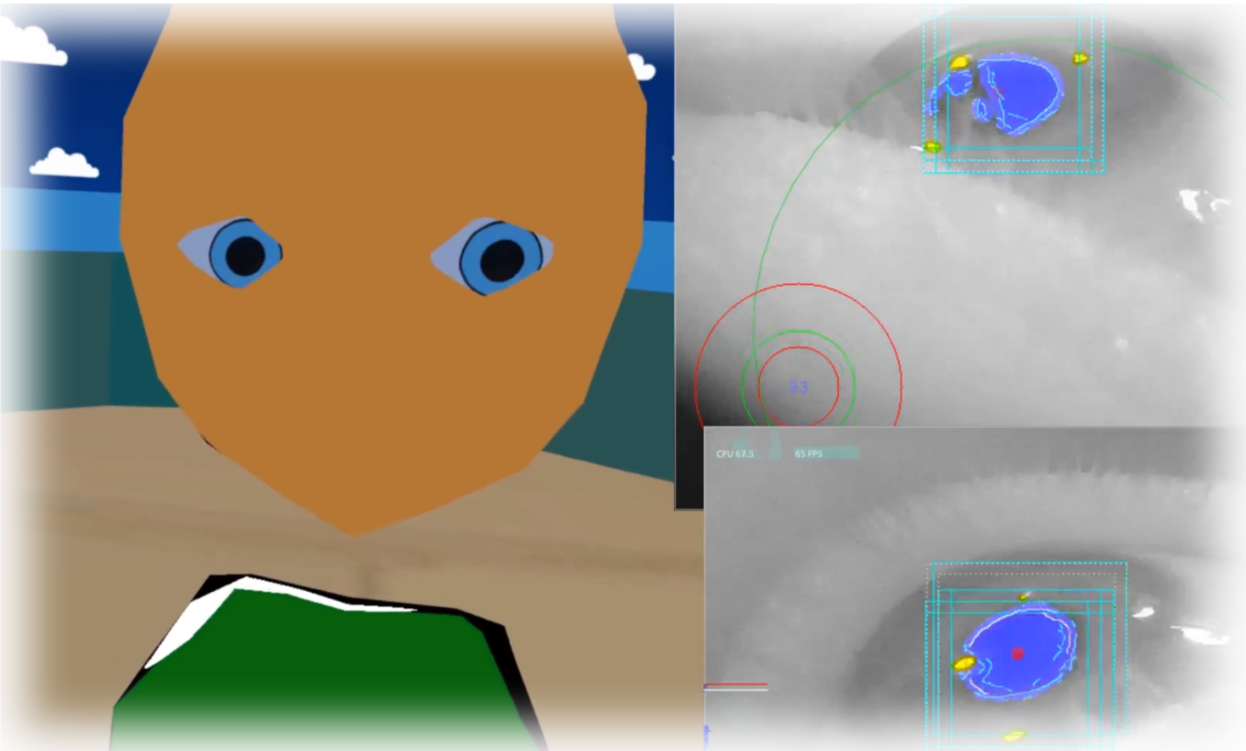
The Effect of Eye Tracking on an Object-Based Task in a Shared Virtual Environment

Master's Thesis

MTA181031



Aalborg University
Department of Architecture, Design, and Media Technology
Medialogy

School of Information and Communication Technology

Medialogy
Rendsburggade 14
9000 Aalborg
Telephone 0045 99 40 71 30
http://www.create.aau.dk

AALBORG UNIVERSITY
STUDENT REPORT

Title:
The Effect of Eye Tracking on an Object-Based Task in a Shared Virtual Environment

Theme:
Master's Thesis

Project period:
10^{th} Semester, Spring 2018

Project group:
MTA181031

Participants:

Dennis Hultberg Klyver

Jonathan Bernstorff Nielsen

Kristian Rokholm Jensen

Supervisor:
Martin Kraus

Circulation: 5

Page count: 102

Appendix count and type: 6, Bibliography and Appendices

Finished on May 29^{th} 2018

Abstract

Eye-gaze is an important part of nonverbal communication, but its effect in avatar-mediated communication for object-based tasks has barely been researched. Eye tracking is a novel technology that is expected to become important for immersive technologies using head-mounted displays in the near future as it allows for a rich form of media communication. This project tested an object-based cooperative task in a shared virtual environment with different oculesic behaviour including occluded, static, modelled, and tracked eye-gaze. Four preliminary experiments were conducted in order to evaluate the eye tracker's accuracy, and the appropriateness of the task. The study's main experiment was conducted in order to evaluate whether the study's four oculesic behaviour conditions affected the participants in regards to social presence, performance, or perceived naturalness. In the main experiment it was found that among the different conditions, there were in most cases no significant differences, although tracked eye-gaze was considered more natural than occluded oculesic behaviour from the viewpoint of the participants not wearing an eye tracker. Several limitations regarding the technologies used and the task design were discussed and, despite the lack of novel findings, the authors encouraged future studies on the subject after improving either of them.

Preface

Throughout this report, various expressions, methods, etc. will be mentioned multiple times. This section provides an overview of the most important expressions, methods etc., which in turn should make it easier and clearer to read the report.

Definitions

Whenever a new expression, name, or something of utmost importance is mentioned throughout the report, it will be written in italics the first time it is mentioned. If it is an expression that is normally abbreviated, the abbreviation too is written in italics and used henceforth. An expression, name, etc. is only written in italics the first time it is mentioned. After that point, it is written normally.

Virtual Environments

Various expressions for *virtual environments* (*VEs*) will be utilised. For the sake of clarity, the authors have divided virtual environments into two expressions:

1. **CAVE-like VR:** This expression covers all VEs found through the literature review, Section 2, *Previous Work*, where the visuals are projected onto a surface, but is still rendered relative to the user's view such as: *CAVE-environments*, projection technologies, etc. This could, for example, relate to studies such as Murray et al. (2009); Peters et al. (2010); Steptoe et al. (2009).

2. **HMD-based VR:** This expression covers all VEs found through the literature review, Section 2, *Previous Work* and commercially available virtual reality environments, in which the participants wear *head-mounted displays* (*HMDs*). This could for example be studies such as Lutz et al. (2017); Roth et al. (2016b); Elgarf et al. (2017).

Acknowledgements

The authors would like to thank their supervisor, Martin Kraus, for his help and feedback.

Table of Contents

1 Introduction **1**

2 Previous Works **3**

2.1 Literature Search . . . 3

2.2 Motivation . . . 3

2.3 How Eye Tracking Has Been Used in Research . . . 4

2.4 Small Overview of Technology . . . 4

2.5 Theory . . . 5

2.5.1 Media Richness . . . 5

2.5.2 Oculesic Behaviour Properties . . . 5

2.5.3 Dyadic and Triadic Eye-Gaze . . . 6

2.5.4 Eye Tracking in Social Contexts . . . 7

2.5.5 Kinematic and Visual Realism . . . 7

2.6 The Gap in Literature . . . 9

3 Design **11**

3.1 Task Design . . . 11

3.1.1 Goal-Driven Behaviour . . . 12

3.1.2 Interaction Design in Task . . . 12

3.2 Game Design . . . 13

3.3 Avatar Design . . . 14

3.4 Objects . . . 16

3.4.1 Super Shapes . . . 16

3.4.2 Waiting Room . . . 18

3.4.3 Monitor . . . 19

3.4.4 Spatial UI . . . 19

3.5 Environment . . . 20

3.5.1 Teleporter . . . 20

3.5.2 Receiver . . . 20

3.5.3 Walls/Window . . . 20

3.5.4 Black Hole . . . 21

3.6 Tutorial Design . . . 22

3.7 Summary . . . 23

4 Implementation **25**

4.1 Object Interactions . . . 25

4.2 Spatial UI . . . 26

4.3 Object Morphing . . . 26

4.4 Shaders . . . 26

4.5 Calibration Method 27
4.6 Virtual Mirror 27
4.7 UNet Implementation 28
4.8 Pupil Capture Service 30
4.9 Oculesic Behaviour Conditions 31
4.9.1 Occluded Oculesic Behaviour 31
4.9.2 Static Oculesic Behaviour 31
4.9.3 Modelled Oculesic Behaviour 32
4.9.4 Tracked Oculesic Behaviour 32
4.10 Motion Capture System 33
4.11 Developer Interface System 33
4.12 Data Logging 33
4.13 Summary 35

5 Preliminary Experiments **37**
5.1 Preliminary Experiment 1: Eye Tracking Accuracy 37
5.1.1 Motivation 37
5.1.2 Method 37
5.1.3 Procedure 38
5.1.4 Results 39
5.1.5 Discussion 39
5.2 Preliminary Experiment 2: Distance Between Objects 40
5.2.1 Motivation 40
5.2.2 Method 41
5.2.3 Procedure 42
5.2.4 Results 44
5.2.5 Discussion 44
5.3 Preliminary Experiment 3: Testing All Four Oculesic Behaviour Conditions 45
5.3.1 Motivation 45
5.3.2 Method 46
5.3.3 Procedure 48
5.3.4 Participants 49
5.3.5 Results 49
5.3.6 Discussion 51
5.4 Preliminary Experiment 4: Testing All Four Oculesic Behaviour Conditions without Nonverbal Communication 52
5.4.1 Motivation 52
5.4.2 Method 52
5.4.3 Procedure 53
5.4.4 Participants 53
5.4.5 Results 53
5.4.6 Discussion 55
5.5 Summary 56

6 Main Experiment **59**
6.1 Motivation . . . 59
6.2 Task . . . 59
6.3 Setup . . . 60
6.4 Procedure . . . 60
6.5 Participants . . . 60
6.6 Recorded Data . . . 60
6.7 Results and Analysis . . . 60
6.7.1 Error Rates and Completion Times . . . 61
6.7.2 Game Experience Questionnaire . . . 63
6.7.3 Simulator Sickness Questionnaire . . . 63
6.7.4 Naturalness Score . . . 64
6.7.5 Observations . . . 66
6.8 Summary . . . 69

7 Discussion **71**
7.1 Quality of Solution . . . 71
7.2 Quality of Technology Used . . . 72
7.3 Participant Communication Strategies . . . 74
7.4 Appropriateness of Task . . . 74
7.5 Task and Game . . . 75
7.6 Discussion of Experiment . . . 76

8 Conclusion **79**

9 Future Works **81**

Bibliography **83**

Appendices **87**

A Discarded Items **89**
A.1 Avatar . . . 89
A.2 CRT Monitor . . . 89
A.3 Object Placement Ideas . . . 90

B Experiment Items **93**
B.1 Game Experience Questionnaire (GEQ) . . . 93
B.2 Simulator Sickness Questionnaire (SSQ) . . . 93
B.3 Preliminary Experiment 2 Items . . . 93

C Literature Review Search Protocol **95**

D Data **97**
D.1 Experience with VR . . . 97
D.2 Glasses/No Glasses . . . 97

D.3 Video Game Playing Frequency . . . 99
D.4 Participants Knowing Each Other Beforehand . . . 101

E Interaction Design **105**
E.1 Affordances . . . 105
E.1.1 Affordances Used in This Study . . . 105
E.2 Signifiers . . . 106
E.2.1 Signifiers Used in This Study . . . 106
E.3 Feedforward . . . 106
E.3.1 Feedforward Used in This Study . . . 106
E.4 Feedback . . . 106
E.4.1 Feedback Used in This Study . . . 107
E.4.2 Auditory feedback . . . 108
E.5 Constraints . . . 108
E.5.1 Constraints Used in This Study . . . 108
E.6 Mappings . . . 109
E.6.1 Mappings Used in This Study . . . 109
E.7 Compliance . . . 109
E.7.1 Position Compliance (Spatial Compliance) . . . 109
E.7.2 Directional Compliance . . . 110
E.7.3 Temporal Compliance . . . 110
E.8 Direct Interaction . . . 111
E.9 Indirect Interaction . . . 111
E.9.1 Indirect Interaction Used in This Study . . . 111
E.10 Modes and Flow . . . 111
E.10.1 Modes and Flow Used in This Study . . . 102

Chapter 1

Introduction

The technology that drives modern *virtual reality* (*VR*) systems is rapidly evolving, and eye tracking is seen as one of the next big steps (Velloso and Carter, 2016). By adding small sensors inside the *head-mounted displays* (*HMD*s), it is possible to track individual eye movements of the users. This potentially allows for technological leaps such as foveated rendering, but it is also interesting how the eye movement can be translated to digital avatars that represent the users Tobii Pro ((n.d.). Most modern VR systems already provide a high level of kinematic fidelity in terms of position tracking for both the head and the hands, which allows for a rich form of communication as it enables nonverbal communication (Roth et al., 2016b). Adding another layer by also having accurate eye movements using the eye trackers would enhance the richness of the media, but the question is to what extent this will affect users.

VR is growing at a fast pace and is predicted to grow at an even faster pace in the years to come. The gaming focused *PlayStation VR* was announced as being the best selling VR system in 2017, and is expected to reach two million units sold in 2018. The number of sales reflects the popularity of the new wave of VR devices commercially available. Similarly, the gaming industry is also starting to adopt eye tracking in games such as *Assassin's Creed Rogue* (Velloso and Carter, 2016; Smith, 2018; Statista, 2017). With VR system developers announcing eye tracking as the next add-on to their products, the relevance for research related to commercially available VR products and eye tracking has become increasingly relevant.

The subject of eye tracking has been researched in several studies over the years. For example tracked eye-gaze in *virtual environments (VEs)* is expected to improve social presence compared to static eye-gaze (Seele et al., 2017), however it has also been shown that the kinematic realism should be correlated with the visual realism of avatars (Garau et al., 2003; McDonnell et al., 2012; Howard and Steptoe, 2016). Avatars with tracked eye-gaze have also been shown to make social interactions more detailed, for example making it easier to see if people are paying attention or even lying (Steptoe et al., 2010). Eye-gaze has also been used for enabling new forms of interaction. Experimental video games have used eye tracking as an input method to enable quick selection or gaze sensitive AI (Velloso and Carter, 2016). There have also been studies on how eye-gaze increases interaction between collaborators in a *shared virtual environment* (*SVE*) when working on object-based tasks by using gaze direction estimation to determine objects in a 3D space that are being looked at (Garau et al., 2003; Steptoe et al., 2008, 2009; Elgarf et al., 2017). However, none of these studies have used two real collaborators working with modern consumer HMD-based VR technologies.

The purpose of this study is to investigate whether the combination of modern eye tracking systems and HMDs for collaborative object-based tasks will improve users' error rates, completion times,social presence, and the perceived naturalness of avatars' eyes. Furthermore, the study's task is based on previous research in the areas of eye-gaze in social contexts, nonverbal communication, eye tracking, etc. Two participants will simultaneously be immersed in a SVE where they will use nonverbal communication to locate objects placed in the environment. The experiment will compare how completing the task using tracked eye-gaze will compare with more commonly used eye-gaze behaviours.

Chapter 2

Previous Works

This chapter explores literature relevant for this study. The topic of eye-gaze and how it can be used in *shared virtual environments (SVEs)* is researched, and the findings will be presented here. The findings are about how eye-gaze is used in natural settings, as well as how it has been used in *human-computer interaction* research. Theory about eye-gaze and trends in relevant literature will also be explained. The chapter will also touch upon the kinematic and visual realism that is relevant for choosing a style of avatar that is most suited for this study. Lastly, the study's goals and how it stands out from the existing research will be explained.

2.1 Literature Search

The first approach to finding literature for this project was to use the AAU research portal, *AUB*, and search on keywords thought to be related to the project regarding eye tracking and SVEs. Several sources were found to be relevant and used to find more formal keywords for a structured search. A list of repeating keywords found in the sources was created and the most important ones were picked out. These keywords were used for a structured search on the research databases *IEEE*, *ACM*, *Proquest*, and *Compendex*. The keywords were adjusted until a search result came up with a reasonable amount of sources that used the keywords, approximately 50-100. The final search resulted in 56 relevant sources. Further information regarding the literature search can be seen in Appendix C, *Literature Review Search Protocol*. Parts of the found papers were read systematically, which were used for assessing whether they still seemed relevant to the study. If a source was deemed irrelevant at some point during the reading order, it was discarded. The reading order mostly was: the title, the abstract, conclusion, method, discussion, introduction/previous works, and then the entire document.

While reading through these sources, it was found that several papers shared citations for fundamental points during their previous work sections. In some cases, these references were already found during the search, but if not, they were also thoroughly read and added to the literature review.

2.2 Motivation

Normally, communication between humans may be thought of as primarily using written and spoken languages. However, despite the complex thoughts that can be conveyed this way, we still rely on *nonverbal communication* to a high degree. Nonverbal communication in this context refers to things such as body posture, hand gestures, facial expressions, and eye-gaze. It is said that up to 55% of our understanding of each other's feelings comes from facial gestures Mehrabian and Ferris (1967).

Eye behaviour can be used to gauge things such as how a conversation is flowing, signal a search for feedback from the interlocutor, express emotion, or even influence others' behaviour (Lee et al., 2002). This is why much literature based around *virtual environments* (*VEs*) revolve around the concept of *copresence* and *social presence* - i.e. the ability to capture the sense of fully being in the virtual company of another person through a virtual medium. The use of eye tracking allows researchers to capture participants' oculesic behaviour, e.g. eye-gaze, blinks, eyelid movements, pupil size, etc., which can then be simulated and displayed by avatars in *immersive collaborative virtual environments* (*ICVEs*) (Steptoe et al., 2010; Eichert et al., 2017).

2.3 How Eye Tracking Has Been Used in Research

A multitude of previous studies have used eye tracking for various purposes. Some studies used eye tracking for proving that it was possible for people inside ICVEs to discern that other participants, or avatars, were looking at either them or at objects placed inside the SVE. Eye tracking was used for testing at which distances eye-gaze was more discernible than head gaze. In extension to this, it was tested at what distance eye-gaze had higher accuracy than head gaze. Findings show that the ability to discern eye-gaze and head orientation decreases significantly after approximately six metres. As part of testing accuracy of eye-gaze, eye tracking was used in settings where one participant had to look at objects between them and another digital avatar. The participant then had to discern at which objects the avatar was gazing (Roberts et al., 2009; Murray et al., 2009; Steptoe et al., 2009). It has also been studied whether participants' task performance and user experience were affected by avatars expressing tracked oculesic behaviour. Tracked eye-gaze yielded better results than static and modelled eye-gaze however, not significantly so when compared to static eye-gaze. Also, tracked eye-gaze yielded fewer errors and lower completion times and better user experience (Steptoe et al., 2009). Oculesic behaviour has also been found to enable participants to discern between truthful and deceptive behaviour in avatars (Steptoe et al., 2010).

Velloso and Carter (2016) made a survey regarding previous uses of eye-gaze, e.g. eye tracking as an input device for video games. Eye tracking implemented in video games can be used e.g. for emulating computer mouse and keyboard commands. This can, for example, be used for selection purposes inside video game environments where you would normally use a computer mouse. However, it was advised to not use it as a single modality due to the *Midas touch* problem. Velloso and Carter (2016) also mentioned how games can use unwritten social rules in regards to eye-gaze as an interaction method.

Finally, a single study was found to not utilise eye tracking and instead compared a real world scenario with a virtual simulation without eye movement and facial expression cues (Roth et al., 2016a).

2.4 Small Overview of Technology

Eye tracking technology has improved significantly during the last decade or so and more types of oculesic behaviour are thus becoming trackable. Eye trackers are also becoming more commonplace technology. This is also the case in the gaming industry, where for example Assassin's Creed Rogue has gaze-interactions enabled (Velloso and Carter, 2016).

A magnitude of equipment has been utilised in previous research regarding eye tracking in VEs. An eye tracker is a so-called *absolute* input device, which means that it outputs tracked coordinates in the screen frame of reference. Because of the jittery and sudden nature of eye movements, it is infeasible to use for controlling a precise system such as a cursor (Velloso and Carter, 2016). Most eye tracking systems measure the participant's pupil size during interaction. When the diameter size is zero, it usually means that the participant blinks. However, it could also mean that the pupil recognition is lost (Murray et al., 2009). Finally, Steptoe et al. (2010) tracked pupil dilation, eyelid movement, and eye blinks to simulate this behaviour on their avatar.

Some studies utilised ICVE systems, which can be populated with multiple participants, avatars, objects, etc. One ICVE system found during the literature review was the *EyeCVE*. The EyeCVE has built-in binocular eye tracking and was supposedly the first system that allows the user to move around in a 3D space with eye tracking (Steptoe et al., 2010). Finally, an ICVE called *ReaCTor* was used in a study by Vinayagamoorthy et al. (2004). Most studies (Murray et al., 2009; Steptoe et al., 2008, 2009; Roberts et al., 2009), utilised IPTs, whereas others used a normal computer screen. In some studies, optical trackers have been used to capture the participants' head and/or hand movements and then mapped this data to avatars in ICVEs (Murray et al., 2009).

Some eye tracker systems, like the *ASL Model 501* used by Murray et al. (2009), only enable monocular eye tracking. This means, that it is not possible for the researchers to consider and calculate convergence of the eyes. This might mean that it becomes difficult to estimate what people are gazing at, especially in regards to their gaze's depth of field.

Other eye tracker systems, such as the *ViewPoint EyeTracker* from *Arrington Research Inc.* and *SMI Eye-Tracking Glasses 2 Wireless* from *SensoMotoric Instruments GmbH* can track eye-gaze while users are wearing HMDs (Steptoe et al., 2010; Eichert et al., 2017). Both of these eye trackers are binocular, which enable researchers to track both eyes simultaneously. The tracked eye movements can then be used to animate avatar eye movements (Roberts et al., 2009; Eichert et al., 2017).

2.5 Theory

In the following subsections, important theory aspects found during the literature review, which are applicable to the authors' study, are presented.

2.5.1 Media Richness

Daft and Lengel proposed the theory of media richness, in which they describe a medium's capability of transmitting and reproducing the full extent of human communication where *rich* information involves nonverbal communication such as in ICVEs. The theory states that task performance will improve in a medium that is capable of transmitting more rich and relevant information compared to media that are less rich in information, such as text-based media (Steptoe et al., 2010). It has been shown that there seems to be a correlation between the visual fidelity of avatars and peoples' expectations of the behavioural fidelity of them. With rich media, it has been shown that aspects such as *personal space* and the unease of eye contact with strangers can also occur in a virtual world (Steptoe et al., 2009).

Steptoe et al. (2010) attempted to illustrate the theory of media richness in terms of how well people can assess whether someone is lying with focus on the importance of eye-gaze. In one experiment, they had participants tell either truths or lies where the audio and eye-gaze was recorded. Later, in another experiment, these recordings were shown in different levels of richness. The recordings were presented on a digital avatar with tracked oculesic behaviour, no oculesic behaviour, or audio-only. Accuracy of detecting lies was found to be 48% with tracked oculesic behaviour, which was statistically higher than without (39%), and the lowest case was with audio-only (34%).

2.5.2 Oculesic Behaviour Properties

According to Eichert et al. (2017); Velloso and Carter (2016), the human eyes are limited to specific movements and properties which are covered in the following:

- **Fixations** stabilise the retina on an object of interest so that its image falls on the fovea of the eyes. This way, it is possible for the viewer to focus on an object. Measuring fixations can tell you what a person is focusing their gaze on and what catches their attention. A fixation consists of a look lasting for at least 100 ms.
- **Saccades** are the rapid movements made when the eyes move from one focused position to another. Saccades can be used for gesture matching or mode selection.
- **Saccade duration** is the time it takes for the eye to change from one position to another.

- **Saccade magnitude** is the angle the eye needs to rotate in order to focus at a new position. 90% of all natural saccades have a magnitude of less than 15 degrees. In a dyadic conversation, the speaker's maximum saccade magnitude is 27.5 degrees. For a listener in the same condition, this magnitude is 22.7 degrees.

- **Saccade direction** is the direction the eye moves.

- **Saccade velocity** is the non-uniform velocity with which the eye moves in order to reach a new position. The eyes do not have a constant velocity, instead, they accelerate from one position to the maximum velocity, then they decrease and stop at a new position. The higher the saccade magnitude, the higher average saccade velocity becomes.

- **Smooth pursuits** are the eye movements occurring when one follows a moving object with the eyes. Contrary to saccades, smooth pursuits are smooth and match the targets' relative velocity. Smooth pursuits only occur while following moving targets, which makes it possible to identify whether a participant actually follows a moving target on e.g. a screen. Smooth pursuits can be used for calibrating an eye tracker and for selecting moving targets.

- **Compensatory eye movements** are involuntary smooth movements occurring when moving the head while keep fixating at a specific point in one's field of view. Compensatory eye movements could e.g. be used for recognising head nods and hand gestures for NPCs in a VE.

- **Vergence** is the movements focusing the eyes at distant objects. The further away an object is, the more parallel the eyes will be. In a virtual environment, vergence could be used for detecting whether a participant focuses on an object or another player.

- **Optokinetic nystagmus** is an eye movement comprised of smooth pursuits and saccades, which enables people to keep perceiving an object moving through the field of view. Optokinetic nystagmus movements occur while observing e.g. a train passing by.

- **At gaze** is exhibited when the eye is looking directly ahead, thus, it is in *primary position*.

- **Away gaze** is exhibited when the eye is *not* looking directly ahead.

- The **inter-saccadic interval** is the time the eye spends at a focused position.

2.5.3 Dyadic and Triadic Eye-Gaze

According to Vinayagamoorthy et al. (2004); Elgarf et al. (2017); Steptoe et al. (2009, 2010), two types of eye-gaze exist; dyadic and triadic:

Dyadic eye-gaze regards eye to eye or eye to face interactions. Dyadic eye-gaze enables people to read other people's emotions. Dyadic eye-gaze is used for getting feedback during conversation.

Triadic eye-gaze, also known as *joint attention*, regards a person's perception of another person gazing in some direction. Triadic eye-gaze is used for understanding other people's intentions. Also, triadic eye-gaze regards object-oriented interactions with another person.

Some studies focused on dyadic and triadic eye-gaze in ICVE systems. The experimental task could be that one participant had to gauge which object in the VE that the other participant was looking at. Some of these experiments studied whether participants were able to discern which objects in an ICVE participants were gazing at. Most studies focused on training participants in an ICVE using triadic eye-gaze in various social settings with real-world applicability (Murray et al., 2009; Roberts et al., 2009; Steptoe et al., 2009; Elgarf et al., 2017; Eichert et al., 2017).

2.5.4 Eye Tracking in Social Contexts

In the experiment by Steptoe et al. (2010), they asked about the perceived engagement of a virtual avatar under three different conditions. Each condition had a different level of richness for the avatar, ranging from audio only, to a tracked eye-gaze animated avatar with audio. It was found that higher levels of richness increased the perceived levels of engagement. Eye-gaze has been shown to be an important part of initiating, terminating, and gauging interest in a conversation, so the correlation should be a natural consequence of this (Steptoe et al., 2009). In an experiment by Grillon and Thalmann (2008), eye-gaze of virtual avatars was experimented with to find how realism of avatars was affected by eye-gaze behaviour triggered by a user. In this scenario, a virtual avatar would appear either 1) attentive and look straight at the user or 2) bored with a distracted appearance towards the user based on whether the user was looking directly at the avatar. The scenario was tested in regards to four conditions based on the avatar's behaviour: always attentive, always bored, random, or attentive when looked at and bored otherwise. Based on a questionnaire, results from 12 participants suggested a significant difference in how normal they viewed the avatar. The more attentive the avatar was, the more normal it seemed. Though the researchers could not find evidence to suggest that having a reactive behaviour was more normal than always attentive, they did show a difference in how people notice and describe nonverbal behaviour from digital avatars.

One of the more recent studies on how eye tracking affects the quality of communication is by Seele et al. (2017) who claims that most research in the field of virtual reality has been performed on outdated technology. Seele et al. (2017) had couples participate in social verbal interaction games using *avatar-mediated communication*, *AMC* in VR with different oculesic behaviour conditions including tracked eye data. A saccadic oculesic behaviour model was used together with an altered model that also accounted for dynamic fixations. Despite these different models, the participants found no significant changes in quality of communication and perceived avatar realism. However, Seele et al. (2017) also strongly argue that the novelty of the VR technology overwhelmed the participants more than the subtlety of the eye animations, as the participants expressed a very low rating in experience with VR (median of 1 out of 7).

2.5.5 Kinematic and Visual Realism

Studies have investigated the differences in communication that exist between AMC and *video-mediated communication* (*VMC*). Currently, VMC is the more accessible and the higher quality option of the two as it is able to faithfully show how the user looks, while AMC still has some way to go in that regard (Roberts et al., 2009; Steptoe et al., 2010). Not showing the user faithfully enables users to ignore the physical aspect of anyone they might be interacting with. Thus, enabling the users to form impressions about other users, purely on the behaviour that they display in the *immersive virtual environment (IVE)*. AMC not automatically being a faithful representation of the user, allows researchers to take control of relevant and irrelevant nonverbal behaviour, if they should wish to do so (Bente et al., 2007). For example testing oculesic behaviour conditions on participant avatars to determine perceived realism. This control over nonverbal behaviour can be used in the creation of both realistic and abstract avatars alike. Realistic avatars try to simulate/replicate aspects such as looks and/or the verbal or nonverbal behaviour of the user. Abstract avatars, however, are created as representations of humans in general. This general human appearance can make them simple to create and animate (Roth et al., 2016a).

Working with object-based tasks where moving around an environment is required, a VMC based system already has its limitations, as it is dependant on finding a proper viewpoint for the system's camera. Such an object-based task making use of eye tracking would not function reliably if done on a VMC platform since the eye tracking would lose its accuracy when participants would navigate the environment. A similar scenario is mentioned by Roberts et al. (2009) when conducting an eye tracking experiment on a VMC system. An AMC system making use of an HMD for the object-based task, however, places the screen right in front of the participant's eyes at all times, thus it is not as prone to the same movement based tracking discrepancies (Roberts et al., 2009).

Rendering Styles and Uncanny Valley

With the rise of VE's, their avatars have become a subject for research. With the rise of VE's, avatars have become a subject for research. Especially the different styles of avatars and the effect they have on the user of the virtual system have been subject for a number of studies.

A study on different render styles was conducted by McDonnell et al. (2012). It compared ten different render styles, five cartoonish and five realistic, to see how these were perceived by participants. It was found that what participants used to rate the virtual characters could mostly be derived from a still image of the characters. Of the six experiment measures, showing character movement would only change how participants rated the familiarity and appeal of the characters. Especially highly unappealing characters were seen as being even more so when movement had been applied. Also, motion anomalies on realistic characters were found more unpleasant than they were on cartoonish characters. This correlates with earlier findings by Garau et al. (2003) that showed that the more photorealistic an avatar is, the more people demand/expect it to display realistic behaviour.

Latoschik et al. (2017) investigated the *illusion of virtual body ownership* (*IVBO*) for participants who were placed in an SVE within one of two avatar body render styles. One render style was a photorealistic human avatar, while the other was an abstract genderless humanoid avatar, made to look like a wooden mannequin. IVBO refers to the users' acceptance of the virtual body, or parts thereof, as their own. The realistic avatars were shown to produce a higher level of acceptance in the participants in relation to IVBO than the abstract avatar. The realistic avatars showed higher scores in regards to eeriness, while the abstract avatar showed higher scores for attractiveness. Vinayagamoorthy et al. (2004) found that perceived communication quality did not increase when the developed oculesic behaviour model was used for neither a cartoonish nor a realistic avatar. Additionally, it was found that the more visually realistic an avatar is, the more users expect from that avatar in terms of behaviour and human-like qualities (Howard and Steptoe, 2016).

The findings by Latoschik et al. (2017); Garau et al. (2003); McDonnell et al. (2012); Vinayagamoorthy et al. (2004) could be contributed to the uncanny valley relationship. The uncanny valley has long existed as the explanation for why increasingly realistic renderings of humans may have attributes that appear disturbing or off to viewers (McDonnell et al., 2012; Vinayagamoorthy et al., 2004; Howard and Steptoe, 2016). Uncanny valley has been observed within advances in making robots, avatars, movies, etc. or more generally said, in areas where human appearance or behaviour are being simulated or recreated for some purpose. The causes for these feelings have been attributed to higher awareness of minor imperfections in various areas of the realistic objects' behaviour or look (McDonnell et al., 2012).

A study done by Roth et al. (2016a) investigated the social interactions of abstract avatars in an SVE where social and behavioural channels had been limited. The experiment had the participants participate in two real-world and virtual-world tasks, where gaze and facial expression cues had been excluded from the interaction in the virtual world. Roth found that these limited channels can have an impeding effect on social interactions when using abstract avatars. The missing gaze and face expression cues can, however, be partly compensated for since participants would shift their focus to the available channels, such as movement behaviour. Roth's results thus show that using abstract avatars that do not display facial expressions or eye-gaze cues impedes social interactions through AMC, but it does not make social interactions impossible. Further work suggested by Roth et al. includes applying eye-gaze or facial expression to similar tests on abstract avatars to uncover the added effect they may bring. Applying different types of eye-gaze to abstract avatars in an SVE will be the focus of this report and its corresponding experiments.

Levels of Body Tracking

Naturally, the equipment used by older sources considered in this review makes use of less advanced tracking devices, such as tracking of only head, hand(s) or similar body parts (Steptoe et al., 2010; Vinayagamoorthy et al., 2004; Roberts et al., 2009), while more recent sources make use of full body tracking instead (Latoschik et al., 2017; Dodds et al., 2011; Roth et al., 2016a). Full body tracking suits have the benefit of enabling movement for all body parts covered by the suit to be tracked and simulated directly onto the corresponding body parts of a virtual avatar. Although systems that enable tracking of a full body motion tracking suit are effective at tracking motion, they are nonetheless expensive and not commercially common (Roth et al., 2016a).

The currently available consumer VR headsets mostly utilise three point tracking, i.e. tracking six degrees of freedom for the head and the hands. Mainly the *Oculus Rift*, *HTC Vive*, and Playstation VR (VR, n.d.; htc, n.d.; Sony, n.d.a) have been considered as the primary commercially available VR headsets at the time of this report. Since most of these devices are shipped with three tracked objects, this study will use this as the commercially available standard. Having three trackers distributed between the head, and one for each hand, creates some challenges for displaying the tracked movement of users on the avatars used within the SVE. The occurrence of proprioceptive conflict is one of the main concerns for having untracked body parts. It occurs when the felt physical placement of body parts through one's own proprioception does not correlate with the viewed placement of one's virtual body parts (Howard and Steptoe, 2016). A video on the avatar creation process for Oculus VR's avatar software development kit explains their measures taken to avoid proprioceptive conflict. The approach taken is to only display the virtual body parts that can be tracked reliably by the trackers available. In the case of a three point tracking system, only the hands, head, neck, and shoulders are shown to the user. Hands are tracked by the controllers which users have in each hand while the HMD tracks the head placement and rotation which is used to control the neck and shoulders. Furthermore, Howard and Steptoe (2016) explicitly avoided to simulate eye movement or facial expressions that cannot accurately be tracked, in order to avoid the uncanny valley relationship.

Similar approaches to avatar design have become increasingly popular in multiplayer VR games, especially over the recent years (AGAINST GRAVITY, n.d.; Oculus, n.d.; Sony, n.d.b). At the time of this study, no research has been found on this specific avatar design, nor have a commonly used term been found. Therefore, this study will refer to this avatar design as a floating limb avatar.

2.6 The Gap in Literature

One of the more recent studies on how eye tracking affects the quality of communication is by Seele et al. (2017), which claims that most research in the field of VR has been conducted using outdated technology. Seele had couples participate in social verbal interaction games using AMC in VR with different oculesic behaviour models, including tracked eye-gaze. Seele used a saccadic oculesic behaviour model and an altered model that also accounts for dynamic fixations. Despite these different models, the participants found no significant changes in quality of communication and perceived avatar realism. However, Seele also strongly argue that the novelty of the VR technology overwhelmed the participants more than the subtlety of the eye animations helped them.

Instead of working with a social task, Murray et al. (2009); Steptoe et al. (2009) focused on how eye tracking allowed people to judge what digital avatars are looking at using spatial object based tasks. However, as mentioned by Seele et al. (2017), these studies were using already dated technologies that had lower fidelity and less features than modern HMD-based immersive environments. Both Murray et al. (2009); Steptoe et al. (2009) limited their participants by having them standing still and not being able to freely communicate with the other avatar, while Seele et al. (2017) was focused on a social verbal task and the participants were sitting down. To the knowledge of the authors, this project is the first time that an object-based task has been combined with a VE in which the participants are able to move and communicate freely, including tracked eye-gaze, with each other using an HMD-based IPT.

Chapter 3

Design

Based on the literature search, see Chapter 2 *Previous Works*, the authors found a gap in the existing literature. The gap was, to the knowledge of the authors, that no other study had utilised an object-based task, making use of an HMD and focusing on eye tracking, within a SVE. This combination of task and technology should provide the participants the best collaboration possibilities and increase object-based task performance in virtual environments.

This chapter will elaborate upon the design choices made for this study. Some of the design solutions were made and/or changed recursively based on the study's four preliminary experiments, see Section 5, *Preliminary Experiments*. Thus, this chapter should be read as regarding the design decisions made throughout the entirety of the study and not as a single revision of the virtual environment. This chapter also presents information on the design of the experiment task, game elements, interactions, avatar, objects, environment, and the tutorial. The chapter concludes with a summary of the findings and decisions made throughout the design phase in order to highlight what was to be implemented in the study's SVE.

3.1 Task Design

Based on the literature review, it was decided that the task should revolve around two immersed participants trying to use eye-gaze to communicate about the position of objects in a 3D scene. The participants had a role each, these being the *fixer* and *fetcher*, respectively. Only one eye tracker was available for the project, meaning only one person would have tracked eye-gaze. Because of this, the task was made to be asymmetrical, where one participant with the eye tracker, the fixer, had to look at different objects and then tell the other participant which object to pick up. It was important that the participants were not restricted in terms of the conventional movements one can perform in a VR application, such as free room scale movement. To ensure that the fixer participant would not simply move to the objects and pick them up, it was decided to have a virtual window separating the players. On the side of the participant with the eye tracker, a screen was placed which showed them what object to look for and a UI crosshair, which was only visible to that participant, was placed over the object.

In order to mask the true aim of the study and to contextualise the task in a fictional setting, a few more parts to the task were added. After the participants correctly identified an object together, the participant without the eye tracker, the fetcher, had to interact with a teleporter to send the object to the fixer participant. The fetcher participant then had to pick up this object and move it into a black hole. This was narratively described as the solution to closing the black hole and functioned as the end of the base task

Murray et al. (2009); Elgarf et al. (2017); Eichert et al. (2017); Roberts et al. (2009) all had tasks with objects spread out in a similar fashion, where a participant had to guess which object a digital avatar was looking at. The studies were different in how the objects were laid out though, where Murray et al. (2009); Elgarf et al. (2017) had objects aligned on an xy-plane, Eichert used an xz-plane, and Roberts had objects only in the horizontal dimension. This study internally tested a couple of configurations, but settled with having objects placed on a shelf in an xy-plane. The reasoning being that it was deemed to be the configuration that most likely would create a scenario in which tracked eye movement would have a significant effect.

3.1.1 Goal-Driven Behaviour

Goal-driven behaviour covers establishing a new goal, which leads to planned goals (Jerald, 2016: p. 285-287).

Goal-Driven Behaviour in This Study

1. This study used goal-driven behaviour by establishing a clear goal for the interaction; two participants help each other choose, transfer, and use various objects to close a black hole from destroying the virtual world.

2. This goal-driven behaviour tried to reduce the *gulf of execution* as much as possible. The gulf of execution happens when participants know *what* to do, but not *how* to do it. The authors of this study have tried to define how to complete the task as clearly as possible. This was done by using a tutorial explaining the point of the study, how both participants can interact with the environment and complete their part of the experiment, and how to finish it. In addition, to reduce the gulf of execution, the authors used signifiers, constraints, and mappings, as suggested by (Jerald, 2016: p. 285-287).

3.1.2 Interaction Design in Task

The study's task made use of several interaction design techniques that can all be seen in the Appendix E, *Interaction Design*. In accordance to the definition of interaction design, found in this related section of the appendix, interaction design techniques were used in order to make the task aspects easier to understand and use for the study's participants.

Object Interactions

With the base of the task built on participants moving objects from one room to another, and their controllers being the only direct way to influence the environment, some measures were taken to make the controller interactions intuitive. The virtual hands were given a grabbing animation that was mapped directly to the Vive controllers' trigger buttons. This gave instant information of how much the trigger was pushed down. A physical constraint for the controllers was the virtual floor being calibrated too high up in accordance to the real world floor, making it hard to pick up dropped objects. This was solved by making objects reset positions when they were dropped. The reset was accompanied by feedback showing the object disappearing into a small particle explosion. Participants' object interactions was a crucial part of the task. Therefore, object interaction feedback was emitted when either of the participants' controllers collided with an object. This would make the otherwise idle object's shader outline expand to another size and change to another colour while the controller was colliding with the object. This feedback was used to convey when participants were within range to grab an object. When a participant would grab an object, auditory feedback in the form of a grabbing sound was emitted to notify them of the action having been performed. Almost the same feedback was used on all virtual buttons, where their outline colour from yellow to green when the participants' controller collided with them. When a button was pushed, a click sound was played as feedback for the action.

Teleporter Button

For the participants to successfully complete the task, they had to interact with a teleporter or a black hole depending on whether their role was the fetcher or the fixer. The teleporter was designed to afford the fetcher participants the ability to place objects and teleport them to the the room with the fixer participants. The affordances of where to place the objects, and where they were teleported, were implemented by adding the particle field on the teleporter and by using a similar particle field on the receiver in the adjacent room. How to activate the teleporter was indicated by a single button placed on a stand right next to it, designed in the same style as the field and the teleporter model itself. Feedback was used to assist the fetcher participants when they pressed the teleporter button, as it moved into its stand. The teleporter button's vibrotactile feedback was delivered by making the participant's controller vibrate when colliding with the button. The pressing interaction was based on the direct interaction of pressing a button in the real world. When pressed all the way in, the teleporter button activated the teleporter and started a cooldown timer, which was visualised by the button becoming a darker colour. During this cooldown, the button was not affected by controller colliders. The cooldown is further visualised by making the teleporter button move slowly back to its initial position while regaining its original colour. Activating the teleporter caused the correct object to be dissolved, an animation for the teleportation field to play, a teleportation sound to be emitted, and afterwards it would appear in the receiver.

Black Hole

Having picked up an object from the receiver, the last step was for the fixer participant to throw it into the black hole placed in the SVE. The black hole had the affordance of allowing the fixer participant to throw objects into it to complete the goal of the task. A particle effect was created to make the black hole one of the SVE's few animated parts signalling its importance to the participants. When an object was thrown into the black hole, feedback was provided to the participants through visual and auditory channels. Visually, the hole, like the teleporter field, quickly expanded in scale and then returned to its original scale, while auditorily it would play a sound when beginning the expansion.

Signifiers

In relation to the tutorial, described in Section 3.6, *Tutorial Design*, signifiers were assigned to the teleporter and black hole in form of non-diegetic spatial UI name tags. These signifiers were added to make the objects easy to identify by the participants when mentioned in the tutorial for each participant role. The tutorial also contained pictures of the Vive controllers, showing the mapping of the grabbing mechanic to the trigger buttons, which is another example of signifiers added to assist participants in the experience.

3.2 Game Design

The task was designed using theory and principles from game design to make it more compelling and to make the scenario resemble a situation from a real video game. At the same time, it was stressed that the design was meant for an experimental scientific study and not a reason for playing around. Schell (2014) discusses the difference between a game and a normal task, and concludes that the main difference is the individual's approach to the task, whether if it is with a playful mind or not. The design is supposed to reflect the intended purpose so that participants are willful to complete the task, but that they do not get too caught up in playing around.

Fullerton (2008) distinguishes between the formal and the dramatic elements of a game, similar to how Schell (2014) defines four categories of game design: mechanics, technology (formal elements), aesthetics, and story (dramatic elements). The aesthetics and story behind the task were created to give the participants a sense of conflict and a concrete goal rather than an abstract task. The rules and mechanics were created to be simple to understand and execute. Besides communicating freely, the participants can grab and throw objects as the main mechanic. This is what Schell (2014) describes as the *toy*: the main interaction with the system and the main way the players can have an impact on the environment. Therefore, it is essential that this toy is intuitive and engaging to use. The main resources of the experiment was *information* and *time*. As one participant had access to more information, the main task was for them to cooperate and share this information as quickly as possible. The time taken to complete the task was recorded and put on a leaderboard to create a form of intrinsic motivation. If the other participant made a mistake and grabbed the wrong object, there would be a penalty to their overall time This was done to encourage the participants to not recklessly attempt to try all objects quickly, but rather make sure they were confident in their choice beforehand.

3.3 Avatar Design

For this project, an avatar was needed in order to embody the participants in the virtual environment. Also, an avatar makes participants feel as if they are part of the virtual environment and that their virtual limbs move correspondingly to their real-world limbs, which reduces the proprioceptive conflict. The avatar was decided to be a humanoid character for representing a human participant, as this enables a connection between the participant's real-world and visual behaviour and movements (Howard and Steptoe, 2016).

Initially, it was intended to implement a realistically looking avatar, described in Appendix A, *Discarded Items*. Instead, it was decided to use an abstract cartoonish floating limb character which consists of a head, two hands, and a torso. Though the authors had found no studies utilising one such throughout their literature review, there was a range of commercially available VR applications already using abstract floating limb gender neutral characters, such as *Rec Room* (AGAINST GRAVITY, n.d.), *Job Simulator* (Owlchemy Labs, n.d.) and *Racket Fury VR* (10Ants Hill), as seen in Figure 3.1.

(a) Rec Room (AGAINST GRAVITY, n.d.) **(b)** Job Simulator (Owlchemy Labs, n.d.) **(c)** Racket Fury VR (10Ants Hill)

Figure 3.1: Three commercial VR applications utilising floating limb characters.

In order to create naturally moving limbs with a limited number of tracking points, most avatars are made with inverse kinematics. However, despite the use of inverse kinematics, it is highly difficult to accurately model how a user is moving in a VE. Because of this lack of accuracy, limbs such as elbows and knees are usually not implemented. An elaboration on the importance of kinematic realism can be seen in Section 2.5.5, *Kinematic Realism and Visual Realism*. According to Howard and Steptoe (2016), having floating limb characters does not affect a participant's perception, as we as humans are able to perceive a whole from characters only consisting of floating limbs. Additionally, when creating virtual embodiment one of the most important parts is the motion of the avatar, rather than the visuals (Jerald, 2016). Another reason for using an abstract floating limb character was to avoid having the character's appearance influence the participants' perception, e.g. if the participants thought it looked sad. This avatar functioning was similar to that of Bente et al. (2007).

As the authors did not track other facial movements, no mouth was implemented for the abstract avatar. According to Howard and Steptoe (2016), simulating the avatar's facial expressions with mouth movements is a difficult feat. Simulating only mouth movements, on the other hand, is simpler, however this likely induces the Uncanny Valley relationship and thereby breaks participant's feeling of talking to someone real. Thus, the participant's immersion in the VE is broken. For these reasons, the authors decided to not implement mouth movements. Additionally, the authors did not find anything about the importance of having mouth movements throughout the literature review, seen in Section 2, *Previous Works*.

The avatar's head and hands were controlled through tracking. Head and hand tracking matching real-world head and hand movements was used as it makes an avatar seem more realistic, increases social presence and makes participants use their hands significantly more for communicative purposes while being inside SVEs (Howard and Steptoe, 2016). The hands were by default shown as open, but had a grabbing animation which was mapped to how much the user would hold down the trigger button.

The avatar had eyes as this was seen throughout large parts of the literature and in commercial VR applications. Having avatars with eyes increases the social interaction and duration of mutual eye-gaze in virtual environments. This was important in two aspects for this study, as it was based on a collaborative object-based task for measuring the effects of utilising tracked eye-gaze for avatars. Mutual eye-gaze awareness and eye contact is achieved by co-locating the participant's eyes and viewpoint (Howard and Steptoe, 2016). These eyes were separate meshes which allowed for them to rotate individually from the head and each other. The eyeballs and sockets were scaled to closely resemble the size one would expect from a real person.

Next, the abstract floating limb avatar needed some visual characteristics. These were added to the character using shaders. Through the use of a shader, the authors were able to change the avatar's skin colour. However, realistic skin rendering should be avoided. The reason for this is that it is close to impossible to match the real-world's and the virtual environment's skin colours and texture exactly. This slight mismatch may induce the uncanny valley relationship and thus decrease the participant's immersion and believability in the virtual environment. In this project, the skin colour of the avatar was somewhat orange, rendered using a cel shading style. This skin rendering was intended to be somewhat familiar to the participants, however, without being so similar that the uncanny valley relationship would make them lose their immersion in the virtual environment (Howard and Steptoe, 2016).

The abstract floating limb avatar can be seen in Figure 3.2

Finally, this study's avatar had its oculesic behaviour represented through four different eye-gaze control models; *occluded*, *static*, *modelled* and *tracked* eye-gaze. Because only one eye tracker was available for this study, only the fixer participant's avatar would change between these models. The fetcher participant would always be represented with static eye-gaze.

In the first condition, occluded eye-gaze, the participants saw an avatar wearing an HMD, modelled by the authors, in order to occlude their eyes. Furthermore, the lack of eye-gaze forces the fixer participant to focus on other collaborative aspects, such as physical and verbal social interactions throughout the experiment (Roth et al., 2016a). The study's avatar, using the HMD, can be seen in Figure 3.3.

In the second condition, static eye-gaze, the fetcher participant saw an avatar with static eye-gaze. Static eye-gaze simply means that the eyes did not move under any circumstances. The fetcher participant had to discern the fixer participant's head gaze direction in order to pick up and transfer the correct object. Based on (Steptoe et al., 2009)'s experiences, the authors expected the static eye-gaze model to perform worse than tracked eye-gaze, but better than modelled eye-gaze. The study's avatar with visible eyes can be seen in Figure 3.4.

In the third condition, modelled eye-gaze, the fetcher participant saw an avatar with modelled eye-gaze. The mathematically based eye-gaze model controlled the avatar's oculesic behaviour randomly in regards to saccadic movements and gaze direction. Based on Vinayagamoorthy et al. (2004)'s findings, the authors expected modelled

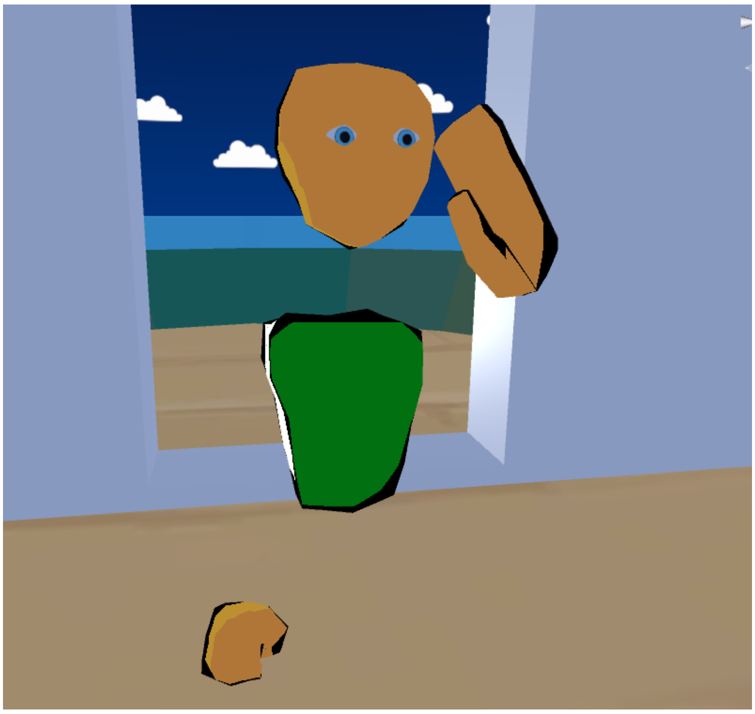


Figure 3.2: The abstract floating limb avatar used throughout the study.

eye-gaze to increase quality of communication. However, modelled eye-gaze was expected to perform the worst in regards to task performance, cause most errors, and be the most confusing eye-gaze model, based on the findings by (Steptoe et al., 2009). The eye model used was the *Realistic Eye Movements* asset by (Knabe, n.d.).

In the fourth condition, tracked eye-gaze, the fetcher participant saw an avatar with eye movement controlled by the other player's tracked eye-gaze. This eye-gaze model was expected to best enable the fetcher participant to discern where the fixer participant was gazing. Also, eye tracking was expected to yield the best task performance i.e. the lowest error rate and completion times, based on the literature (Steptoe et al., 2009; Murray et al., 2009; Vinayagamoorthy et al., 2004; Roberts et al., 2009).

3.4 Objects

For this study, a variety of objects was modelled using *Autodesk Maya*. The objects used for the object-based task and the shelf used for placing the task objects on will be elaborated upon throughout this section.

3.4.1 Super Shapes

For this study, a way of estimating gaze direction for the fetcher participants was needed. For this purpose, it was decided to implement relatively small objects, approximately the size of a handball, in the SVE. In the first iteration, the objects were randomly generated using Autodesk Maya's *Super Shapes* (AUTODESK.Help, 2018) cube and sphere based primitives. Using this feature, the authors could create objects which were unfamiliar to the participants. The intention behind this was to make it more difficult for participants to tell each other to pick up a specific object, such as the cylindrical, pyramid shaped or spherical object, etc. Finally, the objects were modelled to morph between two or three states in order to further avoid participants being able to describe them. The super shaped objects can be seen in Figure 3.5.

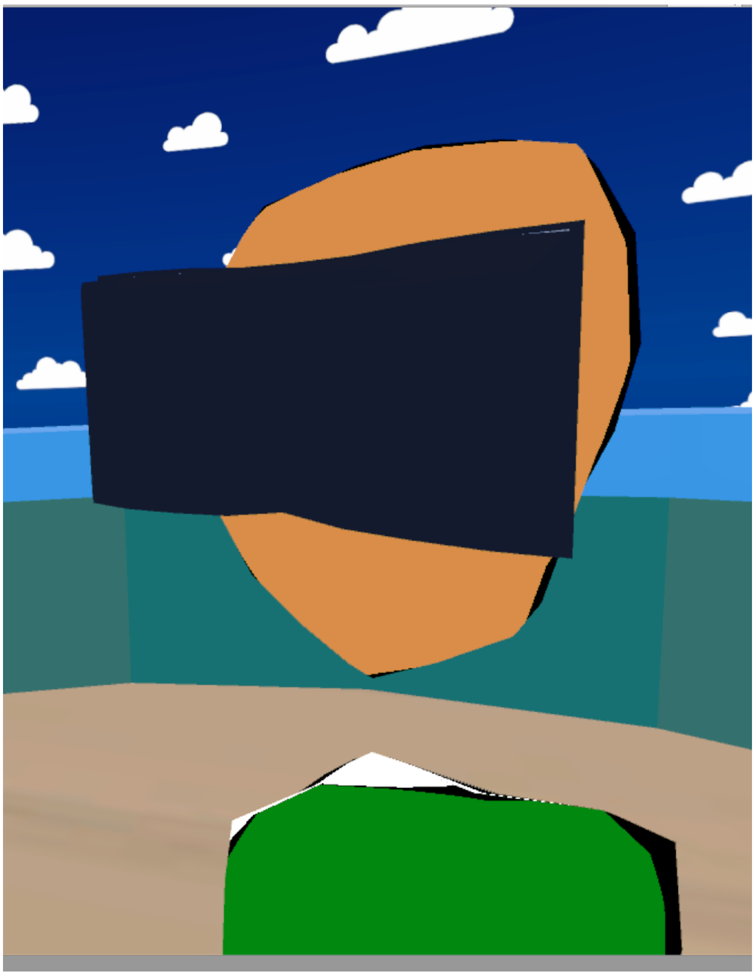


Figure 3.3: The study's avatar with the HMD added.

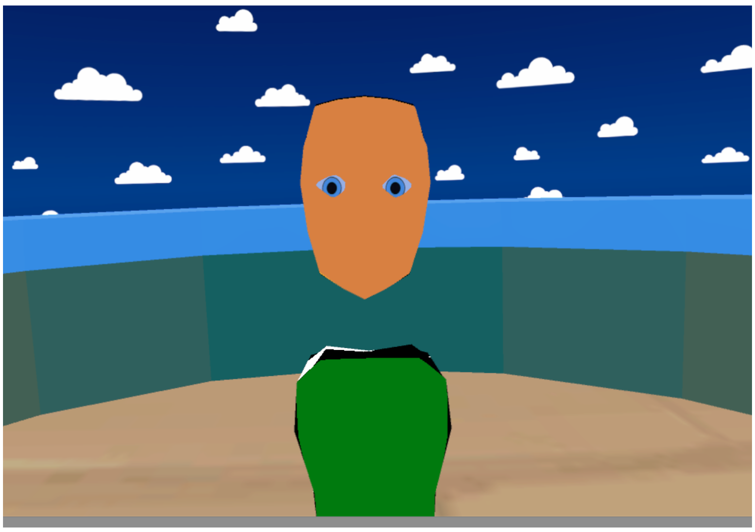


Figure 3.4: The study's avatar with visible eyes.

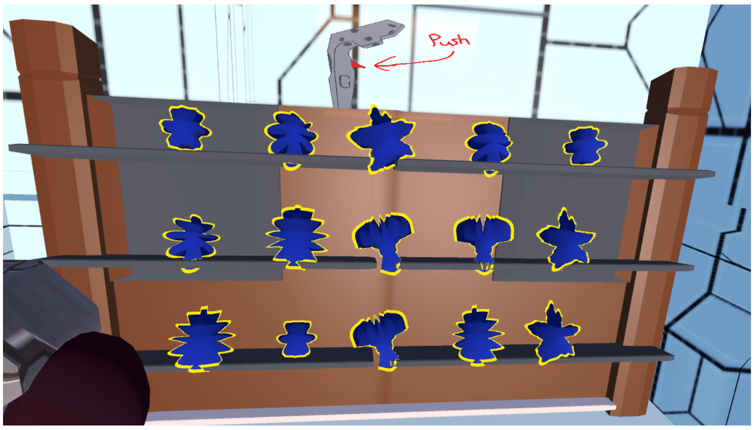


Figure 3.5: Selection of the super shapes used in the study.

During internal testing, the authors were made aware that focusing at the center of the objects was highly difficult due to their special geometric properties. Elaboration on the preliminary experiments is provided in Chapter, 5, *Preliminary Experiments*. Because of this, it was decided to discard the super shapes and instead use standard Unity sphere primitives for the preliminary experiments.

The objects in the SVE had a shader added to them in order to make them clearly visible. It also conveyed the affordance that something would happen when touching them with the participant's virtual hands. This was in the form of an outline shader which was added to the spheres and super shapes. This way it was easier for participants to focus on the objects and especially their centre points.

It was decided to place the objects vertically in the SVE, as seen in studies (Peters et al., 2010; Roberts et al., 2009; Elgarf et al., 2017; Murray et al., 2009). This was done as it was believed that this placement would distort the participants' vision the least, thus providing the best circumstances for discerning precise eye-gaze direction. It was decided to model a shelf for placing the objects in. This enabled the authors to place the objects in a grid pattern. The first two preliminary experiments used a 3x3 grid, while the rest of the experiments used a 5x3 grid.

3.4.2 Waiting Room

Before starting the task, the participants were located in a different VE where they would, for example, complete their calibration. This area was designed to be open and bland, though with some visual features relatively close the viewpoints of the participants, so they had something to frame their position around. Between the main experiment's conditions, the participants had to answer a module of the *Game Experience Questionnaire* (*GEQ*). An explanation of the experiment and its various parts can be found in Chapter 6, *Experiment*. The participants were placed in the waiting room while answering questions which were from the social presence module of the GEQ. The participants could see their virtual selves in a mirror. Using a mirror in a VE increases the participants' awareness of their virtual body/avatar (Dodds et al., 2011). The awareness of one's own virtual self happens as the avatar is embodied through presenting the participants to themselves as an avatar in the virtual environment. In short, the participants are able to inspect their own and each other's avatars from a first-person perspective in a VE. The participants first saw their avatars as they faced directly towards the mirror yielding an egocentric view using the HMD. Finally, mirrors let the participants see their own virtual bodies and thus their IVBO is increased. Being able to see themselves should help the participants' body ownership (Jerald, 2016: p. 301-302) (Latoschik et al., 2017; Howard and Steptoe, 2016).

3.4.3 Monitor

A way for conveying to the fixer which object they should get the fetcher to transfer was needed. For this purpose, an old CRT TV was 3D modelled and used as a monitor. The intention was that the monitor should act as an information screen seen at for example train stations. Just like such screens, the CRT monitor was placed above the fixer's head. Thus, they had to look up at the monitor and away from the fetcher in order to discern which object the fetcher needed to transfer next. The monitor can be seen in Figure 3.6.

Figure 3.6: The CRT monitor used for showing the fixer the current object to have transferred.

3.4.4 Spatial UI

A non-diegetic spatial UI Unity (n.d.b) was added to the fixer's viewport, highlighting the current object they needed the fetcher to transfer. This was done by having a crosshair graphic around the current object. When an object was transferred and the fixer had thrown it into the black hole, the spatial UI changed to highlight the next object. This highlighting would continue until the experiment was concluded. This highlighting system was intended to simplify the collaboration, as the fixer more constantly would gaze on the current object, thus making it simpler for the fetcher to discern which object was being looked at. The spatial UI can be seen in Figure 3.7.

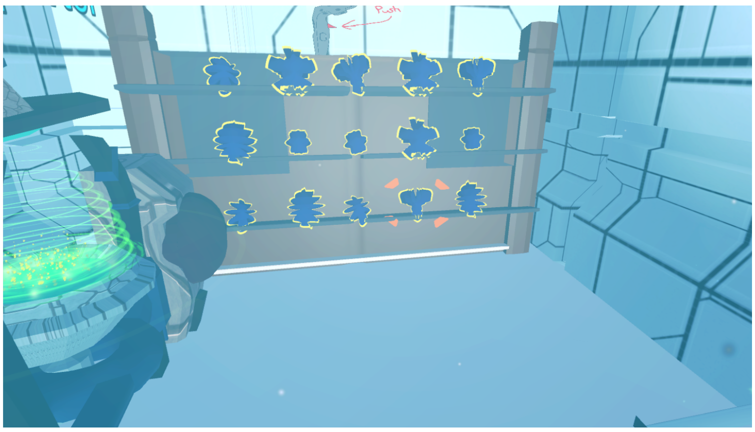


Figure 3.7: The spatial UI added to the fixer's viewport in order to highlight the current object.

3.5 Environment

For this study, it was necessary to have a VE to use as the main SVE. Throughout this section, the SVE's most important aspects, these being the teleporter, receiver, walls, window, and the black hole, will be elaborated upon.

3.5.1 Teleporter

In order to enable the fetcher participants to transfer objects from their part of the SVE to the fixer's, the authors decided to create a teleporter. For an elaboration on the experiment in regards to task, participants, environment, etc., see Section 5.3 *Preliminary Experiment 3: Testing All Four Oculesic Behaviour Conditions*. It was decided to use a teleporter, as this would enable the fetcher to transfer the object to the fixer, while making sure that the fixer could not just reach through the window and pick up the object in question themselves.

The fetcher participant puts the sphere in the teleporter's particle field and pushes its button. If it is the correct sphere, it is dissolved, teleported, and transferred to the fetcher's part of the VE. In case it is the incorrect object, it is declined and pushed away from the teleporter field. The teleporter can be seen in Figure 3.8. 3.8.

Figure 3.8: The teleporter used throughout the study's experiment.

3.5.2 Receiver

In order to enable the fixer to receive objects from the fetcher's part of the room, the authors decided to create a receiver for the teleporter. An elaboration of the experiment in regards to task, participants, environment, etc., can be seen in Section 5.3 *Preliminary Experiment 3: Testing All Four Oculesic Behaviour Conditions*. The receiver can be seen in Figure 3.9.

3.5.3 Walls/Window

In order to create a two-part room inside the SVE, boundaries had to be created. For an elaboration of the experiment in regards to task, participants, environment, etc., see Section 6, *Experiment*. This room separation was created using walls. The room was a simple rectangular room with a dividing wall in the middle. The dividing wall contained a hidden window, which revealed itself when both participants had completed the tutorial. Until then, it was closed and looked like a wall. The overall idea of the dividing wall and window was to hinder the participants from walking into each other's part of the room. Furthermore, the room division should hinder the fixer participant from grabbing the objects from the fetcher's side of the room and put them into the black hole themselves. This was also why the teleporter and receiver were implemented, to force both participants through certain collaborative steps before

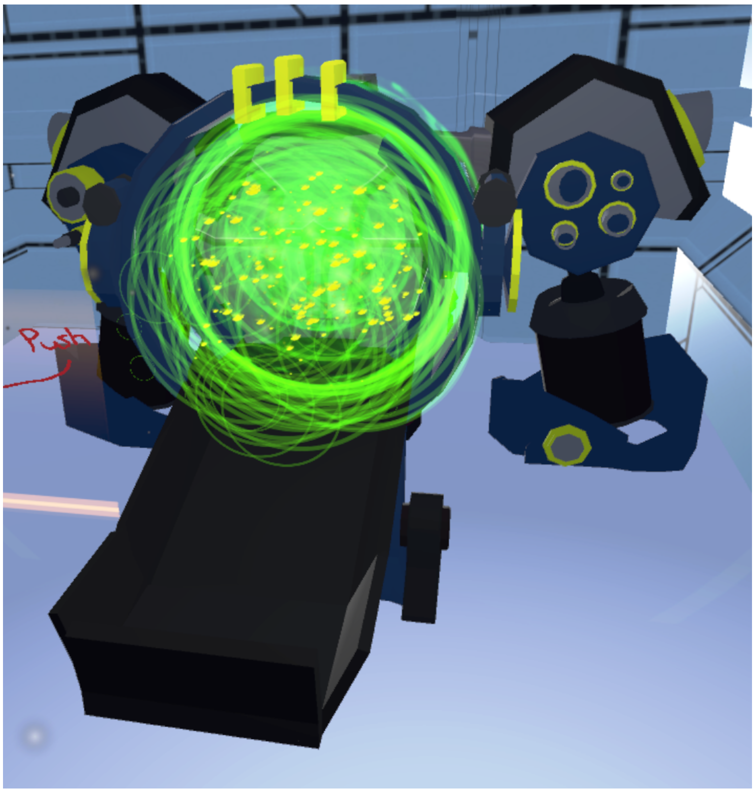


Figure 3.9: The receiver used throughout the study's experiment.

being able to complete the task and thus the experiment. The combined dividing wall and window can be seen in Figure 3.10.

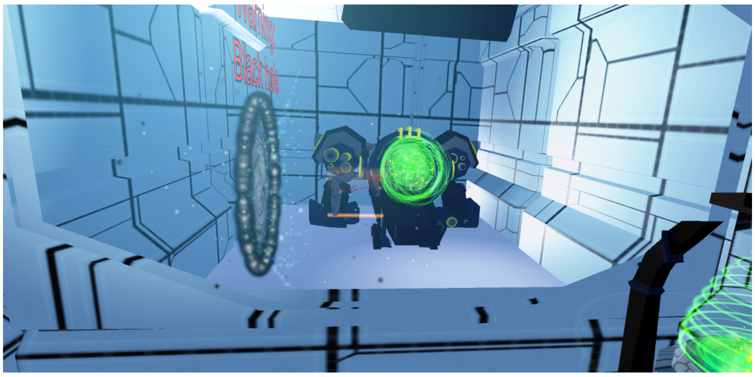


Figure 3.10: The combined dividing wall and window used for dividing the SVE into two rooms and for hindering participants from entering each other's part of the SVE.

3.5.4 Black Hole

A black hole was implemented in order to facilitate collaborative behaviour between the two participants. Whenever the fixer receives an object from the fetcher, they have to put it into the black hole. This interaction is needed in order to keep the black hole from imploding and destroying everything. In reality, the interaction was implemented in order to force the participants to interact in a specific order, ending with the fixer putting the teleported objects into the black hole. The black hole consisted of a particle effect and an audio source, indicating to the participants that it is in the process of sucking in and destroying everything. The black hole can be seen in Figure 3.11.

Figure 3.11: The black hole used for this study.

3.6 Tutorial Design

After both participants had been connected to the waiting room and had confirmed that they were ready to start the task, they were transferred to the SVE. Before they could start the collaboration part of the experiment, they were presented with a tutorial. The participants were placed in their respective sides of the SVE, however the window had not been opened and they could not see each other. This division of the participants was decided as they would receive different instructions due to their task roles.

The dividing wall was closed during the tutorial and was used as a canvas onto which instructional text was written. This way, the participants could read what the task was about and what their role was.

Voice-overs were played for each participant, which informed them of the experiment while looking around the SVE. Thus, this should reduce the risk of information potentially being lost by only informing the participants using text. Voice-overs were also utilised as the participants would be standing in the same physical space during the experiment. Initially, it was considered whether one or two simultaneous facilitators should be used during the tutorial. However, using two facilitators was discarded due to the risk of auditory overload, which makes it difficult to focus on what was being said.

3.7 Summary

Throughout this chapter, an overview of the SVE's design has been provided. Through Section 3.1, *Task Design*, the workings of the participants' task are presented: two participants interact with each other, where they can use eye-gaze and other communication channels to communicate about the position of various objects inside a VE.

Throughout Section 3.1.2, *Interaction Design in Task*, the SVE's various interaction design elements are elaborated upon. The most important of these being how the various objects were designed to ease their use for the fixer and fetcher.

Due to only one available eye tracker, only one participant could wear it during the task. In order to make the task more compelling to the participants, it was decided through Section 3.2, *Game Design*, that a goal was needed for the study: closing a black hole. The SVE also needed simple and understandable rules and mechanics. Finally, the main resources of the task was found to be time and information, in order to enable the authors to conclude whether the eye tracked condition yielded better object-based collaboration.

Section 3.3, *Avatar Design*, explains that this study's avatar was a gender neutral, floating limb, cartoony character, as this was the style utilised commercially at the time of the study. In Section 3.4, *Objects*, it is explained how the study's various objects were created along with their intended use. In addition, it is elaborated that the SVE mainly consists of objects (super shapes) to transfer between the participants, a shelf for storing the objects, a waiting room for before beginning the task, a monitor, and spatial UI for informing the fixer of the current object.

Section 3.5, *Environment*, elaborates on the intended use of the teleporter, the receiver, the dividing wall/window, and the black hole.

In Section 3.6, *Tutorial Design*, it is explained how the participants start in their own part of the SVE with the window closed. Here, the participants are informed about the experiment in its entirety together with their task.

Chapter 4

Implementation

Based on the design decisions presented through Chapter 3, *Design*, it was decided that an SVE with a wall dividing the participants should be implemented. It was found that the implementation needed to take into consideration that only one participant could wear an eye tracker during the task. Game design elements were implemented for creating a goal, understandable rules, and mechanics for making the experience compelling to the participants. Interaction design elements were implemented in order to guide and help the participants through the task. It was decided to include a tutorial for informing the participants about the point of the task as well as their roles. Including the tutorial meant that informative text and voice-overs had to be created and implemented, together with tutorial specific mechanics.

During this chapter, an elaboration on implementation choices and solutions are provided. Some of the implementation solutions are based on literature as well as experiences gained throughout the preliminary experiments, seen in Section 5, *Preliminary Experiments*. Thus, the chapter should be read as a recursive process. The chapter presents information on the game engine and API used for the study. The chapter also presents information on various aspects: the spatial UI, the mirror, networking between two PCs, the various eye gaze controls, and the motion capture system. Other aspects are elaborated on too: the morphing objects, the shaders, the developer interface, and the data logging system. The chapter also elaborates on setup and calibration of *Pupil Labs* and the room scale tracking. Finally, the chapter concludes with a summary of the most important aspects of the implementation for this study.

The prototype was implemented using the *Unity* game engine. An HTC Vive was used as the IPT, running on the *OpenVR* API. The *SteamVR* plugin was used for basic VR development, such as controller management and having a play area indicator.

The eye tracker used was developed by the open source project Pupil Labs and fitted inside a consumer version of the HTC Vive. A plugin was used to make the data captured by the Pupil Labs software transferable to Unity.

4.1 Object Interactions

The most common interaction in the prototype was the participants' ability to pick up and throw objects. This interaction system inherited some functionality from buttons that were developed earlier in the process. Each of the participants' controllers had a component which managed input as well as the objects currently being interacted with. Whenever an object came in contact with a hand, it added itself to a list of objects to that hand. The object also kept notice of whether both hands were interacting with it. Similarly, when a hand left the object, it removed itself from the list. Whenever a button was in a list of either of the hands, it changed its outline to another colour and became slightly bigger. When an object got removed from both lists it reverted to its normal outline. If the participant pressed the trigger button on either controller, the specific controller sent an *Action* message only to the first object in its list, and when the trigger was released it sent an *ActionUp*. Thus, different buttons had different functionality in their Action functions.

The objects that participants could interact with had an Action functionality that added a physics joint to bind the object with the hand. The ActionUp function destroys the joint and applies to the object's physics component the velocity and angular velocity from the controller. Whenever an object was thrown or collided with another object, it began to listen for when the object's velocity was close to rest. At this point, if the object was below a specific height value (about knee height), the object's position and velocity was reset as it was considered to have fallen to the ground.

4.2 Spatial UI

The fixer participant had a non-diegetic spatial UI element, a crosshair, as a guiding element to which object was the currently active one. The spatial UI was placed as a normal UI element in world space over the current object and set to constantly face the participant, in order to always follow the fixer's gaze direction. The crosshair was scaled linearly as it got further away from the participant, so that it remained a constant size in camera view.

4.3 Object Morphing

As mentioned in Chapter3, *Design*, several objects were created using different blendshapes. These objects were given a script that continuously cycled through the blendshapes. Blendshapes work by assigning a weight to each shape, which the final object will then apply to a blend between them. The weights go from 0-100, but are not restrained to a sum of 100; in case the total weight sum exceeds 100, a new shape would be created by adding the others. The script created for this project set the first shape's weight to 100 and the rest to zero. It then linearly decreased the first weight while increasing the next, until the values are completely reversed. This process is repeated with the next shape, looping through all the shapes.

4.4 Shaders

Shaders are programs running in parallel on the GPU, which in most cases are used for graphical purposes. In this study, two different shaders were created, one for aesthetic purposes, and one for creating features for feedback mentioned in Chapter 3, *Design*, though most environmental objects used the *Standard* shader provided in Unity.

The aesthetic shader was a non-photorealistic rendering using cel-shading. A black outline was drawn around characters based on the value from the dot product between the normal of the fragment and the view direction. The more the normal faces away from the camera and becomes perpendicular to the view direction, the closer the value will be to zero. A threshold is set and when the value is lower than that, an outline is drawn by changing the fragment colour to black. The normal diffuse shading is quantised using another threshold. Instead of having lighting that smoothly goes from unlit to lit, a threshold is set to determine whether a single lit or unlit colour should be used. The specular highlights also use a threshold to determine if the specular is strong enough to be drawn as clear white or none at all (n.d., 2018). This shader was used on the avatar as seen in Section 3.3, *Avatar Design*.

For the interactive objects, a shader with two programs was created. The first program created a rigid outline (compared to the more stylistic outline in the previous shader) that could change width and colour to indicate the state of an object, e.g. inactive, active, or selected. The vertex positions are multiplied by a width indicator, which expands the entire shape by a small amount which is then drawn before the other program in a single colour. The second program renders the object at normal size with no lighting. This creates the effect of having an outline, as the first pass peeks out behind the normal pass at a constant size for each side. Both programs share the variables used to create a *dissolve* effect, which happens when the objects are teleported. This effect uses a texture lookup and uses the red channel of the colour returned to determine if any of them are below a moving threshold. If the red value is below this threshold, the entire fragment is discarded. A slightly higher threshold is used to determine the edge of the dissolve. Thus, if the value is between the two thresholds, it is determined to be at the dissolve edge and is given a specific colour, seen in Figure 4.1.

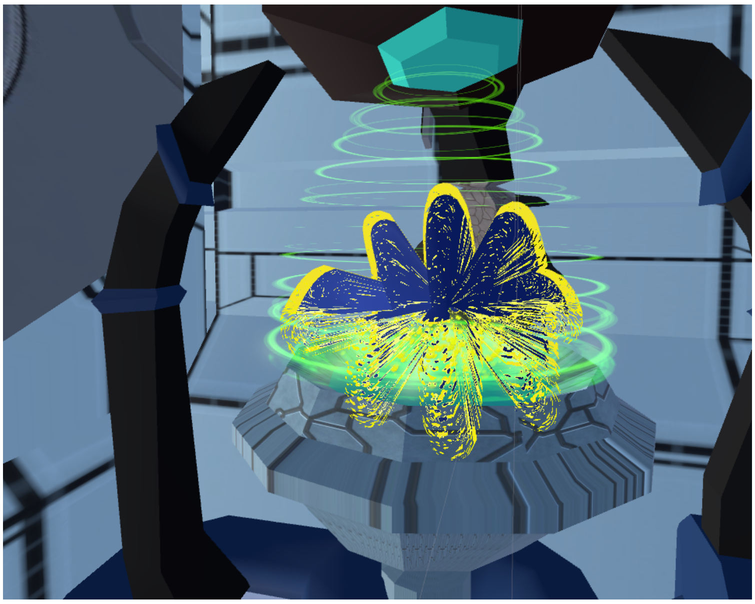


Figure 4.1: Interaction shader shown while being dissolved. The outline, dissolve effect, and dissolve edge are all visible in this example.

4.5 Calibration Method

In order to ensure that the avatars worked for people of different sizes, a certain calibration method was implemented. Before the task began, the user was asked to stand up straight, assume a *T-pose*, and press a button to confirm having done so. At this moment, the height of the user was recorded based on the y-position of the head, while the width was based on the distance between the user's hands. The avatar's components: head, hands, and torso, were then individually scaled. The y-axis was based on the ratio of the user's height over the avatar's original height, and the x- and z-axis were based on the ratio of the distance between the user's and the avatar's original distance between their hands. This moment was also used to move the position of the player so that a custom center point of the tracking space could be defined. The way the lab was set up, two Vive systems were sharing one physical space, with their virtual centres positioned close to each other in the real world. To make better use of this tracking space, the position of the participant at the moment of calibration was considered to be the x and z position of the head. The offset between the virtual center and the player position was then used to move and rotate the virtual space whenever the participants moved to a new location, such that the calibrated center position was positioned properly, see Figure 4.2.

4.6 Virtual Mirror

As part of establishing the participant's self-embodiment of their avatars, they were shown a mirror image of their virtual selves before each playthrough of the task, and before the tutorial. During the early phases of implementation, this was simply a virtual camera that saved its image to a *RenderTexture* which was then displayed on a plane.

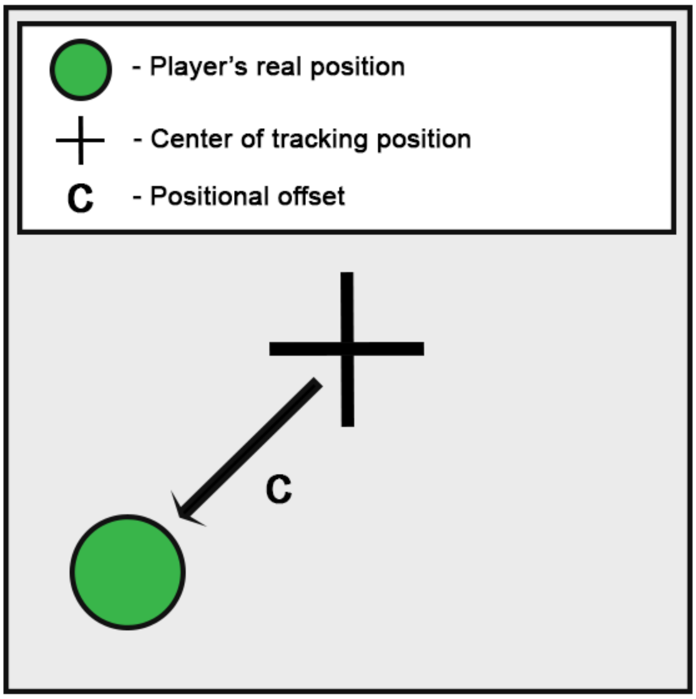


Figure 4.2: The calibration setup: *X* indicates the tracking space's centre, the circle indicates the participant's position and *c* indicates the positional offset vector.

However, the local representation of the user's own avatar would only have the hands being rendered. To solve this, the mirror would instead be mimicking a clone of the participant's avatar which was rotated and scaled to look like the actual avatar being mirrored. The virtual mirror can be seen in Figure 4.3

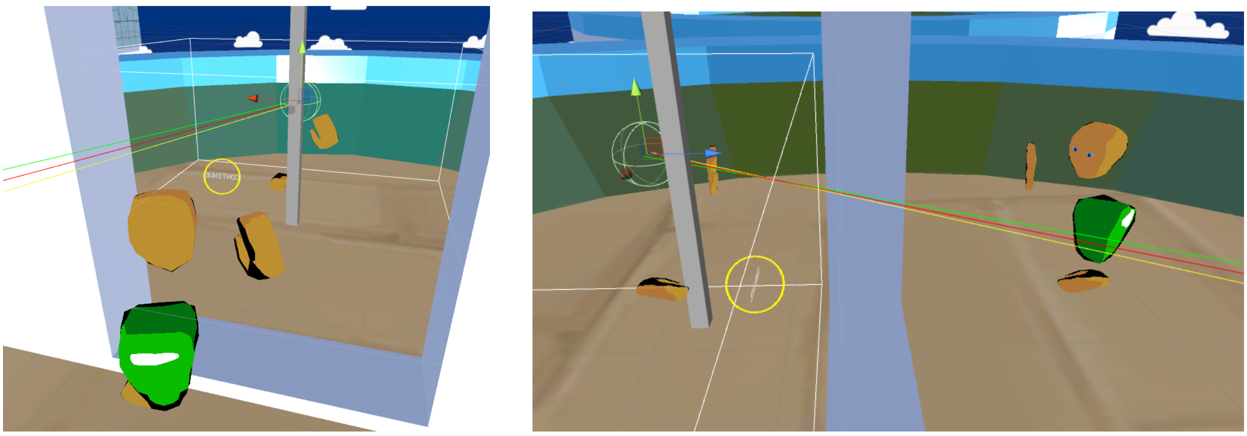


Figure 4.3: The virtual mirror used in the study. The participant's avatar is mostly not rendered, so a copy of it mirrors its movements.

4.7 UNet Implementation

Unity had a high level networking API called *UNet*. This allowed for quick prototyping as many common requirements for networking had already been implemented. However, it also meant that a prototype had to follow the structure defined by the interface. UNet used by default a server authoritative structure running on *User Datagram Protocol* (*UDP*), meaning that the state of the application on the server was the one that determined all actions sent to the clients. Each object that should be affected by networking in the scene was registered with a *NetworkIdentity* component and its corresponding *NetworkId*. By default, each of these objects were said to have *Server Authority*, meaning that they were fully controlled by the server. When a client joined the server, it spawned a *Player* object, and gave that object *Local Authority* to the client it belongs to. Through that player object, the client could request authority over other networked objects.

To make functions in the code that interacted with the networking, the main difference was the use of two attributes: *[Command]* and *[ClientRpc]*. Commands were functions that could only be called on objects with local player authority and were then sent to be executed on the server. The ClientRpc functions could only be called on the server, but were executed on every client, see Figure 4.4. Thus, to have a client perform an action that is reflected on other clients, it was appropriate to create a Command function which called a ClientRpc function that had the desired functionality. Simple variables could be synchronised using the *[SyncVar]* attribute. When a variable with this attribute was changed on the server, it automatically tried to synchronise it with all clients. SyncVar was used, for example, to synchronise the current frame for the avatar hands' grab animation.

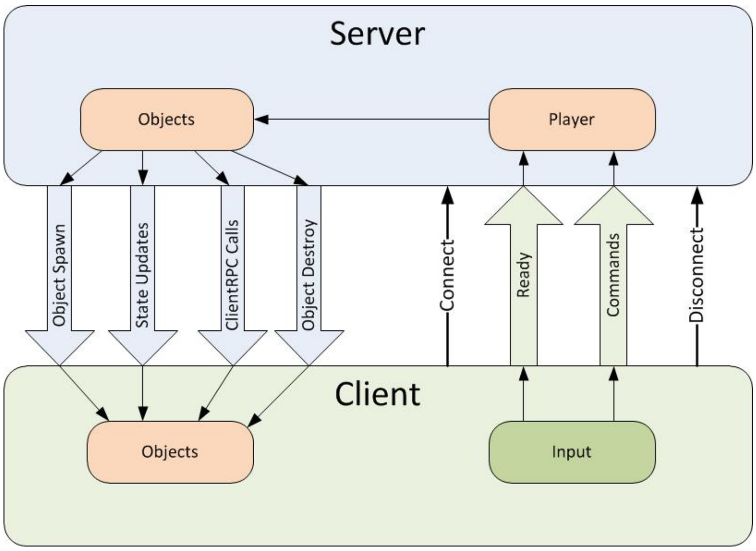


Figure 4.4: The UNet implementation in regards to the communication between server and client, respectively. Figure taken from (Unity, n.d.a)

.

To synchronise the participants' avatar's Transform components (position and rotation) UNet had a simple component called *NetworkTransform*. NetworkTransform managed the values of the Transform component and synchronised them with the clients. For objects with physics simulations, the NetworkTransform's mode could be set to *Rigidbody*, which then also synchronised and interpolated velocity and acceleration. In this study, when a participant tried to grab an object, it was grabbed locally, while the server was requested for authority. When the authority was granted, the player object for the participant who had grabbed the object would then determine the position and rotation of the object on the server, which was then further transmitted to the other clients.

Other parts of the prototype that were networked was the usage of the teleporter and the state of the task, i.e. what object was to be teleported next. When the teleporter's button was pressed, the player object that pressed it asked for authority over the entire teleporter object. Then the function moving the object was called using a Command that called a ClientRpc function. Similarly, when an object was put into the black hole, a Command into the ClientRpc method was called to destroy the object and advance the task state to the next object. To synchronise the current object to be found, the server determined the object and sent the NetworkIdentity as an argument in the ClientRpc method. The way the Vive system worked in Unity was by having a set of local GameObjects that represented each tracked object such as headset, controllers, etc. To ensure that the tracked objects were represented at all times, they did not have a NetworkIdentity, but were instead controlled locally on each client. Instead, when the player object with the avatar spawned on a client, it linked to the Vive objects and ran a local script to follow them and give the avatar movement.

4.8 Pupil Capture Service

The software that ran Pupil Labs' eye tracker was either *Pupil Service* or *Pupil Capture*. For this study, Pupil Capture was used. The software used an image processing algorithm on each video feed from the sensors to calculate various properties of the eyes. The main function of these properties was to determine where the user was looking by finding the iris and pupil and then extrapolating the entirety of the eyeball. The software had two tracking modes, one for 2D and one for 3D. Since this study was about looking into a virtual 3D world, the 3D option was used. This system had an optimal accuracy between 1.5-2.5 degrees (Pupil Labs, n.d.). However, that was when using the *Pupil Pro* headset rather than the HTC Vive add-on. The algorithm then calculated the normal vector of the eye, i.e. the direction the eye is looking. When both eyes were used, the system could calculate convergence, which is the point on the z-axis where the eyes' line of sight intersect (depth). The sensors could be adjusted to run at different resolutions and frame rates, with performance tradeoffs in both. For this study, a resolution of 640x480 was used with a maximum frame rate of 120Hz. In Figure 4.5, an example frame can be seen with markings on the eye to illustrate what the algorithm has found.

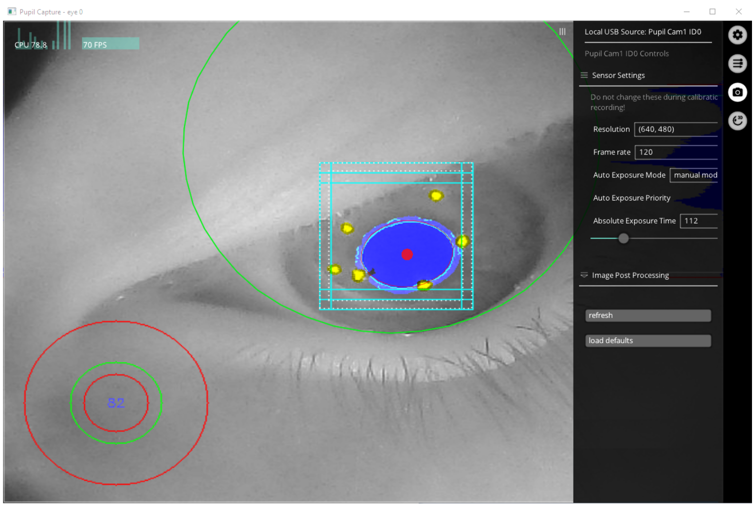
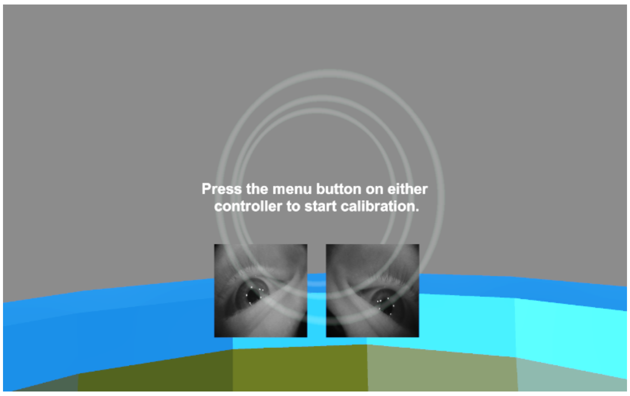


Figure 4.5: The camera feed of the eye sensors with visual representations of the different parts of the algorithm.

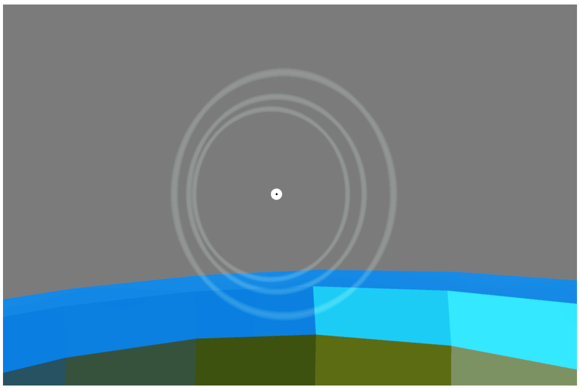
In the bottom left corner, three circles were displayed. The red circles indicated the set minimum and maximum size of the pupil that Pupil Labs was allowed to accept, while the green circle was the size of the pupil in the current frame. The blue colour should only cover the pupil, while the red should be opaque and indicate the center of the pupil. If the blue colour covered too much or too little, there were several parameters of the algorithm that could be tweaked to get a more accurate tracking. The green circle around the eye was the estimation of the entire eyeball. Yellow dots were considered spectral reflections and were to be expected, so these were filtered away by the algorithm (Kassner et al., 2014).

One of the most important properties being calculated was the confidence value. This floating-point value ranged from zero to one and determined how confident the system was that it had properly found the correct values for the other properties. This was useful for determining the quality of the current tracking. A rule of thumb was to have a consistent confidence level of at least 0.6 to ensure useful data (Pupil Labs, n.d.).

Because of how people have different facial morphology, and because the headset was adjusted differently for each person, the eyes were not situated the same place in front of the sensors. Therefore, a calibration process was needed to ensure accurate eye tracking. This process ran on the Pupil Capture service, but also worked with the Unity engine to provide a visual interface. On the visual side, the user was introduced to a series of rings at different depth levels. A small dot was placed on one of the rings and the user was asked to focus on it. This dot moved to different locations and depths on the rings while the participant focused their gaze on it, which allowed for the system to calculate the offsets it needed to map the pupil positions to the gaze positions. The rings and dots for the eye tracking calibration can be seen in Figure 4.6. During this calibration, the dot changed colour based on the confidence level so it immediately could be noticed if the tracking was poor. When both eyes had full confidence the dot was coloured white. While losing confidence on the left or right eye, the dot turned yellow or purple, respectively. If the confidence was low for both eyes, the dot turned red. Before the calibration began, the system displayed the camera feeds of the eyes, so the user could determine if their eyes were placed in full view of the cameras.



(a) The camera feed of the eyes in front of the rings used in the calibration.



(b) The rings with the white dot as the focus marker in the centre.

Figure 4.6: The eye tracking calibration process for the Pupil Labs eye tracking hardware.

4.9 Oculesic Behaviour Conditions

Throughout this section, the four oculesic behaviour conditions utilised in this study are elaborated upon.

4.9.1 Occluded Oculesic Behaviour

For this behaviour, the HMD model was simply added in front of the avatar's face and the eyes themselves were not moved.

4.9.2 Static Oculesic Behaviour

When using static oculesic behaviour, the eyes' target was positioned 100 metres directly in front of the face, making the movement of the eyes nonexistent, compared to the scale of the VE participants were placed in.

4.9.3 Modelled Oculesic Behaviour

For adding modelled oculesic behaviour to the study, the Unity asset Realistic Eye Movements (Knabe, n.d.), was utilised. The asset was developed based on the work of several research papers such as (Lee et al., 2002), among others. The model took many different arguments and calculated the movement of the eyes and, if applicable, head and eyelid movement. The parameters included limiting values, such as maximum horizontal angle and minimum blink time. It also included parameters that change the behaviour such as *nervousness*, *look-at-player ratio*, and *stare-back factor*. These parameters could change for example how often the character changed from staring intensely at a specific object/person to averting their gaze. The eye model could also affect how the eyes move, e.g. by changing the saccade frequency and velocity. By default, the player object was set to the default camera of the scene which works when making the model look at the participant. For the task in this study, the model was set to change its target to the current object that was necessary to advance the task. The parameters were set so that the model would mostly look at the target object with minor random saccades, and infrequently look at the participant to simulate normal oculesic behaviour.

4.9.4 Tracked Oculesic Behaviour

The Pupil Labs software returned several different values that could be used to map eye movement. An often used value is the *GazePosition_3D*, which was the local coordinates for the camera object where the gaze was positioned, including depth. This was tested, but it was found that the convergence factor was often calculated poorly or not at all, defaulting to a depth of 0 which was inside the head. Because of this, this study used the returned normal vectors of each eye. An average vector of the two normals was calculated and a ray was cast along it, see Listings 4.9.4 and 4.1. The target position was placed at the first point the ray intersected with that was not transparent. These two methods were both implemented and compared, and it was found that in most cases their convergence points were at the same local x and y position, only when one eye was failing tracking the average vector became off. The accuracy of this method was tested in a preliminary experiment, elaborated upon in Section 5.1, *Preliminary Experiment 1: Eye Tracking Accuracy*.

```
Vector3 leftEyeDir, rightEyeDir;
leftEyeDir = (PupilData._3D.LeftGazeNormal).normalized;
leftEyeDir = Quaternion.Euler(transform.eulerAngles) * leftEyeDir;
rightEyeDir = (PupilData._3D.RightGazeNormal).normalized;
rightEyeDir = Quaternion.Euler(transform.eulerAngles) * rightEyeDir;
Vector3 averageDir = ((leftEyeDir + rightEyeDir) / 2).normalized;
```

Listing 4.1: Code used for calculating the target gaze position

```
RaycastHit hit;
Physics.Raycast(transform.position, averageDir, out hit, markerDistance);
calculatedLookAt.position = hit.point;
if (hit.collider == null) { calculatedLookAt.position = transform.↩
    position + averageDir * markerDistance; }
```

4.10 Motion Capture System

In order to recreate the subtle movements of a real person controlling an avatar, the *non-playable character* (*NPC*) avatar for the preliminary experiments was created using a motion capture system. In addition, the motion capturing should make the avatar seem more natural and realistic to participants. A system was made which was able to save the position and rotation vectors of a list of GameObjects as a *ScriptableObject*. To save performance, the system had a set rate of 24 Hz for when to record. The ScriptableObjects could be added to another system that could play back the recording on another list of GameObjects. If this list of GameObjects was identical to the recording, it would be a direct playback of the original recording. The recording system was used by an author acting out the movements required using the avatar with eye tracking. This motion capture could be played back on the NPC.

The playback system could blend between two recordings, which was used for looping an animation, transitioning between the end and start of the same animation. The playback system would also change animations, transitioning between the end of one animation and the start of the next. The blend used an *AnimationCurve* to determine the weights for each animation over time. The transition was set to one second and the curve was then normalised over that period of time. As the curve approached one on the y-axis, the weight would shift completely to the second animation.

In the second preliminary experiment, many recordings were to be used, as the avatar was supposed to look at nine different objects placed in five different positions, totalling 45 recordings. To speed up the process of recording these, a system was created that would systematically go through the objects by highlighting them. The actor could then initiate and finish recording using the motion controller which would then highlight the next object. The recording would also automatically be named based on the current highlighted object and its placement.

4.11 Developer Interface System

Initially, there were many options a participant had to choose from before the main task could begin. This was due to various early ideas that later got discarded, such as choosing avatar style, gender, and whether to host the server or join as a client. The system at the time had the participant choose using virtual buttons, see Section 4.1, *Object Interactions*. Eventually, this was discarded and instead a user interface for the authors was created, see Figure 4.7. The authors could choose all of the options required for each test before pressing the *Ready* button. The Ready button was greyed out until the system had properly detected a headset and the two controllers, hereafter the button turned green. Pressing this button applied all the chosen options and began the room scale calibration process.

A similar interface was created for the second preliminary experiment, seen in Figure 4.8. This was used both to test the motion capture and playback system, as well as the different setups of how the objects were placed by allowing researchers to manually scale and move the objects in quantified steps. It automatically gave a warning by changing some of the text to red if the angle between two objects from the NPC's point of view was less than twice the accuracy of the eye tracker.

4.12 Data Logging

For the first preliminary experiment, the accuracy of the eye tracking method explained in Section 4.9.4, *Tracked Oculesic Behaviour*, was recorded. To do this, the participant would look at an object and press a button. At this moment, the vector between the calculated eye target position and the head was saved as well as the vector between

Figure 4.7: The onscreen interface for the developers, where they could define avatar style, participant role, task condition, and mark the participants as ready for the task. This interface was only visible during the initial screen, which only the developers could see.

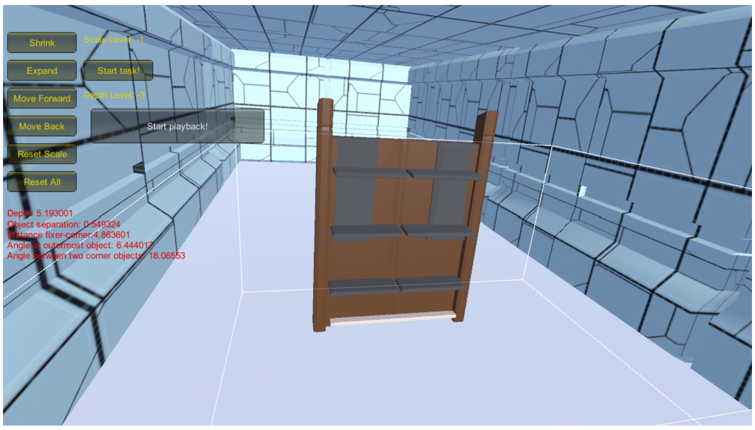


Figure 4.8: The onscreen interface for the authors created for preparing the second preliminary experiment. The red text indicates that the shelf is at a position where the objects placed on it would be too close to each other to discern the tracked eye-gaze.

the center of the object and the head. The angle between these vectors was then calculated. After all the trials of the experiment, the angles were averaged.

The second preliminary experiment would also log different data about the participant's actions. Whenever the participant picked an object, it would log whether it was right or wrong, as well as the distance between the picked object and the correct one. Additionally, some data about the current condition was logged: how far the shelf was from the NPC avatar, how large the distance was between the objects, and whether the NPC was using an occluded or tracked oculesic behaviour condition.

As part of the main experiment, the completion time of the task was recorded as well as the amount of times a wrong object had been picked. Furthermore, the total amount of time the eye tracker was fixating on an avatar's head was also recorded. This was done by checking the collision happening when calculating the target position of the eye movement, which was calculated and recorded for all conditions. If the collision was with the head of an avatar, the time between the current and previous frame was added to the counter. In addition, between each condition of the experiment the participants would be shown a series of questions to answer using a Likert scale. This was done using the virtual buttons, elaborated upon in Section 4.1,*Object Interactions*, which were lined up with numbers from one to five. Each participant then chose a number that would be logged with a question ID given to each question. An example of how the Likert scale questions were shown in the SVE can be seen in Figure 4.9.

Figure 4.9: The Likert scale questions as they were presented in the SVE, arranged in a semicircle around the participant.

4.13 Summary

In order to enable the authors to evaluate the design created throughout Chapter 3, *Design*, the SVE had to be implemented. Through the implementation chapter, several areas of the experiment's development have been elaborated upon.

The networking structure has been explained together with how the system makes use of server authority to handle several objects divided between client and server.

Some of the elements that lay behind the interaction design, like underlying theory behind the shaders, have been explained by the authors.

Through the use of the Pupil Labs software and the Realistic Eye Movements asset, the four different oculesic behaviour conditions were implemented. The tracked eye movements are used in unison with the implemented motion capture system to record animations for an NPC.

This prototype was developed over several iterations, with different purposes and focuses for the experiments conducted with it.

Chapter 5

Preliminary Experiments

Throughout this chapter, the study's four preliminary studies are elaborated upon. They will be covered regarding the motivation, methods, and processes for conducting them. In addition, the results, participants, and discussions resulting from the preliminary experiments are elaborated upon. Before beginning any of the preliminary experiments, a pilot study was conducted using university students. This was done in order to find and iron out various bugs as well as to see whether the overall procedure of the preliminary experiments worked as intended.

5.1 Preliminary Experiment 1: Eye Tracking Accuracy

In this section, the first preliminary experiment will be elaborated upon in regards to motivation, method, procedure, results, and discussion.

5.1.1 Motivation

The first preliminary experiment conducted set out to determine an average accuracy of the Pupil Labs eye tracker and the implemented algorithm for determining where the avatars look. It was quickly found during early implementation that people had slight variations in morphology, which seemed to affect the accuracy of the eye tracker (Kassner et al., 2014). Thus, participants were asked to undergo eye tracker calibration and perform an experiment associated task. This task was carried out in order to estimate the eye tracker's accuracy when adjusting its parameters to the best of the authors' ability.

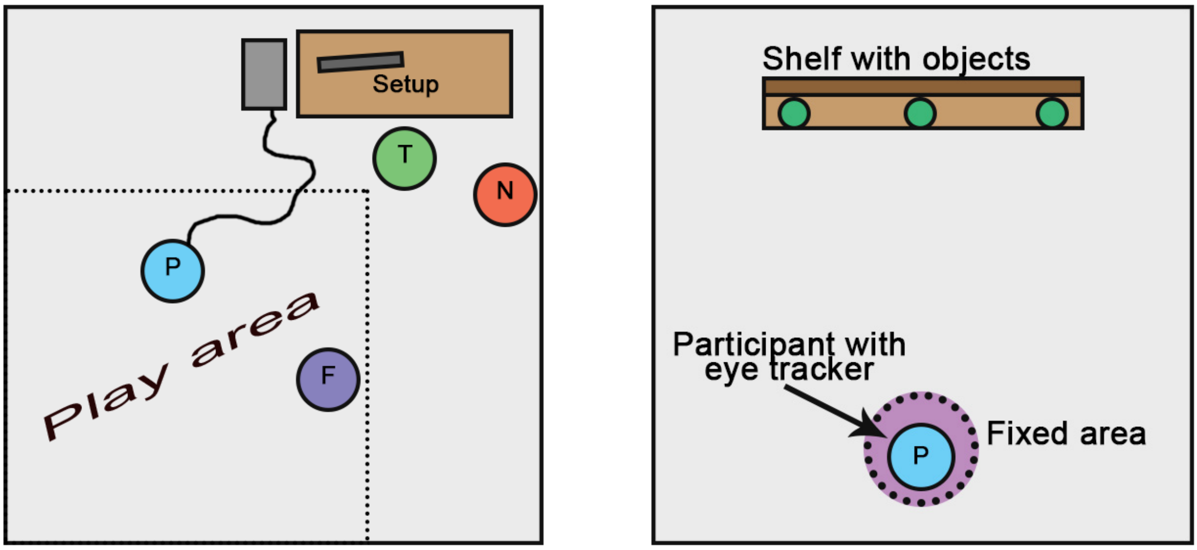
5.1.2 Method

Task

The participants had to put on the HTC Vive with the attached Pupil Labs sensors. In the VE, the participants were placed in an environment similar to the one used in the main experiment. The participants were positioned in one room looking through a window into an adjacent room, consisting of a shelf containing nine blue objects placed in a 3x3 grid. One object at a time would turn from blue to red, at which point the participant was asked to focus their gaze at that object. After focusing their gaze on the red object, the participant had to press the trigger button on the Vive controller. Hereafter, the red object turned blue once again and a random object placed somewhere else on the shelf became red. This was repeated 27 times per participant, as to gather data on how many degrees the eye tracker would deviate from the participant's focus point.

Even though the objects that became red were selected randomly, the selection was still based on a list containing the nine object positions. When an object was chosen, the position was removed from the list, thereby making sure that the object would not appear again within that cycle of nine objects. After having been through the nine available positions, the list was restored and repeated three times to make sure that each position was logged three times.

Setup



(a) The participant (*P*) was placed in the physical room inside the play area. The participant had an HTC Vive with a Pupil Labs eye tracker mounted on their head. In addition, a facilitator (*F*), note taker (*N*) and technician (*T*) was present during the experiment.

(b) The VE in which the participant (*P*) was carrying out the preliminary experiment. The participant was looking straight ahead to a shelf with nine objects in a 3x3 pattern, in which one object at a time was highlighted.

Figure 5.1: The setup for the first preliminary experiment.

The real world setup consisted of the participant (*P*), an experiment facilitator (*F*), a technician (*T*), and a note taker (*N*). The facilitator was responsible for introducing the experiment procedure to the participant and answering any questions they might have about the experiment. The technician made sure that the technical aspects of the experiment ran smoothly, and addressed any problems that might have occured. The note taker's role in this experiment was to observe the behaviour of the participant and to see if they were unsure of anything specific about the procedure or the experiment. When adjustments to the eye tracker software could not achieve good tracking, this was noted so it could be related to outlier performance.

5.1.3 Procedure

Participants were welcomed and thanked for agreeing to helping with the experiment. They had to fill in an informed consent form, thus agreeing to letting the data collected and photos captured during the experiment be used in this study. Participants were introduced to the controllers and the eye tracker calibration procedure. Then, the participants were equipped with the HMD and the technician adjusted parameters in the Pupil Labs software to take participant eye morphology into account. Next step was to perform the experiment described above.

After the experiment, the HMD and controllers were removed from the participants and they were asked whether they had questions or comments for what they had just experienced. Lastly, they were thanked for participating and offered a refreshment.

5.1.4 Results

Eight participants took part in the experiment, an average of 6.70 (SD = 7.48) degrees was calculated as their accuracy when using the eye tracker with the software calibrated by the authors. This average can be reduced since a clear outlier participant was detected during the test, as seen in Figure 5.2.

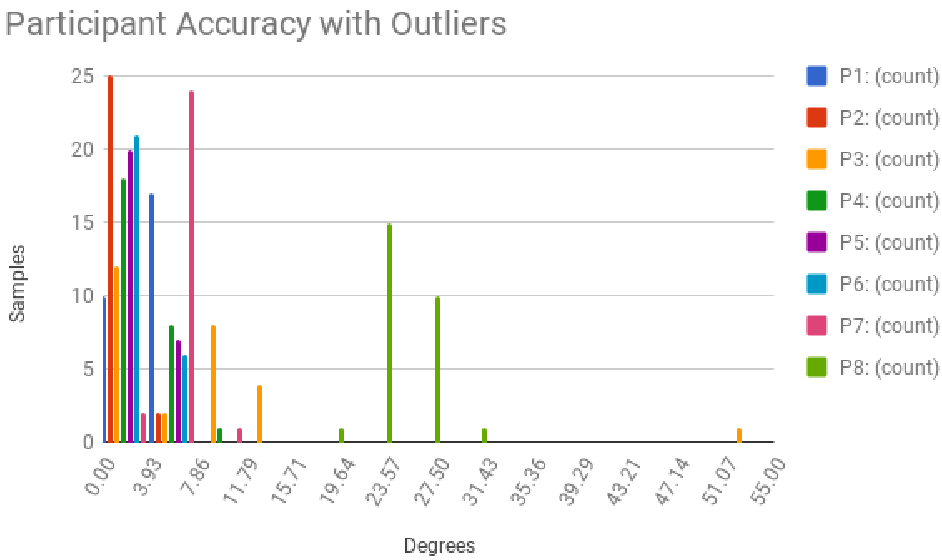


Figure 5.2: The accuracy of all eight participants, shown in a histogram. Participants 8 and 3 were considered outliers by the authors.

This particular participant had an average accuracy of 23.28 degrees, and removing this participant's data reduced the experiment's average accuracy to 4.33 (SD = 4.27) degrees. Furthermore, a second participant's data showed a deviating average accuracy, fluctuating between values similar to and triple that of other participants. Similar deviations were not found in any of the other participants' data. This participant's data were therefore also excluded from the final accuracy average. The final accuracy, excluding outliers, was 3.63 (SD = 1.57) degrees, using the data points seen in Figure 5.3.

This accuracy was not quite the expected 1.5 to 2.5 degrees that Pupil Labs document with their non-HMD products using their 3d algorithm (Pupil Labs, n.d.). As this study utilised the HTC Vive with the Pupil Labs eye tracking add-on installed, the difference between equipment utilised might be the reason behind the differing eye tracking accuracies.

5.1.5 Discussion

The adjustment process of the eye tracker's image processing parameters generally yielded good results. However, it quickly became clear that the parameters could not be adjusted to support the use of glasses within the HMD, as the frame of the glasses would partly occlude the eyes, and the eyes themself would appear slightly distorted by the glass lenses. Therefore, participants using glasses were asked to take them off before equipping the HMD. Participant 7 and 8's elevated results, in relation to the other participants, might be caused by them having to remove their glasses. Glasses having an effect on eye tracker accuracy made personal information questions about participant glass usage relevant in future experiment questionnaires.

When adjusting the eye tracker parameters for participant 8, it proved difficult to get reliable pupil feature

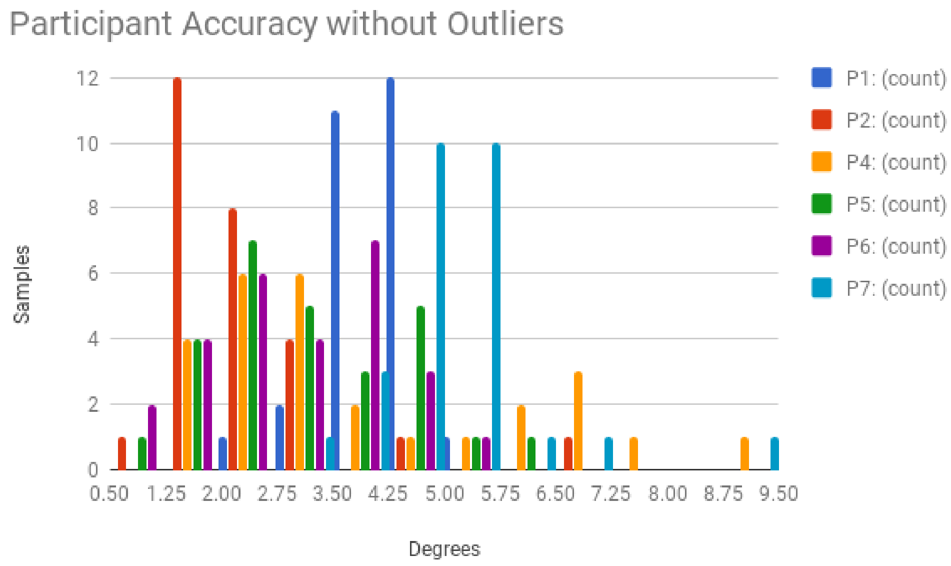


Figure 5.3: A histogram of participants' average accuracy during the first preliminary experiment, where outliers have been filtered out.

detection. A considerable amount of time was spent on attempting to get better detection, however this was not achieved, so calibration and the experiment was completed with a relatively poor detection. The logged accuracy for this participant was sometimes 20 degrees higher than the other accuracies that were recorded, therefore this participant was considered an outlier.

Because of how the sensors were placed inside the HMD, the eyelids would partially block the eyes from the sensors the more the eyes were looking up/down or to the sides. This meant that there potentially was a difference in accuracy based on whether the participants were rotating their heads or rotating their eyes when looking at an object. This also meant that there likely was a difference in the accuracy between objects that were placed differently. The middle object, for example, was more likely to have participants looking straight ahead compared to objects in the corners. All participants were told to look at the objects in a way that felt natural to them, whether this was by rotating their heads or eyes, respectively.

5.2 Preliminary Experiment 2: Distance Between Objects

In this section, the second preliminary experiment will be elaborated upon in regards to motivation, method, procedure, results, and discussion.

5.2.1 Motivation

A second preliminary experiment was planned to see if participants could gauge which objects an NPC was looking at when they were placed in the role of the fetcher. This was to further clarify the optimal setting in which the use of tracked eye movement makes a significant difference in task performance and social presence.

There was a tradeoff between the inter-object distance, the object-fixer distance, and the angle between two objects from the fixer's point of view. This relationship is illustrated in Figure 5.4.

The optimal setting would have the objects placed close enough so that they were within reach by the fetcher, but also such that the eye-gaze of the fixer could be more easily discerned. Proudlock and Gottlob (2007) found

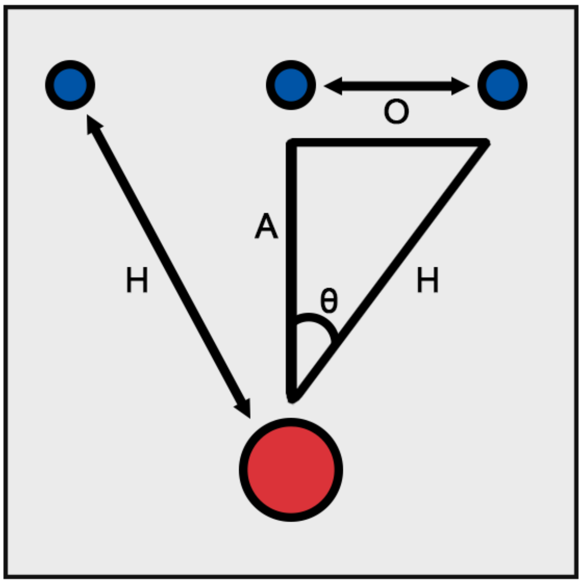


Figure 5.4: The trigonometric relationship between the angle (θ), the inter-object distance (O), the object-fixer distance (H), and the shelf-fixer distance (A). This relationship was used in this study for positioning the objects on the shelf.

.

that an angle larger than 10 to 15 degrees eye movement usually was accompanied by head movement. If this accompanied head movement happens to be too great, it might make the participants judge object positions on head movement rather than the eye movement shown on the NPC. A too small angle would result in objects being placed within the eye tracker's accuracy offset, making it more possible that participants would perceive an object adjacent to the one sought after as the correct one instead. It is important to note that the angle between objects was smaller the further from the center of the shelf the objects were placed, assuming they were placed equally apart. This is important if more objects were to be placed on the horizontal axis. As the angle gets smaller, it will become more difficult to discern eye-gaze and therefore the shelf should not be too wide.

5.2.2 Method

Task

One participant was placed inside the VE besides a shelf containing the objects of interest. An NPC looked into the room via a window. The NPC had two conditions: occluded eye-gaze and recorded tracked eye-gaze. It was randomised for each participant which condition they would start with. The NPC looked at a series of five random objects, one object at a time. The participant had to point out what object they were looking at. When an object was chosen by the participant, the next object in the series would be looked at by the NPC. After the five objects had been looked at, the condition was changed and five new objects were looked at. After both conditions were completed, the shelf shrunk with the objects' relative positions being constant, meaning that the objects' absolute distance between them became smaller. The NPC was randomly assigned a new condition and the experiment was repeated with the new shelf size. The experiment ended when the shelf with the objects had shrunk four times, which in turn made the participants select 50 objects, being either correct or incorrect.

Setup

The experiment setup required one participant to perform the task and a facilitator to introduce and guide the participant through the task. In this case the facilitator was also functioning as the technician. Therefore, the facilitator also handled the Pupil Labs software calibration and technical issues that arose. The note taker took notice of relevant observations regarding the experiment, whether it being the equipment or interesting participant behaviour, etc. Equipment used for conducting this experiment include the HTC Vive, a desktop computer with an Intel i7-4770 CPU, and an Nvidia GTX 980 Ti. Eye behaviour was recorded using the Pupil Labs eye tracker embedded in the HTC Vive HMD.

During the experiment, the shelf had nine objects placed upon it and five different scale levels while the distance between the NPC and shelf was fixed. The upper scale limit was based on not exceeding the 15 degree limit, described by Proudlock and Gottlob (2007) and was dubbed scale level 0. The lower limit was based on the eye tracker accuracy of 3.6 degrees offset such that there would be approximately twice that in the angle between objects from the fixer's point of view, which was dubbed scale level -5. For each scale level, a three second recording was made of an author looking at each of the nine objects on the shelf with the eye tracker active. These recordings were then used as the NPC animations for both the eye tracker and occluded conditions. The order of the recordings for each scale level was randomised so that it was not the same order for each condition on a scale level. The occluded condition used the same recordings but placed a 3D model of an HMD over the eye area of the NPC avatar to occlude the eyes. This solution made sure that the same recorded movements would be played for both of the conditions, making the eyes being shown the only independent variable.

Real World The real-world setup consisted of a participant (*P*), a facilitator (*F*) and a note taker (*N*). As with the first preliminary experiment, the facilitator's role consisted of introducing the participant to the experiment, procedure, and answering any questions the participant might have had. The facilitator also took care of starting the experiment and dealt with any technical issues that arose during the experiment. The note taker observed the participants and noted relevant observations, participant utterances, tendencies, etc. The real-world setup can be seen in Figure 5.5.

Virtual World The VE consisted of a wall with a window in it, dividing the environment in two parts. On one side, an NPC with recorded oculesic behaviour was looking into the other side of the environment to the objects' place in the shelf. The other side of the environment consisted of a shelf with nine objects placed in a 3x3 grid. In addition, the participant was placed in the same side as the shelf. The participant had to look at the NPC and pick up the object focused on, using eye-gaze direction estimation. The participant started the experiment facing the NPC, which meant that they had to turn around to see the shelf. This was done in order to indicate how to estimate what the NPC was looking at by using their eye-gaze direction. The virtual-world setup can be seen in Figure 5.6.

5.2.3 Procedure

First, participants were greeted and asked to sign a consent form allowing the authors to use any data collected in this report and the accompanying audio/visual production. An introduction to the task was given to participants. The script can be seen in Appendix , *Preliminary Experiment 2 Items*. After the introduction to the experiment, the participant equipped the HTC Vive and controllers and performed the previously mentioned task. Having performed the task, the participants had the equipment removed, were thanked for their participation, and compensated for their troubles in the form of refreshments.

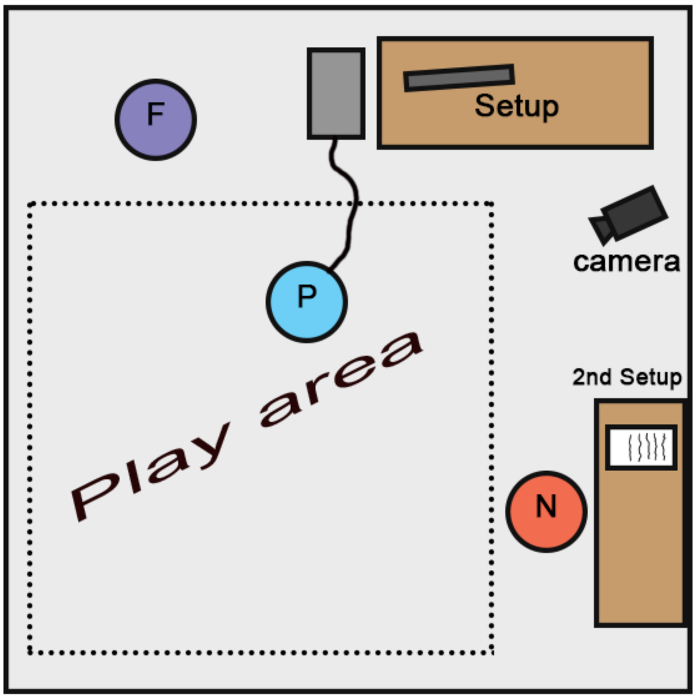


Figure 5.5: The participant (*P*) was placed inside the play area. The facilitator (*F*) stood in front of the participant. The note taker (*N*) sat at a desk next to the play area, noting relevant observations.

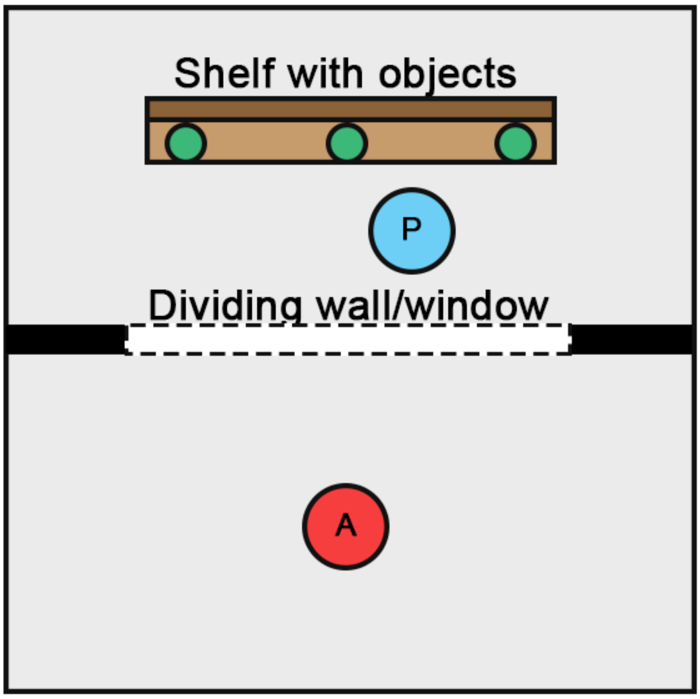


Figure 5.6: The participant (*P*) was placed inside the VE facing the avatar (*A*). The shelf and its objects were placed behind the avatar. A wall with a window in it was placed in the middle of the environment in order to divide it into two parts.

5.2.4 Results

The data gathered during the experiment was analysed and showed that six of nine participants found the tracked eye-gaze condition to be the easiest. The occluded condition performed best in seven of the nine participant cases, while one of the remaining two had it tied with the tracked eye-gaze condition. These scores were based on the amount of objects participants could correctly extract from the nonverbal cues gotten from the two conditions. The nine participants had to go through five objects in each scale level, amounting to 45 objects per scale level when summed between participants. This was repeated for each of the two conditions. The scores in relation to the scale levels showed that the occluded condition outperformed the tracked in all but one case. As seen in Tables 5.1 and 5.2, the scale level -2 yielded an average correct score of 68.89% and 66.67% for the tracked eye-gaze and occluded conditions, respectively. Thus, the scale level -2 was the one where the tracked eye-gaze condition performed the best, while the occluded scored higher in scale level 0.

Scale level	**Correct of total objects**	**Percentage (%)**
0	23	51.11
-1	17	37.78
-2	**31**	**68.89**
-3	21	46.67
-4	20	44.44

Table 5.1: Total amount of correctly grabbed objects for all participants in each scale level in the tracked eye-gaze condition. The row marked with yellow highlights the scale level with the highest accuracy.

Scale level	Correct of total objects	Percentage (%)
0	33	73.33
-1	31	68.89
-2	30	66.67
-3	27	60.00
-4	28	62.22

Table 5.2: Total amount of correctly grabbed objects for all participants in each scale level in the occluded condition. The row marked with yellow highlights the scale level with the highest accuracy.

5.2.5 Discussion

A strategy was quickly formed by all of the participants, which involved following the head movement of the NPC while it was transitioning to the next object in the series. Having only nine objects on the shelf may have assisted in the development of this strategy, since two objects at the most would be placed along the head movement direction. Because the relational head movement between two fixed objects was reported to be the revealing factor for the fetcher's choice of object, minimising this relational movement of the fixer role was presumed to be beneficial for the main experiment.

One participant mentioned that the task felt significantly harder with the eye tracker, than with the occluded condition, since it took more effort to take the combined movements of both the eyes and head into account. This could be a reason for why the eye tracker performed worse than the occluded condition, on both larger and smaller scale levels. It also follows the conclusions from Roth et al. (2016a). The eye tracker performance on the smaller scale levels, -3 and -4, likely could have been caused by the inter-object distance falling below twice the eye tracker accuracy of 3.63 degrees. Thus, allowing for the eye tracker's accuracy offset to be placed more towards an adjacent incorrect object instead of the correct one.

While most participants believed the occluded condition was more difficult, two believed it was easier because of how the HMD protruded slightly from the NPC's face, making it easier to pinpoint the forward vector of the head. There was also concern that the eyes of the avatar were difficult to discern because of the resolution of the display, making the HMD act as a sort of "bigger eye" of which it was easier to determine eye-gaze direction. Since the recorded avatar movements were mostly of someone looking straight at the objects, it was argued that this bigger eye might have been easier to use, since it was looking in the same direction as the individual eyes, but was bigger and protruding from the avatar's face.

The authors also noted that the avatar's eyes during the experiment had a reflection drawn directly into the texture which was not view-dependent. Seen from afar, this could make it difficult to determine where the center of the pupil was, since the reflection could be confused with the sclera. Another possible effect of this was the *Mona Lisa Effect* (Seele et al., 2017). This further meant that light reflections in the eyes were always remaining at the same point, thus making it difficult to estimate eye-gaze direction. A potential fix to this could involve replacing the texture and basing reflections on the viewpoint instead.

Even though eye tracking scored better than the occluded condition in scale level -2, the differences in score were so marginal that they might be contributed to an outlier instead of a concrete tendency. Therefore, this preliminary experiment was judged as not being suited for establishing a hypothesis that may be investigated further. However, as scale level -2 was the only option that might suggest a favourable inter-object distance this was used for future experiments.

5.3 Preliminary Experiment 3: Testing All Four Oculesic Behaviour Conditions

In this section, the third preliminary experiment will be elaborated upon in regards to motivation, method, procedure, participants, results, and discussion.

5.3.1 Motivation

The third preliminary experiment was created based on the experiences from the second preliminary experiment, elaborated on in Section 5.2, *Preliminary Experiment 2: Distance Between Objects*. The results from the second preliminary experiment were so marginal that nothing conclusive could be extracted from them. Because the results were inconclusive, it was not possible to base a hypothesis for the main experiment on them. Consequently, the need for a third preliminary experiment arose. One finding from the second preliminary experiment could be used, however. It was found that scale level -2 was the closest to ideal inter-object distance. Thus, the third preliminary experiment was based on scale level -2. Compared to the second preliminary experiment, this experiment would focus on direct human-to-human communication in an SVE and allow all the subtleties of the nonverbal communication that might arise to be studied.

This preliminary experiment had two major focus points. Firstly, it acted as an experiment to uncover specific tendencies within the main experiment setup. Secondly, this preliminary experiment acted as a way to pilot test the procedure that was planned for the main experiment.

5.3.2 Method

Task

This preliminary experiment investigated four conditions, each one a different approach to display eye behaviour of avatars in the SVE. Two of the conditions had previously been used in Section 5.2, *Preliminary Experiment 2: Distance Between*, namely the occluded and tracked eye-gaze conditions. The remaining two conditions were the static oculesic behaviour and the modelled oculesic behaviour using the Realistic Eye Movements asset.

Two participants performed the task at a time. The participants had two different roles: one being the fixer and the other being the fetcher. The fetcher had to estimate to which object the fixer was referring. Based on this estimation, the fetcher had to transfer the object in question to the fixer using the teleporter. The fixer participant had to communicate to the fetcher which object they needed them to transfer. The fixer participant could do this in two ways, using the monitor above their head or look at the shelf on which the spatial UI crosshair, described in Section 3.4.4, *Spatial UI*, was highlighting the current object. However, the monitor was discouraged to be used by the authors since the crosshair was the more efficient method. The communication could be both verbal and nonverbal, depending on the participants' preferences. When the current object was transferred by the fetcher, the fixer had to pick it up from the receiver and throw it into the black hole. After this was done, another object on the fetcher's shelf was marked and the fixer had to communicate to them to pick up and transfer that object. Each condition continued until five objects were correctly transferred and thrown into the black hole.

Questionnaires Used in the Experiment

For gathering information on the social presence of the participants a *game experience questionnaire*, *GEQ* was prepared. The questionnaire had three modules that each collects data on aspects of the gaming experience. More specifically these modules are the core game module, social presence module, and the post-game module. Because social presence was the only factor relevant for this study, six items from the social presence module were used. This specific module had to be answered directly after a game experience, therefore participants were asked to answer this virtually after having completed each condition. Doing this allowed social presence scores for conditions to be compared. Six items from the social presence module were chosen and used (Game Experience Lab, 2016). These items can be seen in Appendix, B.1, *Game Experience Questionnaire* (*GEQ*).

As a measure for determining potential simulator sickness the prototype might induce in participants, the 16 items for the *Simulator Sickness Questionnaire* (*SSQ*) was used. These allowed for participants to report potential symptoms of simulator sickness they may have felt during the experiment (Kennedy et al., 1993). The entire questionnaire was presented after all conditions had been completed. The items can be seen in Appendix, B.2, *Simulator Sickness Questionnaire* (*SSQ*).

Setup

This preliminary experiment utilised two participants at a time, a facilitator, and two note takers who also acted as technicians. The participants were in the same VE simultaneously in order to collaborate. The facilitator introduced the participants to the experiment, consent form, questionnaire, procedure, HMD and controllers, eye tracking equipment (only for the fixer participant), gave hints in case it was needed, and answered various questions in case the participants had any. After each condition ended, the participants were teleported to a small area in which only a mirror and a button labelled *Continue* was present. The intention behind the mirror was that the participants should be able to see their own embodied avatar and to notice the oculesic eye behaviour condition I applied to their avatar. The problem with the Mona Lisa effect appearing was fixed by creating a new texture without the drawn reflection and basing reflections on the viewpoint instead by using a standard shader.

During the experiment, the shelf had 15 objects placed on it, in a 5x3 pattern and a scale level of -2. A wall in the middle of the VE divided the participants, ensuring that objects could not be transferred that way and thus omitting the essential collaboration.

Real World The real-world setup consisted of two participants (*P*), a facilitator (*F*) and two technicians/note takers (*T/N*). As with the two first preliminary experiments, the facilitator's role was to introduce the participants to the experiment, procedure, equipment, and to answer any questions they might have. The combined note taker and technician roles consisted of two parts: 1) noting observations, utterances, etc. of relevance and 2) helping the participants to equip the HMDs and controllers, start the experiment, calibrate the eye tracker, and solve any technical issues occuring during the experiment. The technicians/note takers sat at each their desk, each with a desktop computer from which the experiment was run. From their desks, the note takers/technicians could follow the experiment on the monitors and by looking directly at the participants. The real-world setup can be seen in Figure 5.7.

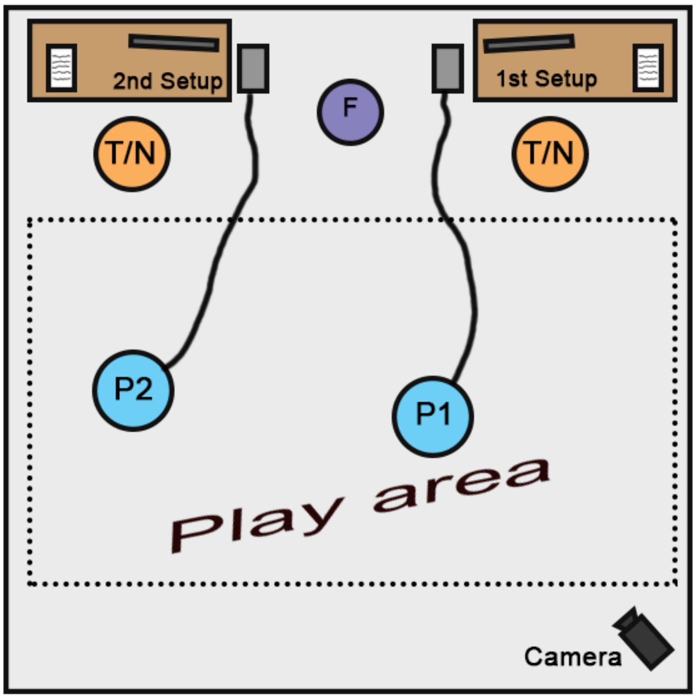


Figure 5.7: The real-world setup for the third preliminary experiment. Two participants (*P*) were placed inside the play area. The facilitator (*F*) stood in front of the participants. The technicians/note takers (*T/N*) sat at each their desk.

Virtual World The virtual world consisted of two areas: The first was the one in which the experiment task took place. The second consisted of a round platform with a mirror in which the participants could see themselves.

The first area of the SVE, was divided by a wall with a window in it, which was used hindering participants from simply dragging the objects from the shelf and into the black hole themselves.

The second area of the SVE with the mirror was used before the experiment started to show the participants their embodied avatars, their tracked movements, and their eye movements. The intention behind this was that

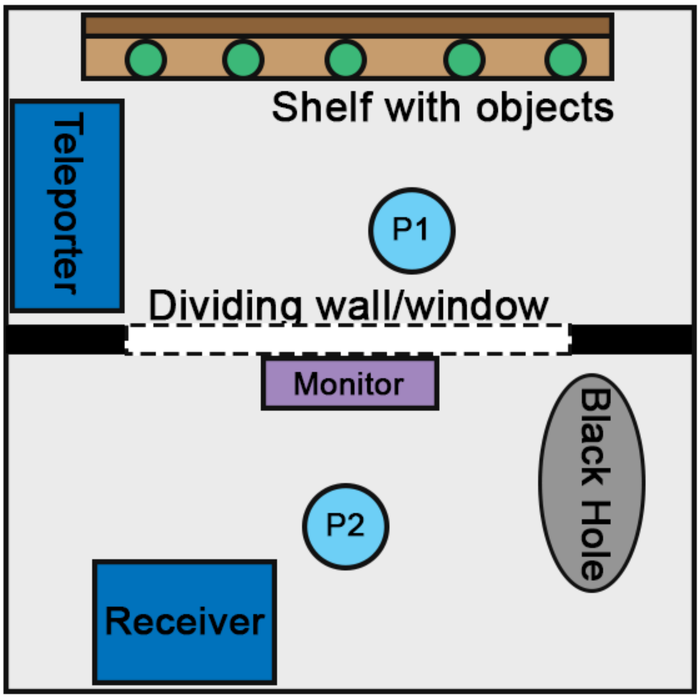


Figure 5.8: The first area of the SVE, used in the third preliminary experiment. The room was divided into two parts: one for the fixer and one for the fetcher participant.

.

the participants should become aware of their virtual embodiment and notice how the experimental condition affected their eye behaviour. The room was also used between conditions, where it had six Likert scale questions added, ranging from one to five, based on the *Social Presence* module from the GEQ. Despite having an external questionnaire, these questions were asked during the experiment as the eye tracker's precision might have suffered severely from pulling the HMD off and putting it back on. This area of the SVE can be seen in Figure 5.9.

5.3.3 Procedure

At first, the participants were greeted and thanked for participating in the preliminary experiment. Then, they were asked to sign an informed consent form allowing the authors to use the captured data for their report and the accompanying A/V production. Next, the participants were asked to sign the first part of a questionnaire, seen in Appendix **??** *Preliminary Experiment 3 Questionnaire*, asking them about their age, how often they played video games, whether they knew their collaborator, and whether their vision was impaired. After this, the participants were introduced to the task, in regards to their individual and mutual tasks, and the HMD and eye tracking equipment. The script used throughout the third preliminary experiment can be seen in Appendix **??**, *Experiment Items*. After the introduction, the participants were equipped with the HTC Vive and its controllers. Before the task was started for the first time, both participants played a tutorial relevant for their role. The aforementioned task was then performed collaboratively. The basic task of moving an object was repeated five times for each of the four conditions, totalling 20 repetitions during the experiment. After completing the experiment, the participants had the equipment removed, answered the second part of the questionnaire consisting of the SSQ, led through a semi-structured interview, were thanked for the participation, and offered a refreshment for their participation.

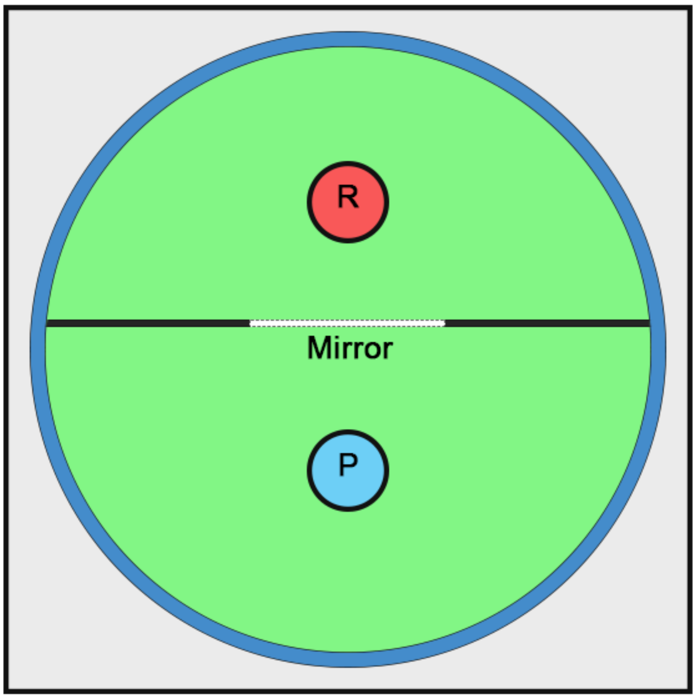


Figure 5.9: The area of the SVE containing the mirror and where the GEQ questions were used in the third preliminary experiment.

.

5.3.4 Participants

For the third preliminary experiment, three pairs, totalling six participants (three male and three female), partook. All participants were university students. The participants were between 23 and 32 years old. Three of the participants stated that they played video games at least once a month. The remaining three participants played video games less frequently. None of the participants were particularly experienced with VR applications. Finally, one participant had impaired vision.

5.3.5 Results

Throughout the third preliminary experiment, various data types were collected. The most important data collected included: GEQ answers (per condition), completion time (for conditions, roles, and in average), and the amount of erroneous grabs (errors for conditions and roles) of items. After analysing the collected data, the authors had collected various measurements, which will be elaborated upon in the following.

Social Presence

Initially, the authors looked at the participants' social presence during the experiment, using the Social Presence module from the GEQ. Studying Table 5.3, it was found that fetchers generally had higher presence for all components over all four conditions. However, the fixer, fetcher, and overall scores were too close to establish any specific tendency based on the social presence measured for this experiment.

Fetchers	**Occluded**	**StDev**	**Static**	**StDev**	**Modelled**	**StDev**	**Tracked**	**StDev**
Empathy	4.83	0.29	4.33	0.87	4.50	0.50	4.83	0.29
Negative Feeling	4.67	0.58	4.33	1.15	4.67	0.58	4.67	0.58
Involvement	4.56	0.19	4.44	0.51	4.78	0.38	4.89	0.19
Fixers	**Occluded**	**StDev**	**Static**	**StDev**	**Modelled**	**StDev**	**Tracked**	**StDev**
Empathy	4.00	0.50	4.17	0.76	4.33	0.58	4.33	0.58
Negative Feeling	2.67	0.58	2.33	0.58	2.00	1.00	3.67	0.58
Involvement	3.89	0.19	4.11	0.19	4.00	0.00	3.89	0.19
Both Roles	**Occluded**	**StDev**	**Static**	**StDev**	**Modelled**	**StDev**	**Tracked**	**StDev**
Empathy	4.42	0.65	4.25	0.88	4.42	0.49	4.58	0.49
Negative Feeling	3.67	1	3.33	1.37	3.33	1.63	4.17	0.75
Involvement	4.22	0.33	4.28	0.39	4.39	0.49	4.39	0.57

Table 5.3: The participants' average scores and standard deviations based on individual roles and both roles accumulated for the third preliminary experiment.

Completion Time and Errors

Next, the authors looked into the participants' completion times and error rates. Figures 5.10 and 5.11 yielded three observations. In general, few errors were made for each condition, thus making it hard to distinguish between condition performance, since one or two errors could potentially make a condition seem considerably worse than the others. Having a larger sample size could make a difference more clear, however with three samples a chance for this seems unlikely.

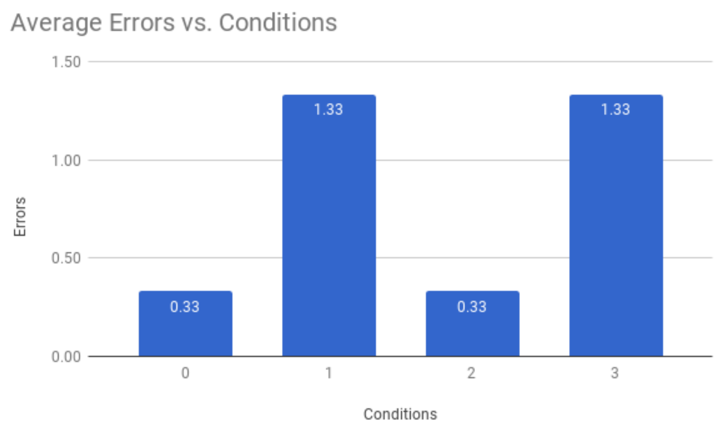


Figure 5.10: The total amount of errors per condition for the third preliminary experiment.

Having a within-subject experiment design meant that one of the conditions had to be the one participants started out with. With limited experience in performing the task during the first condition tried, the start condition's completion time was expected to be the highest and afterwards descending when the task had been learned. Having three samples does not enable every condition to be the first one the groups go through, thus this condition's average completion time logically would be lower than the others. The small sample size thus skewed the completion time in favor of the non-start condition, making it impossible to determine the more effective one in this experiment. Figure 5.11 shows that the occluded condition was the fastest for one to complete the task with, however, this condition was never assigned as the starting condition. The other differences were too small to give away any clear tendency worth investigating.

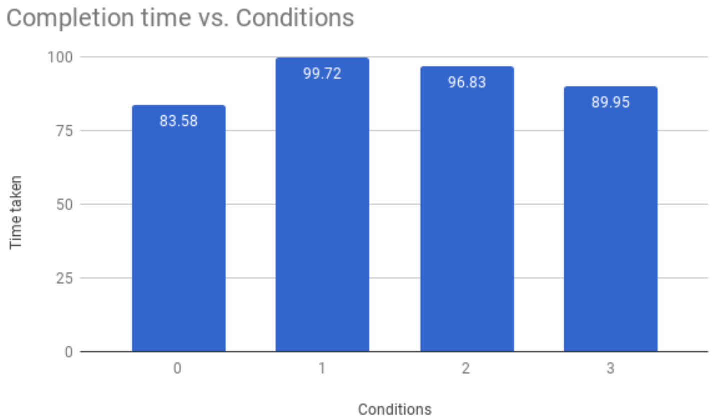


Figure 5.11: Total completion time for the four conditions in the third preliminary experiment.

5.3.6 Discussion

During this preliminary experiment, a couple of tendencies were noticed.

The first tendency regarded the communication between the participants. The fetcher participants were found to use a combination of verbal and nonverbal communication, namely talking and pointing. Usually, upon being told by the fixer what to do, the fetcher would point to an object and ask the fixer verbally whether it was the correct object. If it was indeed the correct object, they would pick it up, put it into the teleporter, and transfer it to the fixer. If it was not the correct object, they would be corrected by the fixer until finding the correct object and then transfer it using the teleporter. Sometimes, the fetcher would not even look at the fixer, but stand with their back to them. The fetcher would then point to the object they thought was the current one. Based on the object the fetcher was referring to, the fixer would confirm it was the correct object or tell them where the object was in relation to that object.

The fixer participants were found to mostly use a combination of talking and pointing. The talking was applied mostly as previously described. Pointing was used as a secondary means of communication, usually when the fetcher became in doubt about which object the fixer was referring to. It was also experienced that a large portion of the fixer participants would tell the position of the current object by stating shelf and place, e.g. *"top shelf, two from the right"*.

Overall, it was found that verbal communication was by far the dominating way of communicating between the participants. The fixer would generally convey precise instructions to the fetcher about what object to pick based on talking alone, thus making it unnecessary for the fetcher to look at the fixer.This ultimately eliminated the reason for the fetcher to turn around to look at the fixer, resulting in them not noticing the fixer's eye movements. Depriving the fixer of the ability to communicate verbally during the experiment could be a possible workaround for this.

The monitor above the fixer's head confused some of the fixer participants. When this happened, the facilitator would usually tell them to ignore it and use the spatial UI instead. It was experienced that this would usually happen if the participants did not notice the spatial UI on the objects themselves. Having the spatial UI crosshair being a shade of green and slightly transparent was determined as a possible cause for this. Making the crosshair completely opaque and changing its colour to one contrasting the blue objects might make it more noticeable to the participants.

All tendencies in the results were shared by all conditions, for both social presence and task performance, meaning that the different conditions did not seem to have any difference in their effect. The authors postulated a couple of reasons based on experiment observations why this was the case.

Firstly, as mentioned previously, some of the fetcher participants stood with their back to the fixer participants for efficiency, thus making them unable to see the fixer's eyes and eye movements. Secondly, due to the distance between the participants, the eye area of the avatars were somewhat limited. Consequently, the resolution of the eyes might have been too low for all participants to clearly notice the different oculesic behaviour conditions. Thirdly, the eye tracking equipment might not have been able to capture and reproduce the fixer's eye movements sufficiently clear for the movements to be clearly visible on their avatar.

Finally, the authors were unable to find any tendencies suggesting a difference between conditions in the performance and social presence results. Therefore, it was concluded that no tendencies relevant to this study were found in this preliminary experiment.

5.4 Preliminary Experiment 4: Testing All Four Oculesic Behaviour Conditions without Nonverbal Communication

In this section, the fourth preliminary experiment will be elaborated upon in regards to motivation, method, procedure, participants, results, and discussion.

5.4.1 Motivation

During the third preliminary experiment, elaborated upon in Section 5.3, *Preliminary 3: Testing All Four Eye-Gaze Conditions*, no difference was found between any of the four conditions: occluded, static, modelled, and tracked eye-gaze. Because of this, it was decided to conduct a fourth preliminary experiment. Based on observations from the third preliminary experiment, it was decided to alter the procedure by telling the fixer participants that they could not communicate verbally, whereas the fetcher was allowed to. The authors assumed that allowing the fetcher participants to communicate verbally would not matter, since they could only use this for getting confirmation by the fixer, and telling them how to indicate objects. The fixer participants, however, would be more limited, as they would otherwise be able to state e.g. shelf and position of the current object, as observed during the third preliminary experiment. The intention was to make the fetcher look more at the fixer.

5.4.2 Method

This preliminary experiment's method was exactly as in the third preliminary experiment, with the exception of the fixer participants not being allowed to communicate verbally. Thus, see Section 5.3, *Preliminary 3: Testing All Four Eye-Gaze Conditions* for an elaboration on this experiment's method.

5.4.3 Procedure

This preliminary experiment's procedure was identical to the one used in the third preliminary experiment, with the exception of the fixer participants not being allowed to communicate verbally. Thus, see Section 5.3, *Preliminary Experiment 3: Testing All Four Oculesic Behaviour Conditions* for an elaboration on the experiment's common procedure.

5.4.4 Participants

Similar to the third preliminary experiment, this fourth preliminary experiment had three pairs, all of which were male, totalling six participants. All participants were university students and were between 23 and 25 years old. Five of the participants stated that they played video games daily. The remaining participant played video games on a weekly basis. All of the participants were somewhat to highly experienced with VR applications. Finally, two participants had impaired vision.

5.4.5 Results

For this preliminary experiment, data similar to the third preliminary experiment was collected which is elaborated upon in Section 5.3, *Preliminary Experiment 3: Testing All Four Oculesic Behaviour Conditions*. Based on the collected data, the authors had collected various measurements which will be elaborated upon in the following.

Social Presence

The authors used the Social Presence module from the GEQ to obtain data regarding the participants' social presence during the experiment. Studying Table 5.4, it was found that the fixers had higher social presence than the fetchers for tracked eye-gaze. It was also found that fixers had lower social presence than the fetchers during the static and modelled eye-gaze conditions. Overall, the differences were minor and a clear tendency could not be found.

Fetchers	Occluded	StDev	Static	StDev	Modelled	StDev	Tracked	StDev
Empathy	4.67	0.29	4.67	0.29	4.67	0.29	4.33	0.29
Negative Feeling	2.00	0.87	1.67	0.76	2.33	1.04	2.17	1.15
Involvement	5.00	0.00	4.83	0.29	5.00	0.00	4.67	0.58

Fixers	Occluded	StDev	Static	StDev	Modelled	StDev	Tracked	StDev
Empathy	4.50	0.50	4.67	0.29	4.67	0.29	4.67	0.29
Negative Feeling	2.17	1.26	2.33	0.76	2.17	0.76	2.33	1.15
Involvement	5.00	0.00	5.00	0.00	4.83	0.29	4.83	0.29

Both Roles	Occluded	StDev	Static	StDev	Modelled	StDev	Tracked	StDev
Empathy	4.58	0.42	4.67	0.26	4.67	0.26	4.50	0.32
Negative Feeling	2.08	0.76	2.00	0.77	2.25	0.82	2.25	1.04
Involvement	5.00	0.00	4.92	0.20	4.92	0.20	4.75	0.42

Table 5.4: The participants' average scores and standard deviations based on individual roles and both roles accumulated for the fourth preliminary experiment.

Completion Time and Errors

Next, the participants' completion times and errors rates are shown in Figures 5.12 and 5.13. These figures yielded four findings. Firstly, static eye-gaze had the lowest amount of errors and lowest completion time. Secondly, tracked eye-gaze had the highest amount of errors and highest completion time. Thirdly, occluded and modelled eye-gaze had the same amount of errors and slightly differing completion times. Lastly, the start condition was always much slower than the other conditions, no matter the start condition.

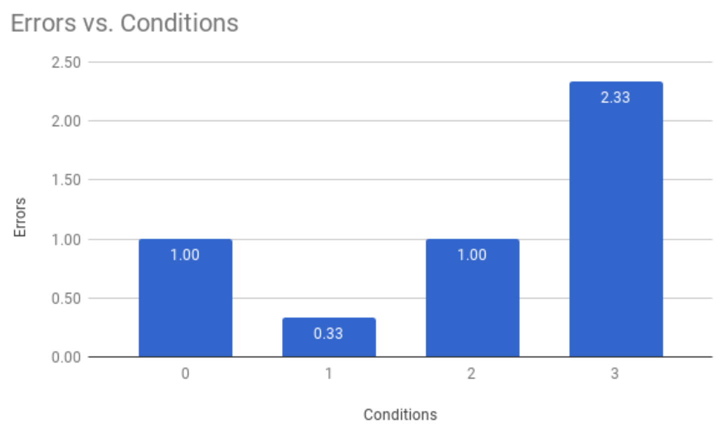


Figure 5.12: Total amount of errors per condition for the fourth preliminary experiment.

As for the third preliminary experiment, it was experienced that the participants' completion times decreased the more experience they got when the task had been learned. Due to this and the experiment's small sample size it was impossible to determine the more effective condition. Figure 5.13 shows that the static condition was the fastest to complete the task with. The other differences were also too small to reveal any tendencies worth investigating.

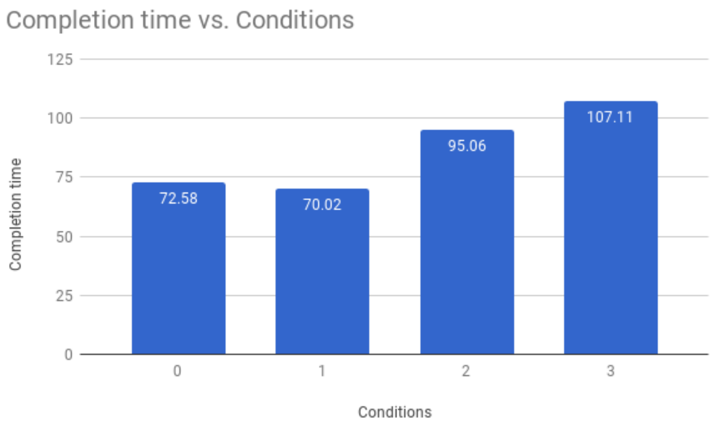


Figure 5.13: Total completion time for the four conditions in the fourth preliminary experiment.

5.4.6 Discussion

Almost all participants were found to adopt a strategy of pointing towards the current objects. Some participants also stretched their arms, had their arms down their sides, or bend their knees to indicate that the current object was at the upper, middle, or lower shelf, respectively. A lesser utilised strategy was to rotate the head to clarify the direction.

Multiple fetcher participants mentioned that they did not notice the eye movements of the fixer participant during the experiment. This could be because the participants found that pointing, gesturing, and other hand movements were more effective; it made the fetcher participants focus more on the hand and body language than on the avatar's face. Particularly, one participant mentioned that they would only have focused on the eyes if they did not have hands and were paralysed from the neck down. From the third preliminary experiment, it was found that verbal communication was the dominant means of communication. From the fourth preliminary experiment it was found that hand and body movements were the dominant means of communication. According to the participants, they would only have used the eyes if these means of communication were unavailable.

During this preliminary experiment, it was also found that the participants looked more at the avatars' faces compared to the third preliminary experiment. The reason was assumed to be that the fetcher participants could not stand with their back to the fixer and get information on which objects to pick up verbally. Instead, the fetcher participants had to look at the fixer participants to see where they were pointing and/or otherwise gesticulating the placement of the current objects. In addition, an often observed strategy among participants was that the fetcher participant would look at the fixer participant, either pointing at or grabbing an object, looking back to the fixer to see whether it was the correct object they grabbed. If it was the correct object, it was put into the teleportation field and transferred to the fixer's side of the SVE.

Another nonverbal means of communication for the fetcher participants was to look at the head position and movements of the fixer's avatar. These movements could be used by the fetcher for estimating gaze direction and by the fixer to indicate whether something was correct or incorrect by nodding or shaking their head, respectively. In addition, this method was used when the fetcher verbally asked the fixer *"is it this one?"*, after which the fixer would nod or shake their head, depending on whether it was the correct object.

Some participants experienced problems with handling the objects. Firstly, the participants were not able to pick up objects with both hands, which would result in them dropping the object on the floor. Secondly, some fixer participants experienced that the objects collided with the receiver's collider when they did not lift the objects enough, thus resulting in the object being dropped and hitting the floor. Finally, due to the hand controllers occasionally losing tracking, the participants experienced that they dropped an object suddenly and/or that their hand would just float away.

As the authors found no tendencies regarding completion time and error rates for the four conditions, it was decided to include an extra Likert question in the main experiment about whether the participants noted the eye movements of the fixer's avatar throughout the main experiment. A few participants mentioned that they had noted differences in the avatar's eye movements between conditions. Since this was not reflected in the results, it was decided to focus the additional Likert scale question on naturalness, as the Realistic Eye Movements asset used for the modelled condition was focused on creating natural eye movements. Furthermore, using eye tracking for reproducing real eye movements was expected to make the avatar eyes seem as natural as possible. This additional question was carried over to the main experiment, as it was suspected that it might yield a significant difference between some of the four conditions.

Finally, both the third and this preliminary experiment acted as a way to pilot test the procedure that was planned for the main experiment.

5.5 Summary

During the first preliminary experiment, the authors focused on establishing the accuracy of the eye tracking equipment used together with the HTC Vive HMD. This accuracy was found to be 3.63 (SD = 1.57) degrees. This accuracy was used in the design of the future preliminary experiments and the main experiment as a measure for the minimum distance between items placed on the shelf relative to the fetcher participants' point of view. In order to provide the participants with a good chance to discern between the various objects, it was decided to try and place the objects with the double distance between them, this being 7.26 degrees apart according to the fetcher participants' point of view.

The second preliminary experiment was about enabling the fetcher participants to best see and discern the objects from each other, testing between five scale levels. The authors found that the participants had the highest average scores for scale level -2 during the tracked eye gaze condition. Consequently, it was decided to include this size of the shelf, and thus the inter-object distances, for the further experiments.

The focus of the third preliminary experiment was to test all four eye-gaze conditions in an experiment scenario resembling the main experiment as much as possible. This third preliminary experiment focused on examining whether there were any tendencies between the error rates, completion times, and social presence for any of the four conditions. Unfortunately, no tendencies signalling a difference between the conditions were found which was why it was decided to create and conduct a fourth preliminary experiment.

The final preliminary experiment was based on the third preliminary experiment. However, for this preliminary experiment only the fetcher participants were allowed to communicate verbally. All kinds of nonverbal communication were allowed between both participants, however. It was found that participants quickly substituted and adopted gesticulation and body language as an alternative to verbal communication. Unfortunately, no differences between the error rates, completion times, or social presence were found for any of the four conditions. A final attempt to find some study relevant difference between the individual eye behaviour conditions then resulted in the creation of a Likert item about naturalness of eye movement that would be used in the main experiment.

In summary, the four preliminary experiments meant that the main experiment for the study would not place objects closer to each other than 7.26 degrees relative to the fetcher's viewpoint, the shelf on which the objects were placed had a size defined by scale level -2, and would collect data for error rates, completion times, social presence, and perceived naturalness of eyes.

Chapter 6

Main Experiment

Throughout this chapter, the study's main experiment is elaborated upon. The experiments' motivation, hypotheses, participants, setup, task, procedure, data analysis, and findings will be covered. Parts of the data analysis are only mentioned briefly in this chapter and are else elaborated upon in Appendix D, *Data*. The reason that some data is not shown in this chapter is that no apparent correlation was found for this data and what the authors wanted to investigate. However, the data has still been found sufficiently relevant to the project to be presented in the appendix. Finally, the chapter will answer whether the various oculesic behaviour conditions differ in regards to parameters including social presence, error rates, and completion times.

The evaluation of the SVE created throughout this study will answer whether the tendencies found during the preliminary experiments, elaborated in Chapter 5 *Preliminary Experiments*, appear in a larger scale experiment. Finally, a summary on whether the oculesic behaviour conditions are noticeable and how they affect the object-based collaborative task in the SVE is provided.

6.1 Motivation

The main experiment was performed to confirm the hypotheses based on the fourth preliminary experiment. By performing a similar experiment with a larger sample size, it would be possible to conduct proper statistical tests on the data to confirm different results. In the third and fourth preliminary experiments, no tendencies were found in neither performance nor social presence. However, qualitative data based on observations and semi-structured interviews showed that participants did notice a difference in the different oculesic behaviour conditions. Because of this, the main experiment added an extra statement between conditions, asking participants to rate, on a scale from one to five, how natural they would judge the eye movements of the fixer's avatar. The overall hypotheses of this experiment were that the tracked oculesic behaviour condition would appear more natural than the other conditions: occluded, static, and modelled eye-gaze. The null hypothesis was that the tracked condition would appear equally or less natural than each of the other oculesic behaviour conditions. In Equation 6.1, the alternative and null hypotheses for the three other oculesic behaviour conditions are presented. *T*, *M*, *S*, and *O* refer to tracked, modelled, static, and occluded eye-gaze, respectively.

$$\begin{array}{lll} H1_A : C_T > C_M & H2_A : C_T > C_S & H3_A : C_T > C_O \\ H1_0 : C_T \leq C_M & H2_0 : C_T \leq C_S & H3_0 : C_T \leq C_0 \end{array} \tag{6.1}$$

These hypotheses were used as a foundation for evaluating the study and whether eye tracking indeed made a difference in a SVE.

6.2 Task

The task was exactly the same as in the fourth preliminary experiment. Two participants shared a VE, separated by a window. The participants were to cooperate in order to transfer objects from one participant to another by being assigned asymmetrical roles. The fixer participant was equipped with the eye tracker and was not allowed to speak, while the fetcher participant started with all the objects on their side, but did not know which object to transfer.

6.3 Setup

The setup of the main experiment was exactly the same as the fourth preliminary experiment, elaborated upon in Section 5.4, *Preliminary Experiment 4: Testing All Four Oculesic Behavior Conditions without Verbal Communication*. Two computers installed with an HTC Vive system were connected to each other in a lab with an approximately 7x4m shared play space. During the experiment, two note takers each sat at a computer while the facilitator was ready in case any problems occurred. A single camera was stationed in the corner of the room for capturing relevant comments, occurrences, etc. The real-world setup is identical to the ones of the third and four preliminary experiments and can be seen in Figure 5.7.

6.4 Procedure

The procedure was identical to the one from the fourth preliminary experiment, elaborated upon in Section 5.4, *Preliminary Experiment 4: Testing All Four Oculesic Behaviour Conditions without Verbal Communication*. However, as mentioned previously, one additional statement was added between conditions to ask for how natural the eye movements of the fixer participant seemed. Additionally, the facilitator's script was slightly changed to encourage participants to spend more time looking in the mirror. Before concluding the experiment, the facilitator would also ask if the participants noticed the differences between conditions and how they would explain these differences. The change in procedure was also reflected in the VE by only showing the mirror to the participants after all the Likert items were answered. For the next ten seconds the participants did not have the option to continue, thus forcing them to spend time at the mirror increasing the chance of making them notice their avatar's eyes and oculesic behaviour.

6.5 Participants

22 participants were taken from an opportunity sample at Aalborg University. The participants were between 21-29 years old with one at 48. 16 participants were male, while six were female. 81.8% played video games on a weekly basis or more. 54.5% used either glasses or contacts, all of these were nearsighted. Participants' previous experience with VR was almost equally distributed. 4 (experienced) was the most answered, while 5 (very experienced) was the least answered.

6.6 Recorded Data

Data logging functioned the same way as in the fourth preliminary experiment, but with the extra added Likert item. The way the data logging was carried out can be seen in Section 4.12, *Data Logging*.

6.7 Results and Analysis

In this section, the main experiment results will be elaborated upon and analysed. The focus will be on communication and cooperation strategies, the GEQ's social presence module, the SSQ, the naturalness score, and various observations made during and after the experiment.

6.7.1 Error Rates and Completion Times

The amount of errors was in general observed to be low; on average, less than one error during all conditions for each participant pair. Because of this, the time penalties felt negligible for comparison and were not used for analysis.

The average completion time for each condition is shown in Figure 6.1. It was seen that the static oculesic behaviour had a noticeably higher completion time than other conditions.

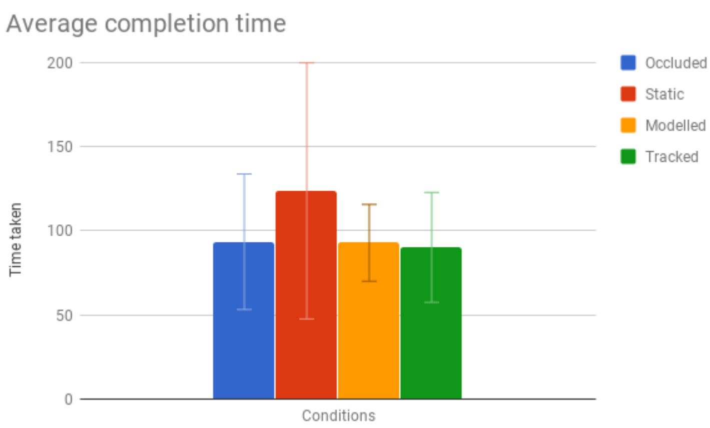


Figure 6.1: Average completion time in seconds arranged for each condition.

The authors decided to perform a Friedman test on the average completion time for all conditions to see if any significant differences between the conditions could be found. From Table 6.1, it can be seen that the results yielded a p-value of 0.66, which is above the threshold 0.05. Thus, there were no significant difference between the average completion time for any of the four oculesic behaviour conditions.

χ^2	**p-value**
1.5818 (3, N = 11)	0.66352

Table 6.1: Friedman test results for average completion time in seconds arranged for each condition.

Figure 6.2 shows that regardless of the condition, there is a significant decrease in completion time going from the first to the second attempt.

The authors decided to perform a Friedman test on the average completion time for each of the participants' four attempts to see if there was a significant improvement from the attempts. From Table 6.2, it can be seen that the results yielded a p-value of 0.00001, which is below the threshold 0.05. Thus, there were significant differences between the average completion time for at least one of the participants' four attempts.

χ^2	**p-value**
26.7818 (3, N = 11)	0.00001

Table 6.2: Friedman test results for average completion time in seconds arranged for each of the participantsø four attempts.

Because of the algorithm for randomising the order of the conditions, there was not an even distribution of which condition was the first one. Figure 6.3 shows the average completion time, excluding the first trials, for each condition. Studying the figure, a clear difference of the average completion time compared to Figure 6.1 was seen.

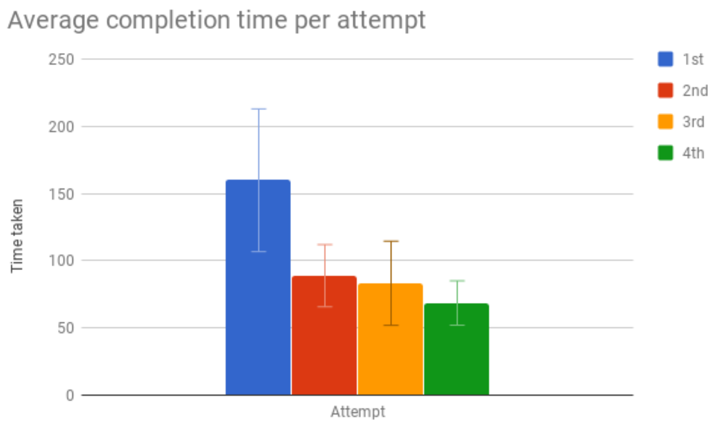


Figure 6.2: Average completion time arranged chronologically for each attempt. A second degree polynomial trendline has been added to show the trend.

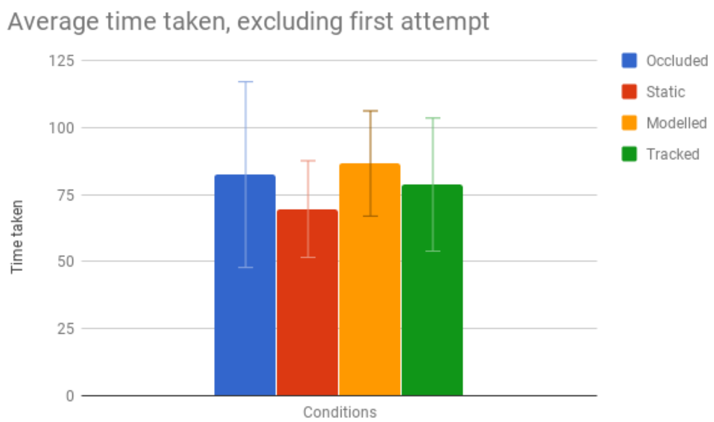


Figure 6.3: Average completion time for each experimental condition.

6.7.2 Game Experience Questionnaire

The social presence scores gathered were used to calculate the three components, *Empathy*, *Negative Feeling*, and *Behavioural Involvement* of social presence for each of the conditions. The scores for each component were found by averaging the participant scores given to the components' relevant items. From the empathy component the statements, *"I felt connected to the other"* and *"I found it enjoyable to be with the other"* were used. Negative Feeling was based on the statements, *"I felt jealous about the other"* and *"I was influenced by the other's mood"*. Behavioral Involvement was based on the last two statements, *"I paid close attention to the other"* and *"What I did affected what the other did"*. The social presence scores for each of the components can be seen in Table 6.3.

Both Roles	Occluded	StDev	Static	StDev	Modelled	StDev	Tracked	StDev
Empathy	4.32	0.85	4.61	0.43	4.55	0.58	4.59	0.43
Negative Feeling	2.07	0.61	2.09	0.55	2.27	0.61	2.27	0.81
Involvement	4.73	0.51	4.73	0.55	4.70	0.50	4.77	0.40

Table 6.3: Average scores for the empathy, negative feeling and involvement components of the GEQ.

Further investigation of the component differences over conditions yielded no significant results, as all the components, when run through a Friedman test, yielded the values shown in Tables D.1 and D.2 in Appendix D, *Data*. Even when results were separated in the two different participant roles, no further significant differences were found.

6.7.3 Simulator Sickness Questionnaire

After having completed the experiment, the participants were asked to report any simulator sickness related symptoms that they may have experienced during the experiment. This data was divided into participant roles so differences and similarities between roles could be compared, as seen in Figures 6.5 and 6.4.

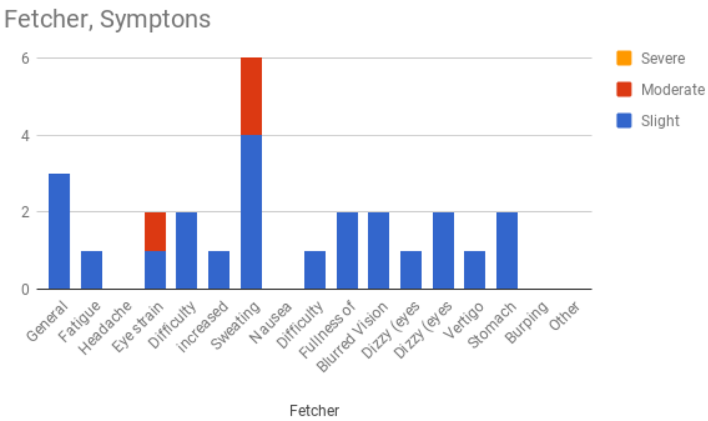


Figure 6.4: Distribution of simulator sickness symptoms reported by participants in the fixer role. Answers are stacked.

Of the 11 participant pairs, five fixer participants and only two fetcher participants reported slightly blurred vision. This difference could be contributed to chance, making the fixer participants more susceptible to this than the fetchers. Another possible reason for this difference could have been contributed by the eye tracker functioning optimally when the HMD lens depth distance is set to the largest. Because of this, this distance was left constant during the experiment and could have made the VE appear more blurry to some of the fixer participants. Having three nearsighted fixer participants combined with depth distance not being ideal could explain the increased amount of blurred vision reports by the fixer participants. It especially holds true when any fixer participants wearing glasses had to take them off to achieve ideal eye tracking. Lens depth distance being larger than ideal, together with nearsightedness, could also explain the other symptoms that fixers reported more frequently than the fetchers: difficulty focusing and eye strain.

The sweating symptom was common among the 22 participants since five of them reported that they began sweating slightly and an additional four experienced a moderate amount of sweating. Reasons for the sweating could have been that participants experienced an aspect of simulator sickness. However, this most likely is attributed to a combination of the slight physical activity that lies in the base of the task, together with the hot weather experienced on the day of testing with temperatures reaching 27 degrees celsius.

Other reported symptoms were minimal in occurrence, but they could suggest that some aspect of the experiment's VR design could be improved upon in order to decrease the amount of experienced simulator sickness.

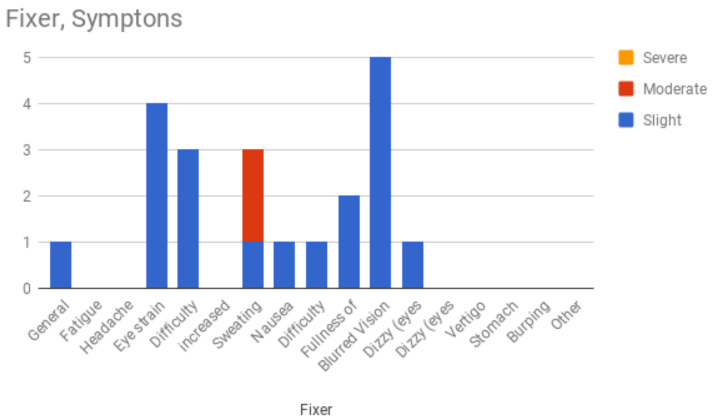


Figure 6.5: Distribution of simulator sickness symptoms reported by participants in the fetcher role. Answers are stacked.

6.7.4 Naturalness Score

The statement that was included in this study's experiment in order to probe participants for how natural they found the oculesic behaviour conditions was phrased as follows: *"The fixer's eye movements were very natural"*, where 5 meant an extreme agreement, and 1 meant no agreement at all. This statement of naturalness yielded an average score of 2 (SD = 1.31) for the occluded, 2.73 (SD = 1.32) for the static, 3.41 (SD = 1.10) for the modelled, and 3.36 (SD = 0.90) for the tracked condition, on a Likert scale ranging from one to five. The count of each participant score for every condition can be seen in Figure 6.6. Based on the figure and the condition averages, both the modelled and tracked conditions had been placed high in the score spectrum, static was placed in the middle, and occluded in the lower end. The differences in placement suggested that a significant difference could exist between tracked and occluded, or the tracked and static conditions. Therefore, data analysis on these scores was conducted.

Before data analysis could commence on the experiment statement that probed participants about the felt *"naturalness of the eye movements"* for the fixer role, the related data was checked for normality using a Shapiro-Wilk normality test in *R Studio*. This test's null hypothesis assumes that the sample came from a normally distributed population. Therefore, with a significance level of .05, p-values for this experiment's conditions were considered normally distributed if they were above that value. P-values for the different conditions, seen in Table 6.4, showed

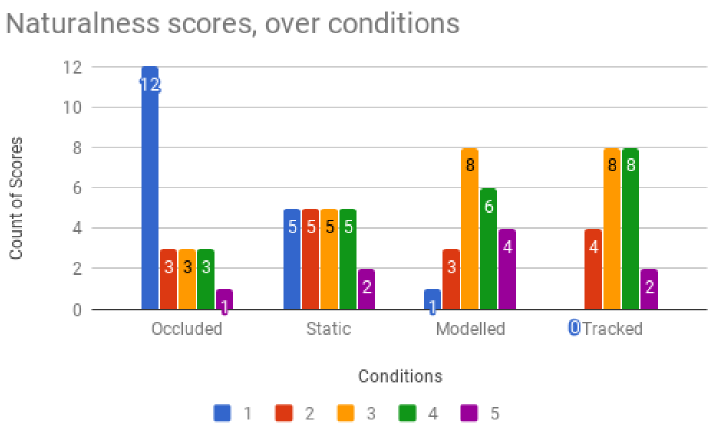


Figure 6.6: Distribution of naturalness scores over conditions.

that the modelled condition was the only one where the Shapiro-Wilk null hypothesis could not be rejected. Otherwise, the p-values were found to be under the chosen significance level, meaning that the test's null hypothesis was rejected since evidence suggested that the data was not from a normally distributed population. Thus, having three of four groups being not normally distributed meant that further testing would be done with non-parametric statistical tests.

Condition	p-value
Occluded	0.00015
Static	0.04
Modelled	0.06
Tracked	0.014

Table 6.4: Shapiro-Wilk results for conditions.

Having established that the naturalness data was not normally distributed, the next step was to check if any significance could be found between the different conditions. A Friedman test was decided upon to check for this. The test produced a p-value of 0.00000289, with a $\chi^2(3) = 23.696$, suggesting that a statistically significant difference existed between two or more of the conditions used. In order to uncover where this difference existed, a post hoc test in the form of a Wilcoxon signed-rank test was used on selected condition combinations. These specific combinations were based on the experiment's hypotheses: tracked and occluded, tracked and static, and tracked and modelled. The Wilcoxon signed-rank test was chosen since it supported the use of ordinal data together with dependent variables.

The p-values calculated by the post hoc test, seen in Table 6.5, showed that no significant difference existed between the tested relationships since no p-values lower than .05 were found. The post hoc test arriving at no significant difference suggested that this should be found between other conditions.

Condition	p-value
Tracked vs. Occluded	0.07
Tracked vs. Static	0.82
Tracked vs. Modelled	0.99

Table 6.5: Wilcoxon results for both roles.

Having found no significant difference when analysing all of the naturalness data available, it was decided to look at the same data, but divided into the different participant roles. Dividing data into fetcher and fixer data halves the amount of samples in regards to this statement to 11 samples per role. Because of this, both sample groups were assumed to not be normally distributed. The Friedman test showed that both groups had at least one significantly large difference relationship between two conditions, with a fixer p-value of 0.047 ($\chi^2(3) = 7.98$) and a fetcher p-value of 0.00072 ($\chi^2(3) = 16.95$).

The Wilcoxon signed-rank test produced the values seen in Table 6.6. These results showed, based on the six relationships tested in relation to the tracked oculesic behaviour condition, that a significant difference was only found between the tracked and occluded condition and only for the fetcher role.

Conditions	Fixer Sample p-value	Fetcher Sample p-value
Tracked vs. Occluded	0.49	**0.014**
Tracked vs. Static	0.72	0.73
Tracked vs. Modelled	0.99	0.98

Table 6.6: Wilcoxon results for the two participant roles. Significant p-values are bolded.

Table 6.6 showed that for the fetcher role tracked eye-gaze, applied to the fixer, was perceived as more natural than having the eyes occluded by an HMD. Thus, the $H3_0$ hypothesis was rejected, however this could not be done for the $H1_0$ nor $H2_0$ hypotheses.

6.7.5 Observations

Throughout the experiment, the authors made various observations. The most interesting and common will be elaborated on in the following.

Communication/Cooperation Strategies

During the experiment, various communication and cooperation strategies between the participants were noted by the authors. The most dominant strategy observed was carried out by the fetcher participants. When they started the task, they would initially look at the fixer in order to estimate where the fixer was looking and/or pointing. Next, the fetcher would either point to or touch a certain object. The fetcher would then look back at the fixer to get confirmation whether it was the correct object they pointed to. If it was incorrect, the fetcher would start the process over until finding the correct object. Conversely, if it was the correct object, the fetcher would transfer the object to the fixer using the teleporter. This strategy was utilised by all participants.

The second most dominant strategy was pointing. Seven of the 11 fixers would point towards the current object. The fixer participants were observed to stretch out their arms like they would in the real world, with the one difference of being that they were unable to use their index finger for pointing towards objects.

Most participants were observed to move the controllers in the direction of an object, in order to indicate its direction. This technique was also used by moving the controllers either left or right, depending on the position of the object currently highlighted by the fetcher. When the fetcher was at the correct object, the fixer made a sudden forward movement to indicate that the fetcher should stop and choose that object.

Nodding was used by the fixer participants for confirmatory reasons in regards to the fetchers, e.g. that they had picked the correct object and then had to transfer it. Nodding was observed to be used by six fixer participants.

Head and/or hand shaking were also widely utilised, exclusively by six fixer participants, for indicating at which shelf and position the current object was placed.

Some fetcher participants verbally encouraged fixers to do their best in the task. That was experienced by fetcher participants saying "textit"let's do it", *"quick, we're gonna beat the highscore"*, *"faster, faster, faster"*, etc. Other encouragements, such as the fetcher saying *"awesome"* every time the fixer had thrown an object into the black hole, were also observed. These encouragements were made by four fetcher participants in total.

Waving was observed to be widely utilised by both fetchers and fixers for communicating and cooperating. Waving occured when participants had answered the Likert scale items between each condition, in the beginning of the task and to get each other to start the task. Waving was also used for indicating to the fetcher that the object was at another shelf, or at another position on the shelf. A total of seven pairs waved during the experiment.

Four fixer participants were observed to exaggerate their movements to clearly indicate to the fetcher which shelf the object was on, as well as its position.

A total of 14 strategies was observed to be used more than once for communicating between both participant roles, whereas, seven of these have been elaborated upon in this section.

Use of Eyes

During the experiment, various communication and cooperation strategies between the participants were noted by the authors. The eyes were the least utilised communication method. However, since this study was about eye-gaze and eye tracking in an SVE, this subsection will look into how much the participants noticed and used the avatar eyes, respectively.

During the experiment, all participants noted the change from visible eyes to the occluded condition. For the other conditions, 15 participants explicitly stated that they saw that the avatar eyes changed. Three of these participants used the avatar's eyes as part of finding the correct object, while four said that they did not notice changes in the fixer's eyes. Figure 6.7 shows the distribution of when the avatar eyes were not noticed, noticed, and actively used, respectively. These measures are listed for the fetcher and fixer participants.

From the figure it can be concluded, that more participants noticed avatar eyes than not. It can also be seen that the fetcher participants generally noticed the fixer avatar's eyes more than the fixer participants, which might make sense as the fetchers had to pay attention to the fixer's avatar throughout the task. Finally, the figure shows that even though avatar eyes were noticed by 15 participants, they were only used by three participants, and another three participants did not explicitly state whether they noticed the eyes. One reason might be due to the resolution of the avatars' eyes, which makes it difficult to discern where the eyes are looking. It should be noted, however, that the three participants actively using eye movements only used them for estimating gaze direction, they did not use them instead of e.g. pointing. One participant mentioned that they actively tried to use the gaze direction, because they thought it was the aim of the experiment, not because it was the best method of communication. It should be noted that not all participants paid attention to the avatar's eyes in all conditions, some noticed it once, others twice, etc.

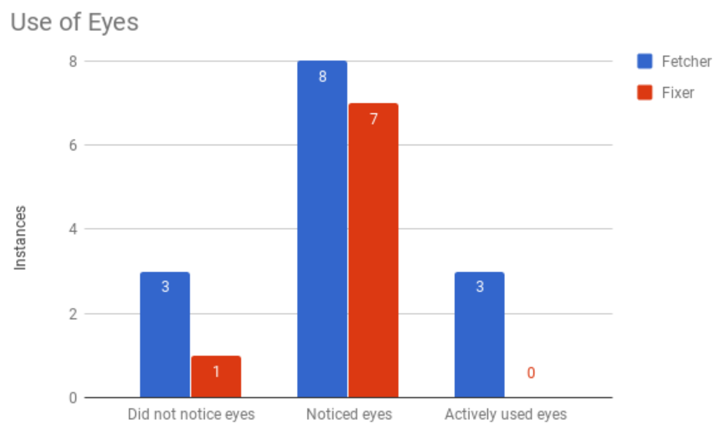


Figure 6.7: The fetcher and fixer participants' use of the fixer avatar's eyes throughout the study's main experiment. Three fixer participants did not state whether they noticed the eyes.

Experience with VR

As this study investigated the use of tracked eye-gaze in SVEs and the use of eyes had already been evaluated, the authors turned their attention to see whether a correlation between participants' experience with VR and their corresponding results existed.

Firstly, the authors looked into the average amount of errors the participants made during the experiment. VR experience was rated on a scale ranging from one to five. Table 6.7 shows the average amount of errors and completion times the participants had during the experiment.

VR Experience (Averages)			
Number of participants	**VR experience**	**Errors**	**Completion time (s)**
4	1	0.5	118.10
4	2	0.25	121.41
5	3	0.3	89.41
7	4	0.25	85.50
2	5	0.25	83.18

Table 6.7: The participants' previous VR experience, their average amount of errors and average completion times during the experiment.

.

From the table it can be seen that the average amount of errors only differed slightly. The average amount of errors according to VR experience can be seen in Figures D.1 and D.2 in Appendix D, *Data*. Table 6.7 showed that the participants' completion times decreased with VR experience. Interestingly though, participants rating their VR experience with a 2/5 had slightly higher completion times than those rating their VR experience as 1/5. From VR experience 3 to 5 their average completion time decreased slightly the more experienced the participants were, showing that previous VR experience might have benefitted them for this study's experiment.

Glasses/No Glasses

As this study utilised HMDs for enabling the participants to enter the SVE, the authors decided to investigate whether people wearing glasses and/or using contact lenses would differ from participants with normal vision. The authors wanted to investigate whether people with impaired vision had differing amounts of errors and completion times compared to participants with normal vision. Ultimately, the authors did not find anything significant when investigating data correlated to this. Figures D.3 and D.4 in Appendix D *Data* display the average amount of errors and completion times according to whether the participants wore glasses.

Video Game Playing Frequency

Next, the authors wanted to investigate whether the frequency with which the participants played video games affected their results and/or completion times. However, the authors did not find anything of significance, since there were few participants who played video games less than weekly. Thus, evaluating using more participants might yield completely different results. Table D.5 and Figures D.5 and D.6 in Appendix D, *Data* display the correlation between video game playing frequency and the average amount of errors and completion times.

Participants Knowing Each Other Beforehand

Finally, due to this study's experiment being a collaborative task between two people in an SVE, the authors wanted to investigate whether participants knowing each other beforehand had any impact on their average error rates and completion times. The authors did not find anything of significance. Table D.6 and Figures D.7 and D.8 in Appendix D, *Data* display the correlation between beforehand knowledge among the participants and the average error rates and completion times.

6.8 Summary

A total of 22 participants, 11 pairs, partook in the evaluation of this study's main experiment. During the experiment, the participants went through an object-based collaborative task in an SVE together with another participant. In the object-based collaborative task, the participants had to communicate, mainly nonverbally, about transferring specific objects from one participant's to another participant's side of the SVE.

The findings indicate that there was a significant difference between tracked and occluded eyes for the fetcher participants in terms of how natural they perceived their partner's eyes, thus rejecting $H3_0$ in favour of $H3_A$.

Participants were observed to communicate nonverbally about specific objects by looking at each other, pointing to the specific objects, moving their hands, nodding their heads, exaggerate their movements to clarify object positions, etc. Additionally, the fact that the fixer participants were not allowed to communicate verbally only seemed to complicate the collaboration slightly and only during the initial stages of the experiment.

Participants were also observed to notice the avatar's eye movements, however, only three fetcher participants used the eye movements actively to locate the specific objects.

Participants' previous experience with VR seemed to have a negligible effect on error rates. However, it did decrease their completion times the more experience they had.

How often participants played video games and whether they knew their collaborator beforehand yielded no significance in regards to error rates and completion times. Thus, these measures were not addressed during this chapter.

Chapter 7

Discussion

Based on Chapter 6, *Experiment*, the previously presented results will be discussed in this chapter, especially in regards to the quality of the solution, the quality of the technology used, participant communication strategies, appropriateness of the task, and the task and the game. The chapter will focus on the most interesting and surprising findings the authors made during the evaluation of the SVE created throughout this study. Also, technological limitations will be elaborated upon. Since few significant results were found in the experiment, this chapter will focus on what could be done to modify the scenario to increase the probability of getting significant results.

7.1 Quality of Solution

One of the main things that could affect how eye-gaze is perceived and used in the study's VE is the quality of the avatars used. For this study, a simplistic avatar created by the authors rather than experienced computer graphics artists was used. The eyes of the avatar did not have details such as eyelids or eyelashes. Furthermore, the eyes were rather small, which made it somewhat difficult to gauge their gaze direction. There was an argument as to why a cartoony style was used for the avatar, but this still allows for different designs, such as one with bigger eyes. An elaboration on the design of the avatar can be seen in Section 3.3, *Avatar Design*.

Various oculesic behaviours, such as pupil dilation and eyelid movement were not used in this study, even though Pupil Labs supports these. Not including them was due to a mix of avatar compatibility, such as lack of fidelity, and problems with the Pupil Labs functions in general. For example the blink functionality worked by recognising dips in the confidence value of the eyes. The functionality did not support individual eye blinks, and would often register false positives and false negatives, hence it was deemed too unreliable to be included.

Compared to most other studies, this study allowed the participants to move around the physical room which would be reflected by their avatar in the virtual world. Because of this, the avatar's movements would quickly become complex. Since there were only three tracking points, the rest of the body had to be simulated using inverse kinematics. Due to the complexity of using inverse kinematics, the avatar used in this study was using floating limbs, similar to many other modern VR applications, such as Rec Room (AGAINST GRAVITY, n.d.) and Job Simulator (Owlchemy Labs, n.d.). In order to not dip into the uncanny valley relationship, the avatar was rendered in a simple and unrealistic cartoony style. While this might not have affected the participants' perception of their embodiment, it might have affected how much they naturally would have looked into each other's eyes. Had the avatar been more photorealistic, both visually and kinematically, the social presence factor might have favored tracked eye movement (Roth et al., 2016b; Garau et al., 2003).

In other similar studies, such as (Latoschik et al., 2017; Seele et al., 2017), the experiment would include different avatars for both genders, some even having multiple for each. There was one pair of participants who had trouble identifying each other in the virtual environment. It was observed that the two participants greeted each other in a similar manner at the same time, shortly after the mirror room. This might be because they believed that the other participant's avatar was still a reflection, since they were identical in appearance.

Even though the modelled oculesic behaviour condition was placed high in terms of naturalness, it might have been ranked higher had it not been for its eyes rotating 180 degrees at specific times during the experiment. The occurances of this were reported by some of the fetcher participants directly after an object had been teleported to the receiver. At these times the fixer avatar's eye would rotate 180 degrees into the head of the avatar. This bug was caused by the model's 'look at' function trying to keep eyes pointed at the correct object when it was teleported behind the fixer's avatar. The angle at which the eyes still followed an object was adjustable and was set to 360 degrees. This value was set to 360 degrees since a lower angle was shown to not track the object reliably which could have had an impact on naturalness. No participants mentioned afterwards that this had an effect on their given naturalness score, however, it cannot be proven that it had no effect. Some participants mentioned that this bug made them more aware of the eyes' movements.

7.2 Quality of Technology Used

The results gathered throughout this study reflect the current technology available, and may not be applicable once the technology improves. The eye tracker used was the open source Pupil Labs add-on for the HTC Vive, elaborated upon in Section 4, *Implementation*. As an add-on, this configuration was not meant as the primary way of using the Pupil Labs sensors. Inside the HMD, the sensors were placed at the edge of the lenses pointing upwards to see the eyes, as seen in Figure 7.1.

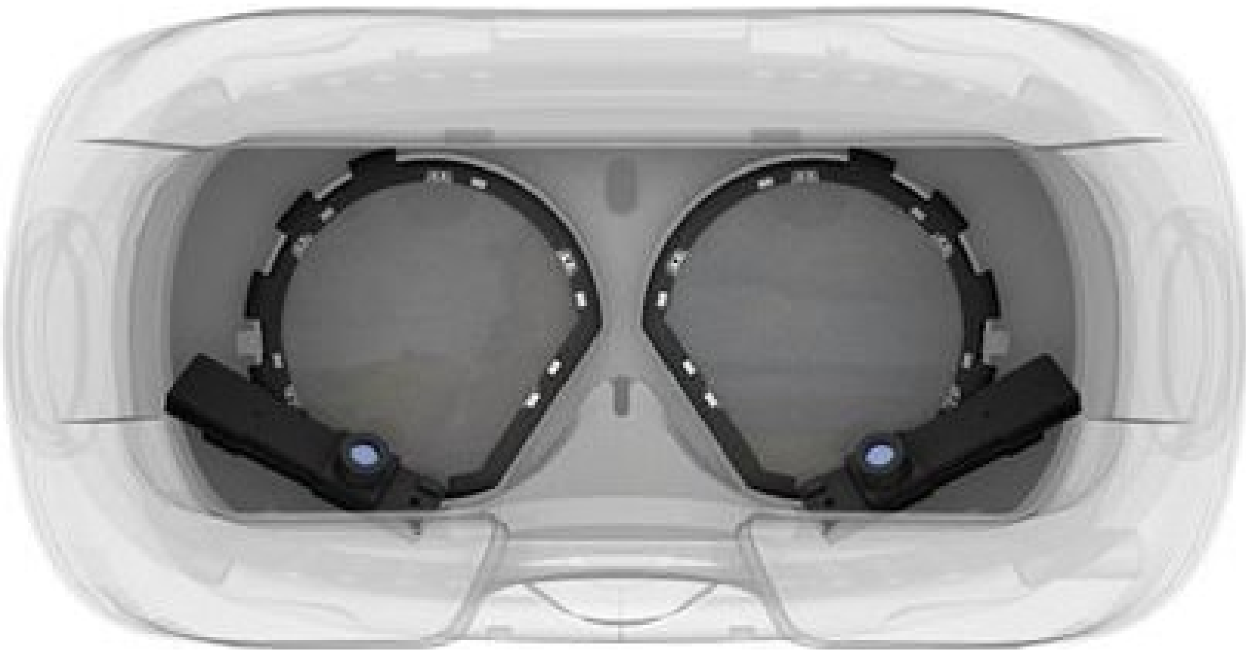


Figure 7.1: The Pupil Labs sensors placed inside the HTC Vive HMD. The sensors are placed in the bottom, looking upwards at the eyes (n.d., n.d.b).

This meant that the eye tracking software would not get a full view of the eyes, which sometimes caused a problem when the eyes were looking too far away from the centre as both the eyelids and eyelashes would block the sensors, as seen in Figure 7.2. When participants performed quick head movements, such as nodding, the HMD would often shake enough so that the eyes would not appear still relative to the sensors. Thus, the tracking often became worse during these movements. In hindsight, this might have been solved by having a head nodding animation that could be activated by fixer participants when needed. Though this goes against many fundamental principles of VR interaction design, it could be a temporary solution until the technology becomes more reliable.

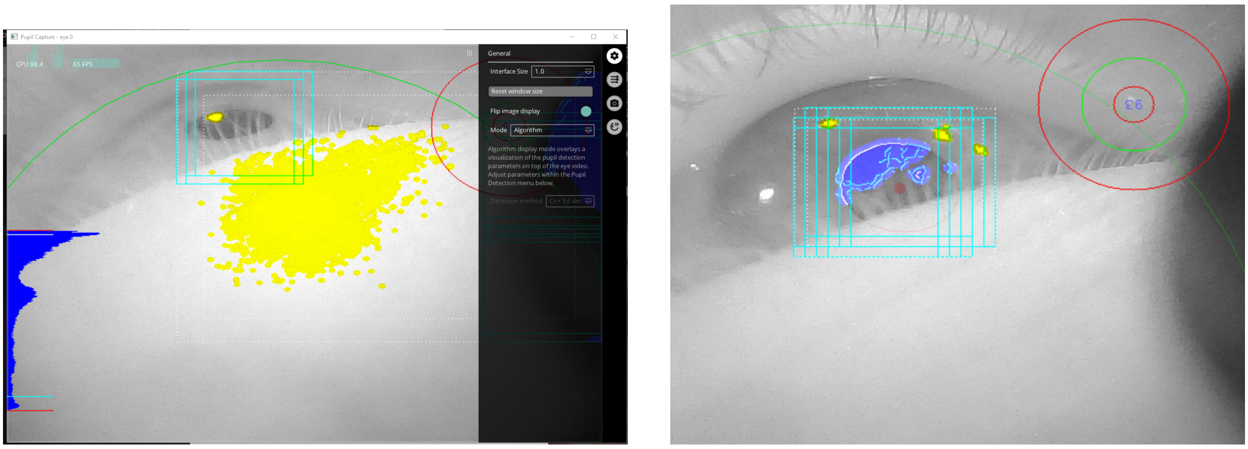


Figure 7.2: The eye as seen by the Pupil Labs sensors inside the HMD. When looking at extreme angles, the eyelids and eyelashes sometimes block the sensors.

The Pupil Labs system also did not use the *pupil centre corneal reflection* (*PCCR*) technique (Kassner et al., 2014) used by many other state of the art eye trackers, such as Tobii Pro ((n.d.). The PCCR technique illuminates near-infrared lights onto the eyes and uses the reflections on the pupil and the cornea as part of discerning the eye-gaze direction. In the near future, companies such as Oculus VR are planning on integrating eye tracking sensors natively into their HMDs which allows for better positioning of both sensors and illuminators, as seen in Figure 7.3. This should allow for much more stable and accurate eye tracking.

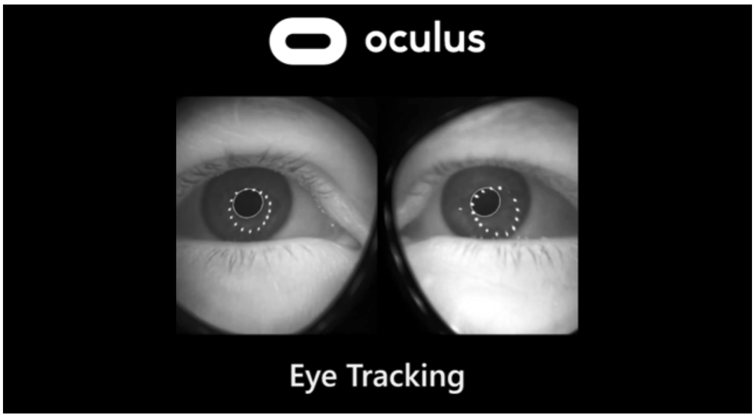


Figure 7.3: Proposed eye tracking technology by Oculus VR. The sensors are placed to see the eyes from the front, and a circular reflection pattern can be seen. The figure is taken from (Abrash, 2016: 00:15:38).

The IPT used for this study was the HTC Vive system. All current HMD-based systems have difficulties with screen quality, as screens need to be very small and also very close to the user's eyes. This made the pixel density a noticeable concern, as the image resolution gets spread over the entirety of the user's field of view. Boger (2017) argued that HMDs should be measured in *pixels per degree* (*PPD*), i.e. how many pixels were visible for each degree at the fovea. According to Boger (2017) the Vive HMD had 11 PPD, compared to the theoretical perceivable limit of 60 PPD. Looking at a digital avatar's eyes through an HMD like the HTC Vive meant that there would be noticeably fewer details compared to when looking at someone in the real world. For this study, the limitations of the resolution meant that after a distance of about two metres, the eyes of another avatar barely had enough detail for gaze direction estimation. With an increased resolution, it would potentially become easier to estimate gaze direction, which in return could yield more significant results in similar studies.

While Pupil Labs normally supports glasses and contact lenses, the add-on for the Vive HMD did not leave enough room for the frame of glasses to fit inside the HMD without blocking the sensors. This affected some users, as they were asked to either perform the task without glasses, or to switch roles with their partner, however it did not seem to affect their performance or social presence.

7.3 Participant Communication Strategies

quickly emerged, elaborated upon in Section 6.7.5, *Observations*.

When designing the task, the authors worried that the use of eye-gaze would be disregarded in favour of directly pointing at objects. In fact, this was not the most dominant strategy, since even using a hand to point was not always accurate enough for instantly recognising the correct object. Instead, most fixer participants had a strategy where they used different gestures to mark the row and then column of the object. When the fetcher pointed at an object and waited for confirmation, the fixer participant performed more gestures to indicate a direction from that point to find the correct object. One fetcher participant stated that they used the fixer participant's eye-gaze to get a general direction of the object, and then used the directional gestures from there. However, it is unknown whether other pairs used this, even subconsciously.

Most participant pairs quickly established one of the common communication strategies, though their use of gestures varied. In one case, the participant pair found it difficult to assess agreement about the object. The fetcher participant repeatedly asked for confirmation regarding the object they selected, and eventually came up with a specific gesture. The fetcher participant told the fixer to raise one arm for indicating 'yes' and the other arm for 'no'. Most fixer participants would nod to sign agreement, but one fixer participant used their hands' grab animation twice to signal a 'yes'. One participant's initial strategy was to draw the shape of the object using their hands. This, however, proved unsuccessful and they resorted to the more common gestures quickly after.

7.4 Appropriateness of Task

While there has been studies showing the positive effects on social presence during social tasks, such as Seele et al. (2017), this is the first study using modern HMD and eye tracking-based SVEs focusing on object based tasks. While the task used for this study was based on tasks used in previous studies, there is some reasoning that another task design might be more suited for finding a scenario where tracked oculesic behaviour would increase performance. The task for this study placed the participants a few metres apart. This distance likely diminished social presence, and made it more difficult to see the other's facial details, compared to if the participants were placed e.g. one metre from each other. However, placing the participants too close would not create a plausible scenario in which one person would need the assistance of the other to fetch objects. It was considered during the design phase to give the fixer a constant active physical task which would limit the use of their hands, such as holding down a button to keep the window open. This could potentially force the participants to focus more on eye-gaze, as well as make it more plausible for them to be closer together. An analogy was made with the real-life situation of a surgeon whose hands are constantly busy, and so must rely on an assistant to fetch specific tools.

Having two participants competing rather than cooperating was also considered as an alternative task. As seen in this study, participants would often make exaggerated movements to increase the understanding of their nonverbal communication. Since eye movements are subtle, putting two participants in a situation where they will try to give away as little nonverbal information as possible might create a scenario where tracked eye-gaze would be more noticeable. The developers of the VR game Rec Room detailed how they used tracked oculesic behaviour in a virtual poker game. In this scenario, the players were deliberately trying to not give away any nonverbal information and keep their head still. At the same time, their eye-gaze was visible as they switched between looking at their cards and their opponents (Tobii Gaming, 2017; Fajt, 2017). An example of this might serve as a more fitting use for eye tracking in VR applications.

Because of the resources available for this study, the task was designed around only one participant being equipped with an eye tracker. This scenario was also the case in other studies, such as Seele et al. (2017), where they assumed that a participant's social presence only depended on their partner's eye-gaze condition, not their own. However, enabling both participants to wear eye trackers allowed for the design of a symmetrical task which might be able to put a stronger focus on looking at each other's eyes. Steptoe et al. (2009) performed a study in which a participant would interact with two confederates both wearing eye trackers. The confederates took turns in guiding the participant towards solving a spatial object-based task. One of the common uses of eye-gaze in a conversation is to initiate and terminate communication, and in group settings can be helpful in directing people towards who is currently speaking. Tracked eye-gaze could have an effect when working in groups larger than two, as it would better help establish who is speaking and where the group's shared attention lies.

7.5 Task and Game

During the development of the prototype there were some discussion about the differences between game and task. Both concepts inherently required the user or player to solve some sort of obstacle using a set of tools within a specific rule set, as described in Fullerton (2008). However, a game is usually considered to be an activity one approaches with a playful mind. As the prototype was developed for a scientific study, the aim was to create a task where the participants wanted to complete their goals as efficiently as they could, rather than playing around. Since manipulation of the objects in the scene was considered the main toy of the task, the authors were slightly worried that the participants would spend time playing with the objects, rather than solving the task. There were two participants who tried to perform a "trickshot" by throwing an object towards the black hole from behind the back rather than a normal throw as a playful way of showing off in front of their partner. Besides that, all participants seemed to take the task seriously only deviating shortly from it when greeting their partner.

Despite the seriousness of the task, most participants expressed positive emotions during and after the experiment. Many participants were laughing and would explicitly state that they thought it was a fun experience. This could be because of how the task is designed around the idea of cooperative play and how it can be joyful to complete a task together with someone (Schell, 2014: p. 218). The asymmetry of the task also adds a layer of exploration where each participant had to not only learn their own role, but discover how to work with the participant in the other role (Schell, 2014: 40). Having one participant have more information, while being unable to communicate verbally, encourages the participants to work together by other means of communication. (Schell, 2014: p. 218) believed that cooperative games are enjoyed because they give people an opportunity to experience other people's behaviour, especially in difficult and stressed situations. The task in this study allowed this as participants would have to quickly establish new strategies in an unfamiliar setting as fast as possible. The task also allowed the participants to witness failures when a participant would make an independent mistake, such as dropping an object, the other participant would be able to see and sometimes laugh at this. Another factor that could make the task more enjoyable was the novelty of VR. Though approximately half of the participants were at least somewhat experienced with VR, being inside an SVE with someone they know was even more novel.

7.6 Discussion of Experiment

While the results showed that most participants *noticed* the eye movement of the fixer participants, these results did not reflect to which degree. During post-experiment interviews it was revealed that several participants were confused about what each condition changed. Some participants mentioned that they might have seen the eyes move sometimes, but also that it was not something they paid much attention to. Some participants did not notice any eye movements during the tracked condition, but conversely two participants claimed to have seen eye movement during the static condition. During the third and fourth preliminary experiments, there was less focus from the facilitator on pointing out the eye behaviour between conditions, and the only explicit mentions of changes between eye behaviours was with the occluded condition.

The initial design of the SVE was based upon a task that would have the fixer participants look at the monitor placed above them, recognise the object shown, and then deliver instructions to the fetcher about the placement of that object. After the introduction of the non-diegetic spatial UI crosshair, this monitor became functionally obsolete. However, at this point in time it was too deeply integrated in the prototype's code and tutorial to remove. Therefore, the monitor remained a part of the scene and tutorial but not one that participants should make use of. During the experiment, fixer participants were introduced to the monitor first and afterwards the UI crosshair, however several participants would still try to base the task on recognising objects on the monitor while ignoring the crosshair. The facilitator instructing the participants to use the crosshair instead confused many of them. It was not known how much of a problem keeping the monitor would be, but it seemed to have influenced completion times for the first attempt to varying degrees.

During the final experiment, it was observed that fetcher participants sometimes had trouble using the teleporter button. The button was set to move on its local x-axis when the Vive controller collided with it, and vibrotactile feedback was delivered alongside with this. When reaching a specific position on the x-axis, the button would activate the teleporter. It was, however, observed that when discovering the button moving together along with the vibrotactile feedback participants would think that the teleporter had been activated. When this happened, the facilitator had to make participants aware that the button had not been pressed all the way in, whereafter they would press the button correctly the next time. Even though this was not found that intuitive by participants they did not list it as a frustration when asked during the post experiment debriefing.

Frustrations aired by the participants during the final experiment mostly involved a spatial UI element placed in the scene displaying where the trigger button was located on the HTC Vive controller. The sign mistakenly still had an active collider that would cause the fixer to drop their grabbed object when the object accidentally collided with the sign. However, midway through the number of experiments the collider was deactivated, thus solving the issue.

Unity's networking system introduced small bugs where objects being teleported would quickly interpolate from the teleporter's to the receiver's position rather than move instantly between positions. This bug was attributed to the networking component controlling object position automatically interpolating from one position to the other. The one or two participants that noticed it said that it looked weird, but did not think anymore of it.

Because the main findings of this study were based on participant agreement with a statement about naturalness, the degree to which these findings are biased can be discussed. From the results of Section 5.2, *Preliminary Experiment 2: Distance Between Objects*, the experiment virtual setup has been biased towards tracked eye-gaze to some degree. This bias came from placing objects at the inter-object distance shown to produce a better result for the eye tracker i.e. a distance where the eye tracker scored 68.89% correct, compared to the occluded score of 66.67%. The attempt to skew the experiment results in favor of the eye tracker was made, thus enabling the best circumstances for the eye tracker, since it was still at the bleeding edge of technology. It should, however, be mentioned that the inter-object distance only yielded a result 2.22% better than the occluded condition, which had also been tested in the same preliminary experiment. Thus, it was theorised how much this specific inter-object distance added to how natural fetcher participants perceived the eyes of the fixer. This slightly decreased distance compared to earlier experiment designs would have made eye movements that were unaccompanied by head movements more possible by decreasing the angle which Proudlock and Gottlob (2007) talks about. This potential increase in eye movements alone could have had an effect on the perceived naturalness by making the differences between conditions more noticeable.

Chapter 8

Conclusion

This study set out to investigate the use of eye tracking in a SVE, based on commercially available VR technology with an eye tracking add-on. A literature review, Section 2 *Previous Works*, was carried out, which found studies that had been conducted on how tracked oculesic behaviour affected interaction in social tasks and spatial object-based tasks. However, no study found had used modern room-scale VR technology when testing object-based tasks, which then became the focus of this study. The literature review also touched upon how different types of avatars would affect the effect gained from having tracked eye-gaze.

A task was designed based on some of the literature reviewed and general game design elements. This task aimed to find a plausible scenario in which two participants would work together where tracked eye-gaze would have a significant positive impact on performance and social presence. Unlike some of the tasks seen in previous studies, this study had two participants working together using a range of nonverbal communication methods. Several internal and preliminary experiments were conducted, elaborated upon in Chapter 5 *Preliminary Experiments*, to assess the accuracy of the eye tracker, the optimal distance between objects in regards to discerning eye-gaze, and to gauge how appropriate the task was. These tests were part of the iterative process of implementing the final task.

The main experiment was conducted on 22 participants, or 11 participant pairs, to see if a statistical significance could be found for either task performance, social presence, or the perceived naturalness of the oculesic behaviour during four different conditions. These four conditions were occluded eyes, static eyes, modelled eyes, and tracked eyes. The only significant finding was that tracked eyes were found to be significantly more natural than occluded eyes, and only for fetcher participants. Several considerations for the technology used throughout the study were discussed along with the quality and design of the implemented software. It was concluded that the technology used was not of high enough fidelity to create a significant effect based on eye-gaze, in the specific object-based task that was used in the experiment. Future work should possibly try to come up with another scenario in which to test, or use better technology, such as a higher resolution screen or a more accurate eye tracker.

Chapter 9

Future Works

Throughout the evaluation of the prototype, elaborated in Section 6, *Main Experiment*, no significant results were found, except for tracked oculesic behaviour appearing more natural than occluded oculesic behaviour from the fetcher's point of view. This chapter will go into depth with some of the considerations the authors made regarding how the study could be improved to find more significant results.

Eye Tracking

The eye tracker used for this study was the Pupil Labs add-on for the HTC Vive system. The base of the Pupil Labs system was not designed to be fitted inside an HMD, and as such its tracking quality suffered from being used in this context. Seeing as other eye tracking solutions were being developed directly for use in a VR context, this was an area in which this study could likely improve by using another eye tracker with higher fidelity when used inside an HMD (Tobii Pro, (n.d.; Howard and Steptoe, 2016).

HMD

The HMD used in this study was the HTC Vive released in 2016 with two lenses with a resolution of 1080x1200px spanning a field of view of 110°. As of writing this report, HTC had released a *Pro* version of the HMD with an increased resolution of 1440x1600px per lens. Similar headsets with similar or higher resolutions were also being developed. Having an increased resolution would allow for more detail when inspecting the eyes of an avatar, which might also make tracked eye-gaze have a more profound effect (n.d., n.d.c,n).

Avatar Design

Using different designs for the avatar might change the effect of tracked eye-gaze. The design used in this study did not have a high level of detail and the eyes were scaled to be of a size closely resembling that of a real person. It was shown in other studies how different levels of realism can affect how oculesic behaviour is perceived, though these used mostly the binary categories: *realistic* or *abstract* avatars (Garau et al., 2003; Roth et al., 2016b; Latoschik et al., 2017). Therefore, it could be interesting to test the effect of a more varied set of avatar designs. These designs could, for example, change the size of the eyes, the level of oculesic detail, or the level of abstraction, among others.

Task

While this study focused on the effect of oculesic behaviour in object-based tasks, future work could focus on changing the type of task. Social tasks had already been studied, such as Seele et al. (2017), but not much had been found on object-based tasks. The task used for this study attempted to create a plausible scenario, which could still be modified in various ways in an attempt to make the participants more reliant on oculesic behaviour when performing it. A direct way would be to force participants to rely less on their hands and gestures, and instead use their gaze more. Other approaches could be to make the task more intimate, by having the participants closer to each other or have them perform a different task closer together. Furthermore, there had not been found studies on how oculesic behaviour affected competitive scenarios in which two or more participants must complete a task while competing against others. The eye tracking company *Tobii*, for example, worked with game developers *Against Gravity* to show how oculesic behaviour could be displayed in a game of poker (Tobii Gaming, 2017; Fajt, 2017).

Summary

During this study, it was shown how participants would not always pay attention or even be aware of how another person's avatar's eyes behaved. Results from the main experiment, Section 6, *Main Experiment*, further showed no significant results, besides the naturalness of eye movement between the tracked and occluded conditions. This chapter presented several suggestions for future works on how to increase the likelihood of having oculesic behaviour create a significant effect on either performance, social presence, or naturalness. These suggestions included using slightly better technology, i.e. eye trackers or screens, change the visual presentation of the avatar and its eyes, and to change the task itself to put oculesic behaviour closer to the focus of the task.

Bibliography

10Ants Hill. 10Ants Hill. *RACKET FURY TABLE TENNIS VR*. URL http://10antshill.com/racketfury/. [Online; accessed May 09th, 2018].

Abrash, 2016. Michael Abrash. *Oculus Connect 3 Opening Keynote: Michael Abrash*, 2016. [Video online], available: https://www.youtube.com/watch?v=AtyE5qOB4gw&feature=youtu.be&t=15m38s, accessed May 21st, 2018.

AGAINST GRAVITY, n.d. AGAINST GRAVITY. *REC RROM*, n.d. URL https://www.againstgrav.com/. [Online; accessed April 18th, 2018].

AUTODESK.Help, 2018. AUTODESK.Help. *polySuperShape*, 2018. URL https://knowledge.autodesk.com/support/maya/learn-explore/caas/CloudHelp/cloudhelp/2018/ENU/Maya-Modeling/files/GUID-EEF70E46-D226-4020-AF9E-44175C5E3838-htm.html. [Online; accessed May 12th, 2018].

Bente et al., 2007. Gary Bente, Felix Eschenburg and Nicole C Krämer. *Virtual gaze. A pilot study on the effects of computer simulated gaze in avatar-based conversations*, 2007.

Boger, 2017. Yuval Boger. *Understanding Pixel Density & Retinal Resolution, and Why It's Important for AR/VR Headset*, 2017. https://www.roadtovr.com/understanding-pixel-density-retinal-resolution-and-why-its-important-for-vr-and-ar-headse [accessed May 20th].

Dodds et al., 2011. Trevor J Dodds, Betty J Mohler and Heinrich H Bülthoff. *Talk to the virtual hands: Self-animated avatars improve communication in head-mounted display virtual environments*. PloS one, 6(10), e25759, 2011.

Eichert et al., 2017. Nicole Eichert, David Peeters and Peter Hagoort. *Language-driven anticipatory eye movements in virtual reality*. Behavior research methods, page 1–14, 2017.

Elgarf et al., 2017. Maha Elgarf, Slim Abdennadher and Menna Elshahawy. *I-Interact: A Virtual Reality Serious Game for Eye Contact Improvement for Children with Social Impairment*, 2017.

Fajt, 2017. Nick Fajt. *Eye Tracking in Rec Room*, 2017. https://www.againstgrav.com/blog/2017/2/28/eye-tracking-in-rec-room[accessed April 20th].

Fullerton, 2008. Tracy Fullerton. *Game design workshop: a playcentric approach to creating innovative games*. CRC press, 2008.

Game Experience Lab, 2016. Game Experience Lab. *Game Experience Lab Main Page*. [Website], 2016. Available at http://www.gamexplab.nl/index.php?page=home [Accessed: May 19th 2016].

Garau et al., 2003. Maia Garau, Mel Slater, Vinoba Vinayagamoorthy, Andrea Brogni, Anthony Steed and M Angela Sasse. *The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment*, 2003.

Grillon and Thalmann, **2008**. Helena Grillon and Daniel Thalmann. *Eye contact as trigger for modification of virtual character behavior*, 2008.

Howard and Steptoe, **2016**. Mike Howard and Will Steptoe. *Oculus Avatars: Maximizing Social Presence*, 2016. [Video online], available: https://youtu.be/X6XOwtcscnY, accessed April 17th, 2018.

htc, **n.d.** htc. *Room-Scale VR*, n.d. URL https://www.vive.com/eu/product/. [Online; accessed April 15th, 2018].

Jerald, **2016**. J. Jerald. *The VR Book: Human-centered Design for Virtual Reality*. ACM books. Association for Computing Machinery, 2016. ISBN 9781970001150. URL https://books.google.dk/books?id=IgVyjwEACAAJ.

Kassner et al., **2014**. Moritz Kassner, William Patera and Andreas Bulling. *Pupil: an open source platform for pervasive eye tracking and mobile gaze-based interaction*, 2014.

Kennedy et al., **1993**. Robert S Kennedy, Norman E Lane, Kevin S Berbaum and Michael G Lilienthal. *Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness*. The international journal of aviation psychology, 3(3), 203–220, 1993.

Knabe, **n.d.** Tore Knabe. *Realistic Eye Movements*, n.d. URL https://assetstore.unity.com/packages/tools/animation/realistic-eye-movements-29168. [Online; accessed May 09th, 2018].

Latoschik et al., **2017**. Marc Erich Latoschik, Daniel Roth, Dominik Gall, Jascha Achenbach, Thomas Waltemate and Mario Botsch. *The effect of avatar realism in immersive social virtual realities*, 2017.

Lee et al., **2002**. Sooha Park Lee, Jeremy B Badler and Norman I Badler. *Eyes alive*, 2002.

Lutz et al., **2017**. Otto Hans-Martin Lutz, Charlotte Burmeister, Luara Ferreira dos Santos, Nadine Morkisch, Christian Dohle and Jörg Krüger. *Application of head-mounted devices with eye-tracking in virtual reality therapy*. Current Directions in Biomedical Engineering, 3(1), 53–56, 2017.

McDonnell et al., **2012**. Rachel McDonnell, Martin Breidt and Heinrich H Bülthoff. *Render me real?: investigating the effect of render style on the perception of animated virtual humans*. ACM Transactions on Graphics (TOG), 31 (4), 91, 2012.

Mehrabian and Ferris, **1967**. Albert Mehrabian and Susan R Ferris. *Inference of attitudes from nonverbal communication in two channels*. Journal of consulting psychology, 31(3), 248, 1967.

Morph 3D, **2017**. Morph 3D. *MCS Male*, 2017. URL https://assetstore.unity.com/packages/3d/characters/humanoids/mcs-male-45805. [Online; accessed April 20th, 2018].

Murray et al., **2009**. Norman Murray, Dave Roberts, Anthony Steed, Paul Sharkey, Paul Dickerson, John Rae and Robin Wolff. *Eye gaze in virtual environments: evaluating the need and initial work on implementation*. Concurrency and Computation: Practice and Experience, 21(11), 1437–1449, 2009.

n.d., **2018**. n.d. *Cg Programming/Unity/Toon Shading*, 2018. [Online], available: https://en.wikibooks.org/wiki/Cg_Programming/Unity/Toon_Shading, [accessed May 12th, 2018].

n.d., **n.d.a**. n.d. *TECH SPECS*, n.d. https://www.pimaxvr.com/en/8k/#page6 [accessed May 24th.

n.d., **n.d.b**. n.d. *HTC Vive Add-On*, n.d. https://docs.pupil-labs.com/#htc-vive-add-on [accessed May 10th.

n.d., **n.d.c**. n.d. *VIVEPRO*, n.d. https://www.vive.com/eu/product/vive-pro/ [accessed May 24th.

Oculus, **n.d.** Oculus. *Toybox*, n.d. URL https://www.oculus.com/experiences/rift/1083042371786607/. [Online; accessed April 18th, 2018].

Owlchemy Labs, **n.d.** Owlchemy Labs. *PREPARE. TO. JOB.*, n.d. URL https://jobsimulatorgame.com/. [Online; accessed April 20th, 2018].

Peters et al., **2010**. Christopher Peters, Stylianos Asteriadis and Kostas Karpouzis. *Investigating shared attention with a virtual agent using a gaze-based interface.* Journal on Multimodal User Interfaces, 3(1-2), 119–130, 2010.

Proudlock and Gottlob, **2007**. Frank Antony Proudlock and Irene Gottlob. *Physiology and pathology of eye–head coordination.* Progress in retinal and eye research, 26(5), 486–515, 2007.

Pupil Labs, **n.d.** Pupil Labs. *pupil docs Introduction*, n.d. URL https://docs.pupil-labs.com/#introduction. [Online; accessed May 09th, 2018].

Roberts et al., **2009**. David Roberts, Robin Wolff, John Rae, Anthony Steed, Rob Aspin, Moira McIntyre, Adriana Pena, Oyewole Oyekoya and Will Steptoe. *Communicating eye-gaze across a distance: Comparing an eye-gaze enabled immersive collaborative virtual environment, aligned video conferencing, and being together*, 2009.

Roth et al., **2016a**. Daniel Roth, Jean-Luc Lugrin, Dmitri Galakhov, Arvid Hofmann, Gary Bente, Marc Erich Latoschik and Arnulph Fuhrmann. *Avatar realism and social interaction quality in virtual reality*, 2016.

Roth et al., **2016b**. Daniel Roth, Kristoffer Waldow, Felix Stetter, Gary Bente, Marc Erich Latoschik and Arnulph Fuhrmann. *SIAMC: a socially immersive avatar mediated communication platform*, 2016.

Schell, **2014**. Jesse Schell. *The Art of Game Design: A book of lenses.* AK Peters/CRC Press, 2014.

Seele et al., **2017**. Sven Seele, Sebastian Misztal, Helmut Buhler, Rainer Herpers and Jonas Schild. *Here's Looking At You Anyway!: How Important is Realistic Gaze Behavior in Co-located Social Virtual Reality Games?*, 2017.

Smith, **2018**. Chris Smith. *TECHVirtual reality gaming defies criticism as its popularity surges*, 2018. KnowTechie, [Online] accessed May 23rd, 2018, available: https://knowtechie.com/virtual-reality-gaming-defies-criticism-as-its-popularity-surges/.

Sony, **n.d.a**. Sony. *Playstation VR*, n.d. URL https://www.playstation.com/en-nz/explore/playstation-vr/. [Online; accessed April 18th, 2018].

Sony, **n.d.b**. Sony. *The Playroom*, n.d. URL https://www.playstation.com/en-us/games/the-playroom-ps4/. [Online; accessed April 18th, 2018].

Statista, **2017**. Statista. *Virtual Reality (VR) - Statistics & Facts*, 2017. Statista, [Online] accessed May 23rd, 2018, available: https://www.statista.com/topics/2532/virtual-reality-vr/.

Steptoe et al., **2008**. William Steptoe, Robin Wolff, Alessio Murgia, Estefania Guimaraes, John Rae, Paul Sharkey, David Roberts and Anthony Steed. *Eye-tracking for avatar eye-gaze and interactional analysis in immersive collaborative virtual environments*, 2008.

Steptoe et al., **2009**. William Steptoe, Oyewole Oyekoya, Alessio Murgia, Robin Wolff, John Rae, Estefania Guimaraes, David Roberts and Anthony Steed. *Eye tracking for avatar eye gaze control during object-focused multiparty interaction in immersive collaborative virtual environments*, 2009.

Steptoe et al., 2010. William Steptoe, Anthony Steed, Aitor Rovira and John Rae. *Lie tracking: social presence, truth and deception in avatar-mediated telecommunication*, 2010.

Tobii Gaming, 2017. Tobii Gaming. *Valve, Against Gravity and Tobii show the future of social VR.*, 2017. Tobii Eye Tracking Blog, [Online] accessed May 21st, 2018, available: https://blog.tobii.com/tobii-valve-and-against-gravity-show-the-future-of-social-vr-in-rec-room-84789e6b9180.

Tobii Pro, (n.d.). Tobii Pro. *How do Tobii Eye Trackers work?*, (n.d.). Statista, [Online] accessed May 20th, 2018, available: https://www.tobiipro.com/learn-and-support/learn/eye-tracking-essentials/how-do-tobii-eye-trackers-work/.

Unity, n.d.a. Unity. *Remote ACtions*, n.d. URL https://docs.unity3d.com/Manual/UNetActions.html. [Online; accessed May 12th, 2018].

Unity, n.d.b. Unity. *User Interfaces for VR*, n.d. URL https://unity3d.com/learn/tutorials/topics/virtual-reality/user-interfaces-vr. [Online; accessed May 10th, 2018].

Velloso and Carter, 2016. Eduardo Velloso and Marcus Carter. *The emergence of eyeplay: a survey of eye interaction in games*, 2016.

Vermeulen et al., 2013. Jo Vermeulen, Kris Luyten, Elise Van Den Hoven and Karin Coninx. *Crossing the Bridge over Norman's Gulf of Execution: Revealing Feedforward's True Identity.* CHI '13 CHI Conference on Human Factors in Computing Systems, page 1931–1940, 2013. URL https://uhdspace.uhasselt.be/dspace/bitstream/1942/14759/1/VermeulenLuytenVandenHovenConinx{_}chi2013.pdf.

Vinayagamoorthy et al., 2004. Vinoba Vinayagamoorthy, Maia Garau, Anthony Steed and Mel Slater. *An eye gaze model for dyadic interaction in an immersive virtual environment: Practice and experience*, 2004.

VR, n.d. Oculus VR. *oculus rift*, n.d. URL https://www.oculus.com/rift/#oui-csl-rift-games=robo-recall. [Online; accessed April 18th, 2018].

Appendices

Appendix Chapter A

Discarded Items

Throughout this study, the authors 3D modelled and downloaded various objects for use in the SVE. The initial intention of these items was to populate the SVE with objects which should either convey a certain feeling or help the participants complete the task. A subset of items was modelled in order to create a realistically looking male and female avatar.

A.1 Avatar

Initially, the authors implemented an asset for Unity called textitMorph3D (Morph 3D, 2017), which is a fully modelled 3D male avatar. However, this model was discarded, as it proved difficult to make the avatar's legs, upper body, and feet move naturally, as none of them were tracked. The feet and legs could have remained stationary, but it was opted not to do so, as it would look too unnatural to the participants. Thus, legs and feet were not used in the study. Instead, the authors considered implementing a realistically looking floating limb avatar. This meant that only the avatar's head, torso, and hands with no connecting limbs were visible, thus focusing on only displaying the tracked body parts. A realistically looking pair of male and female hands and heads was modelled. The hands were created with underlying skeletons, using joints and then had their meshes skinned to the respective joints, while the head used blendshapes to morph between different expressions. This way, it was possible to create realistically looking and moving limbs, on which the meshes would act somewhat realistic. However, it was decided to discard the realistic avatars for two reasons: Firstly, the participants might assume a certain gender according to the look of the avatar's head or hands, which might not correlate to the head and/or hands implemented by the authors. Secondly, the more realistically looking the avatars would be, the higher the risk would be of inducing the uncanny valley relationship, which is elaborated upon in Section 2.5.5, *Rendering Styles and Uncanny Valley*. The third reason for not implementing realistically looking floating limb avatars was that it was neither found during the authors' literature review, seen in Chapter 2, *Previous Works*, nor in any of the current commercially available VR applications in which realistically looking avatars had been implemented.

The modelled limbs were: heads and hands, as can be seen in Figure A.1. Also, the figure shows the *Morph3D*, character, the authors' 3D modelled character, and the male and female versions of the abstract avatar used for the study.

A.2 CRT Monitor

The purpose of the monitor was changed during the project. Instead of it being the primary way for the fixer to see the current object to transfer, it became a secondary way of doing so. The reason for the change being that based on the discussion from the second preliminary experiment it was experienced that the monitor might have hindered the collaboration between the participants, elaborated upon in Section 5.2, *Preliminary Experiment 2: Distance Between Objects*. The reason was that head movement seemed very dominant compared to eye movement. By having the fixer constantly make big head movements back and forth from the monitor and the shelf, there was concern that the eye movements would be less noticeable. The new primary system for highlighting objects for the fixer can be seen in Section 3.4.4, *Spatial UI*. The monitor was still above the fixer's head, so they could use this for getting the

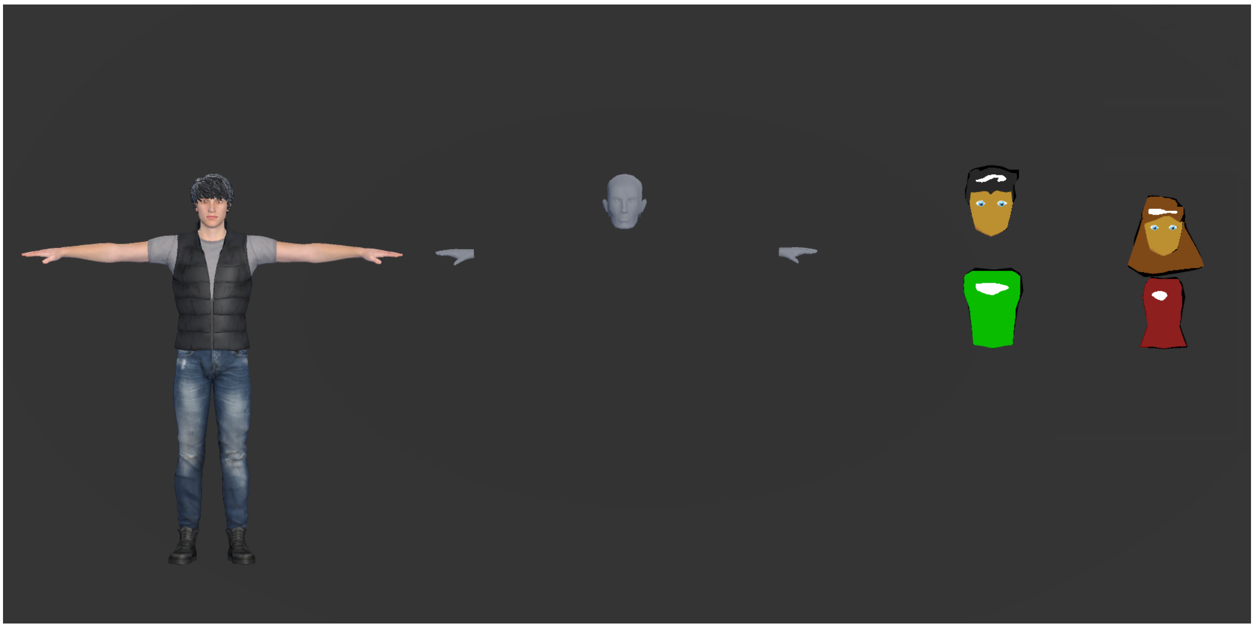


Figure A.1: Unused ideas for player avatars. From left to right: the Morph3D male photorealistic avatar, realistic disembodied avatar, and male and female variations of the avatar used in the study.

same information as provided by the spatial UI, should they so desire. The monitor can be seen in Figure 3.4.3.

A.3 Object Placement Ideas

Another idea for how the objects should be placed in the SVE was to have them floating around in the air in a ring. The ring could rotate so the objects would never be stationary and the participants would have to rely on various ways of communicating, such as using eye-gaze direction estimation, pointing and speaking, among others, however this was not used in any experiment.

In the beginning of the study, the objects that the fetcher had to pick up were placed horizontally on a table. The table was 3D modelled by the authors and was intended to look futuristic, yet abstract. The table can be seen in Figure A.2.

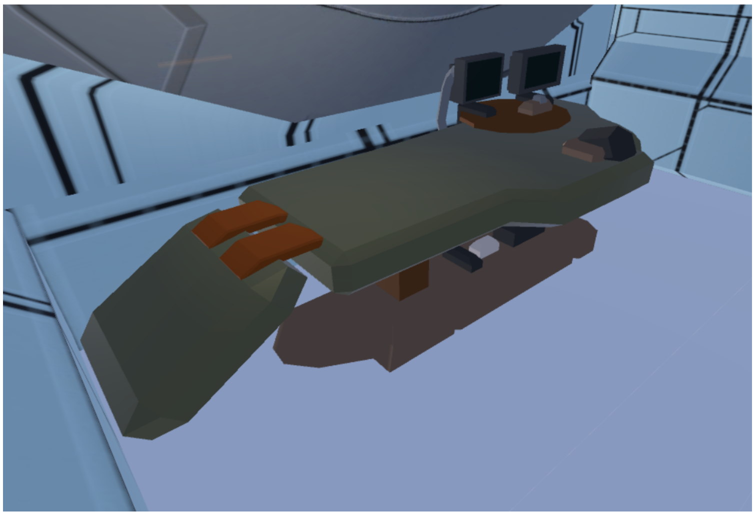


Figure A.2: The table on which it was originally thought to place the objects.

Appendix Chapter B

Experiment Items

Throughout this section, various items in regards to the preliminary experiments and the main experiment are shown. Especially questionnaires and facilitator script(s) are shown.

B.1 Game Experience Questionnaire (GEQ)

The game experience questionnaire used in the third and fourth preliminary experiments, as well as the main experiment is shown in the following

Empathy

1. *"I felt connected to the other"*
2. *"I found it enjoyable to be with the other"*

Negative Feeling

3 *"I felt jealous about the other"*

4 *"I was influenced by the other's mood"*

Behavioral Involvement

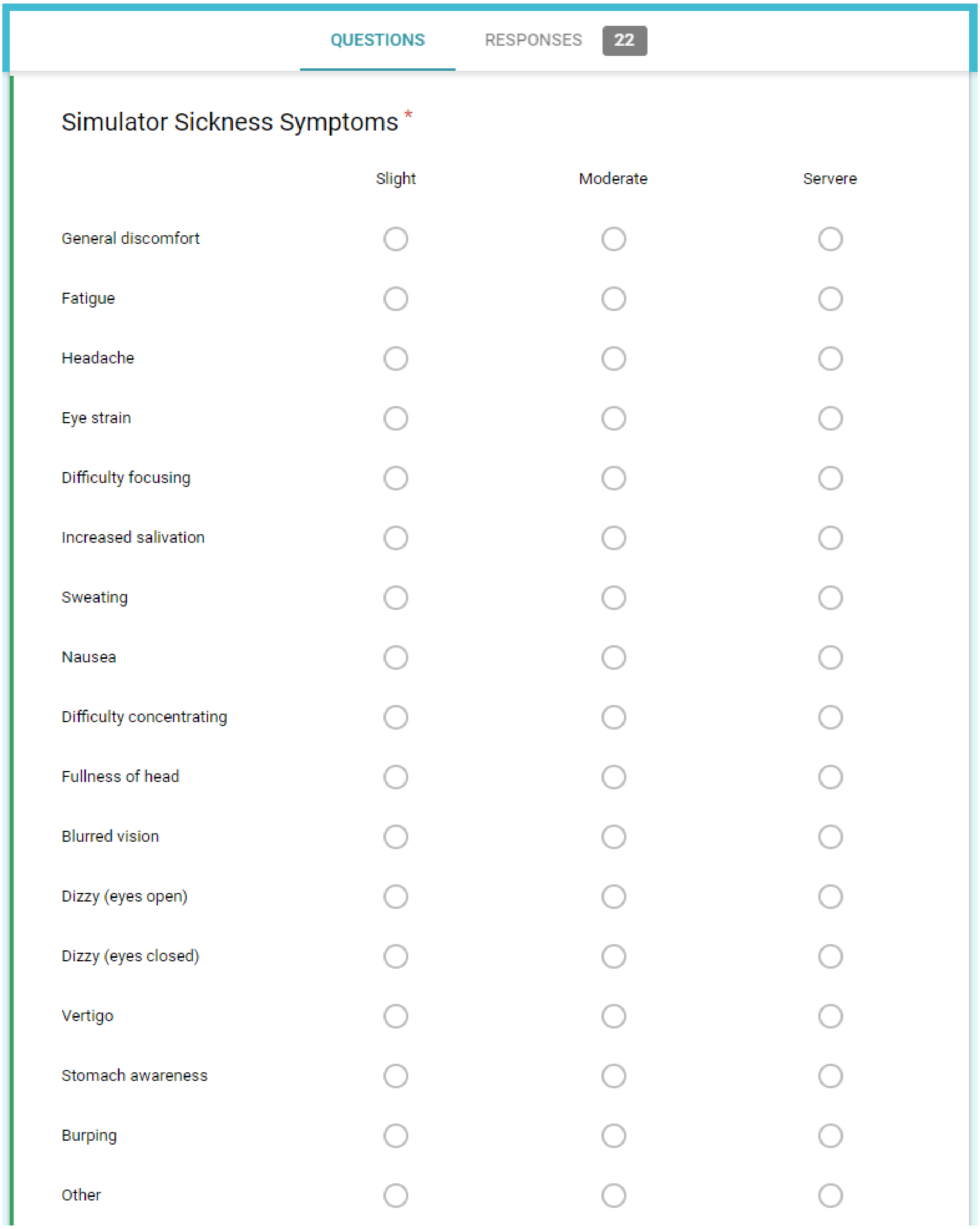
5 *"I paid close attention to the other"*

6 *"What I did affected what the other did"*

B.2 Simulator Sickness Questionnaire (SSQ)

The simulator sickness questionnaire used in third and fourth preliminary experiments, as well as the main experiment can be seen in Figure B.1.

B.3 Preliminary Experiment 2 Items



QUESTIONS RESPONSES 22

Simulator Sickness Symptoms *

	Slight	Moderate	Servere
General discomfort	○	○	○
Fatigue	○	○	○
Headache	○	○	○
Eye strain	○	○	○
Difficulty focusing	○	○	○
Increased salivation	○	○	○
Sweating	○	○	○
Nausea	○	○	○
Difficulty concentrating	○	○	○
Fullness of head	○	○	○
Blurred vision	○	○	○
Dizzy (eyes open)	○	○	○
Dizzy (eyes closed)	○	○	○
Vertigo	○	○	○
Stomach awareness	○	○	○
Burping	○	○	○
Other	○	○	○

Figure B.1: The game experience questionnaire in the third and fourth preliminary experiments, as well as the main experiment.

Appendix Chapter C

Literature Review Search Protocol

Search Protocol

1. Define your research subject and describe the specific focus of the performed search:

The initial idea for this project is to work with the effect of realistic eye movement in a modern HMD-based virtual reality application using an abstract floating limb avatar. An eye tracker will be used to transfer eye movements onto digital avatars for a series of collaborative object-based tasks. In the tasks, two people will have to communicate to find specific objects in a shared virtual 3D environment. One scenario will have a prerecorded character cooperate with a participant and another scenario will have two immersed participants work together. It is hypothesised that the extra modality, oculesic behaviour, in non-verbal communication will improve aspects such as task performance and visual attention. The controlled variable will be how the eyes of the digital avatars move. Tests will be conducted with occluded, static, modelled (based on a realistic eye-gaze model) and tracked eye-gaze.

2. List the aspects that your subject contains and the search terms for each of the aspects:

Virtual Environments	Eye Behaviour	Collaboration	Avatar Behaviours
Virtual real* Immersiv* Virtual environment*	Eye track* Eye gaz* Eye mov* Ocul*	Collaborat* Communicat* Socail*	Avatar mediat* Behav*

3. Selection of relevant sources:

Source (databases, search engines, sources hand searched, persons/organizations contacted...)	Provider (which provider you accessed the source through)	Reason for selection of source (subject coverage, accessibility, key source…)
IEEE		The database focuses on topics in the field of computing,
ACM		The database focuses exclusively on topics in the field of computing.
Proquest		Thought it covered the field of computing.
Compendex		The database is on the topic of engineering.
Proquest		The data is focused on technology.
AUB		Provides sources in various fields.

Appendix Chapter D

Data

Friedman test output from running social presence scores, all data:

Measure	chi-squared	df	p-value
Empathy	3.8532	3	0.28
Negative Feelings	6.0227	3	0.11
Behavioural Involvement	3.4	3.4	0.33

Table D.1: Friedman measures for social presence for both participant roles for this study's main experiment.

Output from running fetcher social presence scores through a Friedman test:

Friedman test scores for the fixer participants' social presence scores:

D.1 Experience with VR

As this study used HMDs and an SVE, the authors decided to investigate whether participants' previous experience with virtual reality had any correlation to their results throughout the experiment.

Figure D.1 shows that previous VR experience only had a marginal influence on average amount of errors made by the participants in this experiment. The figure also shows that, on average, participants with the lowest previous VR experience had the highest amount of errors, which might be expected. VR experience 2, 4, and 5 all the lowest average error amount of 0.25.

Figure D.2 shows that the more previous VR experience that participants had, faster they completed the experiment. Interestingly, VR experience 2 yielded higher completion time than VR experience 1.

D.2 Glasses/No Glasses

As this study utilised HMDs for enabling the participants to enter the SVE, the authors decided to investigate whether people wearing glasses and/or using contact lenses would differ from participants with normal vision. The first focus point was to investigate whether participants with impaired vision had their eyes tracked differently compared to participants with normal vision. Secondly, the authors wanted to investigate whether people with impaired vision had differing amounts of errors and completion times compared to participants with normal vision.

Measure	chi-squared	df	p-value
Empathy	3.5806	3	0.3105
Negative Feelings	4.83	3	0.185
Behavioural Involvement	2	3.4	0.5724

Table D.2: Friedman measures for social presence for the fetcher participants for this study's main experiment.

Measure	chi-squared	df	p-value
Empathy	55.385	4	0.2364
Negative feelings	1.5	3	0.68227
Behavioural Involvement	12.222	4	0.8744

Table D.3: Friedman measures for social presence for the fixer participants for this study's main experiment

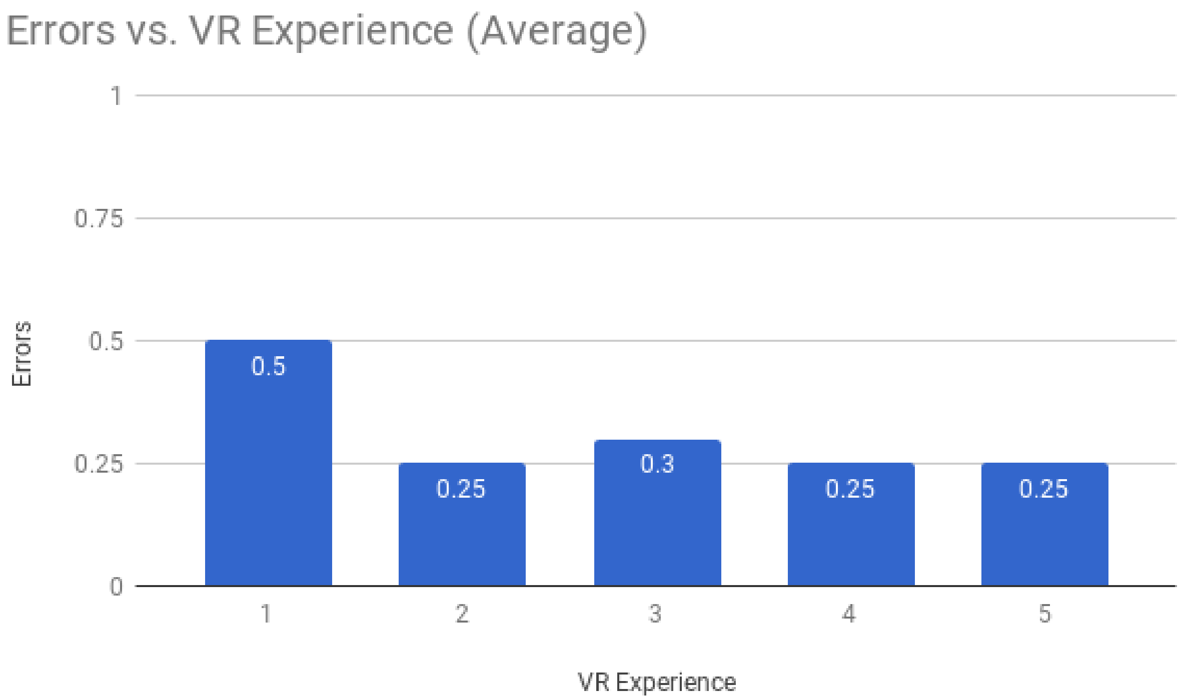


Figure D.1: Average error rates arranged according to VR experience.

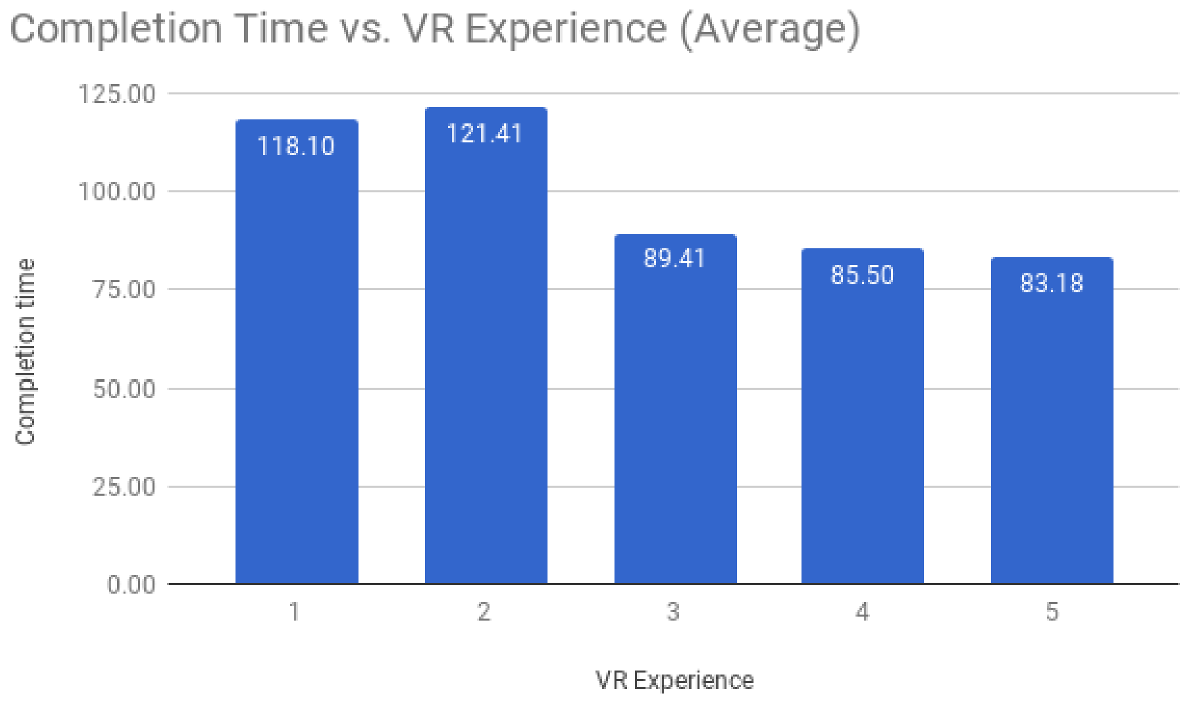


Figure D.2: Average completion time in seconds arranged according to VR experience.

Table D.4 and Figures D.3 and D.4 show the distribution between the average amount of errors and completion times for participants with impaired vision and normal vision, respectively. The amount of errors was somewhat equally distributed, however, participants with normal vision had more errors with the modelled oculesic behaviour. Considering the average completion times, the two participant types had somewhat equal results, however, both with the slowest completion times for occluded oculesic behaviour. In summary, participants wearing glasses and/or using contact lenses did not seem to perform differently from the study's other participants.

D.3 Video Game Playing Frequency

Next, the authors wanted to investigate whether the frequency with which the participants play video games affected their results and/or completion times. Table D.5 and Figures D.5 and D.6 show the distribution between video game playing frequency, average amount of errors and completion times. No participants chose the option "Other", which might indicate that they play video games occasionally, although this might be with more than a month's frequency.

From the results it is seen that the average amount of errors only differ slightly. Interestingly, however, is that people playing video games on a weekly and monthly basis had fewer errors than people playing at a daily basis. The authors postulate this might be because participants playing on a daily basis might be more prone to the trial and error approach, whereas people playing less regularly might be more considerate about their actions. It might also be because of chance due to a low sample size.

The results also show that participants playing daily and/or weekly have almost the same completion times, which are significantly lower compared to participants playing on a monthly or less frequent basis. The authors propose that the reason for this result is that participants playing often are familiar with various control schemes

Impaired Vision	Condition	Errors	Completion Time (s)
Yes	Occluded	0.1	103.63
Yes	Static	0.4	137.71
Yes	Modelled	0.4	90.95
Yes	Tracked	0	91.01
No	Occluded	0.08	85.21
No	Static	0.5	112.20
No	Modelled	1	94.55
No	Tracked	0	89.48

Table D.4: Participants with impaired vision versus participants with normal vision in regards to average amount of errors and completion times in seconds.

.

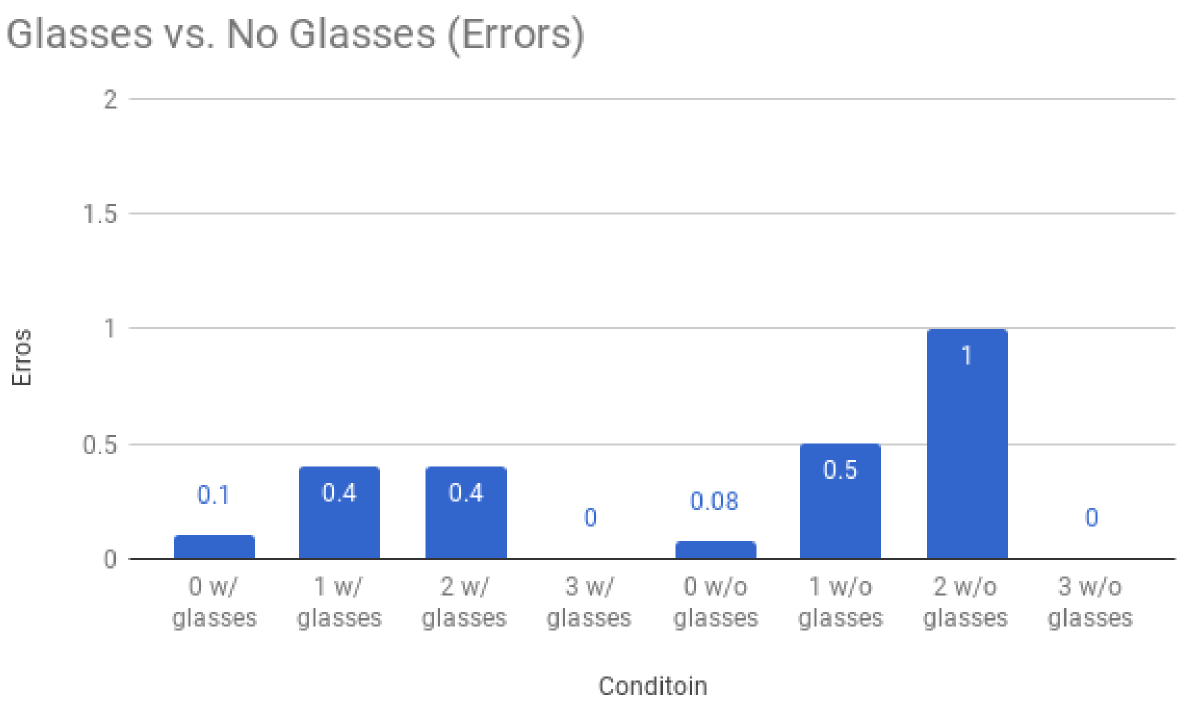


Figure D.3: The participants' average amount of errors in regards to the conditions and whether the participants have impaired vision. 0 = occluded, 1 = static, 2 = modelled, 3 = tracked

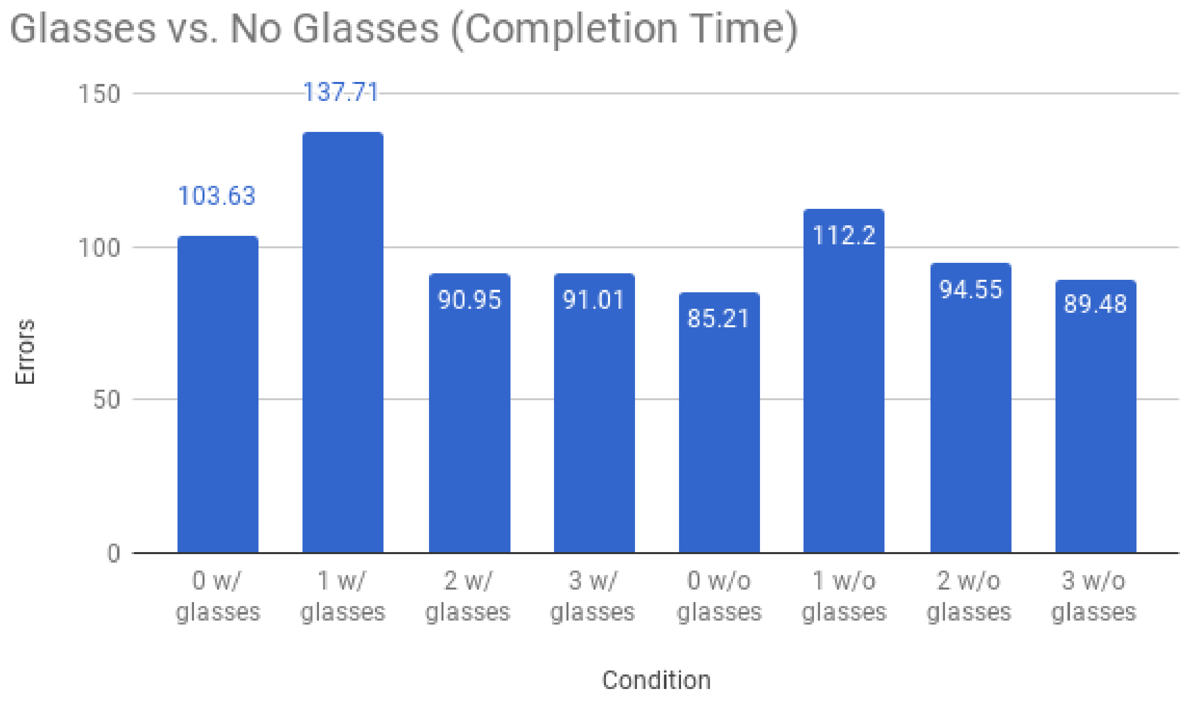


Figure D.4: The participants' average completion times in seconds in regards to conditions and whether the participants had impaired vision. 0 = occluded, 1 = static, 2 = modelled, 3 = tracked

and game mechanics, of which some are implemented in this study's SVE.

D.4 Participants Knowing Each Other Beforehand

Finally, due to this study's experiment being a collaborative task between two people in an SVE, the authors wanted to investigate whether participants knowing each other beforehand had any impact on their average amount of errors and completion times.

Table D.6 and Figures D.7 andD.8 show that whether participants knew each other beforehand did not influence the average error rate. However, it was found that completion times for participants knowing each other beforehand were higher compared to participants not knowing each other beforehand. This result was not what the authors expected before the experiment, however, the result might be because of the limited sample size of only two parti-

Number of participants	Video Game Playing Frequency	Errors	Completion time (s)
13	Daily	0.35	91.34
5	Weekly	0.25	91.13
1	Monthly	0.25	157.35
3	Less than monthly	0.33	134.05
0	Other	0	0

Table D.5: The frequency with which participants play video games in regards to average amount of errors and completion times.

.

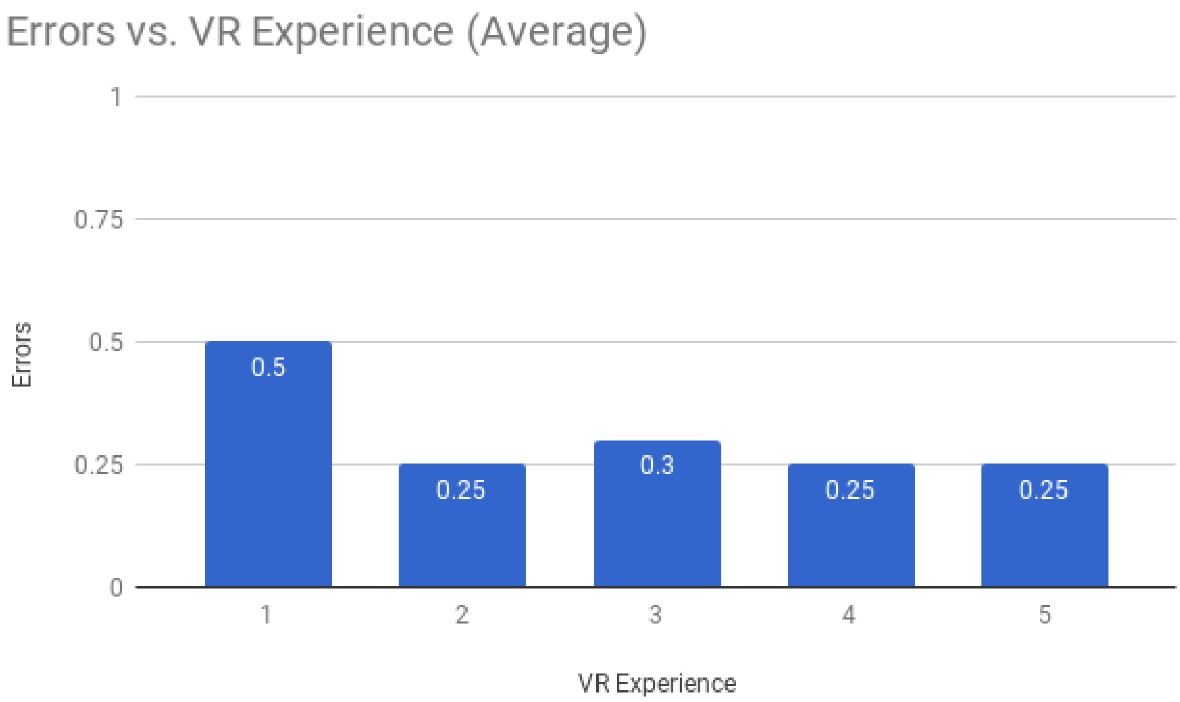


Figure D.5: The participants' video game playing frequency compared to their average error rates.

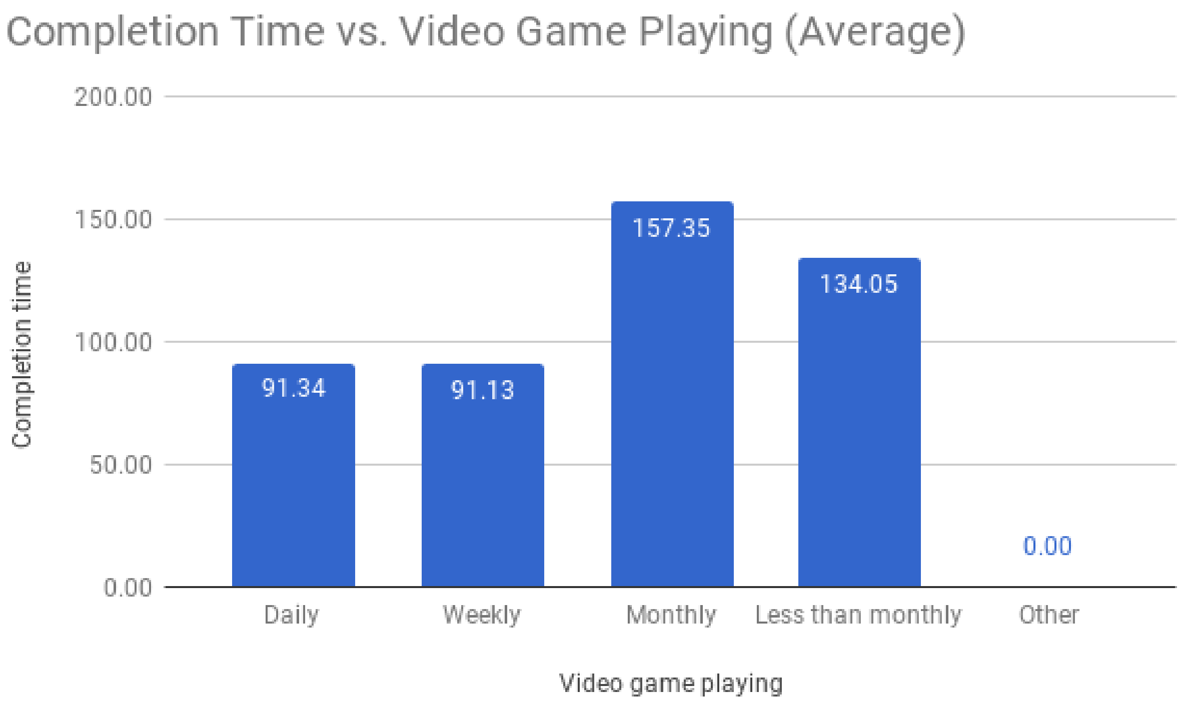


Figure D.6: The participants' video game playing frequency compared to their average completion times in seconds.

Beforehand Knowledge	Errors	Completion Time (s)
Yes	0.32	102.59
No	0.31	88.97

Table D.6: The distribution of the average amount of errors and completion times in seconds for participants knowing each other beforehand versus participants who did not know each other beforehand.

.

cipant pairs not knowing each other beforehand. Thus, the results might differ from the ones collected during this study, if it was tested on more participants who did not know each other beforehand.

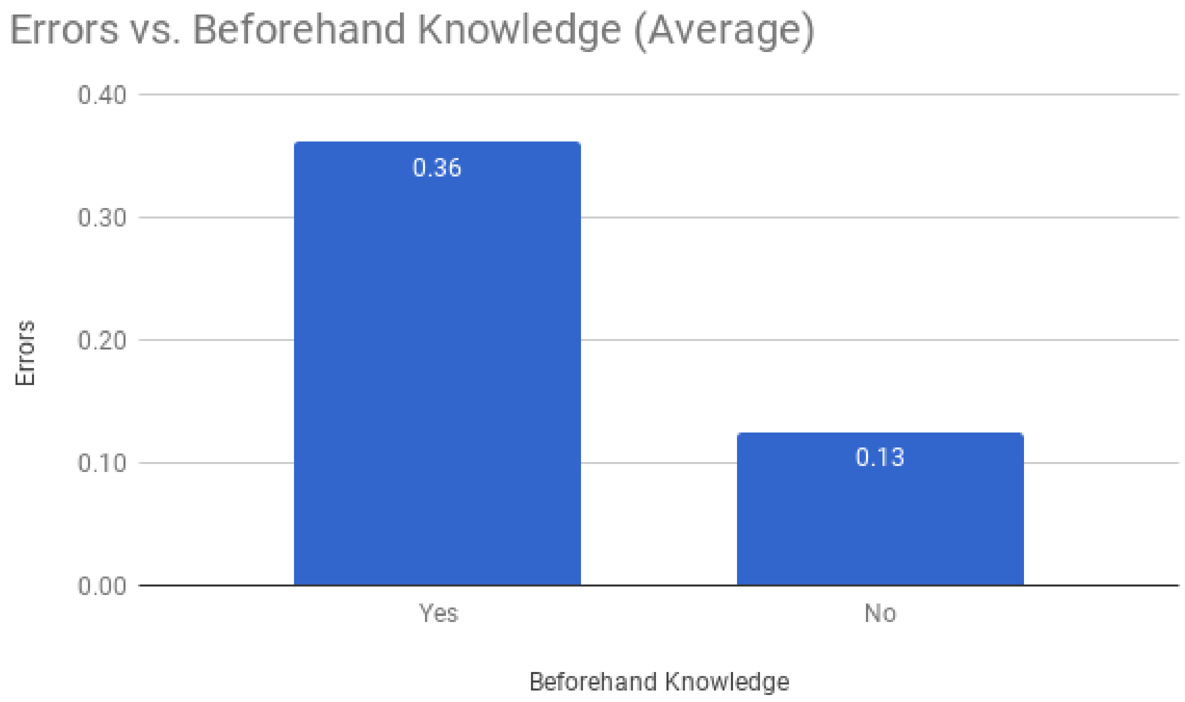


Figure D.7: The participants' beforehand knowledge of each other compared to their average error rates.

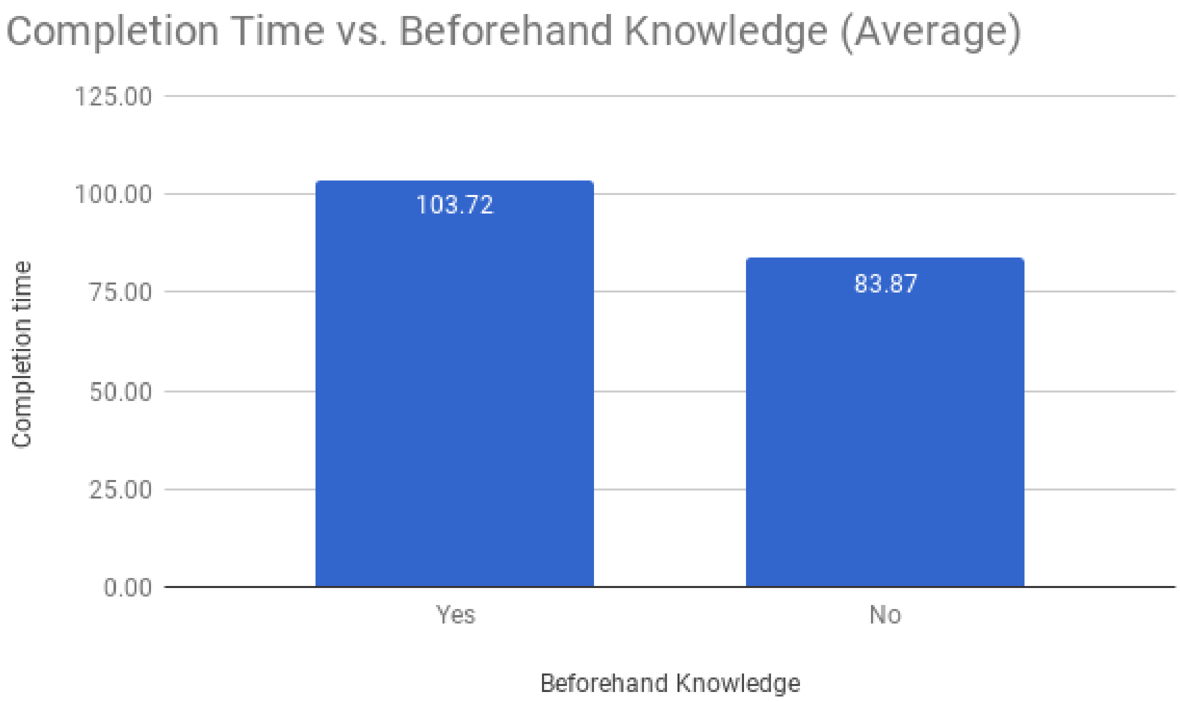


Figure D.8: The participants' beforehand knowledge of each other compared to their average completion times in seconds.

Appendix Chapter E

Interaction Design

According to (Jerald, 2016: p.275), interaction is the part of communication, which occurs between a user and an application through the use of input and output devices. Well designed interactions in an application increase performance and comfort while reducing human as well as hardware limitations. Thus, the interaction designer strives to make an application's interactions intuitive and effective (Jerald, 2016: p.275).

Throughout this subsection, a selection of this study's interaction design elements are elaborated upon. This is done in order to emphasise which aspects have been of focus while designing the study's task, in order to make it as usable for the authors and enjoyable for the participants as possible. Three overall aspects will be elaborated upon in Sections 3.1.2 *Interaction Design*, **??**, *VR Interaction Concepts*.

E.1 Affordances

An affordance is the relationship between user capabilities and object properties. Affordances can differ between users of the same object. For example, a door affords the ability to be opened and closed, but only for a user able to reach the handle (Jerald, 2016: p. 278-279).

Perceived affordances inform the users that something will happen, but not *what*, before a certain action has occurred (Vermeulen et al., 2013).

E.1.1 Affordances Used in This Study

- The **teleporter** affords that the participants can place objects in the teleporter field, remove, and teleport them. The teleporter field, funnel, and pylons have been added in order to emphasise the teleporter look to the participants.
- The **teleporter button** affords the ability to activate the teleporter by pressing it.
- The **receiver** affords the ability, together with the field and ramp, to receive an object and have it roll down the ramp when received.
- The **black hole** affords to ability to throw objects into it and have them disappear. The intention is that it should hint to the fixer to put in the transferred objects in order to fix the black hole. In addition, this interaction is explicitly told to the fixer during the tutorial.
- The participants' **hands** afford the ability to select, pick up, and throw objects inside the SVE, as well as using the teleporter button.
- The **shelf** on the fetcher's side of the SVE affords that objects can be placed in and removed from it.
- Finally, this study has also used some perceived affordances. These signalise to the participants that *something* will happen from interacting with the objects in question, but it is unknown *what* will happen from the interaction.
- Objects **changing colour** when the participants move their hands into touchable distance.

- The **particle effects** in the teleporter afford that if the participants place an object in the teleporter field it will be transferred. In addition, this is explained to the participants through the tutorial. Particle effects are also present in the receiver and the black hole.

E.2 Signifiers

A signifier is a perceivable indicator of an object's purpose, function, structure, and/or behaviour, such as signs, images, labels, handles etc. In VR, something known as *accidental signifiers* covers objects that look opposite according to its function. This could for example be that an object looks like it can be interacted with, when in reality this is not the case (Jerald, 2016: p. 279-280).

E.2.1 Signifiers Used in This Study

1. The **teleporter sign** signifies that objects can be teleported using this machine.
2. The **knobs on the CRT-monitor** on the fixer's side of the SVE. These knobs represent a misleading signifier (which is a signifier not representing an affordance) , as they look interactable without this being the case. However they only represent an affordance for people tall enough to reach them (Jerald, 2016: p.279).
3. The **tutorial photos and graphics** signify controller button assignments for the participants' tasks throughout the experiment, etc. The controller button assignment signifier can be seen in Figure E.1.
4. The **text for the black hole** for the tutorial signifies to the participants what they are looking at. This should signify the importance of the black hole to the study and that they should be interacted with in some way. The text also makes it easier to find the objects when they are mentioned in the tutorial, and afterwards they serve as a reminder for the participants.

E.3 Feedforward

Feedforward is displayed before a user performs an action and conveys what will result from said action. Compared to perceived affordances, feedforward communicates *what* will result from performing a certain action (Vermeulen et al., 2013).

E.3.1 Feedforward Used in This Study

1. **Tutorial photos** informing about controller button assignments, as they state what will happen when pressing the buttons in question.

E.4 Feedback

Feedback can be provided during or after a user has performed an action and informs the user that they have just performed an action and sometimes the result. When using feedback in VR, it is paramount that it happens in real-time. If not, the user's immersion in the environment is broken and/or they might become motion sick (in case of visual feedback, e.g. when the user moves their head). A user's physical movement is transformed into visual, auditory, and haptic feedback.. Feedback for all senses is difficult to provide using VR. It is especially difficult to

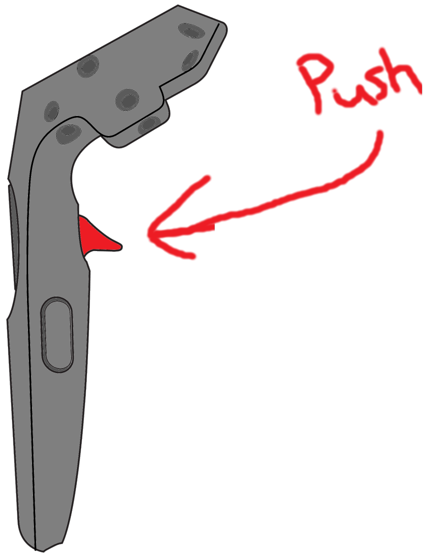


Figure E.1: The controller button assignment signifier displayed to the participants during the tutorial.

make simulated haptic feedback feel like real-world feedback (forces). It is possible to substitute (some) sensory inputs, for example through auditory feedback, having an object becoming highlighted or have a controller vibrate whenever the user performs certain actions (Vermeulen et al., 2013) (Jerald, 2016: p. 281).

Having too much feedback can result in sensory overload, which ruins the user's perception and understanding of the interaction in question. Crucial feedback should always capture the user's attention, making it important to balance the intrusiveness of the feedback. Essential information could be displayed within the head-up display, but it should be as simple as possible, as information displayed in a heads-up display generally reduces the user's awareness of the VR environment. For auditory feedback, too many beeps and overlapping audio sources should be avoided, as this can cause the users to ignore all auditory feedback and thus miss essential information. It is considered good practice to enable users to turn down or turn off auditory feedback as they wish (Vermeulen et al., 2013) (Jerald, 2016: p. 281).

E.4.1 Feedback Used in This Study

In this study, mainly three feedback modalities have been used; tactile, visual and auditory.

Tactile feedback is provided

1. After **having pressed the teleporter button** with an object placed in the teleporter field, which should let the participant know that their actions were successful. While pressing the teleporter button, a short vibrotactile pattern is played for the duration of the press. Also the button visually moves inward.

2. When **grabbing an object**, using either of the controller's trigger buttons, a short vibrotactile pattern is played. This serves the purpose of letting the participants know that they have properly grabbed the object in question.

Visual feedback

1. The **teleportation field plays an animation** when the correct object is being teleported. This confirms that the participants are transferring the correct object and that the transfer process is underway.

2. The **black hole plays an animation** when the correct object is thrown into it. This confirms to the fixer that they have thrown in the correct object and the the black hole is accepting it.

E.4.2 Auditory feedback

1. When the **participant tries to transfer an object** a corresponding sound effect is played. This way, the participant knows whether the object in question was the correct one.

2. When **an object in the teleporter is being dissolved** a sound effect is played indicating to the fetcher that the object in being transferred to the fixer.

3. An **audio effect is played for when an object is being absorbed** by the black hole. This way, the fixer knows that the object has been absorbed and that they can continue with the experiment, by having the fetcher transfer more objects.

4. A **celebratory sound** effect is played when the participants have completed the task.

E.5 Constraints

Constraints generally refer to one of two things; either 1) the technical constraints of the hardware or software being used, such as the degrees of freedom available, or 2) the rules/mechanics of the VR environment (Jerald, 2016: p. 280-281).

E.5.1 Constraints Used in This Study

1. The virtual space was minimised to match the physical space available. The physical space was shared with another HTC Vive setup, so there could potentially be two HMD users in the same physical space simultaneously. To avoid collisions, the virtual space was thus minimised.

2. The headset used throughout this study, an HTC Vive, and its controllers both have six degrees of freedom as they are able to track translation (up/down, left/right, and forward/backward) and rotation (roll, pitch, and yaw) (Jerald, 2016: p.307-308).

3. For this study, the objects are automatically repositioned on the shelf in case they hit the floor. This way, the participants do not have to bend down and pick them up. This requirement arose as the floor of the laboratory in which the experiment took place was placed higher than the virtual floor. Thus, the participants might scrape the laboratory floor with the HTC Vive controllers while trying to pick up objects from the floor.

4. The teleporter rejects an incorrect object when a participant tries to teleport it, by catapulting the object off the teleporter.

5. The teleporter only allows one object at a time, rejecting any subsequent objects. This way, the participants cannot put all objects into the teleporter and press the button to complete the task faster and easier.

E.6 Mappings

A mapping is a relationship between objects. Mappings between controls, actions, and intended purposes make it possible to learn the relationship between controlling objects and the associated actions they cause. An example of a good mapping in VR is between hand-trackers and virtual hands, which are good for pointing, gesticulating, etc. (Jerald, 2016: p.282).

E.6.1 Mappings Used in This Study

1. The avatar's hands are tracked using the HTC Vive controllers' built-in motion trackers and ensures that the participants' real-world hand movements are translated and displayed in the SVE in real-time.

2. The avatar's head is tracked using the HTC Vive's built-in motion tracking and ensures that the participants' real-world head movements are translated and displayed in the SVE in real-time.

3. The avatar's position inside the SVE is based on the player's head position. This means, that the avatar's torso do not follow the participants' position exactly. Instead, the avatar's torso is simulated by having it quickly move towards a fixed position underneath the avatar's head and rotate towards the same y-rotation of the head.

4. Between pushing the teleporter button and the avatar's hands. When the participants touch the teleporter button with their hands, the button moves along the hand until it is pushed far enough to activate the teleporter.

5. Between touching an object and the outline shader changing colour. This illustrates to the participants that they then are able to do something with the object in question.

6. The mapping between holding an object while pressing the controllers' trigger buttons and the animation of the hands, which simulates a grabbing action of a real-world object.

7. The fixer's avatar's eye movements while using eye tracking. This mapping is reinforced by the fixer being able to use the non-diegetic spatial UI crosshair, making it easier to gaze at the next object the fetcher should pick up and transfer.

E.7 Compliance

"Compliance is the matching of sensory feedback with input devices across time and space" (Jerald, 2016: p. 282). Compliance makes it feel like one is interacting with a single object (Jerald, 2016: p. 282).

E.7.1 Position Compliance (Spatial Compliance)

An example on this compliance type is when the proprioceptive and visual sense of a hand's position match, which can be when a user puts on an HMD and has to pick up the hand controllers; they have to be able to see the controllers in their correct location in order to be able to pick them up (Jerald, 2016: p. 282).

Position Compliance Used in This Study

1. The avatar's hands can be seen at all times while being inside the SVE. This way, the participants are able to pick up their virtual hands by picking up the real-world controllers (Jerald, 2016: p. 282).

2. The avatar's position and movements inside the SVE match their real-world positions. This means, that if they take a step to the right they will experience that their avatar simultaneously will take a step to the right of equal length (Jerald, 2016: p. 282).

3. The avatar's height inside the SVE and the participant's height scales according to each other. This way, the participant should be more immersed and get an increased feeling of being connected to their avatar (Jerald, 2016: p. 282).

E.7.2 Directional Compliance

Virtual and real (input) devices should move and rotate correspondingly, as this yields correspondence between what is seen and felt. This means, that if the user makes a motion for moving an object forwards, the object in the VR environment should move correspondingly. Furthermore, when rotating a virtual object with a physical input device, the virtual object should be rotated around the same axis as the physical input device (Jerald, 2016: p. 282-283).

Directional Compliance Used in This Study

1. Movement of objects. When grabbing an object inside the SVE, it is moved exactly according to how the participant moves their controllers in the real world. This means, that if the participant e.g. lifts their controllers, the objects are lifted accordingly. Additionally, if the participant rotates the controller 170 degrees clockwise, then their avatar's hands too are rotated 170 degrees clockwise along the same axis. This is also the case if the participant is holding an object inside the SVE while performing the rotating motion (Jerald, 2016: p. 282).

2. Body movements of the avatar match those of the participant. If the participant rotates their hands, moves along a certain axis, lifts something, etc., then their avatar matches these actions (Jerald, 2016: p. 282).

E.7.3 Temporal Compliance

Sensory feedback should be synchronised to the actions being performed. This is especially important in regards to visual feedback, which should be immediate as the users might else become motion sick from the interaction in question. Other forms of feedback should also be immediate, as users otherwise might become frustrated and/or interrupt the action prematurely (Jerald, 2016: p. 283-284).

Temporal Compliance Used in This Study

1. Visual feedback in form of hands moving, the mirror image, the collaborator's movements, including eye-gaze, etc. are transmitted directly to the participants. This way, it is possible to give instant feedback on the participants' physical actions and their correlations in the SVE. Additionally, this should reduce the participants' risk of motion sickness (Jerald, 2016: p. 283-284).

2. Haptic feedback in form of controller rumble happens when an object is picked up and when the teleporter button is pressed. This way, the participants instantly get feedback on the fact that they have touched/activated something. Also, they get perceived affordances indicating that something will happen soon (Jerald, 2016: p. 283-284).

3. Auditory feedback is emitted when an object is teleported and put into the black hole. It is also emitted from the black hole as an ambience sound effect, voice-over from the tutorial, and a fanfare is played when the participants finish the task. All the auditory feedback signals to the participants, that something paramount for the task has just taken place (Jerald, 2016: p. 283-284).

E.8 Direct Interaction

Direct interaction is the impression of handling an object directly rather than through an intermediary. The most direct kind of interaction happens when the user interacts with a physical object using their hands. Well-designed hand-held tools that directly affect an object are only slightly less direct, as they become an extension of this user's body when they learn how to use them probably (Jerald, 2016: p. 284).

1. This study uses direct interaction when participants use their virtual hands for grabbing, releasing and throwing objects, as well as pressing the teleporter button (for the fetcher).

E.9 Indirect Interaction

An "Indirect interaction requires more cognition and conversion between input and output." (Jerald, 2016: p. 284). Thus, input often have to go through an intermediary, where the output in not necessarily of the same type. An extreme example on this is if one uses a database; here, it is necessary to go through the aforementioned steps after having completed the search in question (Jerald, 2016: p. 284-285).

E.9.1 Indirect Interaction Used in This Study

1. An indirect interaction is used when the fixer has to look at the monitor, or the spatial UI HUD, in order to get information regarding which object to get the fetcher to transfer to them. The fixer has to look at the object information, convert it into a visual imagery of the object in question, and in some way communicate to the fetcher which object they need to transfer. This communication can be nonverbal through eye-gaze, pointing, etc.

2. The fetcher has to understand the fixer's communication regarding the object to transfer. These cues can be nonverbal, such as eye-gaze, pointing, head nodding, etc. The fetcher then needs to understand these communicative cues in order to select the correct object to transfer to the fixer.

E.10 Modes and Flow

When designing an interaction that requires an object to be manipulated, it should be ensured that it is selected first and manipulated afterwards, this is known as a so-called object-action interface design. The reason for this is twofold: firstly, before starting to use an object, the users should think about using it (possibly only subconsciously). Secondly, because users focus on the object-action sequencing, i.e. first choosing the object, and then choosing what to do with it. The transition between selecting and manipulating an object should be effortless to the users, e.g. by them only having to think about the interaction with the object in question subconsciously. Ultimately, this should result in users interacting with objects effortlessly (Jerald, 2016: p. 301-302).

E.10.1 Modes and Flow Used in This Study

1. The flow of interaction should be introduced when participants pick up and transfer objects from the fetcher's to the fixer's side of the SVE. Firstly, the fetcher has to analyse to which object the fixer is referring, pick it up, put it in the teleporter field, and finally pressing the teleporter button. However, the more this interaction is repeated, the easier it should become to the participant in question.

2. The flow of interaction should be introduced when the fixer is looking at the monitor, or using the spatial UI crosshair. Firstly, the fixer has to analyse which object should be transferred, and then they should indicate this to the fetcher. The interaction should become easier the more it is repeated.