all questionnaires were calculated, and all recorded input data was analysed.

## 3.6 Summary

In this chapter, we reintroduced our research questions before presenting the apparatus and methods we used to investigate them. Based on the previous work discussed in Chapter 2, we created a virtual environment capable of inducing a sense of presence. This environment served as the setting for a mixed reality game that would prompt players to interact periodically with the real world. Experimental participants used the AV system that we developed to carry out these interactions, while control participants had to peep. This between-group experiment allowed us to directly compare VR to AV in terms of aspects of the user experience (presence and workload). We also developed three versions of a typing test so that we could compare typing performance under VR, AV, and desktop (baseline) conditions. In the following chapter, we present the results we obtained using these methods.

---

<sup>19</sup>This information was recorded only for remuneration purposes, and not associated with any data yielded by the participant.



## Chapter 4

# Results

The aim of this research was to investigate the effects of Augmented Virtuality (AV) on presence, workload, and input performance compared to pure Virtual Reality (VR). The results are promising, and show that in a context where users of HMDs will occasionally need to interact with the real world, AV preserves presence in the virtual world and is less frustrating than VR, and ameliorates the hindrance to providing input caused by visual cut-off.

Presence was measured using the Igroup Presence Questionnaire (IPQ). Workload was measured using the NASA Task Load Index (NASA-TLX). Input performance (via a standard keyboard) was assessed by measuring gross typing speed, net typing speed, overall accuracy, and time to first key press (TTFKP). Scores from participants in the experimental group (AV) were compared to scores from participants in the control group (VR) in order to answer our research questions.

In total, 24 participants signed up to participate in the study. However, one participant from the control condition withdrew from the experiment before any data could be collected due to severe motion sickness. All of the remaining 23 participants were students at the University of Cape Town between the ages of 19 and 25 (mostly Computer Science students). Three participants were female and 20 were male. All participants had played 3D games before<sup>1</sup>.

Many statistical tests make assumptions about the underlying distribution of the data. If these assumptions are not met, the results of the tests can be skewed. For this reason, all data were first tested for parametricity using Shapiro-Wilk tests (Shapiro and Wilk, 1965) as this allowed us to determine which statistical tests would be most appropriate to use. Parametric data were then tested for homogeneity of variance (homoscedacity) with Levene's tests (Levene, 1960) before being subjected to either an independent-samples t-test (which assumes homoscedastic data) or a Welch's t-test (Welch, 1947) (designed for heteroscedastic data). Non-parametric data were analysed using Mann-Whitney U tests (Mann and Whitney, 1947). The t-test and Mann-Whitney U test were used, for each factor of each measure, to see if there was a significant difference between the scores of the AV and VR groups by confirming or denying the null hypothesis (that there was no relationship between group and score for that factor). Effect size (where significant differences were found) for t-tests was calculated using Cohen's d (Cohen, 1988). Effect size for Mann-Whitney U tests ( $r$ ) was calculated according to the following formula (Fritz, Morris, and Richler, 2012):

$$r = \frac{Z}{\sqrt{N}}$$

---

<sup>1</sup>Participants were simply asked whether or not they had experience playing 3D games, as we had no valid reason to assume that a level of experience beyond this binary would have an effect on this particular experiment.

where  $r$  is the effect size,  $Z$  is the z-score calculated by the Mann-Whitney U test, and  $N$  is the sample size. Effect sizes were interpreted as shown in Table 4.1.

TABLE 4.1: Interpretation of Effect Sizes (Cohen, 1988; Sawilowsky, 2009; Fritz, Morris, and Richler, 2012).

Effect Size	Cohen's $d$	$r$
Very Small	0.01	-
Small	0.20	0.10
Moderate	0.50	0.30
Large	0.80	0.50
Very Large	1.20	-

In this chapter, the results of these statistical analyses of the gathered data for presence, workload, and input performance are presented. Results indicate that the experiment was a success, revealing several significant differences in our chosen measures between the control and experimental conditions. We will now present the results obtained from each measure in detail.

## 4.1 Presence

Presence was measured using the Igroup Presence Questionnaire (IPQ) (Schubert, Friedmann, and Regenbrecht, 2001). The IPQ consists of three subscales and one general item:

**PRES** General Presence: the “sense of being there”.

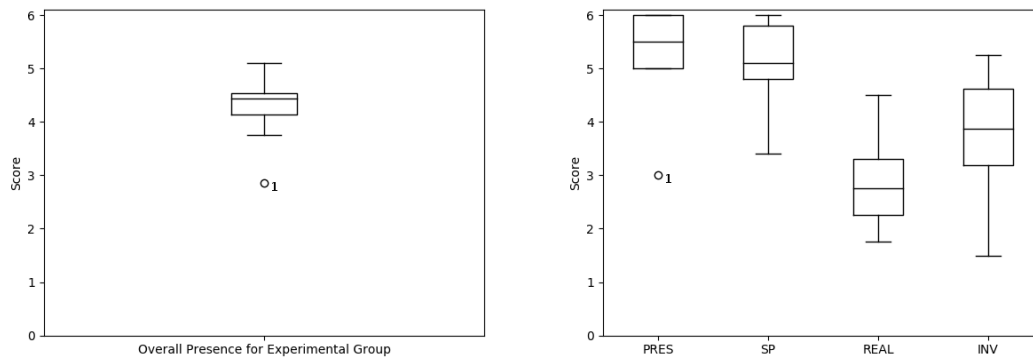
**SP** Spatial Presence: the sense of being physically present in the virtual environment.

**INV** Involvement: the attention devoted to the virtual environment and the involvement experienced.

**REAL** Realism: the subjective experience of realism in the virtual environment.

The IPQ measures each subscale using seven-point likert-type items. SP, INV, and REAL were assessed with multiple items, and the scores presented for them are the average scores of these items. PRES was assessed with only one item, and is presented directly.

To get an overview of the presence data, an overall presence score was calculated for each participant by taking the mean of their scores for each of the presence factors. A Tukey's test (Tukey, 1977) was then performed on these overall presence scores for each group to identify any outliers that might be skewing the data. One participant ( $O_1$ ) from the experimental condition was identified as a significant outlier (see Figure 4.1a). The interquartile range of average presence scores for experimental participants was 0.6, and the value of the first quartile was 4.07, resulting in a lower fence of 3.17.  $O_1$ 's average presence score of 2.85 fell below this lower fence. Subsequent Tukey's tests were conducted for each presence factor for the experimental group (Figure 4.1b). The interquartile range for PRES scores was 1.00, and the value for the first quartile was 5.00, resulting in a lower fence of 3.5. At 3.00,  $O_1$ 's SP score fell below the lower fence for PRES. Furthermore,  $O_1$  had the lowest scores in the experimental group for SP and REAL. An examination of the notes taken during the experiments showed that  $O_1$  had asked, after completing the experiment, “Is



(a) Outlier in overall presence scores (average of presence factors) for the experimental group. Higher is better.

(b) Box plots per presence factor for the experimental group.  $O_1$  is a significant outlier in PRES, and has the lowest scores for SP and REAL. Higher is better.

FIGURE 4.1: Box plots showing a significant outlier ( $O_1$ ) in presence data from the experimental group.  $O_1$  in subfigure A and B is the same participant.

TABLE 4.2: Descriptive statistics for responses to the IPQ for control (VR) and experimental (AV) conditions: mean (M), median (Mdn), variance (Var), standard deviation (SD), and sample size (N).

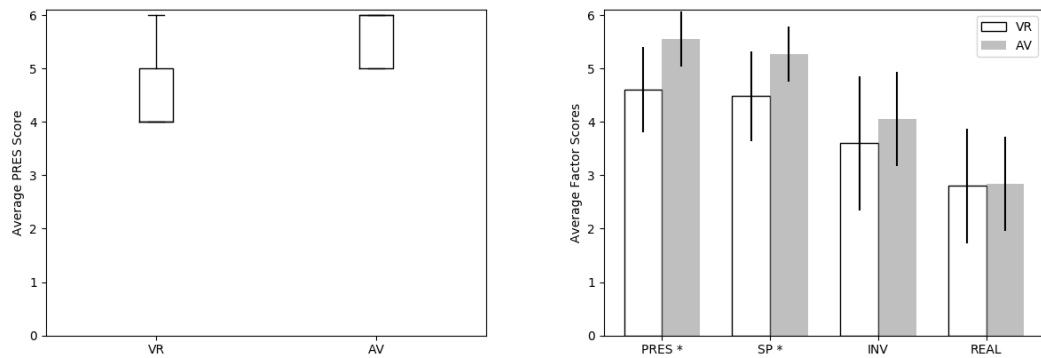
Condition	Subscale	M	Mdn	Var	SD	N
VR	PRES	4.60	4.00	0.71	0.80	10
	SP	4.48	4.60	0.79	0.84	10
	INV	3.60	4.13	1.77	1.26	10
	REAL	2.80	2.88	1.30	1.08	10
AV	PRES	5.55	6.00	0.27	0.52	11
	SP	5.27	5.20	0.27	0.52	11
	INV	4.05	4.25	0.77	0.88	11
	REAL	2.84	2.50	0.78	0.88	11

it meant to be so blurry”. The HMD had shifted on  $O_1$ ’s head during the experiment, severely decreasing the quality of the image being viewed. For this reason, we felt that the cause for this numerical outlier was extraneous, as it resulted from improper usage of the HMD, and the presence data for  $O_1$  were therefore excluded from all subsequent analyses.

Descriptive statistics for the gathered presence data are presented in Table 4.2. Average scores for the four factors measured by the IPQ are depicted in Figure 4.2. A total of 23 participants took part in the experiment. However, responses from two of the participants were removed from the presence data. One participant in the control condition submitted an incomplete response to the IPQ<sup>2</sup>. Data from  $O_1$  were also removed, as discussed previously.

The data for the general item, PRES (Table 4.3, Figure 4.2a), were not normally distributed ( $W = 0.796, p < .01$ ). The median scores for the control and experimental group were 4 and 6, respectively. A Mann-Whitney U test revealed a significant

<sup>2</sup>The participant accidentally skipped one page of the questionnaire, and the error was unfortunately not detected by the experimenter. The page in question contained questions pertaining to each subscale, and the remainder of the response was therefore unusable.



(a) Box plots for PRES scores of control and experimental conditions. The difference is significant at  $p < .05$ .

(b) Mean scores for SP, INV, and REAL scores of control and experimental condition.

FIGURE 4.2: Overall scores for IPQ items for each condition. Higher is better. Scores for PRES (A) were not normally distributed, while scores for SP, INV, and REAL (B) were. PRES, while not parametric, was included in (B) to facilitate comparison to the other factors. Factors in (B) marked with \* show a significant difference between AV and VR at  $p < .05$ .

difference in PRES scores between the control and experimental condition ( $U = 21$ ,  $Z = -2.543$ ,  $p < .05$ ), with mean ranks of 7.6 and 14.09, respectively. The effect size was large ( $r = 0.55$ ). Participants in the control group were less likely to feel like they were present in the virtual world than those in the experimental group.

The data for SP (Table 4.4, Figure 4.2b) were normally distributed ( $W = 0.91$ ,  $p > .05$ ). A Levene's test revealed that the data were heteroscedastic ( $F = 1.07$ ,  $p < .05$ ). Therefore, a Welch's t-test was performed, finding a significant difference between SP scores in the control and experimental conditions ( $t(15) = -2.52$ ,  $p < .05$ ), with participants in the experimental group reporting a higher mean level of SP (5.27) than those in the control group (4.48). The effect size was large, bordering on very large ( $d = 1.14$ ). On average, participants in the experimental group experienced significantly more spatial presence than those in the control group.

The data for INV (Table 4.4, Figure 4.2b) were normally distributed ( $W = 0.91$ ,  $p > .05$ ) and homoscedastic ( $F = 2.41$ ,  $p > .05$ ). Although the mean score for the experimental group (4.05) was slightly higher than that of the control group (3.6), an independent-samples t-test revealed that this difference was not significant ( $t(19) = -0.91$ ,  $p > .05$ ).

REAL scores were practically identical across groups (see Figure 4.2b), with a mean score of 2.8 for the control group and 2.84 for the experimental group. As with INV, the data (Table 4.4) were found to be both normally distributed ( $W = 0.96$ ,  $p > .05$ ) and homoscedastic ( $F = 0.36$ ,  $p > .05$ ), with no significant difference between the mean scores ( $t(19) = -0.922$ ,  $p > .05$ ).

## 4.2 Workload

Workload was measured using the NASA Task Load Index (NASA-TLX)(Hart and Staveland, 1988). The NASA-TLX consists of six subscales. Together, these subscales yield an overall measure of workload (WORK):

TABLE 4.3: Mann-Whitney U tests for non-normally distributed presence data, and tests for prerequisite assumptions: Shapiro-Wilk test for normality ( $W, p$ ), sample sizes ( $N$ ), mean ranks, U value, Z value,  $p$  values and effect size ( $r$ ). \* indicates significance at  $p < .05$ .

	Shapiro-Wilk		N(C)	N(E)	Mean Rank (C)	Mann-Whitney		Z	$p$	$r$
	W	$p$				U	Mean Rank (E)			
PRES *	0.796	0.001	10	11	7.6	14.09	21	-2.543	0.016	0.5549281

TABLE 4.4: T-tests for normally distributed presence data, effect sizes, and tests for prerequisite assumptions: Shapiro-Wilk test for normality ( $W, p$ ), Levene's test for homogeneity of variance ( $F, p$ ), t-test ( $t, p$ ), and effect size where significance was found. \* indicates significance at  $p < .05$ .

	Shapiro-Wilk		Levene's Test		T-test		Effect Size (d)
	W	$p$	F	$p$	t	$p$	
SP *	0.911	0.058	1.066	0.315	-2.516	0.021	1.143
INV	0.91	0.055	2.413	0.137	-0.914	0.372	-
REAL	0.956	0.44	0.361	0.555	-0.092	0.927	-

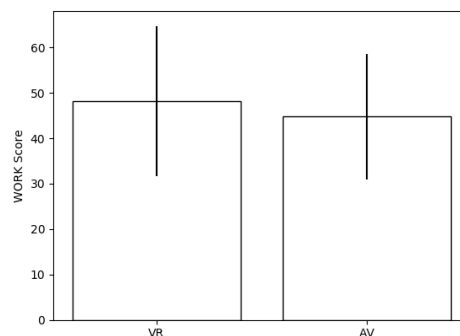


FIGURE 4.3: Mean scores for WORK for control (VR) and experimental (AV) groups. Lower is better. The maximum possible score is 100.

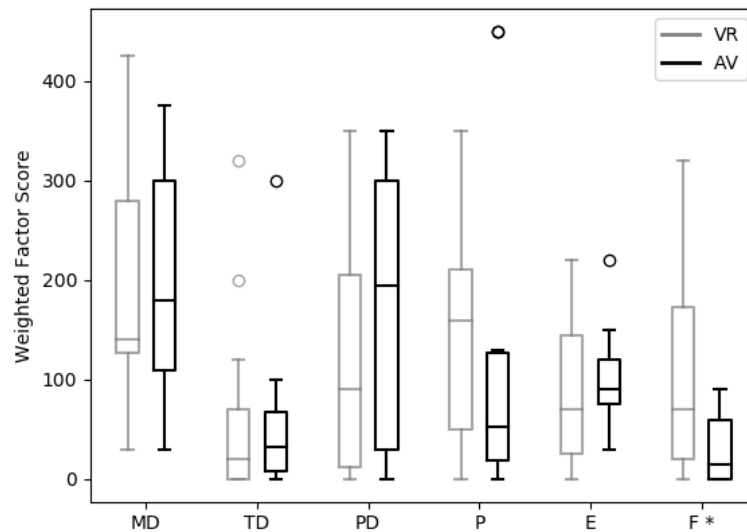


FIGURE 4.4: Weighted scores of factors comprising workload for control and experimental conditions. Lower is better. The maximum possible score is 500. Factors marked \* show a significant difference between AV and VR condition at  $p < .05$ .

**MD** Mental Demand: how much mental and perceptual activity was required.

**TD** Temporal Demand: how much time pressure was felt.

**PD** Physical Demand: how much physical activity was required.

**P** Performance: how successful the participant felt in accomplishing the specified tasks.

**E** Effort: how hard the participant had to work to achieve their level of performance.

**F** Frustration: how insecure, discouraged, stressed, and annoyed the participant felt.

Each subscale of the NASA-TLX is measured using a rating scale combined with a weighting procedure. The final WORK score is the weighted average of the subscales as described by Hart and Staveland (1986).

Descriptive statistics for the gathered workload data can be found in Table 4.5. The overall WORK score per condition is depicted in Figure 4.3, and the scores comprising it are shown in Figure 4.4. The sample sizes for the control and experimental condition were 11 and 12 respectively. While a Tukey's test did reveal several numerical outliers in the workload data (see Figure 4.4), no single person was an outlier on more than one subscale, and an examination of the notes taken during experiments did not reveal any reasons that these data were not valid. Therefore no data were excluded. There were two very extreme outliers for P, as shown in Figure 4.4. There is a small chance that these outliers may be due to the fact that in the NASA-TLX questionnaire, the scale for P is reversed (see Appendix D). This fact

TABLE 4.5: Descriptive statistics for responses to NASA-TLX for control (VR) and experimental (AV) groups: mean (M), median (Mdn), variance (Var), standard deviation (SD), and sample size (N).

Condition	Subscale	M	Mdn	Var	SD	N
VR	MD	191.36	140	14505.46	120.44	11
	TD	64.55	20	11227.27	105.96	11
	PD	123.64	90	17375.46	131.82	11
	P	148.64	160	13335.46	115.48	11
	E	89.09	70	7054.09	83.99	11
	F	105.91	70	10059.09	100.30	11
	Workload	48.21	49.33	274.580	16.57	11
AV	MD	195.42	180	14570.27	120.71	12
	TD	59.58	32.5	6797.54	82.45	12
	PD	168.75	195	19364.21	139.16	12
	P	117.08	52.5	26133.90	161.66	12
	E	100.83	90	2462.88	49.63	12
	F	30	15	1109.09	33.30	12
	Workload	44.78	49	190.92	13.82	12

TABLE 4.6: T-tests for normally distributed workload data, effect sizes, and tests for prerequisite assumptions: Shapiro-Wilk tests for normality ( $W, p$ ), Levene's tests for homogeneity of variance ( $F, p$ ), t-tests ( $t, p$ ), and effect size where significance was found.

	Shapiro-Wilk		Levene's Test		T-test		Effect Size (d)
	W	p	F	p	t	p	
MD	0.936	0.149	0.016	0.901	-0.081	0.837	-
WORK	0.952	0.324	0.019	0.893	-0.32	0.975	-

is even pointed out to would-be experimenters/participants in the original NASA-TLX manual (Hart and Staveland, 1986), suggesting that it may have caused problems before. We too pointed out the reversal to our participants, making any misunderstandings unlikely, and even upon removal of these outliers, the difference in P between the control and experimental groups was still statistically insignificant ( $p = 0.51$ ).

The scores for WORK and MD (Table 4.6) were normally distributed ( $W = 0.95, p > .05; W = 0.94, p > .05$ ) and homoscedastic ( $F = 0.02, p > .05; F = 0.02, p > .05$ ). Independent-sample t-tests revealed that there were no significant differences between the mean scores for the control and experimental conditions for either WORK ( $t(21) = -0.32, p > .05$ ) or MD ( $t(21) = -0.08, p > .05$ ).

The data for TD, PD, P, and E were not normally distributed, and subsequent Mann-Whitney U tests did not reveal any significant differences between the control and experimental conditions (see Table 4.7 for details).

The data for F (Table 4.7) were also not normally distributed ( $W = 0.79, p < .01$ ). There was a large difference in the median F scores of the control and experimental groups: 105.91 and 30, respectively. A Mann-Whitney U test revealed that this difference was significant ( $U = 30.5, Z = -2.217, p < .05$ ), with the control group having a higher mean rank (15.23) than the experimental group (9.04). The effect size was moderate ( $r = 0.46$ ). This indicates that participants in the control group were more frustrated than those in the experimental group.

TABLE 4.7: Mann-Whitney U tests for non-normally distributed workload data and tests for prerequisite assumptions: Shapiro-Wilk tests for normality ( $W$ ,  $p$ ), sample sizes ( $N$ ), mean ranks, U values, Z values,  $p$  values and effect sizes ( $r$ ). \* indicates significance at  $p < .05$ .

	Shapiro-Wilk		N(C)	N(E)	Mean Rank(C)	Mann-Whitney			$r$	
	W	$p$				Mean Rank(E)	U	Z		$p$
TD	0.691	<0.001	11	12	10.68	13.21	51.5	-0.898	0.379	-
PD	0.854	0.003	11	12	11.27	12.67	58	-0.495	0.651	-
P	0.839	0.002	11	12	13.55	10.58	49	-1.048	0.316	-
E	0.901	0.026	11	12	10.27	13.58	47	-1.177	0.26	-
F*	0.79	<0.001	11	12	15.23	9.04	30.5	-2.217	0.027	0.46228

### 4.3 Typing

Input performance was assessed via examination of typing on a standard keyboard. Typing on a keyboard was the exercise chosen to assess input performance for several reasons. Typing is considered a difficult, ‘high bandwidth’ task, but it is easy to measure very accurately. Additionally, typing is a ubiquitous form of input that is not easily replaced without a loss of performance, and we believe that typing in VR may still be necessary for some years to come. Typing input was measured in the following ways:

**ACC** Overall Accuracy (%): a combination of Minimum String Distance error rate and Key Strokes Per Character, as defined in Section 2.4.3.

**GTS** Gross Typing Speed (words per minute): the number of words typed in one minute.

**NTS** Net Typing Speed (words per minute): a combination of GTS and ACC, as defined in Section 2.4.3.

**TTFKP** Time to First Key Press (seconds): amount of time it took to start typing once the test began.

Each participant completed two one-minute typing tests. One baseline test using a conventional desktop monitor, and one treatment test, either in VR (for the control group), or AV (experimental group). In each test, participants were presented with strings of random words that they had to type as quickly and accurately as possible. The raw input was captured by the software for later analysis, where the number of words, keystrokes, and errors (fixed or otherwise) were used to calculate the ACC, GTS, and NTS.

The initial sample size for the typing data was 23, with 11 participants in the control group and 12 participants in the experimental group. Of these, 7 were touch typists (3 in the control group and 4 in the experimental group). There were no significant differences between the control and experimental groups for participants who were able to touch type (see Appendix H). For this reason, the remainder of the input analysis focuses on the remaining 16 participants who were not able to touch type. There were 8 participants each in the control and experimental groups.

The data analysis for input performance is split into three subsections. Section 4.3.1 compares the baseline typing data of the control and experimental groups. Section 4.3.2 compares the baseline and treatment data within each group. Section 4.3.3 compares the treatment data of the control and experimental groups.

TABLE 4.8: Descriptive statistics for baseline typing data for control (VR) and experimental (AV) groups: mean (M), median (Mdn), variance (Var), standard deviation (SD), and sample size (N).

Condition	Subscale	M	Mdn	Var	SD	N
VR	ACC	94.61	95.55	11.77	3.43	8
	GTS	38.50	39.00	59.14	7.69	8
	NTS	33.50	35.00	50.00	7.07	8
	TTFKP	2.00	2.05	0.15	0.39	8
AV	ACC	91.23	91.52	35.53	5.96	8
	GTS	43.75	41.50	69.36	8.33	8
	NTS	36.88	36.50	176.70	13.29	8
	TTFKP	1.69	1.65	0.10	0.31	8

TABLE 4.9: T-tests for baseline typing data, effect sizes, and tests for prerequisite assumptions: Shapiro-Wilk tests for normality ( $W, p$ ), Levene's tests for homogeneity of variance ( $F, p$ ), t-tests ( $t, p$ ), and effect size where significance was found.

	Shapiro-Wilk		Levene's Test		T-test		Effect Size (d)
	W	p	F	p	t	p	
ACC	0.932	0.262	0.913	0.355	1.391	0.186	-
GTS	0.953	0.531	0.038	0.848	-1.31	0.211	-
NTS	0.985	0.99	2.673	0.124	-0.634	0.536	-
TTFKP	0.952	0.528	0.547	0.472	1.757	0.101	-

### 4.3.1 Baseline

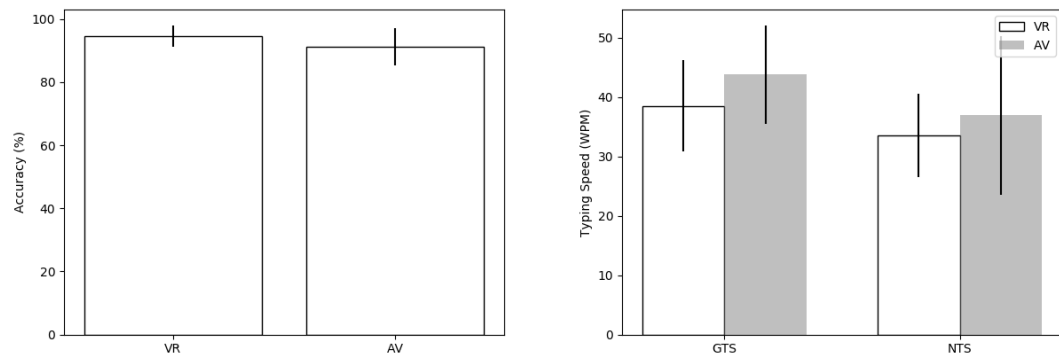
The control and experimental groups both completed a baseline typing test using a traditional desktop setup (see Section 3.1.4) before being exposed to their group's treatment (either AV or VR). Descriptive statistics for the gathered baseline input data are presented in Table 4.8. There were no outliers.

The baseline data for ACC, GTS, NTS, and TTFKP were normally distributed (see Table 4.9 for details). The mean ACC and TTFKP of the control group (94.61; 2.00) were higher than that of the experimental group (91.23; 1.69). Independent-samples t-tests revealed that the differences in neither ACC ( $t(15) = 1.396, p > .05$ ), nor TTFKP ( $t(15) = 1.76, p > .05$ ) were significant. The mean GTS was higher for the experimental group (43.75) than the control group (38.5), but an independent-samples t-test showed this difference to be insignificant ( $t(15) = -1.31, p > .05$ ). Mean NTS for the experimental group (36.88) was also slightly higher than that of the control group (33.40), but this difference was also found to be insignificant ( $t(15) = -0.63, p > .05$ ). Baseline typing data are depicted in Figure 4.5.

### 4.3.2 Within Subjects

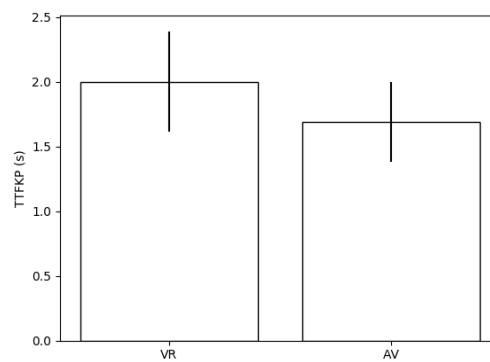
Participants in both the control and experimental groups performed a baseline typing test as well as a treatment typing test (in either AV or VR depending on their group). This subsection presents a within-subjects comparison of the differences between baseline and treatment results within each group.

For both the control (Table 4.10) and experimental (Table 4.11) groups, the mean scores for ACC, GTS, and NTS were significantly lower in the treatment compared to the baseline, while TTFKP was significantly higher in the treatment than in the



(a) Mean baseline accuracy for control and experimental groups. Higher is better.

(b) Mean baseline typing speed (NTS and GTS) for control and experimental groups. Higher is better.



(c) Mean baseline TTFKP for control and experimental groups. Lower is better.

FIGURE 4.5: Mean baseline typing performance for control (VR) and experimental (AV) group. The differences between the control and experimental group are insignificant for all measures.

TABLE 4.10: Paired-samples t-tests for within-subjects typing data in the control group: mean difference between treatments, t-test ( $t$ ,  $p$ ), and effect size. \*\* indicates significance at  $p < .01$ .

	Mean Difference	T-test		Effect Size (d)
		$t$	$p$	
ACC **	-25.524	-5.482	0.001	2.38
GTS **	-24.75	-14.853	<.001	3.077
NTS **	-29.625	-7.018	<.001	3.874
TTFKP **	1.71338	4.183	0.004	2.32

TABLE 4.11: T-tests for within-subjects typing data in the experimental group: mean difference between treatments, t-test ( $t$ ,  $p$ ), and effect size. \*\* indicates significance at  $p < .01$ .

	Mean Difference	T-test		Effect Size (d)
		$t$	$p$	
ACC **	-7.075	-3.756	0.007	0.924
GTS **	-21.375	-8.576	<.001	3.311
NTS **	-23.25	-5.295	0.001	2.218
TTFKP **	2.648	5.519	0.001	2.696

baseline. However, the effect size for ACC in the control group (2.38) was more than double that of the experimental group (0.92), with participants in the control group losing an average of 18.45% more accuracy than those in the experimental group. The effect size for NTS was also much higher in the control group (3.87) than in the experimental group (2.22), with the control group losing 6.38wpm more than the experimental group. Effect sizes for GTS and TTFKP were very large and similar for both the control and experimental groups.

### 4.3.3 Between Groups

In this subsection, the differences between treatment scores (VR and AV) for both groups are presented. Descriptive statistics for the treatment typing data for both the control and experimental groups can be found in Table 4.12. The data for all measures is presented in Figure 4.6. One numerical outlier for TTFKP in the AV condition was identified, taking much longer to begin typing than the average (see Figure 4.6c). We consider this a valid data point, as the participant in question was using the system as intended, and therefore did not remove this outlier from the data.

Data for ACC, GTS, and NTS were all normally distributed and homoscedastic (Table 4.14). Independent-samples t-tests were therefore conducted for each measure (Table 4.14).

The mean ACC was 69.09 for the control condition, and 84.15 for the experimental condition. The t-test revealed this difference to be significant ( $t(14) = -2.46$ ,  $p < .05$ ). The effect size was very large ( $d = 1.131$ ). This difference is visible in Figure 4.6a. Participants in the experimental group were able to type significantly more accurately (just over 15% more accurately) using the AV system than those in the control condition.

Mean GTS scores (Figure 4.6b) for the control and experimental conditions were 13.75 and 22.38, respectively. The difference was significant ( $t(14) = -2.66$ ,  $p < .05$ ),

TABLE 4.12: Descriptive statistics for treatment typing data for control (VR) and experimental (AV) groups: mean (M), median (Mdn), variance (Var), standard deviation (SD), and sample size (N).

Condition	Subscale	M	Mdn	Var	SD	N
VR	ACC	69.09	70.30	218.24	14.77	8
	GTS	13.75	13.50	70.21	8.38	8
	NTS	3.88	2.50	66.98	8.18	8
	TTFKP	3.71	3.32	0.94	0.97	8
AV	ACC	84.15	84.33	81.78	9.04	8
	GTS	22.38	22.00	13.98	3.74	8
	NTS	13.63	15.00	43.13	6.57	8
	TTFKP	4.34	4.02	1.83	1.35	8

TABLE 4.13: Mann-Whitney U tests for non-normally distributed typing data and tests for prerequisite assumptions: Shapiro-Wilk tests for normality ( $W, p$ ), sample sizes (N), mean ranks, U values, Z values,  $p$  values and effect size ( $r$ ).

	Shapiro-Wilk		N(C)	N(E)	Mean Rank(C)	Mean Rank(E)	U	Z	$p$	$r$
	W	$p$								
TTFKP	0.858	0.018	8	8	7.25	9.75	22	-1.05	0.328	-

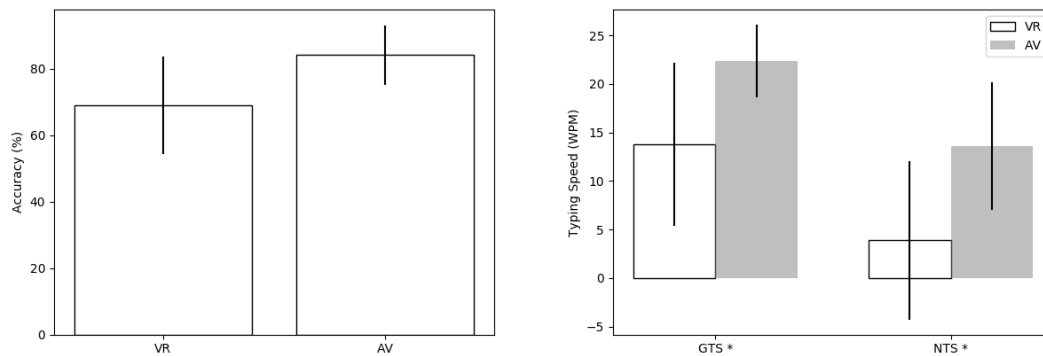
and the effect size very large ( $d = 1.42$ ), with participants in the control group typing approximately 9wpm slower than those in the experimental group.

The difference between mean NTS scores for the control (3.88) and experimental (13.63) conditions (Figure 4.6b) was also found to be significant ( $t(15) = -2.63, p < .05$ ). Once again, the effect size was very large ( $d = 1.41$ ). Some participants from the control condition scored negative values for NTS. While not necessarily an accurate reflection of reality, this indicates that their typing was so inaccurate, and their typing speed so low, that time spent trying to fix already committed errors would result in still more errors, requiring more time to fix.

Data for TTFKP (Table 4.13, Figure 4.6c) were not normally distributed ( $W = 0.86, p < .05$ ). The median TTFKP values for the control and experimental groups were 3.32 and 4.02, respectively. A Mann-Whitney U test, however, revealed this difference to be insignificant ( $U = 22, Z = -1.05, p > .05$ ).

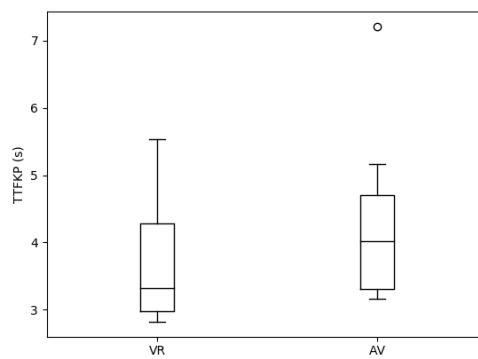
TABLE 4.14: T-tests for normally distributed typing data, effect sizes, and tests for prerequisite assumptions: Shapiro-Wilk tests for normality ( $W, p$ ), Levene's tests for homogeneity of variance ( $F, p$ ), t-tests ( $t, p$ ), and effect sizes where significance was found. \* indicates significance at  $p < .05$ .

	Shapiro-Wilk		Levene's Test		T-test		Effect Size ( $d$ )
	W	$p$	F	$p$	t	$p$	
ACC *	0.941	0.357	2.207	0.16	-2.46	0.027	1.314
GTS *	0.94	0.349	3.972	0.066	-2.659	0.019	1.421
NTS *	0.913	0.132	0.45	0.513	-2.628	0.02	1.272



(a) Mean treatment accuracy for control and experimental groups. The difference is significant ( $p < .05$ ). Higher is better.

(b) Mean treatment typing speeds (NTS and GTS) for control and experimental groups. Higher is better. \* indicates significance at  $p < .05$ .



(c) Box plot comparing treatment TTFKP scores of control and experimental groups. Lower is better. The difference is insignificant.

FIGURE 4.6: Treatment Typing Performance

## 4.4 Summary

Overall, participants in the experimental group (AV) experienced a higher level of presence, as measured by the IPQ, than those in the control group (VR). In particular, the mean scores for General Presence (the “sense of being there”) and Spatial Presence were significantly higher for participants who used the AV system than for those who used pure VR. There was no significant difference between the groups in terms of Involvement or Realism.

There was no significant difference in overall workload, as indicated by the NASA-TLX, between participants in the control and experimental condition. An examination of the factors that contribute to overall workload showed that participants in the control condition felt significantly more frustrated than those in the experimental condition.

There were no significant differences in input performance for the baseline typing treatment (traditional desktop setup), with both groups scoring similarly for overall accuracy, gross and net typing speed, and time to first key press. Both groups performed significantly worse on all measures in their respective treatments (AV and VR), than they did in the baseline. However, the effect sizes for overall accuracy and net typing speed were much higher for participants in the control group (participants in the control group lost more speed and accuracy moving from baseline to treatment than those in the experimental group). Finally, a between-groups comparison of treatment input performance revealed that there was no significant difference in time to first key press between the groups, but that participants in the experimental condition performed significantly better in terms of overall accuracy, gross typing speed, and net typing speed than those in the control condition.

## Chapter 5

# Discussion

In this chapter, we discuss the results presented in Chapter 4. Section 5.1 discusses the results obtained using the Igroup Presence Questionnaire (IPQ) (Schubert, Friedmann, and Regenbrecht, 2001) in relation to our first research question: is the overall level of presence experienced over an extended session in an immersive virtual setting higher in Augmented Virtuality than in Virtual Reality when the session has to be interrupted by interactions with the real world?

Section 5.2 discusses our findings regarding workload (as measured by the NASA Task Load Index (NASA-TLX)(Hart and Staveland, 1988)) in relation to our second research question: is it easier to interact with real-world objects in AV than in VR when users have to remove the HMD?

Finally, Section 5.3 discusses the results generated by our input capture and analysis software with respect to our third research question: is typing performance better in AV than in VR?

### 5.1 Presence

We previously defined presence as *virtual spatial telepresence* (Section 2.2): the subjective experience of “being there” in a computer-generated virtual environment. One of the primary objectives of this research was to investigate the impact that Augmented Virtuality (AV) would have on presence compared to pure Virtual Reality (VR) when users have to interact with real-world objects as they might in home or office contexts. Participants played an immersive 3D game while wearing an HMD, and had to interact with real-world objects several times in order to complete the game successfully. The participants were split into two groups, a control group who would play the game in VR and would have to interact with the real world with no assistance from the system, and an experimental group who would be able to use our AV system to help them navigate the real world. To measure the participants’ level of presence during the experience, the IPQ was administered immediately after they had completed the game.

TABLE 5.1: Average scores per presence factor from the IPQ database compared to scores from this study.

	IPQ Database Average	Measured Average
PRES	3.35	5.07
SP	3.25	4.88
INV	3.00	3.82
REAL	2.00	2.82

The IPQ database (Igroup, 2016) contains presence data obtained with the IPQ from 619 participants across 6 different studies. The average presence factor scores from this database are shown in comparison to the scores obtained in our experiment in Table 5.1. This comparison shows that presence scores for our study were relatively high overall, particularly for general presence (PRES) and spatial presence (SP) factors. The average PRES score for all participants was 5.07 out of 6, and the average SP score was 4.88 out of 6. This indicates that the virtual world we created as part of our apparatus, and the hardware used to deliver it, were successful at eliciting a sense of presence. We expected this to be the case, as we designed the game and virtual environment used in our apparatus to take advantage of qualities known to positively influence presence and the constructs involved in its formation (as discussed in Sections 2.2.3 and 3.1.1).

Average Involvement (INV) scores (3.82 out of 6) were lower than those for PRES and SP. This is a departure from the relationship between INV, PRES, and SP that is suggested in the IPQ database, where INV scores were on par with PRES and SP. This makes sense in the context of this study given the items of the IPQ that measure INV (see Appendix I). Three of the four items ask about the participant's level of awareness of and attention to the real world. Under normal gaming circumstances, one would expect these to be low if the user was present in the virtual world, but in our case, participants were *specifically* required to attend to the real world at several points in time. This likely explains why the INV scores were relatively low.

The average scores for the level of realism experienced (REAL) were the lowest of all the factors, at 2.82 out of a possible 6. As shown in Table 5.1, this is in line with previous findings, as scores for REAL are, on average, lower than the scores for the other factors. While we did attempt to create a graphically and physically realistic virtual environment with the resources available to us, achieving the kind of realism that modern gamers are used to was not within the scope of this work.

Statistical tests revealed several differences in presence factor scores between the control and experimental groups. We will now consider these differences.

### 5.1.1 General Presence and Spatial Presence

General Presence (PRES) is measured by the IPQ with only a single item and measures the extent to which users had a sense of "being there" in the computer-generated world. PRES has a high loading on the other factors measured by the IPQ, but particularly on Spatial Presence (Schubert, Friedmann, and Regenbrecht, 2001). The IPQ measures Spatial Presence (SP) using five items, two of which are very similar to the PRES item, while the others focus on the user's perception of the virtual world as a surrounding space and their actions within it. Because of the similarity of these two factors, and their high level of covariance, we consider them here together.

The results of our experiment show that participants in the experimental condition experienced greater levels of PRES and SP than those in the control condition. In both cases, this difference was significant, and the effect size large. Average PRES scores were 4.6 for the control group and 5.55 for the experimental group. It is interesting to note that in the SUS questionnaire (Slater, Usoh, and Steed, 1994) whence the PRES item in the IPQ is taken, only a score of 5 or higher on a 0-6 scale would have been counted toward the presence total. Taking a similar approach here would mean that, at least as far as PRES is concerned, participants in the experimental condition were present, and those in the control condition were not. Average scores for SP were 4.48 for the control group and 5.27 for the experimental group. Participants in the experimental group felt more present than those in the control group.

This result agrees with previous findings. Both McGill et al. (2015) and Budhiraja et al. (2015) found that the use of their versions of inferred minimal blending AV led to higher levels of presence than if users had to peep. This confirms that when real-world interactions are required of a user, their sense of presence in the virtual world is better preserved through the use of minimal blending AV than if the user must lift off the HMD.

### 5.1.2 Involvement

Involvement (INV) is the measure of how aware of and attentive to the real/virtual environment participants were (see Appendix I for the items of the IPQ measuring INV). Higher scores for INV suggest that participants pay more attention to the virtual world, and less to the real world. Although statistically insignificant, participants in the experimental condition reported a slightly higher average level of INV (4.05) than participants in the control condition (3.6). This suggests that the experimental group may have been less aware of and attentive to the real world than the control group. While both groups were given the same instructions, and were both required to pay attention to the real world, participants in the experimental condition received fewer stimuli from the real world through the AV system than participants in the control condition who lifted the HMD off to see the real world. This may explain the small difference between the scores, as experimental participants had fewer real-world stimuli to attend to. Although previous work regarding the effects of AV on user experience did not go into detail regarding the factors that contribute to presence (see Section 2.5), our INV scores are lower than expected based on the relationship between factors demonstrated in the IPQ database (Table 5.1). The fact that participants in both groups of our experiment had to pay attention to the real world to an equal extent explains why the INV scores were low for both groups compared to the other factor scores and not significantly different from one another. We argue that lower INV scores, relative to PRES and SP, are a necessary consequence of, and perhaps even a requirement for successful interactions outside of the virtual environment – someone who was fully involved with the virtual world would be completely unaware of the real world, and therefore unable to interact with it.

### 5.1.3 Realism

The final factor measured by the IPQ was Realism (REAL). The items measuring this factor (see Appendix I) focus on how realistic the virtual world seemed compared to the real world. There was almost no difference in REAL scores between the control and experimental group. The experimental group had a mean score of 2.84, less than one percent higher than the control group's 2.80. This difference is insignificant ( $p = 0.93$ ). This result is expected, since the virtual world itself was identical in both conditions.

## 5.2 Workload

We defined workload as the subjective mental and physical cost of performing a task. Our second research objective was to find out if the workload associated with performing peripheral real-world tasks while using an HMD could be reduced through the use of minimal blending AV. To this end, participants in our experiment had to perform several real-world interactions while using an HMD to play an immersive

3D game. These real-world interactions involved locating and interacting with clues placed in the real world that were necessary to solve puzzles in the virtual world. Participants in the control condition had to lift off the HMD every time they had to interact with the real world, and participants in the experimental condition used our implementation of minimal blending AV.

To measure workload, the NASA Task Load Index (NASA-TLX) (Hart and Staveland, 1988) was administered to participants after they had completed the game. The NASA-TLX (Appendix D) is a subjective self-report questionnaire that measures overall workload (WORK) according to six factors: Mental Demand (MD), Temporal Demand (TD), Physical Demand (PD), Performance (P), Effort (E), and Frustration (F).

The average WORK score over all participants was 46.49 out of a possible 100. The biggest contributor to the overall score was MD. This is likely because of the genre of the game that participants played – puzzle games are more mentally demanding, and less demanding in other areas, than other genres. The smallest contributor to workload was TD. This is also easily explained, as no time constraints were placed on participants while playing the game. PD and P contributed similarly to overall workload, and were higher than E and F, which also contributed to overall workload to a roughly equal degree.

Overall WORK was slightly higher for the control condition than it was for the experimental condition, but not significantly so. In the following subsections, we discuss the scores for the individual factors that contributed to workload, with F being the only factor that was significantly different between groups.

### 5.2.1 Mental and Physical Demand

The differences between MD and PD were insignificant, but in both cases, slightly higher for the experimental condition than for the control condition. For participants in the experimental condition, the novelty of using the AV system made real-world interactions slightly more mentally demanding than the intuitive “peeping” resorted to by participants in the control condition. We believe that the main contributor to MD for both groups was the game that the participants played. The game required several clues to be interpreted so that puzzles could be solved. This activity is inherently mentally demanding, and both groups had to solve the same puzzles in the same way. This likely explains the fact that participants reported higher levels of MD than any other workload-related factor, and that there was no significant difference in MD between the groups.

The result for PD is unexpected. We assumed that having to lift the headset, and in some cases hold it up (see Section 5.2.4), would be more physically demanding than the AV alternative.

### 5.2.2 Temporal Demand and Effort

Scores for TD and E were almost identical in both conditions. The result for TD is expected, as there were no time constraints for participants in either condition. The use of the AV system required no more effort than peeping. We interpret this result to mean exactly that: it takes roughly the same amount of effort to use the AV system to interact with real-world objects as it does to remove the HMD. This is an acceptable result, as in both cases, the amount of effort that was reportedly required was low.

### 5.2.3 Performance

P measured how successful participants thought they were in completing the tasks that were required of them. Looking at Figure 4.4, it seems that P was far higher for participants in the control group than for those in the experimental group, especially if the effect on the data of the outliers in the experimental group are considered. However, even with the removal of these outliers, the difference in P scores between the groups is not significant ( $p = 0.051$ ). All participants from both groups successfully completed the game and the required real-world interactions. Our results show that participants in the experimental condition felt that they were at least as able to successfully complete their required tasks as those in the control condition, and suggest the possibility that they were more successful.

### 5.2.4 Frustration

Finally, there was a significant difference in Frustration scores between groups – participants in the control condition reported being more frustrated than those in the experimental condition. This difference was the biggest contributor to the difference in overall workload scores between the groups. This is understandable, as each real-world interaction required users to lift the HMD off their head, sometimes multiple times. This is disruptive and uncomfortable, especially for users wearing glasses which sometimes get stuck in the headset. Some adjustment was then required to return the HMD to its proper and comfortable position. Furthermore, because of the front-heavy design of the HMD, it would sometimes fall back onto a users face while they were trying to see the real world. Some participants therefore had to hold the HMD in place with one hand, which meant they had to use only one hand to execute desired real-world tasks. The use of the AV system left both of the user's hands free, facilitating real-world interaction, and meant that the HMD could be left in its most comfortable position on the face without ever having to be readjusted. We believe that being able to leave the HMD in place throughout the experiment was what made the experience less frustrating for our experimental participants. Although the NASA-TLX has been used to evaluate AV systems in previous work, only the overall workload score was reported (McGill et al., 2015). However, our result agrees with survey data collected by McGill et al. (2015) in which users reported that having to lift off the HMD was frustrating. Furthermore, both McGill et al. (2015) and Budhiraja et al. (2015) found that their participants ranked their respective versions of inferred minimal blending AV highly in terms of preference and satisfaction compared to VR with peeping. Our results regarding frustration can at least partially account for this, as we have shown that real-world interaction is less frustrating in AV than in VR, and AV is therefore likely to be preferred by users.

## 5.3 Typing

The final objective of our research was to investigate the impact on typing performance (defined as accuracy, typing speed, and time to first key press (TTFKP)) of minimal blending AV compared to pure VR. To do this, all participants first completed a baseline typing test using a normal computer monitor. This served as a point of comparison to the typing performance achieved under treatment conditions. Subsequently, participants in the control condition performed the same typing test, but while wearing an HMD and unable to see their hands or keyboard. Participants in the experimental condition used our implementation of minimal blending

AV, which allowed them to see their hands and the keyboard while they were wearing the HMD. Our results pertaining to typing performance are divided into three sections:

1. Between groups comparison of baseline typing performance.
2. Within-subjects comparison of baseline to treatment typing performance (for each condition).
3. Between groups comparison of treatment (AV vs VR) typing performance.

### 5.3.1 Baseline Typing Performance

Our results revealed that there were no significant differences between the control and experimental group under normal typing conditions. Scores for both groups were similar for accuracy, speed, and TTFKP, with the control group being slightly more accurate and slightly slower, on average, than the experimental group. However, results from t-tests show that both samples can be assumed to come from the same population – the differences were not significant enough to assume that one group had better typists than the other. We were therefore justified in using these baseline results to compare against results obtained under treatment conditions for both groups, assuming that any differences in the subsequent tests are a result of our interventions rather than individual differences. The average gross and net typing speeds for all participants were 41.13 WPM and 35.19 WPM respectively. The average accuracy for all participants was just under 93%. A survey of over 3000 typists shows that these figures are almost exactly average, falling within 0.3 standard deviations from the mean (Ostrach, 1997).

### 5.3.2 Within-subjects Typing Performance

In order to see how typing performance changed moving from normal typing conditions to each of our treatments (VR and AV), we performed a within-subjects analysis of typing performance for each group. Participants in both groups performed significantly worse in their respective treatment condition than in the baseline. Scores for typing speed and accuracy were significantly lower, and TTFKP was significantly higher.

In the VR treatment, our control participants achieved 69.09% accuracy, a 25.52% decrease from the 94.51% they achieved under baseline conditions. Typing speed (NTS) decreased from 33.5 WPM in the baseline to just 3.88 WPM in the VR treatment (a drop of 29.63 WPM). TTFKP went up 1.71s from 2.0s to 3.71s. It is easy to explain why performance dropped so much for participants in the control condition. None of our participants were touch typists, and without the ability to see the keyboard or their hands, finding the correct keys quickly or at all, and even finding the keyboard itself, were much more difficult tasks than in the baseline condition.

This was an expected result, and similar to that obtained by McGill et al. (2015), who saw a very similar slip in performance for their VR condition compared to a baseline: 26.22% less accuracy, 35.3 WPM less typing speed, and an increase of 1.2s to TTFKP.

In the AV treatment, our experimental participants lost only 7.08% accuracy (from 91.23% in the baseline to 84.15% in the treatment), but lost 23.25 WPM of typing speed (from 36.88 WPM to 13.63 WPM) and gained 2.65s in TTFKP (from 1.69s to 4.34s). While we had hoped for better results, particularly with respect to

typing speed, the loss of performance for our AV implementation is similar to previous findings. McGill et al. (2015) also observed decreases in accuracy (4.56%) and typing speed (20.04 WPM) while TTFKP increased (1.4s). In our case, we believe this loss of performance was due mainly to the slight discrepancy between optical flow and proprioceptive feedback caused by our implementation of the AV system. Although barely noticeable for gross motor tasks such as the real-world interactions described in Section 3.1.2, the difference between the apparent (webcam) and actual eye position of the participant (see Figure 3.16) made participants move more slowly because they were slightly unsure of the precise positions of their fingers.

### 5.3.3 Between Groups Typing Performance

Despite the fact that both our control and experimental participants lost performance in their respective treatments, the effect sizes for these differences were much higher for those in the control condition. A between groups analysis of the data gathered from the VR and AV treatment typing tests revealed significant differences between the groups in terms of both speed and accuracy.

The experimental group, using the AV system, typed significantly more accurately than the control group. On average, the experimental group was able to type just over 15% more accurately than the control group. This is a considerable amount. Such an improvement was to be expected, as the AV system allowed participants to see what they were doing. Previous work yielded a similar result, with AV allowing approximately 20% better accuracy than VR (McGill et al., 2015).

Our experimental group was also able to type significantly faster than the control group under treatment conditions. The experimental group managed to type approximately 9 WPM faster than the control group. This was also expected, for the same reasons as the improvement in accuracy. Previous work yielded a better result, with the use of an AV system leading to an improvement of about 15 WPM over pure VR (McGill et al., 2015).

On average, the experimental group took 0.2 seconds longer to find the keyboard and begin typing than the control group. This difference was not significant. Previous work showed a slight improvement (0.6s) to TTFKP (McGill et al., 2015). We believe that this result would have changed if we had measured time to first *correct* key press – an examination of the raw typing input that was captured reveals that although our experimental group took an average of 0.2s longer to begin typing, the first key they pressed was the correct one 91.66% of the time, compared to only 63.63% of the time for the control group.

McGill et al. (2015) reported that their participants were able to type at an average of 38.5wpm at 90.8% accuracy. Although our results regarding typing performance are not as good, we believe that they are comparable for several reasons. First of all, our sample of typists, under baseline conditions, is statistically almost exactly average. McGill et al. (2015), however, report a well above average mean baseline typing speed of just under 60 WPM for their sample, which is more than one standard deviation above the mean (Ostrach, 1997). This suggests that some of the participants in that experiment were touch-typists, whose performance we have shown to be insignificantly affected by the use of VR/AV compared to a desktop baseline. It is possible that this had an impact on McGill and Boland's results. They also showed that, on average, participants type slower in AV (38.5wpm) than in baseline conditions (58.9wpm), which agrees with our findings. Furthermore, McGill and Boland never tested the effects of minimal blending AV on typing performance, but rather partial blending AV, which shows the user more of reality. Our results are therefore

fairly novel, as we tested an implementation of minimal blending AV, and achieved similar results while further restricting the subset of reality that was shown to users. The data we present is also from the group whose typing performance is most likely to suffer from visual cut-off: those who are unable to touch type. Finally, all of our testing was conducted on a completely standard keyboard, while previous work relied on modified input devices to make up for lacking display resolution.

Walker et al. (2017) proposed a non-AV solution to the problem of typing while wearing an HMD: their implementation consisted of a virtual keyboard assistant overlay displayed by the HMD, combined with automated error correction. Using this system, their participants were able to achieve an average typing speed of 43.7wpm at an accuracy of 97.4%. The majority of participants in their study were touch typists, making a thorough comparison to our results difficult. However, the few touch typists who used our system scored an average of 47.5wpm at 95.08% accuracy, suggesting that our AV system can preserve typing performance to a similar extent as the keyboard assistant for proficient typists.

## 5.4 Summary and Limitations

In this section we briefly highlight our most important findings and discuss factors that may have limited our results.

### 5.4.1 Presence

With respect to the ability of our minimal blending AV to preserve a sense of presence in the home/office context, we believe our results are very promising. That participants in our experimental condition reported significantly higher levels of general and spatial presence than those in the control condition attests to this.

The only aspect of presence that our apparatus was not successful at producing was a sense of realism. Improvements to the virtual environment we created may have ameliorated how realistic it seemed, and may very well have improved other presence factors as well, as they are known to co-vary to a high degree (Regenbrecht, Schubert, and Friedmann, 1998).

### 5.4.2 Workload

Participants in our experimental group were able to complete all of the required real-world interactions successfully while wearing the HMD. Their interactions were as quick and effective as those in the control condition. Our experiment revealed that, although our AV implementation did not significantly reduce the level of workload that participants experienced while interacting with real-world objects, it did significantly reduce the level of frustration that participants felt.

However, we expected better results in this area, and believe that several factors prevented us from attaining them. The game that we designed was inherently more mentally demanding than anything else, and this probably drew attention away from other factors that contribute to workload. Usually the NASA-TLX is administered on a task by task basis, but because presence was our primary concern and we did not want to disrupt the experience too much, we administered the NASA-TLX after a series of tasks, and after administering the IPQ, which may have skewed the data somewhat. Finally, subjective workload methods are known to suffer from susceptibility to high inter-subject variability (Hart and Staveland, 1988). This is evident from the very high standard deviation scores recorded for almost all the

measured workload factors. This susceptibility was compounded by our decision to adopt a between groups design, rather than a within-subjects design that would have ameliorated this effect to an extent.

### 5.4.3 Typing

Our experiment revealed that using an AV system to type leads to better performance than typing in pure VR. Our experimental participants were able to type significantly more quickly and accurately than their control counterparts.

While we were disappointed that typing performance in AV did not reach the level achieved in the baseline condition, this finding is in line with previous work, and still suggests that AV is a promising solution to input problems caused by visual cut-off. We believe that the drop in typing speed we observed in the AV condition (compared to the baseline) was mainly due to the slight mismatch in the position of the users' eyes compared to what was shown to them by the webcam (see Section 3.1.3), and that reducing this somehow may have led to better typing speed.



## Chapter 6

# Conclusions

This dissertation presents our research into the effects of Augmented Virtuality (AV) on various aspects of user experience compared to pure Virtual Reality (VR). In particular, we were interested in the effects on presence, workload, and performance that AV has on Head-Mounted Display (HMD) users in the home context, where interactions with real-world objects and input devices are commonplace. We began by arguing that presence is the primary goal of VR, and the most important feature that distinguishes HMDs from other, less immersive, systems (Section 2.1). We then provided an explication of presence and its theoretical underpinnings (Section 2.2), before highlighting the issue of visual cut-off during HMD usage and the negative impact it can have on presence (Section 2.3). Subsequently, we introduced AV as a potential solution to this problem, discussing several previous works in the area (Section 2.4). While earlier results are promising, they fail to provide a thorough accounting of presence and workload in AV in direct comparison to VR. Furthermore, previous work relies on implementations that are either impractical for potential end users, requiring them to make extensive modifications to their environments, or attempt to solve the problem of visual cut-off for only specific real-world interactions, rather than in general (Section 2.5). To corroborate previous findings, and to address the gaps in the literature, we decided to implement and test our own version of inferred minimal blending AV that did not require any modifications to the user's environment, and compare it directly to VR in terms of presence, workload, and typing performance. Our research questions were as follows:

1. Is the overall level of presence experienced over an extended session in an immersive virtual setting higher in AV than in VR when the session has to be interrupted by interactions with the real world?
2. Is it easier to interact with real-world objects in AV than in VR when users have to remove the HMD?
3. Is typing performance better in AV than in VR?

To answer these research questions, we developed several pieces of apparatus, including a Virtual Environment (VE) designed to be capable of supporting a sense of presence (Section 3.1.1), a unique mixed reality game that allowed us to closely simulate the context in which HMDs are most commonly used (Section 3.1.2), a novel AV system based on the most successful designs from previous work (Section 3.1.3), and several pieces of software to measure typing performance under AV and VR conditions (Section 3.1.4). We then conducted a user experiment (Section 3.2) that allowed us to directly compare AV to VR in terms of presence, workload, and input performance. In the remainder of this chapter, we answer our research questions based on the data gathered in our experiment, remark on interesting and

unexpected results and the implications thereof, and suggest avenues for potential future work.

## 6.1 Presence

Our hypotheses regarding the impact of AV on presence were as follows:

$H_{A0}$  There is no difference between presence scores in AV and in VR.

$H_{A1}$  Presence scores are higher in AV than in VR.

$H_{A2}$  Presence scores are lower in AV than in VR.

We measured presence using the IGroup Presence Questionnaire (IPQ). Our experiment revealed that presence scores were higher for participants in the experimental condition, who used our AV system, than they were for those in the control condition, who used pure VR. In particular, the experimental group reported significantly higher levels of general and spatial presence than the control group. Although there were no significant differences between the groups in terms of involvement or experienced realism, these results are easily explained. Involvement is a measure of how much attention is being paid to the virtual environment. In our case, participants in both conditions were required to attend to both the real and virtual environments to the same extent. The nearly identical scores for experienced realism are trivially explained by the fact that participants in both groups were exposed to exactly the same virtual environment. We therefore reject the null hypothesis ( $H_{A0}$ ) in favour of  $H_{A1}$ , and conclude that AV preserves the sense of presence in the virtual world when HMD users must do something in the real world.

Although our results regarding the higher levels of general and spatial presence observed under AV compared to VR conditions are not new, they are nevertheless interesting because they are, as of yet, not explained in the literature. In Section 6.1.1, we suggest an explanation for this phenomenon. Our results for involvement were unexpected, as they are relatively low compared to the other presence factors. In Section 6.1.2, we consider the possible implications of this finding.

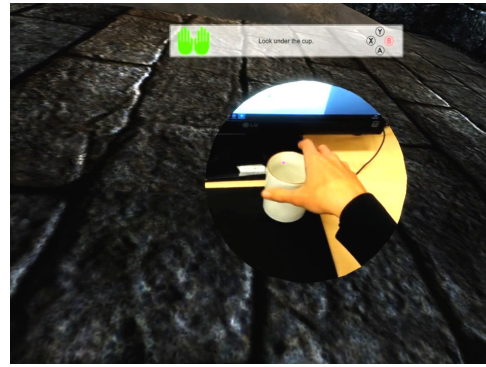
### 6.1.1 General and Spatial Presence

Based on the theory of presence formation presented in Section 2.2.2, we believe that there are two reasons why participants who used AV to perform real-world interactions may have reported higher levels of presence than those in the VR condition.

Let us say we have two users, a control user (A) and an experimental user (B), corresponding to participants in our control and experimental conditions. Both A and B are wearing HMDs and exploring the same virtual environment. We assume that both of these users are present in the VE. Due to the highly immersive nature of the HMD and the VE, they have both formed a strong spatial situation model (SSM) of the VE, organised from an egocentric perspective (the SSM is an egocentric reference frame, or ERF). This ERF has challenged the users' real-world ERF, and through perceptual testing, has become the primary egocentric reference frame (PERF), confirming the medium-as-PERF-hypothesis. In order for these users to continue to feel present, the medium-as-PERF-hypothesis must be continually confirmed by the stimuli that each user is perceiving. Now let us assume that the users need to interact with a real-world object. The HMD of user B has been equipped with a minimal blending AV system, and the HMD of user A has not. User A lifts off the HMD to



(a) The control participant lifts off the headset to locate the cup. All stimuli they are now receiving are from the real world.



(b) The experimental participant uses the AV system to locate the cup. The majority of stimuli being received are still coming from the virtual environment.

FIGURE 6.1: Comparison of what control and experimental participants would see when interacting with the real world.

find and interact with the object of interest, and user B simply reaches out, and a window into the real world around their hand is shown by the HMD. What each user is seeing at this point is shown in Figure 6.1.

For user A, the control participant, all of reality is shown (Figure 6.1a). All audio-visual stimuli from the VE are replaced by stimuli from the real world, and their ERF in the VE is lost. For user B, the experimental participant, only a small subset of reality is shown (Figure 6.1b). Only a fraction of visual stimuli from the VE have been replaced by stimuli from the real world, and auditory stimuli are still coming from the VE<sup>1</sup>. Furthermore, user B maintains their egocentric viewpoint in the VE.

From this point, there are two possibilities that may explain why participants in the experimental condition (like user B) reported higher levels of presence than those in the control condition (user A). The first is that the use of the AV system did not break presence for those in the experimental condition at all. The second is that participants in both the control and experimental conditions suffered breaks in presence when they had to interact with the real world, but it was easier for participants in the experimental condition to re-establish the lost sense of presence. Let us consider these possibilities in more detail.

The stronger the SSM that presence forms from, the more conflicting stimuli are required for presence to be broken (Wirth et al., 2007). Our apparatus was constructed to be as immersive as possible, meaning that our participants were able to form very strong SSMs. This is suggested by our, on average, very high PRES and SP scores. However, user A, when he lifts the HMD to interact with the real world, receives *only* stimuli that conflict with the medium-as-PERF-hypothesis, more than enough to break presence formed from even the strongest SSM. As Slater (2002) puts it: "If (somehow) an actor were receiving signals from only one environment, then by definition that actor is present in that environment". Since user A is receiving stimuli only from the real environment, they must be present there, and not in the virtual environment. User B, on the other hand, has formed an SSM that is just as strong as that of user A, but thanks to our AV system, receives far fewer stimuli that

<sup>1</sup>Note that the Oculus Rift CV1, the headset used in these experiments, has built-in headphones which are therefore removed when the HMD is removed.

conflict with the medium-as-PERF-hypothesis. Not only that, but B is still receiving stimuli that support the medium-as-PERF-hypothesis in both the auditory and visual channels. It is therefore possible that the highly immersive nature of HMDs facilitates the creation of an SSM so strong, that the small number of conflicting stimuli presented by our minimal blending AV implementation are not enough to break presence.

Alternatively, even though user B is exposed to fewer conflicting stimuli than user A, it is still possible that presence is broken for user B every time the AV system is used because the majority of controlled attention is focused on the stimuli from the real world. Post-hoc measurement via self-report questionnaires cannot rule this out, as it does not give us the level of presence experienced from moment to moment (Hartmann et al., 2015). If this is the case, there are still reasons why user B would report a higher level of presence than user A. Because much of the VE is left intact for B (they are still receiving stimuli from the VE that affirm the medium-as-PERF-hypothesis), it could be that it is easier to re-establish the lost sense of presence once the real-world task is complete. This is because, when the real-world interaction is complete and the users return to the virtual world, A has to start the process of presence formation from scratch, forming the SSM of the VE from the presented spatial cues and organising them from an egocentric perspective before the real-world PERF can be challenged. User B, on the other hand, even though they may not have been paying controlled attention to stimuli from the VE while interacting with the real world, was nevertheless receiving stimuli that could maintain the SSM – visual and auditory spatial cues, and their egocentric viewpoint in the VE. Once user B has finished with the real world, they come back to the VE already at the second stage of presence formation, facilitating their return to presence.

For several reasons, we believe that this is more likely than the possibility that presence is not broken at all. First, a hallmark of presence is that for the person experiencing it, their actions are associated with the environment in which they are present. Therefore, if they are successfully interacting in the real world, they must be present in the real world. Secondly, while interacting with the real world using AV, users are likely dedicating controlled attention to the displayed subset of the real world, which could allow it to challenge the currently active medium-as-PERF-hypothesis. However, unlike inset AV, minimal blending AV does not require so much attention that the SSM of the VE is totally lost, and unlike full blending or peeping, not all virtual stimuli are replaced, which would also destroy the SSM of the VE. That AV might allow the SSM/ERF of the VE to survive interactions with the real world, facilitating a *return* to presence, therefore seems more realistic than assuming that presence is not broken at all.

In summary, we believe that presence was better maintained for participants in the experimental condition for one of two reasons. Either the AV system does not provide enough conflicting stimuli (that suggest that the real world is the PERF) to break presence during real-world interaction or, more likely, the AV system allows enough stimulation from the VE that the SSM and inherent ERF of the VE are maintained during real-world interaction, and less cognition needs to take place to get from this point back to presence.

### 6.1.2 Involvement

As we discussed in Section 5.1.2, the scores we observed for involvement were unexpectedly low given the high scores observed for general and spatial presence. We

believe that this is due to the nature of the items that are used to measure involvement, as well as the differences between the context of AV applications and more traditional VR applications.

The items of the IPQ that measure involvement (Appendix I) are set up in a way that suggests an inversely proportional relationship between awareness of the real world and involvement with the virtual world. In the context of VR, this relationship makes sense. If a VR user acknowledges that they “still paid attention to the real environment” and “were aware of their real environment” when they were not expected to, then they were likely not involved with the virtual environment, and likely not present. The purpose of AV, however, is to increase the level of awareness of the real environment that users can have so that they may better interact with it. Particularly in experiments such as ours, where participants were explicitly asked to pay attention to objects in the real world, participants are likely to make the aforementioned acknowledgements. This leads to a lower involvement score.

In the context of VR, a low involvement score is an undesirable result. In the context of AV, however, this is not necessarily the case. Without being aware of, and paying attention to, the real world, users would be unable to interact with it. For AV applications, a lower involvement score, as determined by the IPQ, indicates that the system is doing what it is meant to do. That the virtual world is less involving, and therefore less capable of inducing presence, does not necessarily follow from this added awareness of the real world when the user requires it, but the nature of the items of the IPQ that measure involvement will suggest that this is the case.

There are therefore several considerations that future research into presence in mixed reality (particularly AV) should take into account. First, if an overall presence score (an average of contributing factors) is used that includes involvement in some form, it may be an underrepresentation of the level of presence that was actually experienced because of an artificially low involvement score. Second, if involvement is treated separately from other presence factors, care must be taken in its interpretation. The relationship between involvement and presence in AV is not as straightforward as it is in VR. If these considerations do not stop results from being confounded, it may be necessary for presence measures to be revised specifically for the evaluation of mixed reality applications.

### 6.1.3 Summary

With respect to presence, this research has been a success. We were able to successfully create a sense of presence in our participants, and showed that our implementation of inferred minimal blending AV was able to mitigate damage to presence during real-world interactions. This result corroborates previous findings, adding evidence to the theory that a small subset of reality is all that is required to interact with the real world effectively while preserving presence (McGill et al., 2015). Additionally we have shown that minimal blending AV, when compared directly to a VR baseline, significantly preserves spatial presence as well as the more general “sense of being there”. Finally, we were able to explain the demonstrable power of AV to preserve presence during real-world interactions according to existing theories of presence formation. This link between AV and presence theory has not been previously explored.

## 6.2 Workload

Our hypotheses regarding the impact of AV on workload were as follows:

$H_{B0}$  Workload for performing real-world interactions is the same in AV as it is in VR.

$H_{B1}$  Workload for performing real-world interactions is lower in AV than it is in VR.

$H_{B2}$  Workload for performing real-world interactions is higher in AV than it is in VR.

Our experimental participants were all able to successfully execute the required real-world interactions while wearing an HMD. We measured workload using the NASA Task Load Index (NASA-TLX). While the overall workload reported by participants who had to resort to peeping was higher than that reported by those who used AV, the difference was not significant. However, an examination of the factors contributing to workload revealed that the use of AV made real-world interactions significantly less frustrating. Without significant effects on any other factor contributing to workload, we do not feel that we have enough evidence to reject  $H_{B0}$ . We therefore conclude that AV does not necessarily make real-world interactions easier, but does make them less frustrating.

The most interesting aspect of the workload data that we gathered was that the only contributing factor that was significantly different between the control and experimental condition was frustration. In Section 5.2, we discussed each of the factors that contribute to the overall workload score and suggested reasons why we did not find significant differences between our control and experimental conditions. We believe that our results were skewed somewhat due to our choice to use a puzzle game as part of our apparatus. The game's inherent mentally demanding nature possibly skewed the weightings of the NASA-TLX, resulting in mental demand feeling like a bigger contributor to workload than the other factors. This is a context effect that we did not properly consider before commencing with experimentation. Workload scores for both groups were nevertheless acceptable (overall workload scores of below 50 out of 100 (Hart and Staveland, 1988)), and it is reasonable to assume that, whether peeping in VR or using AV, real-world interactions just aren't that much work.

One useful result that the NASA-TLX did reveal was that participants in our control condition found the experience significantly more frustrating than those in the experimental condition. We discussed possible reasons why this may have been the case in Section 5.2.4. A thorough usability study of AV would likely yield useful insights in this area.

### 6.2.1 Summary

Previous work has focused on the difference in workload between inferred and user-initiated transitions between VR and AV, showing that inferred transitions are preferable in this regard (McGill et al., 2015). We therefore expected that workload for real-world interactions would be significantly lower for inferred minimal blending AV than if participants had to "peep". Our research has shown that the difference in workload between VR and inferred minimal blending AV was not as great as we expected. However, the fact that real-world interactions are so much more frustrating in pure VR than in AV is in line with previous work, and can explain why

implementations of minimal blending AV score higher in terms of preference than alternatives (McGill et al., 2015; Budhiraja et al., 2015).

### 6.3 Typing Performance

Our hypotheses regarding the impact of AV on typing performance were as follows:

$H_{C0}$  There is no difference in typing performance between AV and a VR baseline.

$H_{C1}$  Typing performance is better in AV than in a VR baseline.

$H_{C2}$  Typing performance is worse in AV than in a VR baseline.

We operationally defined typing performance as accuracy, speed, and time to first key press (TTFKP). Although participants were not able to reach the level of performance using AV that they achieved under baseline conditions, they were able to type significantly faster and more accurately than those who used pure VR. Though there was no significant difference in TTFKP between conditions, there is still enough evidence to reject the null hypothesis ( $H_{C0}$ ) in favour of  $H_{C1}$ . We conclude that typing performance is indeed better in AV than it is in VR.

The efficacy of minimal blending AV for typing has not previously been assessed, but our results are similar to previous experiments where slightly more of reality was shown (partial blending as opposed to minimal blending) (McGill et al., 2015). People are able to type better, particularly in terms of accuracy, using inferred minimal blending AV than they can in VR. That minimal blending AV helps with typing as much as partial blending is more evidence that only a restricted subset of reality is required to facilitate even high bandwidth interactions.

However, AV will not have solved the problem of visual cut-off, with respect to typing, until people are able to type as well in AV as they can under normal desktop circumstances in terms of speed as well as accuracy. There are several limiting factors that we believe will have to be overcome before this is possible. While we reduced latency as much as possible, and graphical resolution was as high as we could make it, both could be further optimised, and will improve as the technology improves. We do not believe, though, that these were the main contributors to the loss of typing performance we observed moving from desktop conditions to AV conditions. In Section 3.1.3, we mentioned that our implementation of AV introduces a slight mismatch between visual flow and proprioception due to the relative positions of the various hardware components. Put simply, a user's hands are in a slightly different place compared to where they see them. This offset is small enough that gross motor tasks still feel natural. This is evidenced by the fact that our AV participants were all able to complete the real-world interactions required of them in our puzzle game as quickly and easily as those in the control condition who used their real eyes. However, at small distances, such as those between the centre of a keyboard key and the space between an adjacent key, this offset may have had an impact. We believe that participants had to slow down their typing to make sure that what they were seeing and feeling was in agreement. If indeed the users of HMDs will need to type on standard keyboards in the future, these issue will need to be solved, otherwise alternative methods of facilitating text-entry, such as that presented by Walker et al. (2017) (Section 5.3.3), may be preferable to AV. Although their virtual keyboard assistant is not a general solution to the problem of visual cut-off, it may be the case that typing in VR is enough of a challenge to warrant the use of a targeted solution until AV technology and techniques improve.

## 6.4 Future Work

Our research indicates several potential avenues for future work. With respect to presence, we provided two possible explanations of why AV might preserve it: either that it allows presence never to be broken during real-world interactions, or that it at least provides enough stimuli to facilitate the return to presence once interactions with the real world are complete. A common hallmark of presence in a virtual environment is that the person experiencing it will react to virtual stimuli as if they were real. Leveraging this fact, it would be possible to design an experiment that could reveal whether people are still present in the virtual world during real-world interactions through AV. Let us say that while people are interacting with the real world, the virtual environment provides a stimulus. For the sake of this example, let us assume that the stimulus is meant to give the user a fright. If they do not react to this stimulus, then they must not be present. Conversely, if the person does react to the stimulus – we could observe their behaviour and see if they are visibly startled – then they must have still been present in the virtual environment, even while interacting with the real world. This would be a very interesting result, because another key feature of presence is that the environment in which one is present is the one with which they associate their possible actions. If a person is interacting with the real world, that is where they associate their actions, so that is where they must be present. However, if at this point they can still be made to react to a stimulus from the virtual world, it may suggest that levels of presence in more than one environment are possible. Whether it is even possible to experience degrees of presence in multiple environments is still a point of contention in the literature (Hartmann et al., 2015). Further research into precisely why AV is able to preserve presence could potentially pave the way toward better design and implementation choices for future AV systems, and toward a better understanding of presence itself.

That there was not much difference in the workload of interacting with real-world objects using AV compared to VR was a surprise to us. Our workload data was fairly noisy, most likely due to high individual variability and context effects. A more thorough examination of the effects of AV on workload and its factors would reveal whether deficiencies in our experiment masked a significant difference, or if there really is none to be found. We expect that, at least in the context of home use, workload will not be a major issue for AV, and that usability studies could yield results more useful for further informing design choices.

Finally, although we showed that typing performance is better in AV than in VR, it is not yet as good as it is under desktop conditions, particularly with respect to typing speed. We believe that latency and the small offset between optical flow and proprioception, a side-effect of the relative positions of the eyes and the webcam used to achieve AV, are primarily to blame for this. Future work could explore whether or not this is the case, and if so, how these issues can be solved or if they are something that can be overcome with practice.

## 6.5 Conclusion

For the users of Head-Mounted Displays, the real world must, at some point, impinge on the virtual world. This research has shown that our form of inferred minimal blending AV reduces the negative aspects of these impingements on various aspects of user experience and performance. Not only can real-world interactions be made less frustrating for the user, but the user's sense of presence in the virtual world, a primary goal for HMDs, can be preserved throughout these interactions. Using existing theories of presence formation, we provided a novel explanation as to why AV is able to preserve presence. Furthermore, we showed that input performance can be significantly improved using AV compared to pure VR, though there is still room for improvement. We have also demonstrated that a completely general-purpose solution to visual cut-off, that allows successful interactions with any objects of interest, is possible without users having to make any modifications to their environment or the objects therein.



## Appendix A

# Assets Used

TABLE A.1: Visual Assets Used

Asset	Source
Alduin Dragon	<a href="https://free3d.com/3d-model/alduin-dragon-rigged-76875.html">https://free3d.com/3d-model/alduin-dragon-rigged-76875.html</a>
Game Ready Boats	<a href="https://www.assetstore.unity3d.com/en/#!/content/31917">https://www.assetstore.unity3d.com/en/#!/content/31917</a>
Decrepit Dungeon Lite	<a href="https://www.assetstore.unity3d.com/en/#!/content/33936">https://www.assetstore.unity3d.com/en/#!/content/33936</a>
Golden Dragon Statue	<a href="https://www.assetstore.unity3d.com/en/#!/content/63132">https://www.assetstore.unity3d.com/en/#!/content/63132</a>
18 High Resolution Wall Textures	<a href="https://www.assetstore.unity3d.com/en/#!/content/12567">https://www.assetstore.unity3d.com/en/#!/content/12567</a>
Free Rocks	<a href="https://www.assetstore.unity3d.com/en/#!/content/19288">https://www.assetstore.unity3d.com/en/#!/content/19288</a>
GeoPainter	<a href="https://www.assetstore.unity3d.com/en/#!/content/2389">https://www.assetstore.unity3d.com/en/#!/content/2389</a>
Green Forest	<a href="https://www.assetstore.unity3d.com/en/#!/content/22762">https://www.assetstore.unity3d.com/en/#!/content/22762</a>
Horse Statue	<a href="https://www.assetstore.unity3d.com/en/#!/content/52025">https://www.assetstore.unity3d.com/en/#!/content/52025</a>
Handpainted Keys	<a href="https://www.assetstore.unity3d.com/en/#!/content/42044">https://www.assetstore.unity3d.com/en/#!/content/42044</a>
Lion Statue	<a href="https://www.assetstore.unity3d.com/en/#!/content/34247">https://www.assetstore.unity3d.com/en/#!/content/34247</a>
Nature Starter Kit 2	<a href="https://www.assetstore.unity3d.com/en/#!/content/52977">https://www.assetstore.unity3d.com/en/#!/content/52977</a>
Skybox	<a href="https://www.assetstore.unity3d.com/en/#!/content/4183">https://www.assetstore.unity3d.com/en/#!/content/4183</a>
Low Poly Bowl	<a href="https://www.assetstore.unity3d.com/en/#!/content/66776">https://www.assetstore.unity3d.com/en/#!/content/66776</a>

TABLE A.2: Audio Assets Used

Asset Name	Source
Medium Wind	<a href="http://freesound.org/people/kangaroovindaloo/sounds/205966/">http://freesound.org/people/kangaroovindaloo/sounds/205966/</a>
eerie glow	<a href="http://freesound.org/people/waveplay/sounds/187499/">http://freesound.org/people/waveplay/sounds/187499/</a>
Opening a Fizzy Bottle, C.wav	<a href="http://freesound.org/people/InspectorJ/sounds/335993/">http://freesound.org/people/InspectorJ/sounds/335993/</a>
Thin bell ding 2	<a href="http://freesound.org/people/Khrinx/sounds/333695/">http://freesound.org/people/Khrinx/sounds/333695/</a>
Notification Up 1	<a href="http://freesound.org/people/FoolBoyMedia/sounds/234524/">http://freesound.org/people/FoolBoyMedia/sounds/234524/</a>
01543 flying dragon.wav	<a href="https://freesound.org/people/Robinhood76/sounds/93570/">https://freesound.org/people/Robinhood76/sounds/93570/</a>
Fire-breathing dragon.flac	<a href="https://freesound.org/people/CGEffex/sounds/94102/">https://freesound.org/people/CGEffex/sounds/94102/</a>
Monster Growl.wav	<a href="https://freesound.org/people/ztrees1/sounds/134931/">https://freesound.org/people/ztrees1/sounds/134931/</a>
Dragon Land	<a href="https://freesound.org/people/Cyberkineticfilms/sounds/141148/">https://freesound.org/people/Cyberkineticfilms/sounds/141148/</a>
DRAGON_ROAR.wav	<a href="https://freesound.org/people/JoelAudio/sounds/85568/">https://freesound.org/people/JoelAudio/sounds/85568/</a>
Horn dropped on stone floor	<a href="https://freesound.org/people/kessir/sounds/264202/">https://freesound.org/people/kessir/sounds/264202/</a>
Large piece of wood being dropped 04	<a href="https://freesound.org/people/Jagadamba/sounds/258051/">https://freesound.org/people/Jagadamba/sounds/258051/</a>
Squeaky_Gate_01.wav	<a href="https://freesound.org/people/dheming/sounds/174067/">https://freesound.org/people/dheming/sounds/174067/</a>
Grass Footsteps	<a href="https://freesound.org/people/Motion_S/sounds/221756/">https://freesound.org/people/Motion_S/sounds/221756/</a>
Footsteps, Concrete, A.wav	<a href="https://freesound.org/people/InspectorJ/sounds/336598/">https://freesound.org/people/InspectorJ/sounds/336598/</a>
01-21 Footsteps, sneakers on gravel and leaves.wav	<a href="https://freesound.org/people/SpliceSound/sounds/218304/">https://freesound.org/people/SpliceSound/sounds/218304/</a>
Right Gravel Footstep 5	<a href="https://freesound.org/people/Ali_6868/sounds/384880/">https://freesound.org/people/Ali_6868/sounds/384880/</a>
Clothes Rustling 1	<a href="https://freesound.org/people/Sethroph/sounds/334219/">https://freesound.org/people/Sethroph/sounds/334219/</a>
Jared's Gate to Hell.aif	<a href="https://freesound.org/people/oldedgar/sounds/97976/">https://freesound.org/people/oldedgar/sounds/97976/</a>
Fire Crackling 01.wav	<a href="https://freesound.org/people/kingsrow/sounds/181563/">https://freesound.org/people/kingsrow/sounds/181563/</a>
Bonfire Being Lit	<a href="https://freesound.org/people/samararaine/sounds/186374/">https://freesound.org/people/samararaine/sounds/186374/</a>
Metal Gate	<a href="https://freesound.org/people/Czarcazas/sounds/330793/">https://freesound.org/people/Czarcazas/sounds/330793/</a>
Wooden Log Falling	<a href="https://freesound.org/people/jorickhoofd/sounds/161592/">https://freesound.org/people/jorickhoofd/sounds/161592/</a>
toy mechanism	<a href="https://freesound.org/people/baryy/sounds/163409/">https://freesound.org/people/baryy/sounds/163409/</a>
Wind With Rigging.flac	<a href="https://freesound.org/people/Benboncan/sounds/134699/">https://freesound.org/people/Benboncan/sounds/134699/</a>
wind in the trees	<a href="https://freesound.org/people/cajo/sounds/150174/">https://freesound.org/people/cajo/sounds/150174/</a>
Flock of seagulls.wav	<a href="https://freesound.org/people/juskiddink/sounds/98479/">https://freesound.org/people/juskiddink/sounds/98479/</a>
Ocean Noise - Surf	<a href="https://freesound.org/people/amholma/sounds/372181/">https://freesound.org/people/amholma/sounds/372181/</a>
Door Squeak, Normal, E.wav	<a href="https://freesound.org/people/InspectorJ/sounds/346211/">https://freesound.org/people/InspectorJ/sounds/346211/</a>



## Appendix B

# List of Hardware and Specifications

### B.1 Hardware Specifications

**Processor** Intel Core i5 4690

**Ram** 16Gb DDR3

**Motherboard** ASUSTeK (3 x USB2, 2 x USB3)

**Video Card** Nvidia GeForce 970

**Hand Tracker** Leap Motion Controller

**Head-Mounted Display** Oculus Rift CV1

**Game Controller** Wireless Xbox One Controller

**Webcam** Logitech C930e

### B.2 Software Specifications

**Game Engine** Unity3D 5.6.0f3

**Graphics Driver** Nvidia 382.05

**Operating System** Windows 10 64-bit

**Leap Driver** Version 2.3.2



## Appendix C

# Igroup Presence Questionnaire - Participant Version

Questionnaire 1

For each question/statement below, please circle the number on the scale that best applies to the experience you just had.

---

**How aware were you of the real world surrounding while navigating in the virtual world? (i.e. sounds, room temperature, other people, etc.)?**

-3	-2	-1	0	+1	+2	+3
extremely aware			moderately aware		not aware at all	

---

**How real did the virtual world seem to you?**

-3	-2	-1	0	+1	+2	+3
completely real					not real at all	

---

**I had a sense of acting in the virtual space, rather than operating something from outside.**

-3	-2	-1	0	+1	+2	+3
fully disagree			fully agree			

---

**How much did your experience in the virtual environment seem consistent with your real world experience?**

-3	-2	-1	0	+1	+2	+3
not consistent		moderately consistent			very consistent	

---

**How real did the virtual world seem to you?**

-3	-2	-1	0	+1	+2	+3
about as real as an imagined world				indistinguishable from the real world		

---

---

**I did not feel present in the virtual space.**

-3	-2	-1	0	+1	+2	+3
did not feel						felt present

---

**I was not aware of my real environment.**

-3	-2	-1	0	+1	+2	+3
fully disagree						fully agree

---

**In the computer-generated world, I had a sense of "being there".**

-3	-2	-1	0	+1	+2	+3
not at all						very much

---

**Somehow I felt that the virtual world surrounded me.**

-3	-2	-1	0	+1	+2	+3
fully disagree						fully agree

---

**I felt present in the virtual space.**

-3	-2	-1	0	+1	+2	+3
fully disagree						fully agree

---

**I still paid attention to the real environment.**

-3	-2	-1	0	+1	+2	+3
fully disagree						fully agree

---

**The virtual world seemed more realistic than the real world.**

-3	-2	-1	0	+1	+2	+3
fully disagree						fully agree

---

**I felt like I was just perceiving pictures.**

-3	-2	-1	0	+1	+2	+3
fully disagree						fully agree

---

**I was completely captivated by the virtual world.**

-3	-2	-1	0	+1	+2	+3
fully disagree						fully agree

---

## Appendix D

# NASA Task Load Index

## Questionnaire 2

This questionnaire aims to measure the workload you experienced during the tasks you just performed. It consists of three steps.

Step 1: Your experience will be evaluated using six Rating Scales. Please read the definitions for each of the Rating Scales below before proceeding.

RATING SCALE DEFINITIONS		
Title	Endpoints	Descriptions
MENTAL DEMAND	<i>Low/High</i>	How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
PHYSICAL DEMAND	<i>Low/High</i>	How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
TEMPORAL DEMAND	<i>Low/High</i>	How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
PERFORMANCE	<i>good/poor</i>	How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?
EFFORT	<i>Low/High</i>	How hard did you have to work (mentally and physically) to accomplish your level of performance?
FRUSTRATION LEVEL	<i>Low/High</i>	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

Step 2: For each of the fifteen pairs below, circle the Rating Scale Title that represents the more important contributor to workload for the task you just performed. Refer back to the Scale Title definitions if necessary.

<p style="text-align: center;"><b>Effort</b> or <b>Performance</b></p>	<p style="text-align: center;"><b>Temporal Demand</b> or <b>Frustration</b></p>	<p style="text-align: center;"><b>Temporal Demand</b> or <b>Effort</b></p>
<p style="text-align: center;"><b>Physical Demand</b> or <b>Frustration</b></p>	<p style="text-align: center;"><b>Performance</b> or <b>Frustration</b></p>	<p style="text-align: center;"><b>Physical Demand</b> or <b>Temporal Demand</b></p>
<p style="text-align: center;"><b>Physical Demand</b> or <b>Performance</b></p>	<p style="text-align: center;"><b>Temporal Demand</b> or <b>Mental Demand</b></p>	<p style="text-align: center;"><b>Frustration</b> or <b>Effort</b></p>
<p style="text-align: center;"><b>Performance</b> or <b>Mental Demand</b></p>	<p style="text-align: center;"><b>Performance</b> or <b>Temporal Demand</b></p>	<p style="text-align: center;"><b>Mental Demand</b> or <b>Effort</b></p>
<p style="text-align: center;"><b>Mental Demand</b> or <b>Physical Demand</b></p>	<p style="text-align: center;"><b>Effort</b> or <b>Physical Demand</b></p>	<p style="text-align: center;"><b>Frustration</b> or <b>Mental Demand</b></p>

Step 3: Place an "X" on each of the scales below at the point which matches your experience of the task you performed. Refer back to the Scale Title Definitions if necessary.

**MENTAL DEMAND**



**PHYSICAL DEMAND**



**TEMPORAL DEMAND**



**PERFORMANCE**



**EFFORT**



**FRUSTRATION**



## Appendix E

# Ethics Clearance

UNIVERSITY OF CAPE TOWN  
FACULTY OF SCIENCE RESEARCH ETHICS COMMITTEE

**RESEARCHER ETHICS STATEMENT**

A key function of the Faculty of Science Research Ethics Committee is to screen and approve, or otherwise refuse, all research proposals in the Faculty that relate to human subjects (see definition in section 2 overleaf), including questionnaires involving human participants; this includes proposed research involving students or staff, by UCT researchers or by outside visiting researchers. Research that does not involve human subjects does not need to be submitted to this committee for approval. Research on animals needs to be approved by the Faculty's Animal Ethics Committee; and research that uses biological materials from humans (e.g. fresh tissues, blood or body fluids) needs also to be approved by the Faculty's Biological Safety Committee.

This **researcher ethics statement form** and the appended **informed consent form** should be completed by the actual person undertaking the research ('the applicant'). Place the **tick** provided in the **Yes or No** box, and **type** in details where appropriate. Please read the '**UCT Code for Research involving Human Subjects**' before completing the form:

<http://www.uct.ac.za/downloads/uct.ac.za/about/policies/ethicscode.pdf>

In the case of research that involves a number of researchers, this form should be endorsed and signed by the Principal Investigator (PI). If the applicant is a student, the supervisor must endorse and sign the form and ensure that the student is fully informed of his/her ethical responsibilities. Where the research is part of a project that is being co-ordinated from outside the Faculty of Science, the researcher should fill in the form in relation to her or his part of the larger research project. The turnaround time for a reply is approximately 7 working days.

**E-mail** this completed **form** in the original **MS Word format** to:

The Servicing Officer: Ms Shanaaz Smith, Faculty of Science Research Ethics Committee at [shanaaz.smith@uct.ac.za](mailto:shanaaz.smith@uct.ac.za)

**Expedited Review**

Researchers who use participants only to test the **usability** of programmes and applications (typically from those in the Computer Science Department) *and* are not working with any vulnerable populations (e.g. pregnant women, minors, or prisoners) may apply for **expedited review** by ticking "Yes" for question 2. Applications for expedited review are not necessarily faster, but only require the approval of one member of the Science Research Ethics Committee. '**Usability**' in this context is defined to include learnability, efficiency, memorability, accuracy, ease of use and user experience, typically with an artefact or prototype. However, if such 'usability' research also involves the collection of any privately identifiable or sensitive personal data about participants (e.g. information on disabilities, vulnerabilities, health/medical conditions and/or treatments), then the proposal is **NOT** eligible for expedited review. Projects using human subjects other than for 'usability' purposes will be submitted for review by the full committee.

**Title First Name and Surname of applicant:** Mr. Jacob Clarkson

**Contact e-mail address of applicant:** jacobhclarkson@gmail.com

**Title First Name and Surname of Supervisor or Principal Investigator (if applicable)**

Type here: Prof. Edwin Blake

**Contact e-mail of Supervisor or Principal Investigator:** edwin@cs.uct.ac.za

**Title of research project:** Augmented Virtuality for Head-Mounted Displays

**Purpose of research** – place this tick in the appropriate line: ✓

Honours project

Masters by coursework and dissertation

Masters by dissertation only ✓

PhD thesis

Academic research

Contract funded research

Other research (please specify):

**Under which UCT department do these activities fall:**

Computer Science

**Place this tick ✓ in the box below the appropriate Yes or No response:**

1. Have you read the ' <b>UCT Research Ethics Code for Research Involving Human Participants</b> '? This code is available for download from the UCT web-site's listing of policies – scroll down the alphabetical listing to 'Research', where you will find this specific code <a href="http://www.uct.ac.za/about/policies/">http://www.uct.ac.za/about/policies/</a>	Yes ✓	No
2. Are you applying for <b>expedited review</b> ? Note: you are only eligible for expedited review if you are conducting a <b>usability</b> study.	Yes	No ✓
3. Is your research making use of human participants or subjects as sources of data? <b>Human subject</b> means a living individual about whom an investigator (whether professional or student) conducting <b>research</b> obtains (1) data through intervention or interaction with the individual, or (2) identifiable private information, which includes a subject's opinion on a given topic.	Yes ✓	No
4. Is your research being conducted on property for which you need express permission from owner(s), occupier(s) or manager(s), in which case have you requested and received explicit permission to proceed?	Yes	No ✓
5. Is your research being conducted within a National Park or Nature Reserve (or similar) or any other area for which a permit is required, in which case have you obtained the required permit?	Yes	No ✓

**6. Research focus** (maximum 500 words)

In the space below state your research aim and summarise your key research objectives (or questions); briefly outline your plans for data collection, and indicate the nature/type of information you will be seeking from the participants in your research. Do NOT submit supplementary additional documents. Your proposal will be evaluated on information in this form alone.

Please note that ethics proposals are reviewed by a multi-disciplinary committee and should be written in a manner that does not assume specialist knowledge. For this reason, acronyms/abbreviations should be written out in full the first time they are used, followed by the shortened version in brackets.

Research Focus:

My research focus is on Virtual Reality. Specifically, I aim to create and test a system that allows people to see through a Virtual Reality headset (Oculus Rift) to allow them to easily interact with the real world while wearing the device. This technique is referred to as Augmented Virtuality. I will be comparing user experience and performance in Augmented Virtuality conditions to those in purely virtual conditions (with no AV system).

Research Questions:

1. Is typing performance in Augmented Virtuality better than in pure virtuality?
2. Is it easier to interact with real-world objects in Augmented Virtuality than pure virtuality?
3. Is Presence lower in Augmented Virtuality than it is in pure virtuality?

Plans for Data Collection:

To gather data, a series of user experiments will be conducted. Users will play an immersive 3D game while wearing a Head-Mounted Display. During the course of this game, and in order to proceed through it, participants will have to enter text via a keyboard, and interact with several real-world objects that are in front of them.

All data required to answer research question 1 will be gathered by the game while it is being played. This data will take the form of all keystrokes entered during the game, and the time taken to enter them.

Data required to answer questions 2 and 3 will be gathered via post-experimental questionnaires. Presence will be measured using the ITC-Sense of Presence Inventory. Ease of real-world interaction will be measured through the NASA Task Load Index (NASA-TLX) questionnaire, which is designed to measure cognitive load.

Participants will be comprised primarily of undergraduate and honours Computer Science students from UCT (convenience sampling). The experiment will be advertised around the department via posters. Depending on the response, I may also recruit further participants in person.

Before pilot tests and subsequent power analyses are conducted, it is difficult to say exactly how many participants will be required to yield statistically significant data. I expect that the final number of participants required will be between 30 and 40.

<p>7. <b>Information</b> – place tick in the box below the appropriate response: ✓  Will participants (research subjects) have reasonable and sufficient knowledge about you, your background and location, and your research intentions?  By ticking the ‘Yes’ box, you declare that you have completed and will use the <b>informed consent form</b> appended to this statement, and that you will explain the content verbally to each participant.  Any other information you want to provide to strengthen your application w.r.t. the provision of information may be included in the box below. If your answer is ‘No’, please provide justification.</p>	Yes ✓	No
--	----------	----

Before each participant begins the experiment, they will be given a document to read. This document will contain my name and contact details, as well as my background and location, and the field of this research (user experience in virtual reality). No information about my identity will be withheld. The specific measures to be used in the experiment will not be mentioned to the participant, as this may bias their performance and responses. This is a phenomenon known as the Hawthorne Effect, and the single-blind nature of this experiment aims to minimize it. Note that the title of the research presented to participants will not be the official title of the research, as the control group in the experiment will not be exposed to the Augmented Virtuality Condition. That is to say, all participants will wear the HMD for the experiment, but only those in the experimental condition will have the additional functionality that allows them to “see through” the device via webcam enabled.

<p>8. <b>Consent</b> – place tick in the box below the appropriate response: ✓  Will you secure informed, written consent of all participants in the research?  By ticking the ‘Yes’ box, you declare that you:  • will commit to getting each participant to sign the <b>informed consent form</b> appended to this statement, <b>before</b> you engage with them, and  • will give the participant a copy of the signed form and keep a second copy for yourself.  If your answer is ‘No’, please provide procedure for securing ethical consent of all participants as well as reasons for waiver of prior and/or written consent.  Any other information you want to provide with regard to consent may also be included in the box below.</p>	Yes ✓	No
--	----------	----

<p>9. <b>Permission</b> – place tick in the box below the appropriate response: ✓  Does your research intend to make use of UCT students as participants?  If yes, you must apply to Dr Moonira Khan, Executive Director: Department of Student Affairs (DSA) for approval to conduct this research, but only after you have received ethical clearance from the Faculty of Science Research Ethics Committee. The form for this approval (DSA100) is available at <a href="http://www.science.uct.ac.za/usr/science/research/dsa100.doc">www.science.uct.ac.za/usr/science/research/dsa100.doc</a>  For research involving UCT staff, approval to conduct the research must be</p>	Yes ✓	No
---	----------	----

obtained from the Executive Director: Human Resources.		
--	--	--

<p><b>10. Confidentiality</b> – place tick in the box below the appropriate response: ✓</p> <p>Are you able to offer privacy and confidentiality to participants, if they wish to remain anonymous?</p> <p>The default requirements of the Faculty of Science Research Ethics Committee are to assure that either:</p> <p>(a) study data are de-identified (identifiers are stripped or separated), or</p> <p>(b) data are collected without identifiers (anonymous).</p> <p>If you wish to use the names and organisational affiliations of participants in your research:</p> <p>(i) tick 'No';</p> <p>(ii) provide a reasoned motivation in the box below why you are adopting this approach, indicating why this does not have ethical implications for the participants, and</p> <p>(iii) modify the appended prior informed consent form appropriately so that it reflects a participant's agreement that you may use his or her name and/or affiliation together with the information they provided.</p> <p>If there are any aspects of your research where there might be difficulties or problems with regard to protecting the confidentiality and rights of participants, and honouring their trust, explain this in detail below.</p>	<p>Yes ✓</p>	<p>No</p>
---	------------------	-----------

Student numbers and email addresses of participants will have to be recorded as proof for remuneration, as they will be paid for their participation. This information will not be linked to data gathered from the participant. No other personal information will be recorded, and it will not be possible to trace any data gathered to the participant that yielded it. Participants will be rewarded R50 for their time.

<p><b>11. Potential for harm to participants</b> – tick below appropriate response: ✓</p> <p>Are there any foreseeable risks of legal, physical, psychological or social harm or suffering to participants and/or the environment, which might result from, or occur in the course of, this research?</p> <p>If your answer is 'Yes', outline below what these risks might be and what preventative steps you plan to take to avoid or minimise such harm from being suffered, and include a summary of these risks in the appended prior informed consent form. Residual risks are to be balanced by your response to question 13 below (on the benefits of the research).</p>	<p>Yes ✓</p>	<p>No</p>
---	------------------	-----------

In rare cases, the usage of a Head-Mounted Display such as the Oculus Rift can cause nausea in the wearer. As far as is possible, the virtual environment and navigational controls will be designed to mitigate this risk. Information about this risk will form part of the Informed Consent document, and participants will be told that they are free to halt the experiment at any time if they feel at all uncomfortable.

<p><b>12. Potential for harm to UCT or other institutions</b>          Place tick in the box below the appropriate response: ✓          Are there any foreseeable risks of harm to UCT, or to other institutions, that might result from or occur in the course of the research, for example, legal action resulting from the research; or the image of the university or another institution being adversely affected by association with the research (such as a school being compromised in the eyes of the Department of Education)?          If your answer is 'Yes', give details below (to be balanced by your response to question 13 below).</p>	<p>Yes</p>	<p>No  ✓</p>

<p><b>13. Other conceivable ethical issues</b> – tick below appropriate response: ✓          Are there any other ethical issues that you think might arise during the course of the research? (e.g., with regard to conflicts of interest amongst participants and/or institutions). If your answer is 'Yes', give details in the box below and say what you plan to do to minimise any adverse consequences (to be balanced by your response to question 13 below).</p>	<p>Yes</p>	<p>No ✓</p>

**14. Benefits to Science, to participants, and others**

Note: The core task of research ethics committees is to balance the benefits of research against risks or potential harm that may ensue, as per sections 11, 12 and 13 of this form. In the space below summarise the benefits of your research.

Benefits to Students: I anticipate that none of the participants will ever have used Virtual Reality technology before, as, in its current incarnation, it is very new. Participating in my research will be an opportunity to experience the cutting edge of digital entertainment technology first hand.

Benefits to Science: Virtual Reality is expected to become very popular in the next few years, with the market estimated to be worth approximately \$150 billion by 2020. As such, it is important for us to try to learn about all aspects of this technology. In terms of improving the usability of head-mounted displays using mixed reality techniques, only two research papers, both published in 2015, currently exist. This research is therefore an opportunity to contribute to knowledge in this area.

<p><b>15. Publication of the results</b> – place tick below the appropriate response: ✓ Research projects ideally result in publication of the results. Have you and your Supervisor/PI read and agreed to the principles regarding authorship as set out in the UCT 'Authorship Practices Policy'? This code is available for download from the UCT web-site's list of policies – scroll down the alphabetical listing to 'Research', where you will find this specific code <a href="http://www.uct.ac.za/about/policies/">http://www.uct.ac.za/about/policies/</a></p>	Yes	No
	✓	

**Commitment by Applicant**

**Title First Name and Surname:** Mr. Jacob Clarkson

**Electronic (typed) signature of applicant** - this is not a requirement for a scan of your signature, just type your name here: Jacob H. Clarkson

**Date:** 30/06/2016

**Endorsement by Supervisor or Principal Investigator (as applicable)**

*By signing below, I certify that I have assisted the applicant to identify ethical issues pertaining to his or her research; that I have reviewed this ethics application, including the informed consent form overleaf, and am satisfied that it is accurate and adequately communicates information about the proposed research.*

**Title First Name and Surname:** Prof. Edwin Blake

**Electronic (typed) signature of endorsement** - this is not a requirement for a scan of your signature, just type your name here: Edwin H. Blake

**Date:** 30/06/2016

## Appendix F

# Informed Consent Document

## DEPARTMENT OF COMPUTER SCIENCE

UNIVERSITY OF CAPE TOWN  
PRIVATE BAG X3  
RONDEBOSCH 7701  
SOUTH AFRICA

RESEARCHER'S  
TELEPHONE: 078 348 4450  
FACSIMILE:  
E-MAIL: [jacobhclarkson@gmail.com](mailto:jacobhclarkson@gmail.com)  
URL: [people.cs.uct.ac.za/~jclarkson/](http://people.cs.uct.ac.za/~jclarkson/)



### Informed Voluntary Consent to Participate in Research Study

**Project Title:** Mixed Reality Interaction

**Invitation to participate, and benefits:** You are invited to participate in a research study conducted with virtual reality technology – a head-mounted display. The study aim is to develop an understanding of usability and task performance of a new type of computer game that requires both real world interaction, and interaction in the virtual world. I believe that your experience would be a valuable source of information, and hope that by participating you may gain useful knowledge. You will be rewarded a total of R50 for your participation.

**Procedures:** During this study, you will be asked to perform several tasks while wearing a Head-Mounted Display (Oculus Rift CV1). Afterwards, you will be asked to fill in two questionnaires about your experience.

**Risks:** There is a small chance that you may feel nauseous as a result of wearing the Head-Mounted Display. If you feel at all uncomfortable for any reason during the course of the experiment, you have the right to halt the experiment (see below).

**Disclaimer/Withdrawal:** Your participation is completely voluntary; you may refuse to participate, and you may withdraw at any time without having to state a reason and without any prejudice or penalty against you. Should you choose to withdraw, the researcher commits not to use any of the information you have provided without your signed consent. Note that the researcher may also withdraw you from the study at any time.

**Confidentiality:** All information collected in this study will be kept private in that you will not be identified by name or by affiliation to an institution. Confidentiality and anonymity will be maintained as pseudonyms will be used. Your student number will be recorded in order for you to receive your remuneration. However, this information will not be associated with the data gathered from your participation, and will not be used or published in any way.

#### What signing this form means:

By signing this consent form, you agree to participate in this research study. The aim, procedures to be used, as well as the potential risks and benefits of your participation have been explained verbally to you in detail, using this form. Refusal to participate in or withdrawal from this study at any time will have no effect on you in any way. You are free to contact me, to ask questions or request further information, at any time during this research.

I agree to participate in this research (tick one box)

Yes     No    \_\_\_\_\_ (Initials)

_____ Name of Participant	_____ Signature of Participant	_____ Date
_____ Name of Researcher	_____ Signature of Researcher	_____ Date

## Appendix G

# Tutorial Script

- Thank you for agreeing to participate in this experiment. Before we begin, we are first going to calibrate the headset for you so that the image you see while wearing it is as sharp as possible. After that, I am going to guide you through a short tutorial that will familiarize you with the mechanics, user interface, and controls you will need to understand to play a virtual reality game.
- First, we are going to calibrate the headset for you so that the images you see are as sharp as possible. Put the headset on now and make sure the straps are tight enough to hold it in place securely.
- Now please navigate to the lens calibration section of the menu and follow the on-screen instructions to adjust the lens spacing. Let me know when you have done this.
- Now I am going to start the tutorial. Here you will learn how to play the game that forms the main part of this experiment.
- You should now be standing in a white room. To look around, turn your head from left to right and up and down. Try this now.
- To turn, the easiest thing to do is to turn your whole body by swivelling in your chair. Try this now.
- Do you notice the pink dot in the middle of the screen? That indicates the direction where you are looking, and also tells you information about the object that you are looking at.
- Look at the barrel now. Do you notice how the pink dot has a pink ring around it when you are looking at the barrel? That means that the barrel is an interactive object.
- To interact with an object, you can press the A button on the controller. However, you are currently too far away from the barrel to interact with it. Use the left analogue stick on the controller to move towards it.
- Now that you are close to the barrel, notice that the dot and ring in the center of the screen turn green when you look at it? This means that you are close enough to interact with it.
- Now press the A button on the controller while looking at the barrel to interact with it. This is how you will interact with all objects in the game.
- Finally, lets learn about the objective box at the top of the screen. At any point in time, your goal in the game will be shown in that box. You can toggle whether or not the box is showing by pressing the B button on your controller.

- If the message box shows the green hands icon which you can see now, it means that the instruction pertains to the real world. If there is no green hands icon, then the instruction pertains to the virtual world.
- (Control Condition) If the green hands icon is showing you have to lift the headset off your eyes so that you can see the real world and interact there.
- (Experimental Condition) If the green hands icon is showing, you can reach out with your hands and a window into the real world will be shown, so you can interact with the real world without having to take off the headset. Try doing this now.
- If you want to practice a bit more, or have any questions about the controls, please feel free to ask me. When you are comfortable with the controls and user interface, let me know and we can move on to the game.

## Appendix H

# Statistical Analysis of Touch Typist Performance

This Appendix contains the results of statistical analysis on data collected from participants who were able to touch type (type without looking at the keyboard). As one might expect, there were no statistically significant differences in typing performance between the control and experimental conditions for these participants. Table H.1 shows the descriptive statistics for accuracy (ACC), gross typing speed (GTS), and net typing speed (NTS) that control (VR) and experimental (AV) participants obtained during their treatment typing tests. Table H.2 shows the results of independent samples t-tests on this data. There were no significant differences in typing performance between the two groups on any metric.

TABLE H.1: Descriptive Statistics for Touch Typists

Condition	Subscale	M	SD	N
VR	ACC	90.5367	5.47585	3
	GTS	49.3333	15.30795	3
	NTS	40.6667	8.14453	3
AV	ACC	95.0800	2.73207	4
	GTS	50.0000	13.16561	4
	NTS	47.5000	12.47664	4

TABLE H.2: Results of independent samples t-tests comparing treatment typing performance of control and experimental groups.

	Levene's Test		T-test	
	F	p	t	p
ACC	2.298	.190	-1.466	.203
GTS	.266	.628	-.062	.953
NTS	.229	.653	-.817	.451



## Appendix I

# IGroup Presence Questionnaire Items

TABLE I.1: Items of the IGroup Presence Questionnaire Divided by  
Factor

Factor	Question
<b>General Presence (GP)</b>	In the computer generated world I had a sense of "being there".
<b>Spatial Presence (SP)</b>	Somehow I felt that the virtual world surrounded me.
	I felt like I was just perceiving pictures.
	I did not feel present in the virtual space.
	I had a sense of acting in the virtual space, rather than operating something from outside.
	I felt present in the virtual space.
<b>Involvement (INV)</b>	How aware were you of the real world surrounding while navigating in the virtual world? (i.e. sounds, room temperature, other people, etc.)?
	I was not aware of my real environment.
	I still paid attention to the real environment.
	I was completely captivated by the virtual world.
<b>Realism (REAL)</b>	How real did the virtual world seem to you?
	How much did your experience in the virtual environment seem consistent with your real world experience ?
	How real did the virtual world seem to you?
	The virtual world seemed more realistic than the real world.



## Appendix J

# Minimum String Distance Statistic Algorithm

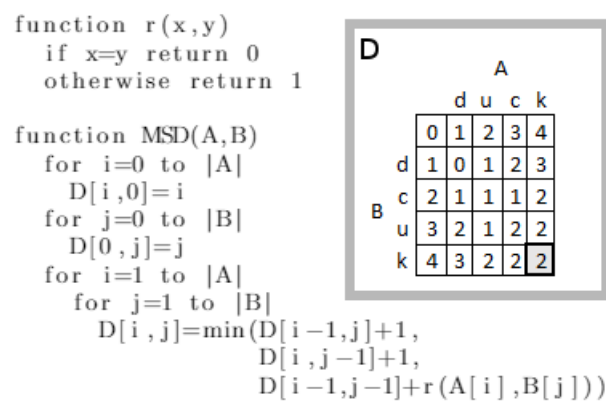


FIGURE J.1: Algorithm to calculate Minimum String Distance between strings A and B using the Matrix D. From (Kruskal, 1983), as presented in (Soukoreff and MacKenzie, 2001).



# Bibliography

- Andersen, Erik et al. (2012). "The impact of tutorials on games of varying complexity". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, pp. 59–68.
- Billinghamurst, Mark and Hirokazu Kato (1999). "Collaborative mixed reality". In: *Proceedings of the First International Symposium on Mixed Reality*, pp. 261–284.
- Biocca, Frank (2003). "Can we resolve the book, the physical reality, and the dream state problems? From the two-pole to a three-pole model of shifts in presence". In: *EU Future and Emerging Technologies, Presence Initiative Meeting*.
- Blizzard (2004). *Blizzard Entertainment*. <http://us.blizzard.com/en-us/>. Accessed: April-2016.
- Bruner, Jerome S and Leo Postman (1949). "On the perception of incongruity: A paradigm". In: *Journal of personality* 18.2, pp. 206–223.
- Budhiraja, Pulkit et al. (2015). "Where's My Drink? Enabling Peripheral Real World Interactions While Using HMDs". In: *arXiv preprint arXiv:1502.04744*.
- Cain, Brad (2007). *A review of the mental workload literature*. Tech. rep. Defence Research And Development Toronto (Canada).
- Chan, Liwei and Kouta Minamizawa (2017). "FrontFace: Facilitating communication between HMD users and outsiders using front-facing-screen HMDs". In: *19th International Conference on Human-Computer Interaction with Mobile Devices and Services, MobileHCI 2017*. Association for Computing Machinery, Inc.
- Clifton, James, Leslie M Scatone, and Emine Fetvaci (2009). *The plains of Mars: European war prints, 1500-1825, from the collection of the Sarah Campbell Blaffer Foundation*. Yale Univ Pr.
- Cohen, Jacob (1988). "Statistical analysis for the behavioral sciences". In: *Hillsdale: Lawrence Erlbaum*.
- Corwin, William H et al. (1989). *Assessment of Crew Workload Measurement Methods, Techniques and Procedures. Volume 1. Process, Methods and Results*. Tech. rep. DOUGLAS AIRCRAFT CO LONG BEACH CA.
- Cummings, James J and Jeremy N Bailenson (2016). "How immersive is enough? A meta-analysis of the effect of immersive technology on user presence". In: *Media Psychology* 19.2, pp. 272–309.
- De Waard, Dick (1996). *The measurement of drivers' mental workload*. Groningen University, Traffic Research Center Netherlands.
- Desai, Amit P, Lourdes Pena-Castillo, and Oscar Meruvia-Pastor (2017). "A Window to your Smartphone: Exploring Interaction and Communication in Immersive VR with Augmented Virtuality". In:
- Digicapital (2016). *Digicapital*. <http://www.digi-capital.com/news/2015/04/augmentedvirtual-reality-to-hit-150-billion-disrupting-mobile-by-2020/>. Accessed: August-2017.
- Edwards, Emily K, Jannick P Rolland, and Kurtis P Keller (1993). "Video see-through design for merging of real and virtual environments". In: *Virtual Reality Annual International Symposium, 1993., 1993 IEEE*. IEEE, pp. 223–233.

- Eggemeier, F Thomas and Glenn F Wilson (1991). "Performance-based and subjective assessment of workload in multi-task environments". In: *Multiple-task performance*, pp. 217–278.
- Ellis, Stephen R (1996). "Presence of mind: a reaction to Thomas Sheridan's "further musings on the psychophysics of presence"". In: *Presence: Teleoperators & Virtual Environments* 5.2, pp. 247–259.
- Foehr, Ulla G (2006). "Media multitasking among American youth: Prevalence, predictors and pairings." In: *Henry J. Kaiser Family Foundation*.
- Fritz, Catherine O, Peter E Morris, and Jennifer J Richler (2012). "Effect size estimates: current use, calculations, and interpretation." In: *Journal of experimental psychology: General* 141.1, p. 2.
- gamefaqs (2015). *Metal Gear Solid V: The Phantom Pain - No text chat in MGO3!!!!* <http://www.gamefaqs.com/boards/718564-metal-gear-solid-v-the-phantom-pain/72609402>. Accessed: April-2016.
- Gregory, Richard L (2015). *Eye and brain: The psychology of seeing*. Princeton university press.
- Gysbers, Andre et al. (2004). "Exploring the book problem: Text design, mental representations of space, and spatial presence in readers". In: *Seventh Annual International Workshop: Presence*.
- Hart, Sandra G (2006). "NASA-task load index (NASA-TLX); 20 years later". In: *Proceedings of the human factors and ergonomics society annual meeting*. Vol. 50. 9. Sage Publications, pp. 904–908.
- Hart, Sandra G and Lowell E Staveland (1988). "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research". In: *Advances in psychology* 52, pp. 139–183.
- Hart, Sandra G and Christopher D Wickens (1990). "Workload assessment and prediction". In: *Manprint*. Springer, pp. 257–296.
- Hart, SG and L Staveland (1986). *NASA Task Load Index (TLX) V1. 0 Users Manual*.
- Hartmann, Tilo et al. (2015). "Spatial presence theory: State of the art and challenges ahead". In: *Immersed in Media*. Springer, pp. 115–135.
- Heeter, Carrie (1992). "Being there: The subjective experience of presence". In: *Presence: Teleoperators & Virtual Environments* 1.2, pp. 262–271.
- Hendrix, Claudia and Woodrow Barfield (1996a). "Presence within virtual environments as a function of visual display parameters". In: *Presence: Teleoperators & Virtual Environments* 5.3, pp. 274–289.
- (1996b). "The sense of presence within auditory virtual environments". In: *Presence: Teleoperators & Virtual Environments* 5.3, pp. 290–301.
- Herring, Susan C et al. (2009). "Fast talking, fast shooting: Text chat in an online first-person game". In: *System Sciences, 2009. HICSS'09. 42nd Hawaii International Conference on*. IEEE, pp. 1–10.
- Hutchison, Andrew (2007). "Back to the Holodeck: new life for virtual reality?" In: *Proceedings of the 2nd international conference on Digital interactive media in entertainment and arts*. ACM, pp. 98–104.
- Igroup (2016). *Igroup Presence Questionnaire (IPQ) Database*. <http://www.igroup.org/pq/ipq/data.php>. Accessed: November-2017.
- Ijsselstein, Wijnand A et al. (2000). "Presence: concept, determinants, and measurement". In: *Electronic Imaging*. International Society for Optics and Photonics, pp. 520–529.
- ISPR (2017). *Presence Defined*. <https://ispr.info/about-presence-2/about-presence/>. Accessed: December-2017.

- Jensen, Carlos et al. (2000). "The effect of communication modality on cooperation in online environments". In: *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. ACM, pp. 470–477.
- Kim, Taeyong and Frank Biocca (1997). "Telepresence via television: Two dimensions of telepresence may have different connections to memory and persuasion.[1]". In: *Journal of Computer-Mediated Communication* 3.2, pp. 0–0.
- Konami (2015). *Konami*. <https://www.konami.com/en/>. Accessed: April-2016.
- Kramer, Arthur F (1991). "Physiological metrics of mental workload: A review of recent progress". In: *Multiple-task performance*, pp. 279–328.
- Kruskal, Joseph B (1983). "An overview of sequence comparison: Time warps, string edits, and macromolecules". In: *SIAM review* 25.2, pp. 201–237.
- Lee, Yongjae and Gerard J Kim (2017). "Vitty: Virtual Touch Typing Interface with Added Finger Buttons". In: *International Conference on Virtual, Augmented and Mixed Reality*. Springer, pp. 111–119.
- Lessiter, Jane et al. (2001). "A cross-media presence questionnaire: The ITC-Sense of Presence Inventory". In: *Presence* 10.3, pp. 282–297.
- Levene, Howard et al. (1960). "Robust tests for equality of variances". In: *Contributions to probability and statistics* 1, pp. 278–292.
- Levenshtein, Vladimir I (1966). "Binary codes capable of correcting deletions, insertions and reversals". In: *Soviet physics doklady*. Vol. 10, p. 707.
- Lichtman, Howard S (2006). "Telepresence, effective visual collaboration and the future of global business at the speed of light". In: *HPL, Human Productivity Lab Magazine*.
- Lin, Jia-Wei et al. (2017). "Visualizing the keyboard in virtual reality for enhancing immersive experience". In: *ACM SIGGRAPH 2017 Posters*. ACM, p. 35.
- Lombard, Matthew and Theresa Ditton (1997). "At the heart of it all: The concept of presence". In: *Journal of Computer-Mediated Communication* 3.2, pp. 0–0.
- Lombard, Matthew and Matthew T Jones (2015). "Defining presence". In: *Immersed in Media*. Springer, pp. 13–34.
- MacKenzie, I Scott and R William Soukoreff (2002). "A character-level error analysis technique for evaluating text entry methods". In: *Proceedings of the second Nordic conference on Human-computer interaction*. ACM, pp. 243–246.
- Mann, Henry B and Donald R Whitney (1947). "On a test of whether one of two random variables is stochastically larger than the other". In: *The annals of mathematical statistics*, pp. 50–60.
- Manninen, Tony (2001). "Virtual team interactions in networked multimedia games". In: *In Proceedings of the 4th Annual International Workshop on Presence (PRESENCE 2001)*.
- McGill, Mark et al. (2015). "A dose of reality: overcoming usability challenges in VR head-mounted displays". In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, pp. 2143–2152.
- Metzger, Paul J (1993). "Adding reality to the virtual". In: *Virtual Reality Annual International Symposium, 1993., 1993 IEEE*. IEEE, pp. 7–13.
- Milgram, Paul and Herman Colquhoun (1999). "A taxonomy of real and virtual world display integration". In: *Mixed reality: Merging real and virtual worlds* 1, pp. 1–26.
- Milgram, Paul and Fumio Kishino (1994). "A taxonomy of mixed reality visual displays". In: *IEICE TRANSACTIONS on Information and Systems* 77.12, pp. 1321–1329.

- Milgram, Paul et al. (1995). "Augmented reality: A class of displays on the reality-virtuality continuum". In: *Photonics for industrial applications*. International Society for Optics and Photonics, pp. 282–292.
- Miller, Sarah (2001). "Literature Review: Workload Measures". In: *University of Iowa Document (ID: N01-006)*.
- Minecraftpsyco. *The Sensorama Machine*. <https://commons.wikimedia.org/w/index.php?curid=47304870>. Accessed: August-2017.
- Minsky, Marvin (1980). "Telepresence". In:
- Mou, Weimin and Timothy P McNamara (2002). "Intrinsic frames of reference in spatial memory." In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 28.1, p. 162.
- Niantic (2016). *Pokemon Go*. <http://www.pokemongo.com/>. Accessed: October-2016.
- Nunez, David (2007). "A capacity limited, cognitive constructionist model of virtual presence". PhD thesis. University of Cape Town.
- Oculus (2016). *Oculus Rift*. <https://www3.oculus.com/en-us/rift/>. Accessed: October-2016.
- Ostrach, Teresia R (1997). "Typing speed: How fast is average: 4,000 typing scores statistically analyzed and interpreted". In: *Orlando, FL: Five Star Staffing*.
- Peña, Jorge and Jeffrey T Hancock (2006). "An analysis of socioemotional and task communication in online multiplayer video games". In: *Communication research* 33.1, pp. 92–109.
- Pino, L. *Oculus Rift*. <http://www.techradar.com/reviews/gaming/gaming-accessories/oculus-rift-1123963/review>. Accessed: August-2017.
- Posner, Michael I (1980). "Orienting of attention". In: *Quarterly journal of experimental psychology* 32.1, pp. 3–25.
- Prothero, Jerrold D et al. (1995). "Towards a robust, quantitative measure for presence". In: *Proceedings of the conference on Experimental Analysis and Measurement of Situation Awareness*, pp. 359–366.
- Rabbit-Duck Illusion*. [https://commons.wikimedia.org/wiki/File:Kaninchen\\_und\\_Ent\\_e.png](https://commons.wikimedia.org/wiki/File:Kaninchen_und_Ent_e.png). Accessed: August-2017.
- Regenbrecht, Holger T, Thomas W Schubert, and Frank Friedmann (1998). "Measuring the sense of presence and its relations to fear of heights in virtual environments". In: *International Journal of Human-Computer Interaction* 10.3, pp. 233–249.
- Reid, Gary B and Thomas E Nygren (1988). "The subjective workload assessment technique: A scaling procedure for measuring mental workload". In: *Advances in psychology* 52, pp. 185–218.
- Sacau, Ana et al. (2005). "The impact of personality factors on the experience of spatial presence". In: *The 8th International Workshop on Presence (Presence 2005)*, pp. 143–151.
- Sawilowsky, Shlomo S (2009). "New effect size rules of thumb". In:
- Schild, Jonas, Joseph LaViola, and Maic Masuch (2012). "Understanding user experience in stereoscopic 3D games". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, pp. 89–98.
- Schubert, Thomas, Frank Friedmann, and Holger Regenbrecht (2001). "The experience of presence: Factor analytic insights". In: *Presence* 10.3, pp. 266–281.
- Schubert, Thomas W (2009). "A new conception of spatial presence: once again, with feeling". In: *Communication Theory* 19.2, pp. 161–187.
- Schuemie, Martijn J et al. (2001). "Research on presence in virtual reality: A survey". In: *CyberPsychology & Behavior* 4.2, pp. 183–201.

- shakko. *Battle of Borodino Panorama*. <https://commons.wikimedia.org/w/index.php?curid=19787538>. Accessed: August-2017.
- Shapiro, Samuel Sanford and Martin B Wilk (1965). "An analysis of variance test for normality (complete samples)". In: *Biometrika* 52.3/4, pp. 591–611.
- Sheridan, Thomas B (1992). "Musings on telepresence and virtual presence". In: *Presence: Teleoperators & Virtual Environments* 1.1, pp. 120–126.
- Slater, Mel (1999). "Measuring presence: A response to the Witmer and Singer presence questionnaire". In: *Presence: Teleoperators and Virtual Environments* 8.5, pp. 560–565.
- (2002). "Presence and the sixth sense". In: *Presence: Teleoperators and Virtual Environments* 11.4, pp. 435–439.
- (2003). "A note on presence terminology". In: *Presence connect* 3.3, pp. 1–5.
- (2009). "Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 364.1535, pp. 3549–3557.
- Slater, Mel, D-P Pertaub, and Anthony Steed (1999). "Public speaking in virtual reality: Facing an audience of avatars". In: *IEEE Computer Graphics and Applications* 19.2, pp. 6–9.
- Slater, Mel and Anthony Steed (2000). "A virtual presence counter". In: *Presence: Teleoperators and virtual environments* 9.5, pp. 413–434.
- Slater, Mel, Anthony Steed, and Martin Usoh (1995). "The virtual treadmill: A naturalistic metaphor for navigation in immersive virtual environments". In: *Virtual Environments '95*. Springer, pp. 135–148.
- Slater, Mel and Martin Usoh (1993a). "Presence in immersive virtual environments". In: *Virtual Reality Annual International Symposium, 1993., 1993 IEEE*. IEEE, pp. 90–96.
- (1993b). "Representations systems, perceptual position, and presence in immersive virtual environments". In: *Presence: Teleoperators & Virtual Environments* 2.3, pp. 221–233.
- Slater, Mel, Martin Usoh, and Anthony Steed (1994). "Depth of presence in virtual environments". In: *Presence: Teleoperators & Virtual Environments* 3.2, pp. 130–144.
- (1995). "Taking steps: the influence of a walking technique on presence in virtual reality". In: *ACM Transactions on Computer-Human Interaction (TOCHI)* 2.3, pp. 201–219.
- Slater, Mel and Sylvia Wilbur (1997). "A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments". In: *Presence: Teleoperators and virtual environments* 6.6, pp. 603–616.
- Soukoreff, R William and I Scott MacKenzie (2001). "Measuring errors in text entry tasks: an application of the Levenshtein string distance statistic". In: *CHI'01 extended abstracts on Human factors in computing systems*. ACM, pp. 319–320.
- (2003). "Metrics for text entry research: an evaluation of MSD and KSPC, and a new unified error metric". In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, pp. 113–120.
- Steam (2016). *Metal Gear Solid V: The Phantom Pain - I miss text chat in mgo*. <https://steamcommunity.com/app/287700/discussions/0/458606248617248420/>. Accessed: April-2016.
- Steuer, Jonathan (1992). "Defining virtual reality: Dimensions determining telepresence". In: *Journal of communication* 42.4, pp. 73–93.
- Sutherland, Ivan E (1965). "The ultimate display". In: *Multimedia: From Wagner to virtual reality*.

- Sutherland, Ivan E (1968). "A head-mounted three dimensional display". In: *Proceedings of the December 9-11, 1968, fall joint computer conference, part I*. ACM, pp. 757–764.
- Suznjevic, Mirko, Ognjen Dobrijevic, and Maja Matijasevic (2009). "Hack, slash, and chat: a study of players' behavior and communication in MMORPGs". In: *Proceedings of the 8th Annual Workshop on Network and Systems Support for Games*. IEEE Press, p. 2.
- Tamborini, Ron and Paul Skalski (2006). "The role of presence in the experience of electronic games". In: *Playing video games: Motives, responses, and consequences*, pp. 225–240.
- Tukey, John W (1977). "Exploratory data analysis". In:
- Tyler, Chrisopher and Michael Kubovy (2004). *The Rise of Renaissance Perspective*. <http://www.webexhibits.org/sciartperspective/raphaelperspective1.html>. Accessed: August-2017.
- Valve (2016). *Counter-Strike*. <http://blog.counter-strike.net/>. Accessed: October-2016.
- Van Baren, Joy and Wijnand IJsselsteijn (2004). "Measuring presence: A guide to current measurement approaches". In: *Deliverable of the OmniPres project IST-2001-39237*.
- View-Master Model G*. [https://commons.wikimedia.org/wiki/File:View-Master\\_Model\\_G.jpg](https://commons.wikimedia.org/wiki/File:View-Master_Model_G.jpg). Accessed: August-2017.
- Walker, James (2017). "Improving everyday computing tasks with head-mounted displays". In:
- Walker, James et al. (2017). "Efficient Typing on a Visually Occluded Physical Keyboard". In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, pp. 5457–5461.
- Weibel, David and Bartholomäus Wissmath (2011). "Immersion in computer games: The role of spatial presence and flow". In: *International Journal of Computer Games Technology* 2011, p. 6.
- Welch, Bernard L (1947). "The generalization of student's problems when several different population variances are involved". In: *Biometrika* 34.1/2, pp. 28–35.
- Welch, Robert B (1999). "How can we determine if the sense of presence affects task performance?" In: *Presence: Teleoperators and Virtual Environments* 8.5, pp. 574–577.
- Wheatstone Mirror Stereoscope*. [https://commons.wikimedia.org/wiki/File:Charles\\_Wheatstone-mirror\\_stereoscope\\_XIXc.jpg](https://commons.wikimedia.org/wiki/File:Charles_Wheatstone-mirror_stereoscope_XIXc.jpg). Accessed: August-2017.
- Wirth, Werner et al. (2007). "A process model of the formation of spatial presence experiences". In: *Media psychology* 9.3, pp. 493–525.
- Witmer, Bob G and Michael J Singer (1998). "Measuring presence in virtual environments: A presence questionnaire". In: *Presence: Teleoperators and virtual environments* 7.3, pp. 225–240.
- Wu, Jen-Her, Shu-Ching Wang, and Ho-Huang Tsai (2010). "Falling in love with online games: The uses and gratifications perspective". In: *Computers in Human Behavior* 26.6, pp. 1862–1871.
- Zhao, Shanyang (2003). "Toward a taxonomy of copresence". In: *Presence: Teleoperators and Virtual Environments* 12.5, pp. 445–455.