

Cinematicraft: Exploring Fidelity Cues in Collaborative Virtual World Interactions

Siddharth Narayanan

Thesis submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Master of Science
in
Computer Engineering

Ivica Ico Bukvic, Chair
A. Lynn Abbott
Nicholas Polys
Pratap Tokekar

December 14, 2017
Blacksburg, Virginia

Keywords: Sensory Fusion, Immersion, Presence, Computer Vision
Copyright 2017, Siddharth Narayanan

Cinemacraft: Exploring Fidelity Cues in Collaborative Virtual World Interactions

Siddharth Narayanan

ABSTRACT

The research presented in this thesis concerns the contribution of virtual human (or avatar) fidelity to social interaction in virtual environments (VEs) and how sensory fusion can improve these interactions. VEs present new possibilities for mediated communication by placing people in a shared 3D context. However, there are technical constraints in creating photo realistic and behaviorally realistic avatars capable of mimicking a person's actions or intentions in real time. At the same time, previous research findings indicate that virtual humans can elicit social responses even with minimal cues, suggesting that full realism may not be essential for effective social interaction. This research explores the impact of avatar behavioral realism on people's experience of interacting with virtual humans by varying the interaction fidelity. This is accomplished through the creation of Cinemacraft, a technology-mediated immersive platform for collaborative human-computer interaction in a virtual 3D world and the incorporation of sensory fusion to improve the fidelity of interactions and real time collaboration. It investigates interaction techniques within the context of a multiplayer sandbox-voxel game engine and proposes how interaction qualities of the shared virtual 3D space can be used to further involve a user as well as simultaneously offer a stimulating experience. The primary hypothesis of the study is that embodied interactions result in a higher degree of presence and co-presence, and that sensory fusion can improve the quality of presence and co-presence. The argument is developed through research justification, followed by a user-study to demonstrate the qualitative results and quantitative metrics.

This research comprises of an experiment involving 24 participants. Experiment tasks focus on distinct but interrelated questions as higher levels of interaction fidelity are introduced.

The outcome of this research is the generation of an interactive and accessible sensory fusion platform capable of delivering compelling live collaborative performances and empathetic musical storytelling that uses low fidelity avatars to successfully sidestep the 'uncanny valley'. This research contributes to the field of immersive collaborative interaction by making transparent the methodology, instruments and code. Further, it is presented in non-technical terminology making it accessible for developers aspiring to use interactive 3D media to promote further experimentation and conceptual discussions, as well as team members with less technological expertise.

Cinecraft: Exploring Fidelity Cues in Collaborative Virtual World Interactions

Siddharth Narayanan

GENERAL AUDIENCE ABSTRACT

The work presented in this thesis explores social interactions and collaboration between users within the context of an immersive game platform. Improving the quality of these interactions is often challenging in terms of creating relatable virtual representations of the user that can also accurately capture user performances and behavioral intentions in real time. This research focuses on changing modes of performance capture to affect the quality of interactions between users. The immersive game platform, Cinecraft, uses a Minecraft style game engine to propose how interaction qualities of a shared virtual space can be used to further involve a user as well as simultaneously offer a stimulating experience. The platform can accurately capture the users' posture, limb movement, facial expressions and lip-synced mouth states and comes with an array of live cinematic production tools. The primary hypothesis of the study is that more natural modes of performance capture would result in a higher quality of interaction. Also, an additional level of intelligence to incorporate voice capture to improve tracking of users' facial performance would yield the highest quality of interactions.

The argument is developed through research justification, followed by a user-study involving 24 participants, to demonstrate the qualitative results and quantitative metrics. The outcome of this research is the generation of an interactive and accessible immersive game platform capable of delivering compelling live collaborative performances and empathetic musical storytelling. This research contributes to the field of immersive collaborative interaction by making transparent the methodology, instruments and code. Further, it is presented in non-technical terminology making it accessible for developers aspiring to use interactive 3D media to promote further experimentation and conceptual discussions, as well as team members with less technological expertise.

Acknowledgments

I would like to express my deepest gratitude to my committee chair and advisor Ivica Ico Bukvic for giving me the freedom to find my own way, always directing me towards clarity and his continued support throughout the project. My committee co-chair, Dr. Lynn Abbott, has also been immensely helpful in his counsel and introduced me to members of his research group to seek new insights. An enormous thank you goes to Dr. Nicholas Polys, for his recommendations on closely related research projects & publications and being readily available for guidance. I am very grateful to my advisor Dr. Pratap Tokekar for his encouragement and support during my master's program. I would also like to thank all my collaborators on previous versions of the project for their contributions, time and expertise.

Contents

List of Figures	viii
List of Tables	x
1 Introduction	1
1.1 Research Problem	2
1.2 Research Questions	3
1.3 Scope of the Thesis	4
1.4 Thesis Structure	4
2 Background	5
2.1 Virtual Environments for Communication	5
2.2 Increasing Avatar Fidelity	7
2.2.1 Avatars and Agents	8
2.2.2 Expressive Avatars	8
2.2.3 Nonverbal communication in face-to-face interaction	9
2.2.4 Constraints on Avatar Fidelity	11
2.2.5 Presence & Co-presence	19
2.2.6 Measurement Approaches	22
2.3 Related Work	25
2.3.1 Operacraft	25
2.3.2 Community Support	26

2.3.3	Findr - Immersion and User Engagement	26
3	System Design	27
3.0.1	Migration to Minetest	27
3.1	Architecture	29
3.1.1	Functional Modifications	31
3.1.2	Modes of Interaction	32
3.1.3	Sensory Fusion	33
4	Experiment	35
4.1	Data Collection	35
4.1.1	Defining the research goals and expectations	35
4.1.2	Defining the independent and dependent variables	36
4.2	Experiment	37
4.2.1	Experimental aims and expectations	37
4.2.2	Experimental design	38
4.2.3	Tasks	40
4.2.4	Piloting	41
4.2.5	Apparatus	41
4.2.6	Procedure	41
4.3	Results	44
4.3.1	Analysis	45
4.3.2	Discussion of Results	48
4.4	Contribution	49
4.5	Conclusion	49
5	Bibliography	51
	Appendix A : Questionnaires	60
A.1	Presence Questionnaire	60
A.2	Co-Presence Questionnaire	61

A.3 Immersive Tendencies Questionnaire	62
Appendix B : Results	63
B.1 : Presence scores	63
B.2 : Co-Presence scores	64
B.3 : Self reported Co-Presence scores	65
B.4 : Empathy scores	66
B.5 : Mutual Awareness scores	67
B.6 : Attention Allocation scores	68
Appendix C : List of body expressions	69
Appendix D : IRB Approval Letter	70

List of Figures

2.1	Summary of avatar requirements	9
2.2	Technical constraints affecting avatar fidelity	11
2.3	Three dimensions of visual fidelity	12
2.4	Impact of degradation on facial and affect recognition	17
2.5	Six conceptualizations of presence	20
2.6	Proposed measurement approaches	22
2.7	Inspiring K-12 students to create stories through a live production of Operacraft	25
3.1	Protocol structure between the Kinect C# application and Minetest	29
3.2	System architecture for the Cinemacraft	31
3.3	Sensory Fusion Design	33
4.1	Interaction 1: Vanilla + Voice	38
4.2	Interaction 2: Vanilla + Voice + Keyboard + Kinect Upper Body Only	38
4.3	Interaction 3: Vanilla + Voice + Kinect	39
4.4	Interaction 4: Vanilla + Voice + Kinect + Sensory Fusion(Audio)	39
4.5	Sample body expressions as part of the experimental task list. A full list of these body expressions is provided in appendix C.	40
4.6	Top Left: Participant 1 - Avatar talking in sensory fusion mode; Top Right: Motion and Audio capture; Bottom Left: Participant 2 - Avatar interacting in virtual world; Bottom Right: Both participants can view the scene in third person	42
4.7	Sensory Fusion Improvements: The user's voice is used to create more expressive avatars synced with their speech	43

4.8	Pairs of participants communicating with each other using embodied interaction and audio	44
4.9	Means of cumulative questionnaire responses for each variable	45
4.10	Means of cumulative questionnaire responses for contributing factors to co-presence	45

List of Tables

3.1	Kinect data for Cinemacraft Avatar Mapping	32
4.1	Mean and standard deviations of count response variables	44
4.2	ANOVA test for Presence scores	46
4.3	ANOVA test for Co-Presence scores	46
4.4	ANOVA test for Self-reported Co-Presence scores	47
4.5	ANOVA test for avatar Empathy scores	47
4.6	ANOVA test for Mutual Awareness scores	47
4.7	ANOVA test for Attention Allocation scores	47
4.8	Correlation Matrix for CO-P and P scores with respect to IT scores	48
B.1	User Presence scores	63
B.2	User Co-Presence scores	64
B.3	User Self reported Co-Presence scores	65
B.4	User Empathy scores	66
B.5	User Mutual Awareness scores	67
B.6	User Attention Allocation scores	68

Chapter 1

Introduction

With the advent of virtual reality (VR) technologies, the discourse around presence in digital environments is developing rapidly. Recent revolution in the area of off-the-shelf immersive technologies has changed the way users interact with games, media, and the arts. Experiments have integrated human interaction into performance, either as stylized body movements [1] or through the use of virtual interfaces [2]. Creative projects have actively adopted the Microsoft Kinect [3] and more recently Kinect HD [4], along with an array of affordable alternative all-in-one consumer-level motion capture devices to explore novel interactions and perceptions. Video games have also served as a rich foundation for artistic expression using immersive devices through game mods, machinima [5], digital puppetry [6][7]. Works of cyber-fiction and digital art have often depicted a fully 3-dimensional and immersive datascape simultaneously accessible by millions of networked users. This virtual world is described as having spatial properties similar to the physical world and its virtual cities are populated by digital proxies of people, called avatars. Here people can interact with each other and with artificial intelligences (AIs) that are visually and sometimes behaviorally indistinguishable from humans. The multi-sensory sophistication of this shared space is such that it supports interpersonal communication on a level of richness interchangeable with face-to face interaction. Researchers have also incorporated immersion through tele-presence[8] and natural interfaces and embodied interactions within the game [9].

This thesis explores the intersection between collaborative virtual environments (CVEs) and mediated communication through sensory fusion. These also encapsulate two of the central goals not only of CVEs, but also of all communication media. First, to enable groups of people to collaborate and interact socially in an efficient and enjoyable way, and second, to foster the illusion that people are together when in reality they are in distinct physical locations. CVEs have the makings of a potentially powerful medium of communication that heralds new opportunities and challenges. It is the inherently spatial property of CVEs that sets them apart from other collaborative media. Though video-conferencing and groupware systems allow users to interact visually, the 3D context of each person's physical

environment is lost. This can pose difficulties in small group interaction, where conversation management can be disrupted by the ambiguous eye gaze cues. The loss of 3D context can be particularly problematic in tasks where the preservation of spatial relationships is essential. CVEs can begin to address these concerns by placing geographically dispersed users in a shared, computer-generated space where they can interact with the environment and with other users represented by avatars. Immersive interfaces can also offer multi-modal, surrounding experiences that can create a strong sense of being inside that artificial space (presence), and sometimes of being there with others (co-presence). As mediators of users' actions and appearance, avatars play a significant role in social interaction in CVEs.

There are two distinct domains in which CVEs are currently being investigated: in online virtual communities and in networked research laboratories. Laboratory based research experiments typically involve smaller groups of participants and highly specialized immersive interfaces including Head Mounted Displays (HMDs) and immersive cave automatic virtual environment (CAVE) systems. Research has typically focused on issues such as the impact of display type or the visual sophistication of the avatars during interactions.

1.1 Research Problem

One of the major drawbacks of CVEs is the relative paucity of avatar expressiveness in comparison to live human faces on video. Avatars in graphical chats vary widely in appearance and can exhibit lively behaviors; however they have been critiqued for serving merely as placeholders and failing to contribute meaningfully to conversation. The avatars used in collaborative laboratory based studies often have limited behavioral capabilities, such as the movement of a single arm for object manipulation despite a high level of visual fidelity. A significant challenge in developing CVEs as a communications medium is the development of expressive avatars capable of contributing to interaction. Although CVEs can offer the benefits of spatial interaction and immersive experience, they remain low-fidelity compared with video-mediated communication (VMC); where VMC portrays objects and events from the real world, CVEs portray an artificial environment populated with artificial representations of people. There are technical challenges as well as theoretical goals to consider when increasing avatar fidelity. These affect both the avatar's static appearance (visual fidelity) and dynamic animation (behavioral fidelity).

In terms of the avatar's appearance, technical restrictions related to rendering and bandwidth mean that there is a tension between realism and real time. VE designers pay particular attention to exploiting the capacity of the human perceptual system to infer information from limited but informative cues. Naturally it is not always admissible to take such shortcuts. The level of realism required depends on task requirements. For instance, insufficient visual and haptic realism in a flight training simulation could result in disastrous consequences. It is arguable that communication is a more forgiving task in that it does not require full photorealism. The ability of humans to decode caricature and cartoons indicates that we do

not require exhaustive photorealism depictions to decipher the human form. Lessons from cartoon animation also indicate that photorealism is secondary to behavior, provided that behavior is convincing. Social psychology research on face-to-face interaction has identified several nonverbal behaviors that serve a communicative function in the expression of emotion and in effective conversation management. These include facial expression, eye gaze, gesture, posture and proxemics (spatial behavior) [10]. Body and facial tracking makes it possible to animate an avatar using motion data from a real person. Achieving convincing avatar behavior, however, introduces additional challenges. Tracking equipment can be expensive as well as intrusive for users. On a theoretical level, it can also be argued whether full tracking is the best way to deliver full avatar control through embodied interactions. Being computer-generated, avatars afford control not only over appearance but also over behavioral expression, thereby potentially avoiding the pitfalls of nonverbal leakage that can occur in both face-to-face and video-mediated communication. However, manual control over a wide range of an avatar's actions using traditional techniques (e.g. keyboard and mouse) would introduce high cognitive load. The problem of driving avatar behaviors that appropriately represent the users can therefore be summarized as the tension between control and cognitive load.

1.2 Research Questions

Given these technical constraints and theoretical considerations, the approach taken in the research presented in this thesis has been to explore the lower boundaries of avatar visual fidelity. The logic used by many VE designers is to exploit minimum cues to obtain maximum results. This research extends earlier studies by investigating whether minimal avatar visual fidelity coupled with increasing avatar behavioral fidelity can contribute to social responses and create positive perceptions of the interaction experience. It comprises of a user-based experiment addressing 2 nested questions:

1. *What is the relationship between the avatar's behavioral fidelity and presence ?*

This question addresses the assumption made numerous researchers, that behavioral fidelity should be prioritized over visual fidelity in the development of expressive avatars. We modify the avatar's behavioral fidelity through different interaction modes for each experimental task varying from standard keyboard + mouse interaction+ audio chat to upper body to full body real time motion capture to study whether improvements in behavioral fidelity benefit the constant low fidelity avatars regardless of their appearance.

2. *Question 2: Does Sensory Fusion increase presence and co-presence?*

The improvements through sensory fusion are explored by adding improved mouth detection through a live audio input layer. When measuring these improvements, a question concerns the research methods employed to study peoples' sense of being

with others in a shared VE. This is addressed through a combination of post-test questionnaires and analysis of user study data.

1.3 Scope of the Thesis

The appearance of avatars can range from abstract to animal-like to humanoid, and from cartoon-like to photorealistic. This thesis is concerned exclusively with voxel game engine based low-fidelity avatars, specifically within the Minetest game engine. Its main focus is on subjective responses to varying levels of avatar behavioral fidelity. The major focus of the user-study is on the perceived contribution of avatars to experiences of interaction. The avatar representation in the study was chosen with the intent of isolating the impact of avatar behavior as far as possible from potentially confounding factors such as real-life relationships, explicit technical knowledge of how the avatars were animated, or communicative conventions derived from long-term use.

1.4 Thesis Structure

Although CVEs potentially support spatial and fully immersive interaction, one significant barrier to interaction is the avatar's limited expressive potential. Chapter 2 covers relevant research and literature. It covers the technical challenges involved in increasing avatar expressiveness, exploring the potential gains of iteratively increasing the behavioral fidelity. This study focuses on presence and co-presence, which are discussed in relation to tele-presence and social presence. Chapter 3 discusses the system design and sensory fusion to address the research questions. It introduces the implementation, modes of interaction and software design. This is followed by the experiment as part of the Institutional Review Board (IRB) study in chapter 4. This features the experimental design, as well as the method of statistical analysis used to analyze the questionnaire data along with the overall findings and the implications that can be drawn from them. It draws conclusions from the findings and proposes directions for continuing research.

Chapter 2

Background

There are numerous application areas for virtual environments (VEs), from simulation to training to the treatment of phobias. This thesis focuses on CVEs as a 3D communications medium. We first explore the potential strengths of CVEs as a communications medium and highlight its effectiveness of graphical embodiments for interaction and collaboration. Next, we look at the creation of expressive avatars for communication purposes, and some technical constraints on the level of visual and behavioral fidelity achievable in current CVEs. Finally, we look at the challenges of defining and measuring this sense in terms of presence and co-presence.

2.1 Virtual Environments for Communication

Anthony Giddens described face-to-face talk as a communications medium [11]. In this thesis, the term communications medium applies exclusively to interpersonal, mediated interaction between geographically dispersed people, or between people and artificial social entities. One of the underlying assumptions behind research in both video-mediated communication (VMC) and CVEs has been that the inclusion of visual information can improve mediated interaction by harnessing our natural ability to read meaning into the human form. Short, Williams and Christie have argued that all attempts at producing visual communications media are primarily directed at remedying what is the most obvious defect of the simple telephone - the fact that one cannot see the other person or group [12]. The question that arises with the advent of CVEs is, what happens when both the environment and the people in it are not portrayals of the real world, but artificial simulations? CVEs are networked, computer-generated environments capable of supporting human-to-human communication by allowing users to interact with the space and with each other via graphical embodiments called avatars. This thesis employs the term ‘collaborative’ in the broadest sense, as “any activity involving a series of tasks within a virtual environment that requires social and co-

operative efforts between users within a group” [13]. In this definition, CVEs include not only environments used explicitly for work-related purposes, but also for social interaction and play. CVE applications can range from conferencing, simulation and training, shared visualization and collaborative design, to social communities and multiplayer games. Avatars play a significant role in all of these contexts because they embody the user in a shared space, opening multiple possibilities for interaction. CVE research is cross-disciplinary, drawing from fields including computer science, psychology, sociology, architecture, urban planning and human-computer interaction. The study of collaboration in CVEs relates closely to the field of computer supported cooperative work (CSCW), which is concerned with investigating how computers can facilitate human interaction. CSCW technology is commonly referred to as groupware, defined by Ellis, Gibbs and Rein as computer-based systems that support groups of people engaged in a common task (or goal) and that provide an interface to a shared environment [14]. Groupware systems differ from single-user applications in that they reflect the activities of multiple users in the environment, therefore actively supporting group communication, collaboration and coordination. The category of ‘synchronous distributed interaction’ includes computer conferencing technologies that combine different configurations of document sharing facilities and live video of participants. Though not explicitly included in the taxonomy, CVEs also belong to this category; like video-conferencing, they differ crucially from face-to-face interaction in that communication is synchronous, but participants occupy distinct physical spaces.

We can adapt Benford et al’s dimensions for computer supported cooperative work systems (CSCWs) that emphasize spatial interaction according to three dimensions as:

1. *Spatiality*: the degree to which participants are provided with a shared and navigable spatial context.
2. *Immersion*: the degree to which participants are provided with a surrounding sensory experience, resulting in a sense of transportation from their physical surroundings to the mediated context.
3. *Fidelity*: the degree to which sensory information in the mediated context is based on information from the real world.

While it is not the aim of this section to compare the relative merits of video and avatar mediated communication, the discussion of some key properties along these three dimensions highlights some of the potential strengths of CVEs as a medium. Video-conferencing portrays participants’ real appearance and actions and is therefore high in fidelity; however, it is experienced on a 2D screen and is therefore low in spatiality and immersiveness. Conversely, Immersive VEs (IVEs) provide a 3D surrounding experience and are high in spatiality and immersiveness, but low in fidelity because portrayals of participants and the environment are synthetic. The general aim of CVE research is to increase fidelity with a view to bridging the gap between virtual and face-to-face interaction.

Fidelity concerns the degree to which objects and events in the mediated space are direct representations of the real world. In the context of group interaction, the degree of fidelity of a CVE hinges on its capacity to portray a convincing context and process for collaboration. This directly affects interaction with shared objects: CVEs enable participants to work with shared access to objects located in the virtual environment, whilst media spaces endeavor to provide participants with the opportunity to work on real, physical objects [15]. Thus the advantage of CVEs is their ability to place objects in a 3D context, but their disadvantage is that the objects are not real. Similarly, human embodiments in CVEs are synthetic and vary in the accuracy with which they mimic the real appearance and behaviors of the person they represent. This ability to couple anonymity with visual expressiveness has been cited as one of the hallmark attractions of online virtual communities [16]. However, the ambiguous relationship between an avatar and the person represented also poses complex challenges in terms of creating expressive embodiments that contribute to the interaction taking place.

CVEs have several properties that make them suited to group interaction:

1. *Spatial*, providing a shared 3D interaction context
2. *Navigable*, allowing users to freely navigate the 3D space
3. *Embodied*, representing users by digital proxies called avatars
4. *Synchronous*, enabling people to interact with each other in real time
5. *Multi-user*, supporting multiple, geographically dispersed users

In summary, this section has discussed some of the potential advantages of CVEs as a communications medium. Their spatiality and immersiveness set them apart from other groupware systems in their ability to provide a surrounding, multi-sensory environment with consistent spatial properties. However, CVEs are by definition synthetic environments and therefore one of the challenges is to increase fidelity while preserving the advantages of spatiality and immersiveness.

2.2 Increasing Avatar Fidelity

The creation of expressive avatars and a full spectrum of interactions is a significant challenge in developing CVEs as a communications medium. This section will begin by defining and classifying avatars. It will then discuss some communication requirements and the challenges these entail. In particular, there are technical restrictions on the amount of visual detail that can be conveyed and on the ability to drive appropriate behaviors in real time. Increasing avatar expressiveness therefore entails a potential trade-off between photorealism and behavioral realism. This section discusses related research studies on the impact of

different aspects of appearance and behavior on people’s social responses to avatars and agents.

2.2.1 Avatars and Agents

Virtual humans are visible, computer-generated humanoid characters used for a wide range of applications. They can function as interface agents, news readers, game characters, digital extras populating film sets and archaeological reconstructions, surrogates for medical training, and as personalized dummies used to try on clothes in virtual shopping applications. By convention, virtual humans are classified in terms of agency, meaning whether the intelligence represented is human or artificial [17], [18]. Where avatars represent real humans engaged in interaction, agents are driven purely by a computer program and can vary widely in sophistication. Some agents have simple, scripted behaviors whereas others such as MIT’s virtual estate agent, Rea, are designed to sustain verbal and gestural interaction with human interlocutors [19]. Research into embodied conversational agents is driven by fundamentally different concerns and is therefore beyond the scope of this thesis, which focuses on human-to-human communication. In the strictest objective sense, agency is binary because the virtual human either represents a human or it does not. Nevertheless, avatars vary in the degree to which their behaviors represent the real actions or intentions of the person represented. Blascovich argues that agency is a continuum ranging from fully artificial at the low end, to fully human at the high end, with the term avatar being reserved for the upper extreme of the continuum. Blascovich stresses that agency is subjective, in that it is the extent to which individuals perceive virtual others as representations of real persons [18]. This thesis primarily focuses on the degree of sentience attributed to virtual humans, which can vary depending on their behavior in the course of the interaction. The question of how attributed sentience improves the visual behavior of virtual humans is addressed in the experiment. The challenge with human-human communication is to drive avatar behaviors that enrich, rather than hinder, communication between remote participants. The following subsection addresses some communication requirements, and is followed by a discussion of some key technical constraints shaping the development of expressive avatars.

2.2.2 Expressive Avatars

Benford et al. have laid out some requirements for avatars in CVEs covering aspects of appearance, behavior and relationship to the real body of the person represented [20]. They argue that these requirements often conflict with each other, and prioritization hinges on interaction context and technical resources. Figure 2.1 illustrates some distinctions between video-conferencing and CVEs.

It is arguable, however, that requirements cannot be easily separated because the avatar’s appearance and behavioral requirements often intertwine. For instance, simple T-shaped

	Requirement to convey	Comment
Appearance	Presence	Must signal person's presence in the VE in an "automatic and continuous" way
	Location	Must signal the person's position and orientation in the VE
	Identity	Avatar's appearance must provide recognisability over time as well as the ability to distinguish between avatars
Behaviour	Availability	Must convey person's availability for interaction, and how busy or interruptible they are
	Activity	Must portray person's current activity and focus of attention
	Expression	Should convey expressiveness through gesture and facial expression

Figure 2.1: Summary of avatar requirements

blocky avatars are sufficient to signal presence and location, and their color can identify them as distinct from other avatars. However, for other functions the avatar is likely to require more visual detail, such as eyes to convey attention and arms to convey simple interaction with objects such as grasping (as implemented by Hindmarsh et al. [15] in their study on collaborative object manipulation). These functions, though challenging, are relatively simple compared to the difficult problem of conveying convincing behavior. In face-to-face interaction people rely heavily on nonverbal cues such as eye gaze, facial expression, posture, gesture and interpersonal distance to supplement the verbal content of conversation. Indeed some argue that nonverbal signals not only constitute a separate channel of communication, but that they often override verbal content [21]; in other words how something is said can be more important than what is said. Thomas and Johnston emphasize that the need to maintain consistency between dialogue and nonverbal expression is equally important in cartoon animation: 'Do not let the expression conflict with the dialogue. Nothing can be more distracting than this' [22]. This points to a need to align the visual behaviors of avatars to the ongoing interaction.

2.2.3 Nonverbal communication in face-to-face interaction

Nonverbal behaviors serve at least two central functions in face-to-face interaction: conversation management and the communication of emotion. Conversation management concerns the use of paralinguistic cues to ensure the smooth flow of conversation. Movements such as eyebrow raises, head nods and posture shifts give structure and rhythm to the conversation and are essential to maintaining a sense of mutual understanding. The communication of emotion is itself integral to the regulation of communication and interaction [23]. In the words of Picard, "emotions not only contribute to a richer quality of interaction, but they

directly impact a person’s ability to interact in an intelligent way”. Emotion is crucial in the communication of understanding, and speakers continually monitor listeners’ body language and facial expression for confirmation that they are being understood.

Facial expression

Within nonverbal communications research, the greatest amount of attention has been devoted to facial expression, possibly because there is considerable consensus that the emotional signals are most efficiently conveyed through the face [24]. Researchers in the Darwinian tradition believe that emotion is the result of evolutionary processes [25] and therefore there are several aspects of emotional communication which are universal across cultures. Ekman [26] and others have agreed on a set 6 primary emotions that can be decoded well above chance from facial expression alone: happiness, sadness, surprise, fear, anger, disgust and contempt. Although these results have been challenged on methodological grounds, primarily because they used static photographs and forced-choice questionnaires, findings suggest that the same six basic emotions can be reliably decoded in moving video as well as computer generated characters [27]. Our experiment also focuses on breaking down these expressions into smaller parts and observe their variability with respect the accuracy of incoming data with and with-out sensory fusion.

Body movements

Gesture, posture and proxemics have received less research attention than either facial expression or gaze. There is evidence [28], [29] that the body can communicate information about emotion on several levels. Ekman and Friesen [28] suggest that while the face communicates information about the nature of an emotion, body movements (acts) convey additional information about the intensity of an emotion. Further, still positions (postures) can communicate information about intensity and sometimes gross affective state along a pleasant/unpleasant dimension. Posture changes at a slow rate, and is therefore more relevant to longer-term aspects of conversation rather than to micro-momentary feedback. Argyle [30] ties posture to the expression of mood and personality. Bull [31] identified a link between postures and certain emotions. Interest is associated with a forward lean and drawing legs back, whereas boredom is associated with a backward lean, lowering of the head or leaning the head on one’s hand, outstretched legs, and turning the head away. In summary, nonverbal behaviors play a central function in face-to-face conversation and the avatars ability to convey such nonverbal cues is likely to affect how they are perceived as well as their contribution to social interaction.

2.2.4 Constraints on Avatar Fidelity

There are key technical constraints affecting the degree of avatar fidelity possible in current CVEs. In this thesis, avatar fidelity is taken to encompass both static properties of avatar appearance (visual fidelity) and dynamic properties of animation (behavioral fidelity). The first consideration with regard to visual fidelity, is the tension between realism and real time [32]. Slater et al. individuate three aspects of realism in VEs: geometric realism, illumination realism and behavioral realism [32]. While all these are desirable in the creation of convincing VEs, they come at the expense of real-time performance. Increased photo-realism introduces computational complexity, resulting in significant and unwanted delays to real-time communication. The second consideration, regarding behavioral fidelity, is the tension between control and cognitive load. Mapping a person’s communicative intentions to their avatar’s behavior presents considerable technical challenges. Full manual control using traditional interfaces like mouse and keyboard to operate the full range of avatar functions introduces high cognitive load; on the other hand, reducing cognitive load through tracking or alternative approaches can result in a loss of control over the full range of avatar behavior.

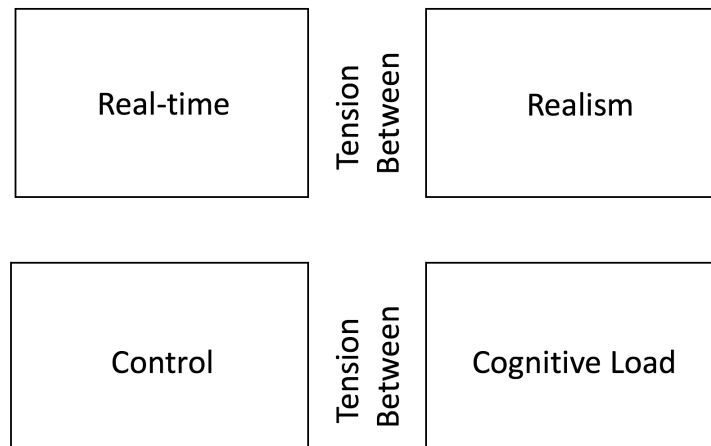


Figure 2.2: Technical constraints affecting avatar fidelity

Constraints on visual fidelity

Schroeder argues that avatar embodiment affects how people relate to each other in CVEs, and that “avatar appearance will influence interaction in all shared VEs, and there is still much research to be done on pinning down this influence” [33]. Findings reported by Nilsson et al. suggest that avatar appearance may not be as important for long-term collaborations, particularly where participants already know each other [34]. Nonetheless, in the context of one-off interactions of interest in this thesis, avatar appearance is likely to have some significance. In terms of appearance, Schroeder points out that “it is not only the shape of

virtual bodies that matters in the experience of virtual worlds, but also the level of detail with which they are represented” [35]. Fidelity concerns not only morphology and photorealism, but also the degree to which the avatar resembles the person represented (referred to by Benford et al. as ‘truthfulness’ [20]).

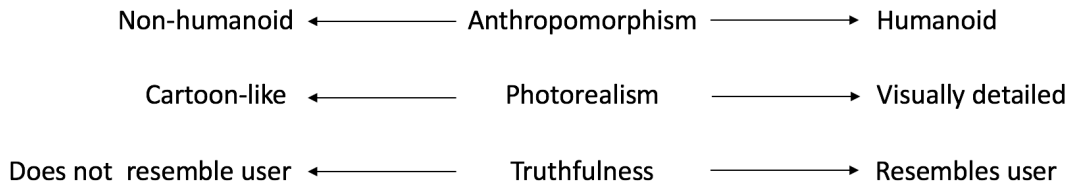


Figure 2.3: Three dimensions of visual fidelity

Avatars can range from simple ‘blockies’ to highly photorealistic forms. Avatar morphology in graphical chats ranges from humanoid to anthropomorphised animals to abstract shapes [36]; the research presented in this thesis is concerned exclusively with 3D, dynamic humanoid avatars. Within this humanoid category, avatars can also vary in terms of their fidelity to the user’s real-life physical appearance. Typically, avatars used for communication purposes are relatively cartoon-like. Cheng, Farnham and Stone suggest that users may prefer to be represented by humanoid avatars that are neither too cartoon-like nor too photorealistic [37]. The reason why highly photorealistic avatars are not used, however, is primarily due to technical constraints on local rendering and network bandwidth. Morningstar and Farmer cite the latter as a particular concern in the design of graphical chats, emphasizing that communication bandwidth is a scarce resource [38]. Similarly Hindmarsh et al. advocate using recognizable but simplistic humanoid avatars for performance reasons: We adopted this approach because we felt that it is the most obvious choice and indeed, is one that has been widely adopted by CVE designers [15]. Their avatars had a head, torso and arms, and were capable of simple behaviors including looking, pointing and grasping objects.

Constraints on behavioral fidelity

Where visual fidelity concerns the static properties of an avatar’s appearance, behavioral fidelity concerns its dynamic properties of animation. The primary focus of this thesis is on the perception of avatar behaviors rather than how they are driven. Nevertheless, the difficult problem of driving appropriate behaviors is of interest because it directly shapes the research problem. Research on nonverbal behavior in face-to-face communication [30] can offer valuable leads on how to improve avatar expressiveness without resorting to full tracking. This subsection discusses the problem of how to implement nonverbal behaviors in humanoid avatars.

Avatars in existing graphical chats have been widely critiqued for their insufficient and sometimes misleading behaviors. Durlach and Slater suggest that CVEs create a web of

relationships connecting people to each other, and individually to their own avatars [39]. The way individuals relate to their own avatar is likely to hinge on how reliably it represents them. In enriching the communicative potential of avatars it is essential not to misrepresent the actions or intentions of users.

The Uncanny Valley

The term ‘Uncanny Valley’ relates to the sense of unease and discomfort experienced when people look at realistic virtual humans. Researchers in immersive virtual reality have defined the term ‘presence’ as ‘the feeling of being bodily in an externally-existing world’ . The Uncanny Valley is relevant to a study of presence because it questions widely held assumptions about the correlation between realism and belief within a virtual world. The eyes of virtual characters are probably crucial elements in this interaction because of their key role in conveying intentional states to other organisms. As immersive environments grow increasingly realistic these may themselves generate a type of Uncanny Valley response thus far only reported when observing virtual humans.

Animated characters give off a number of perceptual cues that suggest they are people. For example, a simple character might move its eyes in a way that we recognize as similar to human behavior, and so we accept it as being a person on some level. At all levels of realism it is possible to create characters that are ‘not quite right’ or that ‘don’ t quite work’ while others are appealing and easily accepted. Highly realistic characters operate within a different set of aesthetic constraints than those exhibiting lower realism. As a character becomes increasingly realistic it is constrained to the physical attributes of a real person and a viewer’s tolerance for abstracted appearance and behavior is reduced.

Current technology allows increasingly graphically realistic characters but often their behavior and movements do not match up to this realism. The high graphical realism gives strong cues that suggest the character is a person and thus raises high expectations for motion and behavior. When the quality of these does not match up, it suggests the character is in fact not real. This creates a perceptual paradox which may generate the sense of the uncanny.

Interaction Fidelity

Interaction designers often strive to design realistic and natural interactions when developing VR applications because naturalness has been associated with increased usability and improved user performance [40]. However, when working with VR systems with limited capabilities, designers often resort to creating or using semi-natural interaction techniques. Cinemacraft is insulated from most of these limitations due to the low-fidelity visuals for the virtual world. In fact, the previous section illustrates how it helps sidestep the uncanny valley.

High-fidelity interaction techniques usually outperform mid-fidelity techniques. This is not surprising given the intuitive notion that usability improves with more naturalness. However, low-fidelity interaction techniques also usually outperform mid-fidelity techniques. This result directly contradicts the intuition that more naturalness is good. Additionally, it indicates that increasing interaction fidelity may produce a U-shaped curve in terms of user performance. As interaction fidelity continues to increase, and the overall degree of fidelity becomes relatively high, user performances will rebound and be comparable, if not better, than those afforded by the low-fidelity techniques.

Bio-mechanical Symmetry

Bio-mechanical symmetry is the objective degree of exactness with which real-world body movements for a task are reproduced during interaction. It consists of three sub-components. First, anthropometric symmetry is the objective degree of exactness with which body segments involved in a real-world task are required by an interaction technique. Second, kinematic symmetry is the objective degree of exactness with which a body motion for a real-world task is reproduced during an interaction technique. Third, kinetic symmetry is the objective degree of exactness with which the forces involved in a real-world action are reproduced during an interaction technique.

Input Veracity

Input veracity is the objective degree of exactness with which the input devices capture and measure the user's actions. It also consists of three subcomponents. First, accuracy refers to how close an input device's readings are to the true values that it attempts to measure. Second, precision concerns a device's ability to reproduce the same results when repeated measures are taken under the same conditions. Finally, latency is defined as the temporal delay between user input and the sensory feedback generated by the system in response to it. Input veracity depends solely on the quality of the input devices and is independent of the user's actions. Consider a Vicon motion capture system for an example. Most Vicon systems offer sub-millimeter accuracy, sub-millimeter precision, and latencies of a few milliseconds. Hence, these systems provide a high degree of input veracity. On the other hand, some tracking devices do not offer the same quality of input data, such as the Microsoft Kinect.

Embodiment

A key difference between today's graphical virtual environments and the text- and VR-based ones is the virtual body with which communicators make themselves present to themselves and others in virtual space. A defining feature of virtual worlds, avatars re-embodiment the communicator who has been disembodied through computer mediation. Their key affordances

are embodiment and presence [41]. Embodiment implies that the communicator can engage in practices of the body (e.g., sit, smile, and dress appropriately), and presence refers to the user's sense that she exists in a given setting, be it virtual or actual. Hoffman et al. [42] point out that presence is the essence of immersive technology. Thus, what we know about the world is embodied and all meaning derives from the experience of our bodies in the world [43]. As a communication or display system, the body emits information (intentionally and involuntarily; consciously and unconsciously; and verbally and non-verbally) that is perceived by other bodies. Thus emotions and mood are communicated through the body.

Even more fundamental than the body's information processing and communicative functions is its role in making the communicator present. It is thus synonymous with the conscious feeling of one's body existing in and being distinct from a prefigured, external world, which can be both real and virtual [44]. As communication devices, our bodies generally work in the background and are thus taken for granted. In computer-mediated communication our bodies seem to become irrelevant and only the presence of our minds matters. However, by emphasizing and problematizing the digital body, virtual worlds offer us an opportunity to become aware of and explore the role of the physical body in communication.

As virtual bodies, avatars promise users the affordances of real bodies and are thus touted as more expressive interfaces that increasingly approach face-to-face communication, even though facial expressions and gestures are contrived and under the user's control [45]. Vasalou et al. [46] highlight that this controlled expressivity creates opportunities for misrepresentation, as does the delay between constructing and actually displaying the avatar because this allows users to be exceptionally strategic in tailoring their avatar to convey a precise message. Galanxhi and Nah's [47] research provides support for misrepresentation in avatar-mediated communication. They found that the users of avatar-enabled chat were less anxious when they engaged in deceptive behavior than their counterparts in text-only chat.

Mapping Person's Real-life Expression to Avatar

Tracking has advantages of reduction in cognitive load and with the use of the marker-less motion capture devices such as the KinectHD, the user no longer requires invasive head and wand trackers to participate. High Fidelity avatars can be expensive in terms of equipment and rendering, however, the low fidelity avatar for the experiment permits real-time rendering within the game. Tracking theoretically allows for the transmission of spontaneous expressions, which Benford et al. cite as a particularly challenging problem in CVEs [48]. The degree to which involuntary expression is desirable is debatable in a medium that is valued for the control it gives users over the appearance and actions they convey to others. If the goal is to replicate each person's real movement, marker-less tracking is the most attractive solution and Immersive systems also reduce the problem of spatial mapping.

Minecraft

Minecraft has become a huge success worldwide since it was first launched in 2009. The basic core mechanics of Minecraft is to make and create in a fashion that replicates LEGO blocks, and it is this simplicity to build from imagination, that attracts a diverse audience. Another key aspect of the experience is the ease of adding modifications and textures to customize your game-play and increasing replay ability. Players use pixelated blocks to create detailed buildings and worlds as well as battle giant spiders and skeletons. Minecraft is compared to LEGO as a brilliant example of a generative toy – a stimulus to the imagination and a chance for people to express themselves creatively. It can be argued that this is precisely the result of the low-fidelity media. Thus, a realistic model of something in Minecraft can only reach a certain level of realism. Minecraft forcibly diminishes the gap between what the pros and the amateurs can accomplish – and in the process, makes things a lot more fun for the amateurs. If it were a more flexible and faithful visual medium it would come to be dominated by the same kinds of high fidelity graphics available in most modern games. The clunky forms and huge pixels give everybody the freedom to free explore and immerse themselves within the virtual world without aesthetic anxieties that come with better visuals.

Prioritizing Aspects of Avatar Fidelity

The previous subsections underlined the technical constraints on avatar fidelity in current CVEs. The tension between realism and real-time limits visual fidelity, and the tension between control and cognitive load poses difficulties for driving high-fidelity behaviors. Given these constraints, this subsection will contextualize the approach taken in this thesis by discussing the current need for trade-offs in developing expressive avatars. It will also present related research suggesting that avatars and agents can elicit social responses even given minimal fidelity. Fraser et al. have stated that “virtual environments” models, avatars, interfaces and so on are often designed with realism in mind’ [49]. The underlying assumption appears to be that more realistic environments and avatars should result in qualitatively better experiences in CVEs. Schroeder argues that this assumption needs empirical validation and lists a series of testable hypotheses, including one that directly concerns avatar fidelity: In relation to the realism of the representation of the other person, the more realistic the appearance of the other person, the higher the co-presence (or social presence) [50]. Benford and colleagues advocate incremental context-driven improvements to fidelity rather than an absolutist drive towards photorealism. Several authors share the alternative assumption that for communication purposes, behavioral fidelity is the higher priority. For instance, Sallnas argues that realistic appearance is secondary to the support of body positioning and pointing necessary in collaborative tasks [51]. Blascovich reasons that because we typically build digital IVEs, including interpersonal ones, using visual media, we tend to think of realism in terms of photographic realism. Although important, photographic realism does not equate with behavioral realism and is, in fact, less important [18]. In a separate paper with Swinth,

he adds that more important than photorealism, and perhaps even anthropomorphism, is an avatar’s behavioral realism. behavioral realism refers to the extent to which avatars and other objects in an virtual environment behave like their counterparts in the physical world [17]. The assumption that visual fidelity is secondary to behavioral fidelity is partly supported by lessons from animation. Disney animators translated films of actors body language and facial expression into simple line drawings and discovered it was possible to achieve effective emotional portrayals in visually simplistic characters, provided the movement was convincing [22]. Katsikitis and Innes’[52] study on line drawings of a smile illustrated that even a cartoon-like representation of an expression can be decoded accurately down to its five phases of development.

Studies on the transmission of nonverbal cues in mediated communication add further support to the argument favoring behavioral fidelity. Ehrlich, Schiano and Sheridan point out that the same bandwidth restrictions constraining CVEs also apply to VMC [27]. They suggest that the standard approach of preserving spatial and color resolution at the expense of temporal degradation is counterproductive. Their experimental findings indicate that preserving motion information is critical to the recognition of facial expression and may compensate for significant losses in image resolution.

	Facial recognition (appearance)	Affect recognition (behaviour)
Associated with	Image quality (spatial and colour resolution)	Visual dynamics (temporal resolution)
Effect of degradation	Robust across visual degradation	Sensitive to temporal degradation

Figure 2.4: Impact of degradation on facial and affect recognition

Considering that the transmission of nonverbal cues can be severely affected by temporal delays and inconsistencies, they suggest that if a bandwidth trade-off is required, one should consider preserving high-fidelity motion information at the expense of image realism, not the other way around [27]. In a separate study on facial affect recognition, Schiano, Ehrlich and Krisnawan compared a low-fidelity robot enacting the six basic emotions with video of human actors enacting the same emotions [53]. Though scores for the robot were lower, the expressions were decoded in a pattern that closely followed the human faces. This further supports the argument prioritizing behavior over accurate appearance in the transmission of nonverbal cues. Bente and Kramer [54] describe a related study on person perception, this time comparing silent video clips of dyadic interactions between human actors with equivalent clips of identically animated agents. Their findings indicate a remarkable correspondence in responses to the video and agent conditions, despite the lower-fidelity appearance of the agents. In summary, technical limitations have forced the need to set priorities in avatar design. In the words of Heeter, Faced with technological limitations which prevent being able to simultaneously simulate all aspects of human perception, the alchemy of presence in VR is in part a science of trade-offs. Which elements are most critical to the experience

of presence? When forced to choose between responsiveness to motion and resolution of images, developers choose responsiveness as the more important factor, based on their own experiences and observations of others [55]. These findings from different media experiences partially support the notion that behavioral fidelity may be more pressing than visual fidelity. This is supported by Tromp et al.'s experiment where higher-realism avatars appeared to raise higher expectations for human like behaviors, suggesting that appearance should remain minimal until behavior is sufficiently sophisticated to satisfy expectations [56].

Exploring the Impact of Minimal Fidelity

The argument for exploring the lower boundaries of fidelity is not born exclusively out of technical necessity. Reeves and Nass document a series of studies suggesting that people respond to media as social actors, and tend to anthropomorphise even the simplest of text-based interfaces [57]. This theory of the medium as social actor is of direct interest to avatar design because it suggests that minimal cues can elicit social responses. Biocca, Harms and Burgoon maintain that “Unlike the physical environment, social communication in virtual environments might be built upon minimal or constrained social cues. Animated characters and even the computer interface itself can generate strong automatic social responses from minimal social cues. Social responses to computer characters for example, are generated even though the user is quite aware that the computer is not an emotional or social agent but a machine” [58]. They later state that “a fundamental question in mediated social presence is why humans respond automatically and socially to virtual representations of other beings” . For Biocca and colleagues, the automatic interpretation of humanoid forms and nonverbal behavior can lead people to attribute a degree of sentience to virtual humans. This tension between automatic social responses and the rational knowledge that virtual humans are artificial entities represents a fundamental and engaging issue that has been addressed in a selection of studies in different research institutions. Virtual humans present promising avenues for social research because they enable the controlled manipulation of specific visual and behavioral variables. However, before they can be employed for social research the underlying premise of whether they elicit comparable social responses to real humans needs to be tested. Bente and Kramer’s study was designed with this goal in mind. Based on their findings they conclude that computer animations can indeed elicit realistic socio-emotional responses. The same underlying question was addressed by Pertaub, Slater and Barker in a series of studies on fear of public speaking, a common and debilitating form of social phobia [59]. The motivation was to explore whether VEs could in principle be useful for the treatment of phobics; before any exposure therapy treatment programs could be developed, it was first necessary to assess whether virtual audiences could evoke the required anxiety responses. [60] also suggest that limited visual feedback from virtual humans can affect social responses even in the absence of two-way verbal exchange, and in spite of a rational awareness that these are artificial entities.

2.2.5 Presence & Co-presence

Presence is a multi-faceted phenomenon [61] whose conceptualization has evolved in part because of technological advancement. As virtual environments became more social and were used for different applications, the conceptual infrastructure of presence grew more elaborate. The objective of introducing various forms of presence is to develop an appreciation for the multi-faceted nature of presence and to highlight their relevance to different technological features. It is the illusion of being in a distant place, that is, being there user's sense of actually flying a plane by interacting with the instruments, even though he is sitting at a computer in an office.

Co-presence Illusion of having access to a remote or distant other that shares the same distant place, that is, being there with others user's sense of actually shaking a customer's hand at the start of a meeting, even though both the user and the customer are in avatar form and the meeting space is virtual. If tele-presence focuses on being there (in a space), then co-presence is the sense of being in a shared virtual setting with remote others [62]. As such, co-presence is conceptualized at the intersection between tele-presence and social presence. It is the virtual equivalent to Goffman's definition of co-presence as collocation of embodied not merely imagined others that become available and accessible to each other. This form of presence is made possible by shared virtual environments. User's also experience **Social presence** in shared virtual environments, which is the illusion of access to a remote or distant other, that is, being with user's sense of knowing another person (i.e., his actual personality and intentions), even though this person is encountered only in virtual space.

Internal and external determinants

Discussions of presence also target the question of how the sense of presence is created and destroyed. IJsselsteijn et al. argue that although research into presence is still at an early stage of development, there is a consensus that presence has multiple determinants [63]. Freeman, IJsselsteijn and colleagues list four classes of presence determinants identified in the literature [64]. The first two are classified as media form variables, relating to properties of the system.

1. The extent and fidelity of sensory information
2. The match between the sensors and the display
3. Content factors
4. User characteristics

Conceptualisation of Presence	Summary of conceptualisation
Social richness	Draws from Short, Williams and Christie's notion that media richer in informative cues enhance social presence
Realism	The degree to which a medium can convey accurate portrayals of objects, events and people. Divided into <i>social realism</i> (the extent to which what is portrayed would be plausible in real life), and <i>perceptual realism</i> (the extent to which events appear realistic).
Transportation	The extent to which a person is 'transported' to another place (e.g. through fiction), the extent to which another place is 'transported' to the person's physical environment (e.g. through film), and the extent to which people are transported to a 'shared space' through mediated interaction.
Immersion	The extent to which a person is perceptually or psychologically immersed as a result of substituting real-world stimuli with stimuli from the medium.
Social actor within the medium	The extent to which people respond to social cues presented by people encountered within the medium, even when it is not appropriate to do so (e.g. responding to television characters).
Medium as social actor	The extent to which the medium itself (e.g. a computer program) is responded to as a social entity.

Figure 2.5: Six conceptualizations of presence

Slater and Steed also propose a number of factors that undermine presence, causing breaks in presence (BIPs) or 'transitions to real' where people's attentional focus is suddenly drawn out of the VE to their physical surroundings [65]. These factors can be either external (sensory information from the physical world intruding or contradicting the VE), or internal (internal inconsistencies in the VE). Given the range of media factors that may impact on the sense of presence, some authors have made a point of conceptually distinguishing between the presence experience itself and its possible determinants.

Immersion

The term immersion is used to describe the extent to which objective characteristics of the technology can provide a surrounding environment by replacing sensory stimuli from the physical world [66][67]. Slater [66] is one of the most vocal proponents of a theoretical perspective of presence as a human response to sensory immersion. Sensory immersion is a technology's ability to create a convincing, immersive environment with which the user can interact. As a technical capability, sensory immersion is thus defined as an objective and quantifiable property of the technology [68]. Although Slater acknowledges that presence and the user's sensory immersion are probably strongly related empirically, he argues that they are theoretically distinct. Presence is the sense of 'being there' that is created when the technology's simulated sensory data and the user's perceptual processing combine to produce a coherent place, in which the user can locate herself and interact with spaces, people and

things. As such, presence is a matter of form. In contrast, psychological immersion, i.e., involvement and emotional engagement, is a matter of content. For instance, users might have a sense of actually being in a virtual concert hall (i.e., presence), but simultaneously experience boredom because the music fails to engage them (i.e., no involvement).

In causal terms, the basic model of presence's antecedents looks something like this: technological features → immersion → realism/sensory fidelity → presence. The technological factors identified in the literature are largely captured in [66] definition of immersion; however, [69] add that having an image to represent oneself in the virtual space produces a greater sense of presence than when one is invisible. Realism or sensory fidelity is the degree to which displays of spatial, auditory and haptic (touch-related) information in the virtual worlds is similar to that in the actual world [70].

With regard to the consequences of presence, a coherent set of dependent variables is difficult to discern. For instance, [71] suggests psychological immersion, i.e., involvement and emotional engagement, as a consequence of presence.

Tele-Immersion

The aim of the 3D tele-immersion is to enhance the experience of geographically distributed interaction in a virtual environment by facilitating digital embodiment of the users through 3D capturing technology. The 3D data, either in a form of 3D video stream, point cloud, or mesh, are transmitted to the remote locations and combined with application data for rendering and interaction. Users can experience geographically distributed 3D tele-immersion through various interaction modes, some of which are listed here.

First-Person Mode: The user interacts with the environment in the first-person perspective, while the remote users see his/her 3D avatar at the corresponding position.

Third-Person Mode: The user observes the scene from a fixed viewpoint relative to his avatar to interact with the data and other users.

Mirror Mode: The user observes a mirrored image of his and remote avatars which can be applied for instructing physical activities. In the Third-Person Mode, the user observes the scene from a third-person view (usually fixed) while interacting with the environment. In this case, it is not possible to preserve the direct connection between the 3D geometry of the real space (i.e., user pointing at objects perceived on the display) and the virtual environment (i.e., avatar pointing at objects). This mode can be utilized when observing the virtual environment on a 2D display where there is a disconnect between the physical space and displayed 3D data. The rendering of the avatar thus provides spatial cues for pointing and interacting with objects in the scene. In the Mirror Mode, the screen represents a virtual mirror with the avatar mirroring user's movements in the physical space. For remote interaction, the avatar of the remote user is projected in such a way as if both users were sharing the same physical space while their movements are also mirrored. This mode is applicable for instructing and teaching movement patterns, such as in rehabilitation, dance,

or fitness training.

In addition to the aforementioned interaction modes where each user has their own avatars occupying the virtual space, users can adopt another person’s viewpoint. Such capability is useful in educational and training scenarios where multiple users may follow an instructor who wishes to point out various features in the observed data.

Another important issue to consider in the real-time interaction over the network is the latencies and jitter in the transmission of video and tracking data [72]. The latency is described as the lag between the time instances when data are sent and received on the other end. Different strategies can be employed to compensate for longer latencies as long as the variability is small. One example includes coordinated interaction between the remote users where at each time instance one of the users is the leader while the others have a role of a follower. On the other hand, the network jitter, which refers to the variability of the latencies between the receiving packets, can cause significant disruption in the remote interaction. The network jitter can be influenced through various quality of service mechanisms that re-route the packets in complex networks. This thesis study uses in-game quantitative metrics to report performance statistics such as jitter, network latency, frame draw rate.

2.2.6 Measurement Approaches

A number of measurement approaches have been proposed, which can be classified according to the time measurement is taken (during or after the experience), and the type of data gathered (subjective or objective). Presence is frequently referred to as a subjective experience [63]; unsurprisingly, presence research has relied extensively on subjective reporting, most commonly on the use of post-experiment questionnaires designed to evaluate people’s sense of ‘being there’ in the mediated environment. Subjective questionnaire measures can combine different approaches [73], including semantic differential techniques using scales anchored to opposing descriptors, as in [74]. Alternatively, Likert scales have been used to measure the degree of agreement or disagreement with a set of statements, as in [75].

	During experience	Post-experience
Subjective	Hand-held slider	Questionnaires
	Breaks in presence (BIPs)	Interviews and Focus groups
Objective	Psychophysiological monitoring	
	Observation of behaviour	

Figure 2.6: Proposed measurement approaches

Task Performance

Sheridan [76] and Hendrix and Barfield [77] suggest objective measures of presence based on task performance in the virtual environment. The problem with this method is that task performance may not necessarily correlate positively with presence, and that factors other than presence might influence task performance. One must find a specific task and show that presence correlates significantly and positively with the performance of that task.

Behavioral Presence

Another way to assess presence in a virtual environment is to measure behavioral presence. Behavioral presence cannot be evaluated using simple questionnaires, and requires a more complex method based on observing the behavior of participants in the real world, reacting to different stimuli in the virtual environment. Held and Durlach [78] suggest a measure of presence based on the ability of the environment to produce a startle response to unexpected stimuli. For example, whether users duck, blink or carry out other involuntary movements in response to threatening events. Slater et al. [79] measure behavioral presence by observing the reactions of the subjects to danger, such as a virtual cliff, or objects thrown towards the participants' head. The problem with behavioral measures is that they may be too complex to clearly identify and measure with clarity. Also, startle-based measurements may only be measuring isolated samples rather than measuring the overall presence created by the environment.

Questionnaires

Two presence questionnaires have received significant attention in the literature: the Witmer and Singer presence questionnaire (PQ) [74], and the Slater-Usuh-Steed questionnaire (SUS) [80]. Witmer and Singer's PQ was developed to elicit subjective presence responses to experiences in IVEs, with a particular focus on investigating the impact of four possible contributing factors to presence: control, sensory factors, distraction and realism. The problem, as discussed by Slater [71], is that the questionnaire confounds measures of individual differences and properties of the VE, making it impossible to separate them. In addition, although they clearly define presence as the subjective experience of being in one place or environment, even when one is physically situated in another [74], their questionnaire contains no items that directly measure this construct. Slater, Usuh and Steed's SUS questionnaire is designed to measure the sense of being there in the VE, as well as two additional aspects central to Slater's definition of presence: the extent to which the VE is experienced as the dominant reality, and the sense of having visited a place as opposed to having simply viewed computer-generated images. This sense of place is particularly central to the experience of presence in VEs. Usuh et al. report on a study designed to test the ability of both the PQ and the SUS questionnaires to distinguish between subjective presence responses to a real-

world environment and its corresponding immersive virtual model [81]. They report that PQ showed no difference between the real and virtual environments, while SUS showed a statistically significant difference. A tradeoff is involved in using post-experience questionnaires. One significant limitation is that subjective reporting only captures post-hoc rationalizations of the experience. This is problematic not only because of demand characteristics [82], but also because of the potential pitfalls of inaccurate recall [83]. Freeman et al. have pointed out that post-test presence ratings are unstable, particularly in the case of naive subjects who lack a lexicon for understanding and describing presence. Slater has similarly argued for a move away from questionnaires in the measurement of presence [84]. In a study, a questionnaire referring to a fictitious construct called colorfulness of an experience was administered to 74 respondents. Reported findings indicate an association between colorfulness and a number of equally arbitrary variables including how late respondents had woken up that day. Slater cautions that questionnaire responses can yield statistically significant but ultimately meaningless results because rather than reflecting how respondents would ordinarily describe their experience, the arbitrary response measure is called into being by the questionnaire.

Slater and Steed propose a “breaks in presence” (BIPs) approach, where participants are asked to signal each time they transition to a state of awareness of their physical surroundings [65]. This method presumes a binary possibility whereby people are either present in the VE or in the physical environment. By the authors own admission, this method fails to capture presence in a third imaginal location. Nevertheless, its advantage is that Slater and Steed’s findings suggest a strong positive correlation between questionnaire-based presence and presence as estimated from the number of BIPs reported. The significant drawback of both the BIPs and slider approaches is their intrusiveness; by requiring participants to continually report on their experience, these methods introduce additional cognitive load and also potentially interfere with the phenomenon of interest, the presence experience itself. Objective approaches have been investigated to address the limitations of both continuous and post-test subjective ratings. Their advantage is that they do not require conscious attention or control and are therefore less cognitively intrusive.

Given the limitations of both subjective and objective measurement approaches, Freeman et al. have proposed the parallel exploration of objective and refined subjective measurement approaches. In particular they propose the use of focus groups to derive improved terminology for rating scales. An aggregate approach combining various measures may be more effective, particularly considering the potentially complex structure of presence. As Slater, Usoh and Steed suggest, presence may consist of two levels: the surface level, which can be consciously articulated, and a deeper level that influences behavior in a basic way [80] and may be better captured by objective means.

Additionally, Chertoff and colleagues presented a questionnaire developed to measure “holistic virtual environment experiences” [85]. The development of their questionnaire was guided by the five dimensions of experiential design: affective (emotion), cognitive (engagement), sensory (immersion), active (personal connection...to an experience), and relational (social)

[86]. The questionnare used as part of this study, also breaks co-presence down into subcomponents.

2.3 Related Work

2.3.1 Operacraft



Figure 2.7: Inspiring K-12 students to create stories through a live production of Operacraft

The OPERAcraft platform [87], a precursor to Cinemacraft was envisioned as an environment to aid creativity and thinking skills and better self-expression, with particular focus on the K-12 education opportunities. It was built as an arts+technology+education platform where students could write a story and libretto, build a virtual set, costumes or virtual character skins, and ultimately control the characters within the virtual setting in a live performance accompanied by live singers and musicians. Many of these affordances are inherent to Minecraft platform users can easily sculpt the landscape, interact with it, and change their own appearance. Others were added as part of the reverse engineering effort, resulting in a mod that is deeply integrated into Minecraft's core. These include character lip syncing based on the singer's input processed through the Pd-L2Ork [88] and forwarded to a FUDI-based parser via a UDP socket embedded inside reverse-engineered version of

Minecraft, audience subtitles and stage cues only visible to the actors, ability to change between discrete arm positions and interpolate between them to provide rudimentary body language, and near-instantaneous scene changes through coordinated character teleportations and scene cross-fades. In an ongoing pursuit of building a compelling real-time machinima production platform, the second generation of OPERAcraft introduced in the fall 2015 as part of the second opera production offers additional affordances, including multiple camera views and cameras that are only visible to the actors, invisible bystanders, as well as stability improvements and optimization that allowed the mod to scale beyond the original limit of five actors.

2.3.2 Community Support

The previous version of the platform was showcased as part of three high-profile exhibitions. The team has used such opportunities to iteratively improve upon and refine the design, as informed by the outcomes demonstrations and real world user feedback. In particular, the prototype was showcased at Virginia Tech's official exhibit at South by Southwest 2016 [89], and as part of ICAT day showcase at the Moss Arts Center in Virginia Tech [90]. More recently, Cinemacraft is also displayed at the Science Museum of South-west Virginia [91]. As a result of the strong response to and interest in the tool, it has also been selected to be integrated in the Virginia Tech Visitor Center. Both exhibits are scheduled to open in the winter of 2017. There has also been strong positive feedback from the Minecraft gaming community and machinima enthusiasts who have expressed particular interest in the realistic posture of the avatars to express intent. This further supports the notions the body movements have a key significance in communication and non-verbal cues.

2.3.3 Findr - Immersion and User Engagement

This study also borrows from a closely related study on Immersion and Engagement in a VR Game, which leverages the Mirrorworlds [92] project to compare the levels of user engagement, task performance and distance travelled across a desktop and a Head Mounted Display platform. An interactive search game was built using the Unity game engine for both a regular desktop version and the Oculus Rift HMD and explores exocentric vs endocentric approaches, the level of confusion factor, and the effect of the virtual avatar representations. The situational awareness of the user during gameplay was also touched upon and the application was evaluated using a total of 18 subjects and the data was collected based on presence questionnaires from the user following the experiment. A notable takeaway from the study is that an increased level of scene realism does not directly correspond to increased user experience and in game task performance. On the other hand, the subjective metrics pointed toward a greater level of engagement on the HMD setup, the objective metrics show superior level of immersion and engagement.

Chapter 3

System Design

One of the main objectives of Cinemacraft is to provide a compelling emotional delivery of storytelling within the context of arts. The system uses live performance capture through the integration of a Microsoft Kinect HD C# application to provide a more immersive and expressive embodied experience in a virtual world through both kinematic data and facial expressions. In many ways it is designed to supplant keyboard controlled arm expressions by providing full body immersion to the extent allowed by the simple skeletal structure of the avatars that lack hands, elbows, and knees, and further enhances expressiveness through facial tracking. Based on previous user tests and audience feedback, the avatar remains compelling despite the minimal character design due to the reported feeling of sentience offered by the user's real-life body motion and facial expressions. As a result, the avatar can show a dynamic range of emotional reactions and responses. In cinematic terms, the avatar no longer appears to be merely acting. Rather, it is the actor who is responding to their projection in and the situational awareness of the virtual environment. Such spontaneous reactions like squinting against a sudden bright light help to humanize characters and make them more compelling than current game characters that seem shallow and with whom we have a hard time forming compelling, coherent relationships [93].

3.0.1 Migration to Minetest

While previous versions of the platform were built on modified builds of Minecraft, the latest version (used in this study) is based on Minetest, an open-source alternative to Minecraft. This switch was made primarily in the interest of future enhancements and openly accessible code.

Minetest also offers several advantages over newer versions of Minecraft. Minetest is more aesthetically similar to the older and more simplistic versions of Minecraft. This works in our favor as it helps sidestep the uncanny valley through low fidelity avatars and representations

while still delivering higher fidelity embodied interactions. The Minecraft code was also not readily modifiable and requires a community-driven effort to decompile JAVA run-time into a human-readable API. As such its forward compatibility is at best cumbersome. The legal implications of modding Minecraft are also not entirely clear, suggesting Microsoft by default owns all modded code, and as a result the distribution of the ensuing deeply integrated mod is difficult if not impossible. Minetest on the other hand, is a strikingly similar yet highly modifiable voxel game engine that covers a majority of Minecraft features. The in-game client server interactions are also better handled in Minetest with modded servers sending textures and other required resources to the clients, unlike Minecraft that may require resources to be downloaded separately. When compared to Minecraft, Minetest has a huge potential vertical size of the virtual world, the total maximum height of the world is about 60 000 blocks (30,000 up, 30,000 down), which allows servers to build sprawling and grandiose structures a few thousand blocks in size (e.g. steep mountain peak 2000 blocks in height). The new Minetest game client for Cinemacraft communicates directly with the external Kinect HD C# motion capture application and retains backward compatibility with vanilla game version, which is the unmodified Minetest game. The game executable for Cinemacraft can function as either a new server or client with backward compatibility.

Minetest Game Design

There are two major parts to the system, the first being a core based on the Irrlicht game engine [94] written in C++. Most of the modifications for Cinemacraft are built on the core, which was the original network multiplayer release of Minetest (Vanilla version). The core comprises of the following components:

1. The Map: Voxel storage + lighting + rendering
2. The Environment: Contains the map and the players, handles the simulation of the world. The environment also controls the first person camera views that had to be modified according the player motions detected from the Kinect.
3. The Client Server logic for the game comprises of all the active server and client objects, network packet handlers and player updates.
4. The main loop: Invokes the client, the server, the environment and the rendering players and GUI using the Irrlicht engine.
5. Wrappers for OS-dependent processes and utilities.

The second major component is the modding API written in Lua that exposes useful core engine functions. While the initial game design had left a lot of implementation open to the API modifications, most of the changes in the game have been made to the core engine due to better game performance. This is because the game architecture currently only supports

a single thread for the Lua API and it briefly halts the execution of the server thread, connection threads and multiple game threads while it completes.

Minetest Protocol

The Minetest protocol is a small layer built on top of the UDP protocol and comprises of four packet types. All packets include a header and all numbers are big-endian. The packet types are split as follows:

1. CONTROL(data) - unreliable control packet
2. ORIGINAL(data) - unreliable small data
3. SPLIT(piece of data) - unreliable piece of large data
4. RELIABLE(CONTROL(data)) - reliable control packet
5. RELIABLE(ORIGINAL(data)) - reliable small data
6. RELIABLE(SPLIT(piece of data)) - reliable piece of large data

In order for the Kinect HD C# application to interface with the Minetest client, the core Minetest engine was retrofitted with a FUDI-compliant protocol [95]. A customized version of the in-game protocol was defined for the connected Kinect HD, which is recognized as a new client by the Minetest server. The Kinect HD C# application first initiates a connection with the Cinemacraft Minetest client using control packets and following which reliable small data packets start flowing in. The following figure illustrates the new protocol structure between the Kinect HD C# application and the Minetest game client.

Section	Protocol ID	Server ID	Channel	Control type	Server Init code	Client ID	Packet Seq num	Data
Size (bytes)	4	2	1	1	1	1	4	96

Figure 3.1: Protocol structure between the Kinect C# application and Minetest

3.1 Architecture

The avatar and scene rendering are performed through the Minetest client on two networked PCs with high-end graphics cards. At each location, a display peripheral can be inserted in the set to project the screen for a larger field of view and immersive experience. Each set is also equipped with a Microsoft Kinect HD device and microphone to capture the actor. The user's motions and expressions are captured in real time using the custom Kinect HD

C# application for simultaneous facial and body tracking. The Kinect HD C# application and Cinemacraft Minetest game are packaged as independent executable files.

A major consideration for our setup was accessibility. For this reason, although earlier prototypes relied on two first generation Kinect HD devices, due to their observed inability to simultaneously do body and face tracking, the current implementation relies on just one second generation Kinect HD device responsible both for kinematic and facial tracking. The actor is tracked using simultaneous body and face motion capture which is calibrated and optimized to transmit skeleton information to each computer. The microphone data is captured through a regular audio chat application while input audio data for sensory fusion is captured through a custom Pd-L2Ork patch which sends UDP packets to the Kinect HD C# application. The Minetest game is capable of parsing remote FUDI-compliant protocol [95] messages. Its simpler version is already found in OPERAcraft and previous versions of Cinemacraft where it was used to coordinate various aspects of the production, including switching camera angles, lip syncing as detected by the singers micro-phones, subtitles, and stage cues. As a result these can be handled remotely through multiple distributed Pd-L2Ork clients [88]. The ensuing UDP based protocol can be seen as a simplified counterpart to the Open Sound Control (OSC) [96]. All communication is relayed through UDP packets between the microphone and Kinect HD C# application and Minetest Clients. On the performer's monitor, participants are able to see the reactions of the other user as well as their own avatars in the virtual world during the interaction, allowing them to monitor how their actions affect both the physical and the virtual world. Because the performance is driven by real-time motion data, the virtual interaction must be synchronized on each performer's monitor, as well as on the server screen. The virtual world in this case is a selection of Minetest game maps that both participants can choose from.

A media layer to manipulate the interactions between audience members, the server and performers can also be readily integrated into the system. Further intelligence can also be incorporated to the motion capture capabilities of the Kinect HD C# application through the addition of sensory fusion layers. This helps to keep all actors involved in the collaborative experience within the virtual world and how it behaves and changes. Due to the need for close to low latency performance, our system runs in real time at a speed of 60FPS on two PCs each equipped with high-end graphics cards, a Kinect HD device + C# application, and high-speed internet. By careful integration and system optimization there is no delay between the remote actor avatar, on-screen user avatar and live actions in the real world. Cinemacraft, handles positions through real-time processing by effectively updating the avatar's motion in game. This allows support for multiple clients that communicate with other users along with out-of-box multiplayer support with chat and other core functionality.

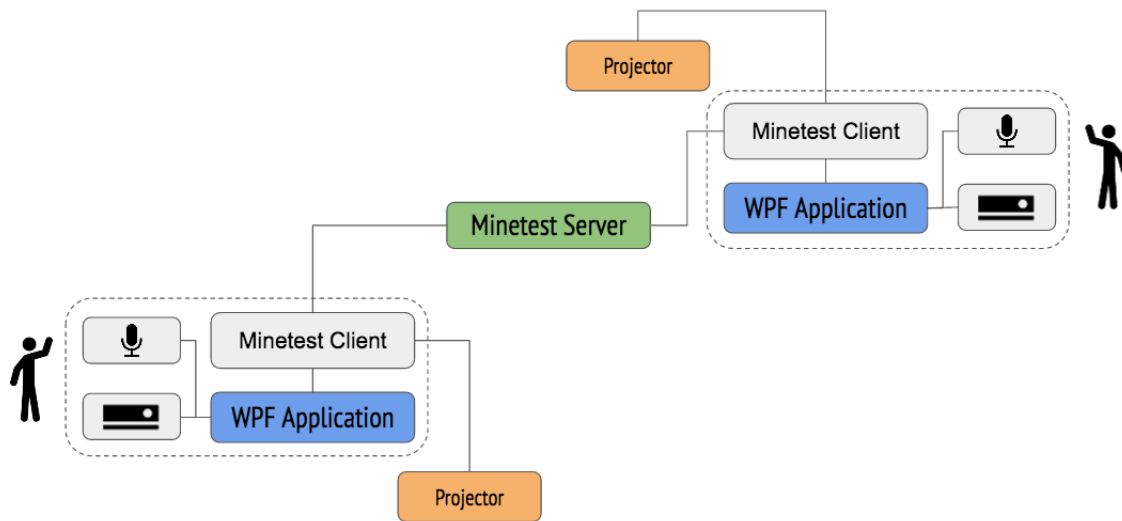


Figure 3.2: System architecture for the Cinemacraft

3.1.1 Functional Modifications

Packet Filtering

A new packet filtering logic was required to check for useful reliable packets arriving from the Kinect HD C# application and process them without delay. Packets are delivered in the order they are sent in and delivery of all reliable packets in the game is forced by acknowledgements as per the protocol. Reliable packets are stored in buffers at the receiving and transmitting ends and the buffer contents are then recursively processed as packets. Additional filtering for Kinect HD C# application packets was written to check for the last complete packet in the socket according to sequence number and packet size, while storing useful packets in the buffer for processing and discarding the incomplete incoming Kinect HD C# application data packets and flushing the socket periodically. This new filtering works alongside the regular in-game client server packet processing.

Mapping of Movement to Game

Minetest adds Irrlicht game engine nodes to render the scene and update player position and speed. This is performed in the environment generation loop. All key-frame animation loops for player limb and head motions are handled in an inner client-server loop. While the initial implementation of the game used the inner loop to update player state information in the game, we also wanted to retain the capability of using the keyboard+mouse to smoothly move the player while the still capturing player limb movements and only overriding the position data. Therefore, the latest game version uses the in-game pipeline and only changes

the position difference with respect to the center reference point as per the Kinect HD C# application and the in-game coordinates.

Kinect User Data	Cinematiccraft Avatar
Client Name	Designated Actor
Arm + leg angles, Arm + leg vectors X & Z	Rotation angles, Orthogonal positions
Shoulder vectors in X & Z	Shoulder position
Torso X, Y and Z coordinates	Body position in 3D space
Vertical angle of rotation	Body Yaw angle
Head Pitch, Yaw and Roll	Head Pitch, Yaw and Roll
Lips, Eye brow selected points	Mouth, eyebrow movement

Table 3.1: Kinect data for Cinematiccraft Avatar Mapping

The Vanilla game version handled in-game rotations for player nodes using Euler angles which create challenges in accurate replication of the limb movements in 3-D space. Therefore, the player limb motions were updated to use Quaternions that map limb motions positions and angles along orthogonal axes. This was implemented through additional Irrlicht game engine function calls. Further modifications were made to all server and client active classes to directly manipulate bone positions using the quaternion data.

3.1.2 Modes of Interaction

Cinematiccraft offers different modes of embodied interaction captured by Kinect HD, namely mirrored, upper torso, full body and full body+sensory fusion. The upper torso mode allows users to act and gesticulate to other players to complement their speech and chat messages and thereby increase the effectiveness of the conversations, while still being able to navigate the expansive landscape outside the range afforded by the area monitored by the Kinect HD using more conventional controls (e.g. keyboard). In a more hybrid setting, a separate user can control the avatar while an operatic singer, for instance, provides only upper body language. Similarly, the mirroring mode has been added to explore illusory experience interactions with the avatar, most notably through the Mirrorworlds research project focusing on the study of integration of physical and virtual mirrored presence [92]. The experimental tasks in the user study utilized these different modes to vary the interaction fidelity between the pairs of participants and affect the avatar's behavior.

These interaction modes also provide an opportunity to draw parity between different approaches to machinima and open new exciting possibilities for sensory fusion, with the introduction of Head-Mounted Displays (HMD) devices like Oculus [97] and Leap Motion [98] and even Haptic Feedback devices [99][100]. Cinematiccraft inherits a battery of OPERAcraft's cinematic tools, empowering users to explore methods of machinima production, including live theatrical play and cinematic production. The virtual audience feature, that enables

audience members to freely roam the scene, or the ensuing world in which the story-telling takes place, offers new research opportunities in the study of perception of story telling, drama, and empathy as a function of vantage point.

Its testing focused primarily on assessing the perceived visual fidelity of mirroring user's interaction. For this test we utilized the mirrored mode where extremity coordinates and directions needed to be reversed or mapped to another plane in order to achieve the desired mirroring and movement.

3.1.3 Sensory Fusion

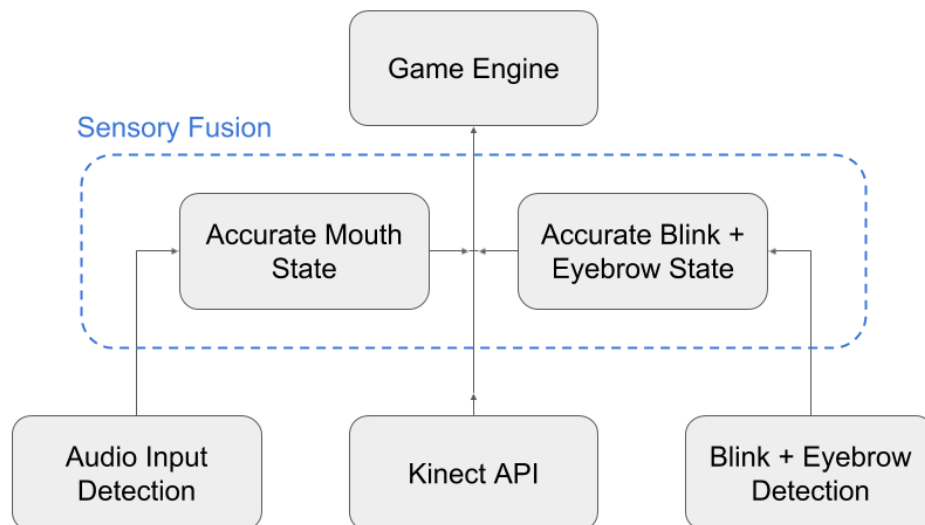


Figure 3.3: Sensory Fusion Design

The emphasis on ease of use and reliance only a single Kinect HD device requires our implementation to essentially stretch the limits of the current Kinect HD API. Despite its improved resolution over the first generation, Kinect HD is still best suited for face tracking in close proximity which limits its ability to track body. In turn, our implementation offers accurate simultaneous full body and facial tracking. We utilize a module of audio input data to more accurately capture the users' mouth states when they make a sound. Here, the sensory fusion allowed us to use voice detection to combine the performer's audio with the facial tracking data and there by improve detection of minor gestures and expressions which may not be otherwise captured due the technical limitations of the two distinct approaches to monitoring user's input. For instance, doing so enabled us to animate mouth motion through captured audio that exceeds the resolution of 60 frames per second, as well as audio centric outliers, such as the cartoon-like quivering of lips in a sung operatic melisma.

We have envisioned a platform with parallel pipelines of Audio Inputs, Kinect API and Computer Vision optimization and learning for improving facial Expressions, with all three working together to further refine the platforms capabilities through sensory fusion.

Implementation

The sensory fusion layer uses a simple switch and thresholds to allow audio data to take precedence over incoming facial motion capture data in the event of audio input is successfully detected. A Pd-L2Ork patch is used to capture the user input sounds through a microphone and translate them into numeric values that were sent to the Kinect HD C# application. The Kinect HD C# application uses a separate socket and thread to read and parse incoming data into a compound packet to be forwarded to the Minetest client. Further, specific sounds and pronunciations can be mapped to unique facial expressions to further enhance the realism of the avatar detection. Therefore, when a threshold for a certain sound is crossed, a certain numeric value is generated through the Pd-L2Ork patch for that loudness and enunciation. This numeric value is then sent to the Kinect HD C# application which maps the appropriate eyebrow state and mouth state using a face matrix and sends the packet to the game. The data is then parsed to check whether the fusion layer needs to be activated, following which the correct mouth avatar texture is loaded. Default values are used as a fail-safe in case a new unknown value for the face matrix is generated that is not found in the game textures. The new version of the Kinect HD C# application also supports a much larger range of facial expressions which are mapped to corresponding textures that change according to the user's expression in the real world.

Additional CV Layer

We have identified problems with Kinect's machine learned library of postures and facial expressions that have resulted in a prevalent number of false positives pertaining to eye winks, eyebrow movement, and eyeglass detection. While we had explored further enhancing face detection with infrared video feed inherent to Kinect HD, the low reflectivity of eye pupils makes the task extremely difficult. These challenges can be addressed through an additional layer of sensory fusion to run a low-latency computer vision algorithm on the facial capture output of the Kinect HD with improved tracking of eye and eyebrow states.

Chapter 4

Experiment

This chapter will focus on user experiments and the choice of methods used to address the research questions. The experiments focused on distinct aspects of fidelity, but shared many similarities in terms of data collection and analysis. Section 4.1 will focus on methods of data gathering. Section 4.2 will detail the strategies used to design and pilot the experiments, as well as the experimental procedures. Section 4.3 will focus on methods of data analysis and results. The data was in the form of quantitative questionnaire data. Methods of analysis for each type of data will be described in turn. The chapter will conclude with tables of data collected in each experiment, and the corresponding results following analysis.

4.1 Data Collection

This section will cover questionnaire responses, the experimental variables and expectation from the data.

4.1.1 Defining the research goals and expectations

All tasks as part of the experiment had a common theme, namely the visual impact of avatar fidelity on the interaction. The impact of behavior fidelity was explored using different response variables through increasing the level of interaction fidelity. The general expectation was that greater the level of interaction fidelity, the more the virtual humans would be seen to contribute to the experience and the more they would elicit presence and co-presence responses from participants. However, one challenge in this area of research is that, just as there exist many questions about the impact of virtual humans, so are there open questions about what constitutes a presence and co-presence response. The first step in designing the experiments was therefore to define the specific research questions in terms of the exact

independent and dependent variables of interest. Questionnaire items used in this study were based on previously published research, supplemented my items developed during the course of the research through the process of piloting.

4.1.2 Defining the independent and dependent variables

The broad purpose of the research was to investigate the impact of avatar fidelity on a selection of responses. Drawing from our research problems, we wish to test a sense of presence (personal presence and co-presence) in the CVE is created by embodying the participants in the virtual environment by means of virtual representations. The hypothesis to test are:

1. Avatars with higher embodied interaction fidelity will enhance the sense of presence and co-presence in a CVE.
2. Sensory fusion for more accurate facial expressions would yield the highest presence and co-presence scores.

Synchronized movements between the user and their avatar have been shown to have a positive effect on both the users cognitive ability and feeling of agency over the virtual avatar [101]. Additionally, the ownership of another person's body, or the "embodiment illusion" can be induced via multi-sensory correlation [102]. However, it's important to investigate such anatomical control systems in more depth, particularly the potential link between motion capture functionalities and embodiment, in this case, in first person. Studies have found that participants' upper body movement being mirrored alone was a strong tool to provoke the illusion of both agency and body ownership towards the virtual body even without full body tracking [103]o test this, we construct response variables from n questionnaire items, each on a 1 to 7 scale with the score adjusted for analysis so that the higher score represented a higher response. The items for each response are detailed in Appendix A and B.

1. **Presence score, P** measures the degree of personal presence experienced by the participant using Slater's presence questionnaire.
2. **Co-presence score, CO-P** measures the co-presence experienced by the user. The Co-presence score is further divided into contributing components adapted from the questionnaire.
3. **The immersive tendencies score, IT** measures the tendencies of individuals to become involved and immersed in the experience. This variable is measured using Witmer and Singer's immersive tendencies questionnaire.

Independent variables

The type of user interaction is varied to measure the expected increase in presence and co-presence:

1. Interaction: Keyboard, Mouse + Inter-user communication: Audio chat
2. Interaction: Kinect for face and upper torso, Keyboard + Inter-user communication: Audio chat
3. Interaction: Full face and body motion + Inter-user communication: Audio chat
4. Interaction: Full face and body motion with Sensory Fusion + Inter-user communication: Audio chat

Witmer and Singer [104] found that the IT predicts, within a given virtual environment, the level of presence felt by participants (as measured by their presence questionnaire). Tromp et. al [105] indicate that they found a positive correlation between personal presence and co-presence in one of their experiments. This small group experiment is described also in Slater et. al [106]. Johns et. al however, have shown that this is may be limited to levels of fidelity [107] and even types of presence experienced by the users [108]. We must therefore check whether there indeed is a positive correlation with the immersive tendencies score using our presence and co-presence questionnaires. It is also important to see if there is a correlation between the P score and the CO-P score since previous research has indicated a positive correlation between personal presence and co-presence.

4.2 Experiment

A Virginia Tech Institutional Review Board approved experiment was conducted as part of this study. A copy of the approval can be found in appendix D. This section will cover experimental expectations, experiment design and procedure.

4.2.1 Experimental aims and expectations

The goal of the experiment was two-fold:

1. To test whether an avatar could contribute to the perceived quality of communication given minimal to high interaction fidelity.
2. The more specific goal was to examine the role of sensory fusion: when the avatar's mouth state was directly related to the conversation, would this improve the quality of communication compared to the visually identical avatar with regular motion capture.

The expectation was that the mouth detection with sensory fusion task would lead to an improvement in perceived communication quality regular mouth detection, based on the logic that its mouth movements were related to an aspect of the conversation taking place.

4.2.2 Experimental design

The experiment investigated avatar behavioral fidelity along the interaction dimension and used a within-group experimental design. The experiment required pairs of participants who did not know each other prior to the experiment. An effort was made to remedy this by randomly allocating participants to each condition using a counter-balanced latin squares methodology to remove any input and ordering biases in the data collection based on their assumptions about what the experiment is about (demand characteristics). 12 pairs of participants were assigned to one of four conditions. The Vanilla build is the unmodified Minetest game version which allows avatar control through only the keyboard and mouse.

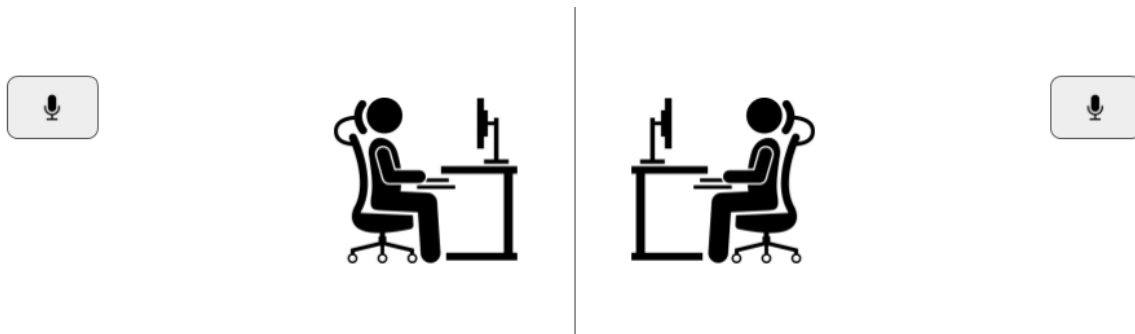


Figure 4.1: Interaction 1: Vanilla + Voice

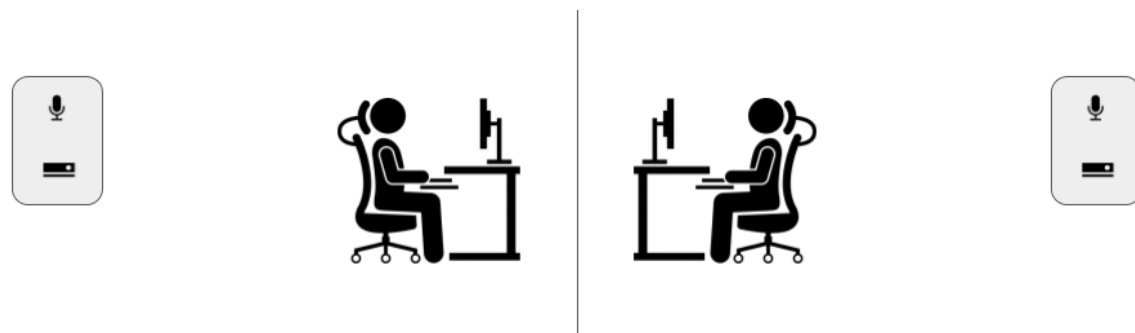


Figure 4.2: Interaction 2: Vanilla + Voice + Keyboard + Kinect Upper Body Only

The conversations took place within the same building over a network link separated by a physical barrier. As mentioned in the previous sections, a deliberate choice was made not to make use of the 3D potential of the avatar and retain the inherent low fidelity presence

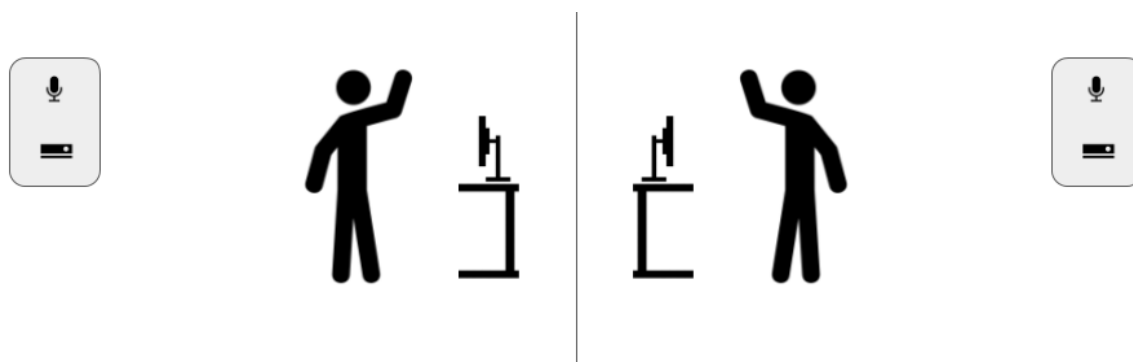


Figure 4.3: Interaction 3: Vanilla + Voice + Kinect

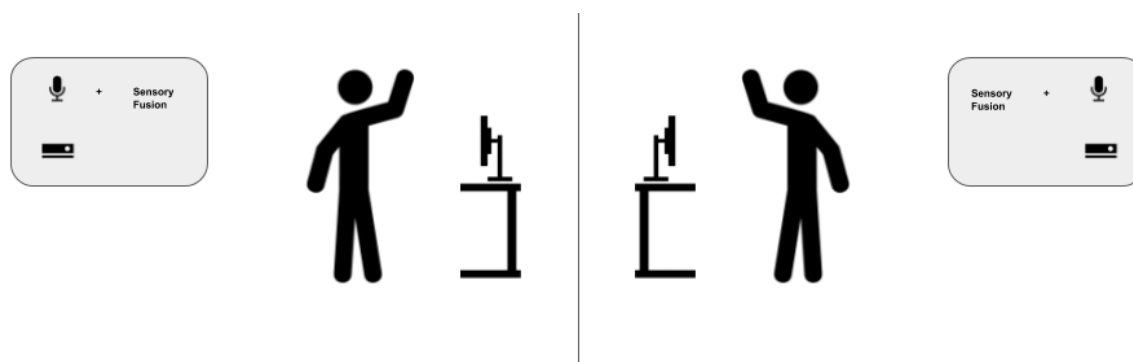


Figure 4.4: Interaction 4: Vanilla + Voice + Kinect + Sensory Fusion(Audio)

of the avatar and the only affect animations. The players could either choose to manipulate the avatar in first person or third person. The parts comprised of randomly chosen participants. They performed a ten-minute role-playing task in which they were randomly assigned to play roles out of a selection of common and most recognizable body expressions and gestures. Literature suggests that conceptualizing users as social actors puts researchers in a better position to “ask with whom an actor is interacting, about what issues, under what conditions, for what ends, with what resources, etc. It is a metaphor that readily expands the scope and scale of the social space of people’s interactions with information, the communication technology and with other people, groups, and organizations” [109]. This approach particularly provides opportunities for advancing our understanding of virtual worlds communication effects. Thus, role-playing various social interaction between the pairs of participants, co-located in the virtual world but separated in the physical world was chosen to be the best task for the experiment.

4.2.3 Tasks

Since the two participants were expected to speak for several minutes and did not know each other prior to the experiment, it was necessary to give them a topic of conversation. The first two sessions were conducted using a simple and contemporary script that the users had to read out to each other, inspired by speech impediment treatment narratives. A notable deficiency that became apparent was that while the scripts seemed interesting by themselves, the conversations between the avatars seemed uninteresting since participants often remained stationary to converse and used minimal head and body motions. Thus the full range of 3D avatar expressions and gestures remained unused even at higher levels of embodiment and interaction fidelity. This led to the adoption of a second script designed as a guessing game where each participant had unknown object placed behind them that was only visible to the other participant. This was done in order to elicit stronger gestures, movements and audio input (for the sensory fusion layer) to generate more expressive avatars. While this led to a significant improvement in avatar expressions, the players still spent a sizable portion of the experiment standing still and the full potential of the full body motion capture remained underutilized.



Figure 4.5: Sample body expressions as part of the experimental task list. A full list of these body expressions is provided in appendix C.

Finally, a set of common and most recognizable body expressions was compiled in the form of a game where each participant must enact the designated body expression from a sheet, for the other person to guess within a stipulated time limit. A full list of these body expressions is provided in appendix C. Users were given identical task sheets for each experimental task and were expected to enact out the expressions without stating or explicitly alluding to the caption on the list. The goal of the game was to guess as many body expressions successfully between them within a stipulated amount of time. Both participants were given 1 task sheet each with a list of these body expressions for each experimental task involving a specific interaction mode. The expectation was that participants' task performance, i.e., the number of body expressions successfully guessed and enacted from their designated lists, would increase with higher interaction fidelity. Finally, the audio sensory fusion layer was expected to give the best results, i.e. the users would be able to guess the most number of body expressions successfully with synchronized mouth and body motions.

4.2.4 Piloting

Conducting pilots was an essential to the iterative process of designing experiment tasks. There were 5 pilot studies conducted in total and a small sample of people were invited to participate in the pilot sessions. The purpose of these sessions was to evaluate the experimental design, procedure, task list and questionnaire items. Piloting helped the experimenters familiarize to the experimental procedure, which was of paramount importance to ensure that a standardized procedure was maintained throughout each experiment. It also allowed us to estimate the number and length of the sessions required to complete the experiment.

4.2.5 Apparatus

The experiment space consisted of two co-joined rooms separated by a physical barrier. Each room contained the projector, PC, Kinect HD, microphone and peripherals for the user and the participants completed questionnaires following each task in the same space. The rooms were equipped with identical equipment as described below. The rooms were purposefully bare in order to avoid providing visual distractions during the conversation. The two rooms in which participants were present were audio channel link through the microphone and a visual link through the Minetest game. The Kinect HD is placed at a sufficiently distance from the participant to ensure that it can capture the entire user skeleton moving in the physical space while also being able to discern the user's facial expressions in sufficient detail. Within the game, the initial position of both participants is facing each with close proximity to replicate a conversation between their avatars. distance between the participants. Participants sat 4 meters from a projector so that as the task list changes with different input modes, they would be able to get up and move within the space without much trouble.

4.2.6 Procedure

Upon arrival, participants were greeted in a reception area by two experimenters (the author and a colleague). One experimenter was assigned to mind each participant for the duration of the session. Participants were explained the experimental procedures and given the task sheets. Participants were informed that all data would be confidential and would only be used for the purpose of data analysis. They were also instructed that they were free to withdraw from the experiment at any time and without giving a reason for withdrawing. Each participant was asked to sit down and the chair height was adjusted so that their face and shoulders were clearly visible on Kinect camera. All applications and the audio channel were pre-configured and running prior to participants' arrival. Participants were then given a few minutes to prepare for their tasks. This included greeting each other and initiating a brief conversation through the audio channel. Once they felt ready to proceed they were reminded of the amount of time they would have to perform their experimental task, and

that at the end of the task the experimenter would return to guide them through the next stage. During the task, the experimenters quietly observed participants. In the interests of a standardized procedure, participants were stopped at the end of the assigned time period regardless of whether the task had been completed. After completing each task, participants filled out questionnaires about their experience.

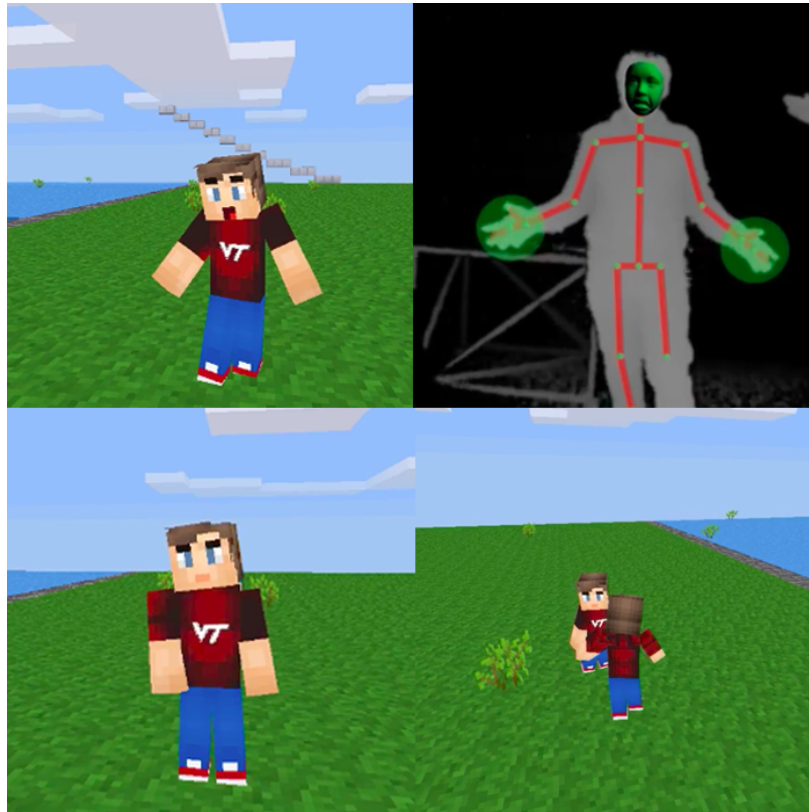


Figure 4.6: Top Left: Participant 1 - Avatar talking in sensory fusion mode; Top Right: Motion and Audio capture; Bottom Left: Participant 2 - Avatar interacting in virtual world; Bottom Right: Both participants can view the scene in third person

Avatars

Participants in each pair were represented by a visually similar avatar as differences in facial geometry and texture mapping could potentially impact on the visual effect of the animations. The only significant change was that a female avatar was used for female participants, and a male avatar for male participants. Each avatar was independently controlled for each user. The avatars are capable of a selection of behaviors such as smiling, frowning, looking sad, shrugging, pointing, waving, jumping, etc.

The participants could either choose to only see the other user's avatar on screen using a

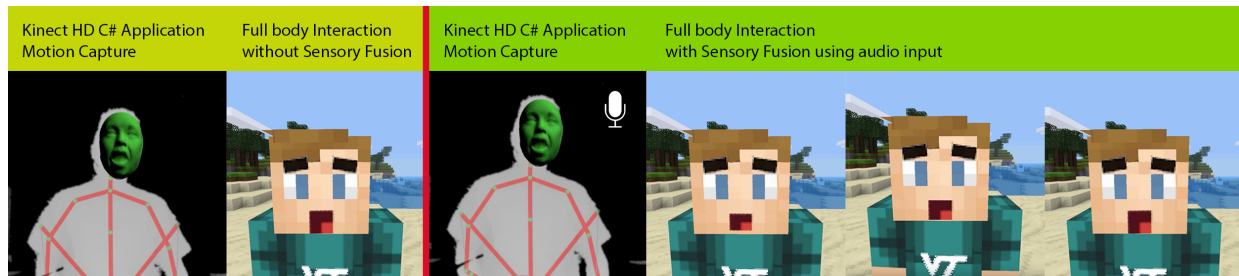


Figure 4.7: Sensory Fusion Improvements: The user’s voice is used to create more expressive avatars synced with their speech

first person view or choose to also view their own avatar in third person. The sense of embodiment into an avatar is constitutive of the sense of presence and affects the way one interacts with virtual elements [110]. It has also been shown that player perspective views support distinctive experiences of immersion for video game play and different perception of the game space [111]. While a first person perspective allows the player to perceive the game through the eyes of the character, observing the world around them up close, giving a clear view of the scenery in front of them. This perspective is believed to provide the most immersive feel for the player [112] [113]. Alternatively, a third person perspective allows the player to observe the main character in action, without giving the player the sense that they actually are the character. This was also observed as part of our experiments, where users preferred to view their own embodied avatars in addition to the other user’s avatar, for better manipulation. While the direct impact of perspective change on the sense of embodiment and presence is beyond the scope of this thesis, it is important to note whether these benefits of third person perspective can be exploited without detrimental consequences on the immersion and the ability to embody an avatar.

Participants

A total of 24 participants were recruited from the campus through an advertising poster campaign. As many as 50% had used some form of immersive technology, either a HMD or Kinect or alternate motion capture device. All participants were familiar with video-mediated communication and video games and 70% of them had had prior exposure to Minecraft or any of the alternate variations like Minetest. Special care was taken to ensure that participants did not know each other.

Short, Williams and Christie argue that “one might anticipate that media effects would be particularly marked when the interactors are relatively unacquainted. While people are still getting to know one another, any small additional piece of information can markedly affect overall judgments; later on in the acquaintance process, small changes in the available information would be expected to have less effect” [12]. The negotiation task, combined with the fact that participants were unacquainted, meant that high demands were likely to be

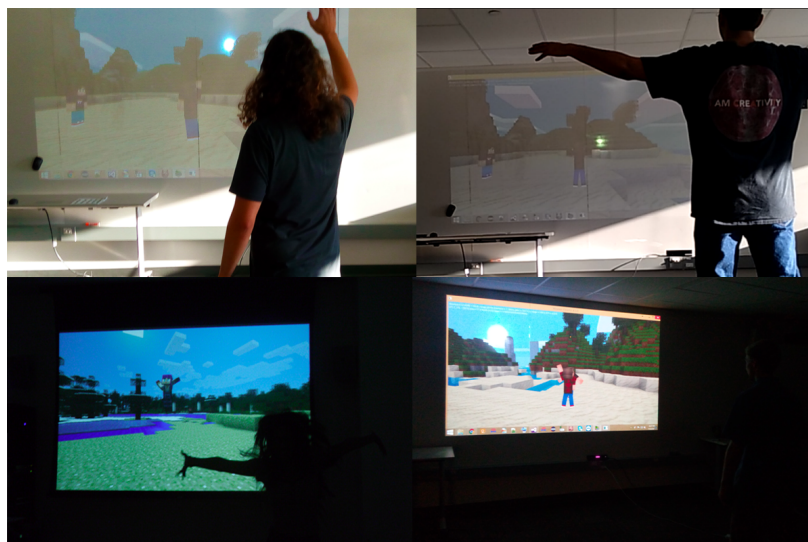


Figure 4.8: Pairs of participants communicating with each other using embodied interaction and audio

placed on the avatar. The original goal was to have a minimum of 10 pairs of participants per condition. Extra sessions were run to complete the 24 set of trials for a complete Latin square matrix along with additional sessions for video documentation.

4.3 Results

This section begins by presenting the findings for the response variables mentioned in the previous section. Table 4.1 shows the descriptive statistics for responses in each questionnaire. Response variables are constructed from n questionnaire items, each on a 1 to 7 scale. The items for each response are detailed in Appendix A and B.

		Keyboard, Mouse	Kinect for face and upper torso, Keyboard	Full face and body motion	Full face and body motion with Sensory Fusion
Presence (n=34)		72.5 ± 8.69	93.54 ± 11.80	160.41 ± 7.56	191.45 ± 9.97
Co-Presence (n=28)	Self-Reported Co-Presence (n=6)	8.25 ± 0.60	12.41 ± 0.92	29.66 ± 2.03	37.83 ± 0.81
	Empathy (n=5)	11.08 ± 0.58	11.41 ± 1.63	18.16 ± 1.00	25.20 ± 0.58
	Mutual Awareness (n=6)	13.66 ± 2.31	16.12 ± 0.53	30.45 ± 1.10	36.16 ± 1.88
	Attentional Allocation (n=3)	7.04 ± 0.69	12.29 ± 1.19	16 ± 0.29	18.04 ± 0.46
	Combined Co-Presence (n=28)	50 ± 5.08	67.33 ± 4.13	123.41 ± 3.95	164.5 ± 4.96
Immersive Tendencies (n=14)				70.16 ± 9.34	

Table 4.1: Mean and standard deviations of count response variables

The means of the raw questionnaire responses illustrates a progressive increase in mean responses across the different interactions for both the response variables.

The Co-Presence scores are further divided into factors that are also scored separately to

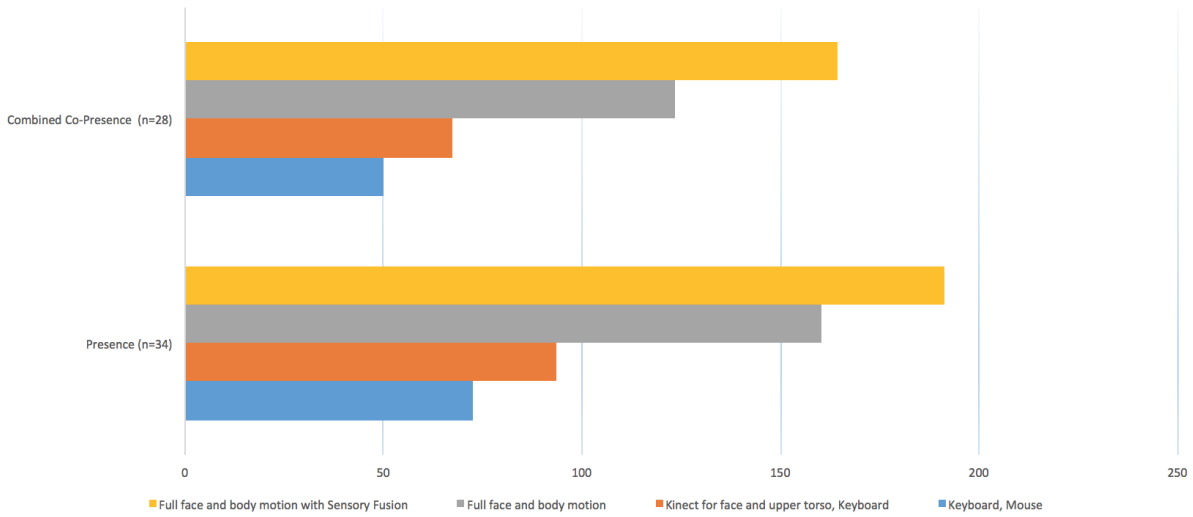


Figure 4.9: Means of cumulative questionnaire responses for each variable

reveal their contributions and trends with the with changing interaction fidelity.

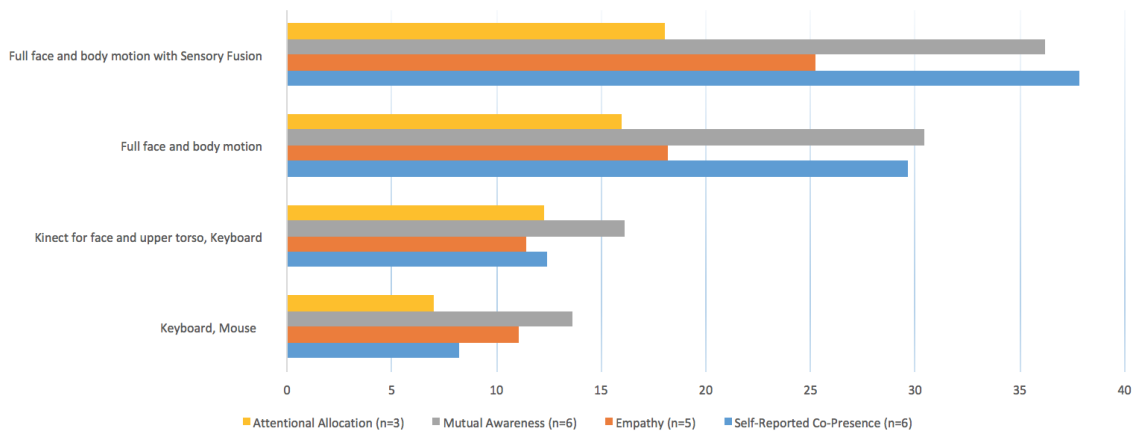


Figure 4.10: Means of cumulative questionnaire responses for contributing factors to co-presence

4.3.1 Analysis

We measured the presence score (P), the co-presence score (CO-P), and the immersive tendencies score (IT) for each interaction mode and performed a one-way Analysis of Variance (ANOVA) between the interaction mode and each response variable score. The Co-Presence was composed of Self-reported Co-Presence, Empathy, Mutual Awareness and Attentional

Allocation to better relate the specific improvements in co-presence to the interaction mode. Please refer to Appendix B for full tabular representation of user scores for each response variable.

Presence

We compared the difference in the P scores between the interaction modes and we found that there was a significant difference at the 0.05 confidence level, with $F(1,24) = 119$, $p < 0.05$. This indicates that participants had a higher P score on the high-interaction fidelity tasks.

ANOVA for P scores							
	<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
Rows		4650.95833	23	202.21558	3.57876906	2.1812E-05	1.68689696
Columns		225823.208	3	75274.4028	1332.19065	5.5979E-61	2.73749231
Error		3898.79167	69	56.5042271			
Total		234372.958	95				

Table 4.2: ANOVA test for Presence scores

Co-Presence

The Co-Presence was composed of Self-reported Co-Presence, Empathy, Mutual Awareness and Attentional Allocation to better relate the specific improvements in co-presence to the interaction mode. Our findings support our hypothesis that increasing level of interaction fidelity showed a positive trend in Co-Presence and Presence scores. A statistically significant difference was observed in the CO-P scores across the interaction modes with $F(1,24) = 119$, $p < 0.05$.

ANOVA for CO-P							
	<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
Rows		1531.625	23	66.592391	12.042918	4.197E-16	1.686897
Columns		198451.46	3	66150.486	11963.002	1.233E-93	2.7374923
Error		381.54167	69	5.5295894			
Total		200364.63	95				

Table 4.3: ANOVA test for Co-Presence scores

A statistically significant difference at the 0.05 confidence level was observed for Self-reported Co-Presence scores with $F(1,24) = 119$, $p < 0.05$.

Similarly, scores for avatar empathy were found to be $F(1,24) = 119$, $p < 0.05$ with an increasing trend with increasing interaction fidelity.

ANOVA for Self-reported copresence:

<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
Rows	40.3333333	23	1.75362319	1.22635135	0.25387189	1.68689696
Columns	14168.8333	3	4722.94444	3302.86993	1.9882E-74	2.73749231
Error	98.6666667	69	1.42995169			
Total	14307.8333	95				

Table 4.4: ANOVA test for Self-reported Co-Presence scores

ANOVA for Empathy Scores

<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
Rows	49.65625	23	2.15896739	2.90375635	0.00034068	1.68689696
Columns	3210.94792	3	1070.31597	1439.54782	4.0365E-62	2.73749231
Error	51.3020833	69	0.74350845			
Total	3311.90625	95				

Table 4.5: ANOVA test for avatar Empathy scores

Mutual awareness scores were observed to be significant with $F(1,24) = 119$, $p < 0.05$, while the difference in attentional allocation scores for the players was also a statistically significant at $F(1,24) = 119$, $p < 0.05$.

ANOVA for Mutual Awareness scores

<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
Rows	111.458333	23	4.84601449	2.61656342	0.00112398	1.68689696
Columns	8603.70833	3	2867.90278	1548.49918	3.3861E-63	2.73749231
Error	127.791667	69	1.85205314			
Total	8842.95833	95				

Table 4.6: ANOVA test for Mutual Awareness scores

ANOVA for Attention Allocation scores

<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
Rows	27.40625	23	1.19157609	3.50332889	2.9496E-05	1.68689696
Columns	1678.78125	3	559.59375	1645.25033	4.3123E-64	2.73749231
Error	23.46875	69	0.34012681			
Total	1729.65625	95				

Table 4.7: ANOVA test for Attention Allocation scores

Immersive Tendencies

Witmer and Singer[114] indicate that their Immersive Tendencies Questionnaire (ITQ) predicts the level of presence as measured by their presence questionnaire in a VE. Since in this experiment we have used a modified presence questionnaire inspired by Slater et al and co-presence questionnaire inspired by networked minds [115] and Nowak questionnaire [116], it is important to see if we can replicate Witmer and Singer’s results with our questionnaires. A correlation analysis was performed on the P, CO-P, and IT variables, and no significant relationships between were observed between them. At a significance level of 0.05, with n=24, we observed a score of 0.41 for P while CO-P was negligible at 0.08 for the Keyboard and mouse mode. Similarly, a score of 0.15 for P and negligible for CO-P was observed for only upper torso embodiment along with Keyboard control, 0.29 (P) and 0 (CO-P) for Full face and body motion and finally, 0.24 (P) and 0.04 (CO-P) for Full face and body motion with Sensory Fusion. The corresponding scores for all modes are presented in the table below.

Keyboard, Mouse				Kinect for face and upper torso, Keyboard			
	<i>IT</i>	<i>Presence</i>	<i>Co-Presence</i>		<i>IT</i>	<i>Presence</i>	<i>Co-Presence</i>
Imm Tend		1		Imm Tend		1	
Presence	0.41682435		1	Presence	0.15608429		1
Co-Presence	0.08974765	0.12492112		Co-Presence	-0.0195071	0.13154025	1
Full face and body motion				Full face and body motion with Sensory Fusion			
	<i>IT</i>	<i>Presence</i>	<i>Co-Presence</i>		<i>IT</i>	<i>Presence</i>	<i>Co-Presence</i>
Imm Tend		1		Imm Tend		1	
Presence	0.29555107		1	Presence	0.24121649		1
Co-Presence	-0.0090208	0.14072581		Co-Presence	0.04315587	0.10145185	1

Table 4.8: Correlation Matrix for CO-P and P scores with respect to IT scores

Since the correlation matrix in our experiments did not show any significant trends i.e no positive or negative correlation of the immersive tendencies score with either the presence or co-presence scores of the participants, we included an additional step as part of our analysis, which was breaking the participant groups into 2 - One with high immersive tendencies scores and the other group with low reported immersive tendencies. Thus we can now find correlation between the between the 2 immersive tendencies participant groups and the co-presence and presence scores.

4.3.2 Discussion of Results

The results show that there was a significant difference in the co-presence scores and presence scores with increasing interaction fidelity. i.e. Interaction modes with the Kinect HD using full body immersion for embodied interactions and additional sensory fusion audio input yielded the highest scores, which was picked up by the co-presence and presence questionnaires. This supports our hypothesis that increasing the avatar’s functionality through a

higher interaction fidelity results in increasing presence. This may be explained by the fact that since the high-collaboration task was more challenging, it required the participants to be more involved in the experience and hence enhanced the sense of personal presence. This might be explained by the fact that full body interaction tasks required the participants to be more involved in the experience and hence enhanced the sense of personal presence. This also supports previous work suggesting behavioral fidelity should be prioritized over visual fidelity in the development of expressive avatars. Our study also shows that improvements in behavioral fidelity benefit the constant low fidelity avatars regardless of their appearance. The Co-Presence and Presence scores were also observed to be the highest in the tasks with sensory fusion, which help us prove the second hypothesis.

4.4 Contribution

This dissertation discusses the impact of avatar behavioral fidelity on user presence and co-presence in a collaborative virtual world. The study presents a methodology for an immersive performance-centric interaction platform to deliver spontaneous avatar expressions using non-intrusive tracking by successfully sidestepping the uncanny valley. It has been shown that there is a strong link between avatar behavioral fidelity and the quality of a performance, along with the difficulties in capturing spontaneous expressions through embodied interactions. The study demonstrates the improvements in interaction fidelity due to the addition of sensory fusion for synchronous mouth movements in accordance to the user's speech. The research has also resulted in the creation of a non-intrusive immersive collaborative platform built using off the shelf hardware, which is readily accessible in the form of drop-in software packages containing Cinemacraft executable and Kinect HD C# application.

4.5 Conclusion

One of the chief attractions of Collaborative virtual environments (CVEs) lies in their ability to combine 3D spatial interaction with a high degree of multi-sensory immersion. They are therefore of particular interest for those collaborative situations, such as remote acting rehearsals, where it is essential to preserve spatial relationships among users. A key barrier to effective communication in current CVEs is the relative paucity of avatar expressiveness as compared to live video. However, increasing the expressive potential of avatars involves significant challenges. In terms of their appearance, the tension between realism and real time means that photorealism comes at the expense of unwanted delays to real-time communication. Visual fidelity must therefore be traded off against available computing resources. In terms of behavior, the tension between control and cognitive load underlines the difficulty of transparently driving avatar behaviors that appropriately represent the user. While full

avatar control through manual keyboard and mouse manipulation would result unnatural interaction and high level of cognitive load, full tracking for all users can be expensive and invasive. Given these constraints, the approach taken in this research was to explore levels of avatar behavioral fidelity using varying interaction modes and low avatar visual fidelity. The platform designed as part of this research focuses on immersive performance-centric interaction inspired by the success of Minecraft and builds on its approach by successfully sidestepping the uncanny valley. The overarching goal was to investigate whether increments in behavioral fidelity could contribute to participants' interaction experience. The study focused primarily on presence and co-presence by combining questionnaires with an analysis of participant responses. Our results so far are promising and we were able to create a high level of immersion by combining multiple interaction techniques into a single system despite relying on a cartoon-like low fidelity environment. Extending sophisticated technology like immersive VR and gesture tracking to easy marker less motion capture our performers could control their avatar with relative ease and accuracy without extended training sessions.

Chapter 5

Bibliography

- [1] Jarrod Ratcliffe. Hand motion-controlled audio mixing interface. *Proc. of New Interfaces for Musical Expression (NIME) 2014*, pages 136–139, 2014.
- [2] Teemu Ahmaniemi. Gesture controlled virtual instrument with dynamic vibrotactile feedback. In *NIME*, pages 485–488, 2010.
- [3] Microsoft. Kinect 360, 2017.
- [4] Microsoft Kinect HD. Kinect kinect hd, 2017.
- [5] Richard Kastelein. The rise of machinima, the artform, 2013.
- [6] Emil Polyak. Virtual impersonation using interactive glove puppets. In *SIGGRAPH Asia 2012 Posters*, SA '12, pages 31:1–31:1, New York, NY, USA, 2012. ACM.
- [7] Robert Hamilton. Sonifying game-space choreographies with udkosc. In *NIME*, 2013.
- [8] Amber Choo, Mehdi Karamnejad, and Aaron May. Maintaining long distance togetherness synchronous communication with minecraft and skype. In *Games Innovation Conference (IGIC), 2013 IEEE International*, pages 27–35. IEEE, 2013.
- [9] Milan Loviska, Otto Krause, Herman A Engelbrecht, Jason B Nel, Gregor Schiele, Alwyn Burger, Stephan Schmeißer, Christopher Cichiwskyj, Lilian Calvet, Carsten Griwodz, et al. Immersed gaming in minecraft. In *Proceedings of the 7th International Conference on Multimedia Systems*, page 32. ACM, 2016.
- [10] David Ed Matsumoto, Hyisung C Hwang, and Mark G Frank. *APA handbook of nonverbal communication*. American Psychological Association, 2016.
- [11] Anthony Giddens. *The constitution of society: Outline of the theory of structuration*, volume 349. Univ of California Press, 1986.

- [12] John Short, Ederyn Williams, and Bruce Christie. The social psychology of telecommunications. 1976.
- [13] Elisabeth Cuddihy and Deborah Walters. Embodied interaction in social virtual environments. In *Proceedings of the third international conference on Collaborative virtual environments*, pages 181–188. ACM, 2000.
- [14] Clarence A Ellis, Simon J Gibbs, and Gail Rein. Groupware: some issues and experiences. *Communications of the ACM*, 34(1):39–58, 1991.
- [15] Jon Hindmarsh, Mike Fraser, Christian Heath, Steve Benford, and Chris Greenhalgh. Fragmented interaction: establishing mutual orientation in virtual environments. In *Proceedings of the 1998 ACM conference on Computer supported cooperative work*, pages 217–226. ACM, 1998.
- [16] Peter Ludlow. *High noon on the electronic frontier: conceptual issues in cyberspace*. MIT Press, 1996.
- [17] K Swinth and Jim Blascovich. Perceiving and responding to others: Human-human and human-computer social interaction in collaborative virtual environments. In *Proceedings of the 5th Annual International Workshop on PRESENCE*, volume 392, 2002.
- [18] Jim Blascovich. Social influence within immersive virtual environments. *The social life of avatars*, pages 127–145, 2002.
- [19] Justine Cassell, Timothy Bickmore, Mark Billinghurst, Lee Campbell, Kenny Chang, Hannes Vilhjálmsón, and Hao Yan. Embodiment in conversational interfaces: Rea. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 520–527. ACM, 1999.
- [20] Steve Benford, John Bowers, Lennart E Fahlén, Chris Greenhalgh, and Dave Snowden. User embodiment in collaborative virtual environments. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 242–249. ACM Press/Addison-Wesley Publishing Co., 1995.
- [21] Joseph P Forgas and Richard Jones. *Interpersonal behaviour: The psychology of social interaction*. Pergamon Press, 1985.
- [22] Frank Thomas, Ollie Johnston, and Frank. Thomas. *The illusion of life: Disney animation*. Hyperion New York, 1995.
- [23] RW Picard. Affective computing mit press cambridge. *MA Google Scholar*, 1997.
- [24] Carroll E Izard. Emotions and facial expressions: A perspective from differential emotions theory. *The psychology of facial expression*, 2:57–77, 1997.

- [25] Randolph R Cornelius. *The science of emotion: Research and tradition in the psychology of emotions*. Prentice-Hall, Inc, 1996.
- [26] Paul Ekman, Wallace V Friesen, and Phoebe Ellsworth. *Emotion in the human face: Guidelines for research and an integration of findings*. Elsevier, 2013.
- [27] Sheryl M Ehrlich, Diane J Schiano, and Kyle Sheridan. Communicating facial affect: it’s not the realism, it’s the motion. In *CHI’00 Extended Abstracts on Human Factors in Computing Systems*, pages 251–252. ACM, 2000.
- [28] Paul Ekman and Wallace V Friesen. The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *semiotica*, 1(1):49–98, 1969.
- [29] Harald G Wallbott. Bodily expression of emotion. *European journal of social psychology*, 28(6):879–896, 1998.
- [30] Michael Argyle. *Bodily communication*. Routledge, 2013.
- [31] Peter Bull. *Body movement and interpersonal communication*. John Wiley & Sons Inc, 1983.
- [32] Mel Slater, Anthony Steed, and Yiorgos Chrysanthou. *Computer graphics and virtual environments: from realism to real-time*. Pearson Education, 2002.
- [33] Ralph Schroeder. *The social life of avatars: Presence and interaction in shared virtual environments*. Springer Science & Business Media, 2012.
- [34] Alexander Nilsson, Ann-Sofie Axelsson, Ilona Heldal, and Ralph Schroeder. The long-term uses of shared virtual environments: An exploratory study. In *The social life of avatars*, pages 112–126. Springer, 2002.
- [35] Ralph Schroeder. *Possible worlds: the social dynamic of virtual reality technology*. Westview Press, Inc., 1996.
- [36] John Suler. The psychology of avatars and graphical space in multimedia chat communities. *Chat communication*, pages 305–344, 1999.
- [37] Lili Cheng, Shelly Farnham, and Linda Stone. Lessons learned: Building and deploying shared virtual environments. In *The social life of avatars*, pages 90–111. Springer, 2002.
- [38] Chip Morningstar and F Randall Farmer. The lessons of lucasfilm’s habitat. in (m. benedikt, ed.) *cyberspace: First steps*, 1990.
- [39] Nat Durlach and Mel Slater. Presence in shared virtual environments and virtual togetherness. *Presence: Teleoperators and Virtual Environments*, 9(2):214–217, 2000.

- [40] Qiong Wu, Pierre Boulanger, Maryia Kazakevich, and Robyn Taylor. A real-time performance system for virtual theater. In *Proceedings of the 2010 ACM workshop on Surreal media and virtual cloning*, pages 3–8. ACM, 2010.
- [41] Tina L Taylor. Living digitally: Embodiment in virtual worlds. In *The social life of avatars*, pages 40–62. Springer, 2002.
- [42] Hunter G. Hoffman, Jerrold Prothero, Maxwell J. Wells, and Joris Groen. Virtual chess: Meaning enhances users’ sense of presence in virtual environments. *International Journal of Human-Computer Interaction*, 10(3):251–263, 1998.
- [43] George Lakoff and Mark Johnson. The metaphorical structure of the human conceptual system. *Cognitive science*, 4(2):195–208, 1980.
- [44] Giuseppe Riva and John A Waterworth. Presence and the self: A cognitive neuroscience approach. *Presence connect*, 3(3), 2003.
- [45] Judith S Donath et al. Identity and deception in the virtual community. *Communities in cyberspace*, 1996:29–59, 1999.
- [46] Asimina Vasalou, Adam Joinson, Tanja Bänziger, Peter Goldie, and Jeremy Pitt. Avatars in social media: Balancing accuracy, playfulness and embodied messages. *International Journal of Human-Computer Studies*, 66(11):801–811, 2008.
- [47] Holtjona Galanxhi and Fiona Fui-Hoon Nah. Deception in cyberspace: A comparison of text-only vs. avatar-supported medium. *International journal of human-computer studies*, 65(9):770–783, 2007.
- [48] Steve Benford, John Bowers, Lennart E. Fahlen, John Mariani, and Tom Rodden. Supporting cooperative work in virtual environments. *The Computer Journal*, 37(8):653–668, 1994.
- [49] Mike Fraser, Tony Glover, Ivan Vaghi, Steve Benford, Chris Greenhalgh, Jon Hindmarsh, and Christian Heath. Revealing the realities of collaborative virtual reality. In *Proceedings of the third international conference on Collaborative virtual environments*, pages 29–37. ACM, 2000.
- [50] Ralph Schroeder. Copresence and interaction in virtual environments: An overview of the range of issues. In *Presence 2002: Fifth international workshop*, pages 274–295, 2002.
- [51] Eva-Lotta Sallnäs. Collaboration in multi-modal virtual worlds: comparing touch, text, voice and video. *The social life of avatars: Presence and interaction in shared virtual environments*, pages 172–187, 2002.
- [52] Mary Katsikitis, Issy Pilowsky, and John M Innes. Encoding and decoding of facial expression. *The Journal of General Psychology*, 124(4):357–370, 1997.

- [53] Diane J Schiano, Sheryl M Ehrlich, Krisnawan Rahardja, and Kyle Sheridan. Face to interface: facial affect in (hu) man and machine. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 193–200. ACM, 2000.
- [54] Gary Bente and Nicole C Krämer. Virtual gestures: Analyzing social presence effects of computer-mediated and computer-generated nonverbal behaviour. In *Fifth Annual International Workshop PRESENCE 2002*, pages 233–44, 2002.
- [55] Carrie Heeter. Being there: The subjective experience of presence. *Presence: Teleoperators & Virtual Environments*, 1(2):262–271, 1992.
- [56] Jolanda Tromp, Adrian Bullock, Anthony Steed, Amela Sadagic, Mel Slater, and Emmanuel Frécon. Small group behaviour experiments in the coven project. *IEEE Computer Graphics and Applications*, 18(6):53–63, 1998.
- [57] Byron Reeves and Clifford Nass. *How people treat computers, television, and new media like real people and places*. CSLI Publications and Cambridge university press, 1996.
- [58] Maia Garau. Selective fidelity: Investigating priorities for the creation of expressive avatars. *Avatars at Work and Play*, pages 17–38, 2006.
- [59] DP Pertaub, M Slater, and C Barker. An experiment on fear of public speaking in virtual reality. *Studies in health technology and informatics*, pages 372–378, 2001.
- [60] Jeremy N Bailenson, Jim Blascovich, Andrew C Beall, and Jack M Loomis. Equilibrium theory revisited: Mutual gaze and personal space in virtual environments. *Presence: Teleoperators and virtual environments*, 10(6):583–598, 2001.
- [61] Taeyong Kim and Frank Biocca. Telepresence via television: Two dimensions of telepresence may have different connections to memory and persuasion. *Journal of Computer-Mediated Communication*, 3(2):0–0, 1997.
- [62] Wijnand IJsselsteijn, Huib de Ridder, Jonathan Freeman, Steve E Avons, and Don Bouwhuis. Effects of stereoscopic presentation, image motion, and screen size on subjective and objective corroborative measures of presence. *Presence: Teleoperators and virtual environments*, 10(3):298–311, 2001.
- [63] Wijnand A IJsselsteijn, Huib de Ridder, Jonathan Freeman, and Steve E Avons. Presence: Concept, determinants and measurement. In *Human vision and electronic imaging*, volume 3959, pages 520–529, 2000.
- [64] Jonathan Freeman, Jane Lessiter, and Wijnand IJsselsteijn. An introduction to presence: A sense of being there in a mediated environment. *The Psychologist*, 14:190–194, 2001.

- [65] Mel Slater and Anthony Steed. A virtual presence counter. *Presence: Teleoperators and virtual environments*, 9(5):413–434, 2000.
- [66] Mel Slater and Sylvia Wilbur. A framework for immersive virtual environments (five): Speculations on the role of presence in virtual environments. *Presence: Teleoperators and virtual environments*, 6(6):603–616, 1997.
- [67] Mel Slater, Martin Usoh, and Anthony Steed. Taking steps: the influence of a walking technique on presence in virtual reality. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 2(3):201–219, 1995.
- [68] Maria V Sanchez-Vives and Mel Slater. From presence to consciousness through virtual reality. *Nat Rev Neurosci*, 6(4):332–339, 2005.
- [69] Kristine L Nowak and Frank Biocca. The effect of the agency and anthropomorphism on users’ sense of telepresence, copresence, and social presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, 12(5):481–494, 2003.
- [70] A Airapetian, V Dodonov, L Micu, D Axen, V Vinogradov, D Akerman, B Szeless, P Chochula, C Geich-Gimbel, P Schacht, et al. *ATLAS detector and physics performance: Technical Design Report, 2*. Number CERN-LHCC-99-015. ATLAS-TDR-015, 1999.
- [71] Mel Slater. Measuring presence: A response to the witmer and singer presence questionnaire. *Presence: Teleoperators and Virtual Environments*, 8(5):560–565, 1999.
- [72] Declan Delaney, Tomás Ward, and Seamus McLoone. On consistency and network latency in distributed interactive applications: A survey part i. *Presence: Teleoperators and Virtual Environments*, 15(2):218–234, 2006.
- [73] Paul De Greef and Wijnand IJsselsteijn. Social presence in the photoshare tele-application. *Proceedings of PRESENCE*, pages 27–28, 2000.
- [74] Bob G Witmer and Michael J Singer. Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and virtual environments*, 7(3):225–240, 1998.
- [75] Mel Slater, Anthony Steed, and Martin Usoh. The virtual treadmill: A naturalistic metaphor for navigation in immersive virtual environments. In *Virtual Environments 95*, pages 135–148. Springer, 1995.
- [76] Thomas B. Sheridan. Musings on telepresence and virtual presence. *Presence: Teleoper. Virtual Environ.*, 1(1):120–126, January 1992.
- [77] Claudia Hendrix and Woodrow Barfield. Presence within virtual environments as a function of visual display parameters. *Presence: Teleoper. Virtual Environ.*, 5(3):274–289, January 1996.

- [78] *Presence: Teleoper. Virtual Environ.*, 19(5), 2010.
- [79] Mel Slater and Martin Usoh. Body centred interaction in immersive virtual environments. In *Artificial Life and Virtual Reality*, pages 125–148. John Wiley and Sons, 1994.
- [80] Mel Slater, Martin Usoh, and Anthony Steed. Depth of presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, 3(2):130–144, 1994.
- [81] Martin Usoh, Ernest Catena, Sima Arman, and Mel Slater. Using presence questionnaires in reality. *Presence: Teleoperators and Virtual Environments*, 9(5):497–503, 2000.
- [82] John V Draper, David B Kaber, and John M Usher. Telepresence. *Human factors*, 40(3):354–375, 1998.
- [83] Jonathan Freeman, Steve E Avons, Don E Pearson, and Wijnand A IJsselsteijn. Effects of sensory information and prior experience on direct subjective ratings of presence. *Presence: Teleoperators and Virtual Environments*, 8(1):1–13, 1999.
- [84] Mel Slater. How colorful was your day? why questionnaires cannot assess presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, 13(4):484–493, 2004.
- [85] Dustin B Chertoff, Brian Goldiez, and Joseph J LaViola. Virtual experience test: A virtual environment evaluation questionnaire. In *Virtual Reality Conference (VR), 2010 IEEE*, pages 103–110. IEEE, 2010.
- [86] Richard Skarbez, Frederick P. Brooks, Jr., and Mary C. Whitton. A survey of presence and related concepts. *ACM Comput. Surv.*, 50(6):96:1–96:39, November 2017.
- [87] Ivica Ico Bukvic, Cody Cahoon, Ariana Wyatt, Tracy Cowden, and Katie Dredger. Operacraft: Blurring the lines between real and virtual. In *ICMC*, 2014.
- [88] I Bukvic. A behind-the-scenes peek at world’s first linux-based laptop orchestra—the design of l2ork infrastructure and lessons learned.
- [89] Virginia Tech. South by southwest 2016, 2016.
- [90] Arts Institute for Creativity and Technology. Icat day 2016, 2016.
- [91] Virginia Tech. Science museum of western virginia, 2017.
- [92] Nicholas F Polys, Benjamin Knapp, Matthew Bock, Christina Lidwin, Dane Webster, Nathan Waggoner, and Ivica Bukvic. Fusality: an open framework for cross-platform mirror world installations. In *Proceedings of the 20th International Conference on 3D Web Technology*, pages 171–179. ACM, 2015.

- [93] Chris Buecheler. Character: The next great gaming frontier?, 2010.
- [94] Nikolaus Gebhardt et al. Irrlicht engine. *Related Pages*, 2010.
- [95] FUDI. Fudi, 2017.
- [96] Matthew Wright, Adrian Freed, et al. Open soundcontrol: A new protocol for communicating with sound synthesizers. In *ICMC*, 1997.
- [97] Facebook. Oculus rift, 2016.
- [98] Leap Motion. Leap motion, 2016.
- [99] Ruofei Du and Liang He. Vrsurus: Enhancing interactivity and tangibility of puppets in virtual reality. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pages 2454–2461. ACM, 2016.
- [100] Hui Liang, Jian Chang, Ismail K Kazmi, Jian J Zhang, and Peifeng Jiao. Hand gesture-based interactive puppetry system to assist storytelling for children. *The Visual Computer*, 33(4):517–531, 2017.
- [101] Elena Kokkinara, Mel Slater, and Joan López-Moliner. The effects of visuomotor calibration to the perceived space and body, through embodiment in immersive virtual reality. *ACM Transactions on Applied Perception (TAP)*, 13(1):3, 2015.
- [102] Lara Maister, Mel Slater, Maria V Sanchez-Vives, and Manos Tsakiris. Changing bodies changes minds: owning another body affects social cognition. *Trends in cognitive sciences*, 19(1):6–12, 2015.
- [103] Tara Collingwoode-Williams, Marco Gillies, Cade McCall, and Xueni Pan. The effect of lip and arm synchronization on embodiment: A pilot study. In *Virtual Reality (VR), 2017 IEEE*, pages 253–254. IEEE, 2017.
- [104] MJ Singer and BG Witmer. Presence measures for virtual environments: Background & development. *Draft United States Army Research Institute for the Behavioral and Social Sciences*, 1996.
- [105] Jolanda Tromp, Adrian Bullock, Anthony Steed, Amela Sadagic, Mel Slater, and Emmanuel Frécon. Small group behavior experiments in the coven project. *IEEE Comput. Graph. Appl.*, 18(6):53–63, November 1998.
- [106] M. Slater, A. Sadagic, M. Usoh, and R. Schroeder. Small group behaviour in a virtual and real environment: A comparative study. pages 37–51, 2000.
- [107] Cathryn Johns, David Nunez, Marc Daya, Duncan Sellars, Juan Casanueva, and Edwin Blake. The interaction between individuals immersive tendencies and the sensation of presence in a virtual environment.

- [108] Saniye Tugba Bulu. Place presence, social presence, co-presence, and satisfaction in virtual worlds. *Computers & Education*, 58(1):154–161, 2012.
- [109] Roberta Lamb, John Leslie King, and Rob Kling. Informational environments: Organizational contexts of online information use. *Journal of the Association for Information Science and Technology*, 54(2):97–114, 2003.
- [110] Konstantina Kilteni, Raphaela Groten, and Mel Slater. The sense of embodiment in virtual reality. *Presence: Teleoperators and Virtual Environments*, 21(4):373–387, 2012.
- [111] Alena Denisova and Paul Cairns. First person vs. third person perspective in digital games: Do player preferences affect immersion? In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 145–148. ACM, 2015.
- [112] Gerald A Voorhees, Joshua Call, and Katie Whitlock. *Guns, grenades, and grunts: First-person shooter games*. Bloomsbury Publishing USA, 2012.
- [113] Laurie N Taylor. Video games: Perspective, point-of-view, and immersion. 2002.
- [114] Bob G. Witmer and Michael J. Singer. Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoper. Virtual Environ.*, 7(3):225–240, June 1998.
- [115] Frank Biocca, C Harms, and Jennifer Gregg. The networked minds measure of social presence: Pilot test of the factor structure and concurrent validity. 01 2001.
- [116] Kristine L. Nowak and Frank Biocca. The effect of the agency and anthropomorphism of users' sense of telepresence, copresence, and social presence in virtual environments. *Presence: Teleoper. Virtual Environ.*, 12(5):481–494, October 2003.
- [117] Frank Biocca, Chad Harms, and Jenn Gregg. The networked minds measure of social presence: Pilot test of the factor structure and concurrent validity. In *4th annual international workshop on presence, Philadelphia, PA*, pages 1–9, 2001.

Appendix

Appendix A : Questionnaires

All 24 participants recorded their responses to each experimental task using a presence questionnaire based on Slater[71] and Witmer's[74] presence questionnaires, a co-presence questionnaire based on the Networked Minds [117] and Nowak's [69] co-presence questionnaires and finally, the immersive tendencies questionnaires [74].

A.1 Presence Questionnaire

Please rate your sense of being in the virtual environment, on a scale of 1 to 7, where 7 represents your normal experience of being in a place.

How much were you able to control events?

How responsive was the environment to actions that you initiated (or performed)?

How natural did your interactions with the environment seem?

How completely were all of your senses engaged?

How much did the visual aspects of the environment involve you?

How much did the auditory aspects of the environment involve you?

How natural was the mechanism which controlled movement through the environment?

How aware were you of events occurring in the real world around you?

How aware were you of your display and control devices?

How compelling was your sense of objects moving through space?

How inconsistent or disconnected was the information coming from your various senses?

How much did your experiences in the virtual environment seem consistent with your real-world experiences ?

Were you able to anticipate what would happen next in response to the actions that you performed ?

How completely were you able to actively survey or search the environment using vision?

How well could you identify sounds?

How well could you localize sounds?

How compelling was your sense of moving around inside the virtual environment?

How closely were you able to examine objects?
How well could you examine objects from multiple viewpoints?
How well could you move or manipulate objects in the virtual environment?
To what degree did you feel confused or disoriented at the beginning of breaks or at the end of the experimental session?
How involved were you in the virtual environment experience?
How distracting was the control mechanism?
How much delay did you experience between your actions and expected outcomes?
How quickly did you adjust to the virtual environment experience?
How proficient in moving and interacting with the virtual environment did you feel at the end of the experience?
How much did the visual display quality interfere or distract you from performing assigned tasks or required activities?
How much did the control devices interfere with the performance of assigned tasks or with other activities?
How well could you concentrate on the assigned tasks or required activities rather than on the mechanisms used to perform those tasks or activities?
Did you learn new techniques that enabled you to improve your performance?
Were you involved in the experimental task to the extent that you lost track of time?
To what extent were there times during the experience when the virtual environment was the reality for you?
When you think back to the experience, do you think of the virtual environment more as images that you saw or more as somewhere that you visited?

A.2 Co-Presence Questionnaire

I often felt as if I was all alone.
I think the other individual often felt alone.
I hardly noticed another individual.
The other individual didnt notice me in the room.
I was often aware of others in the environment.
Others were often aware of me in the room.
I think the other individual often felt alone.
I often felt as if I was all alone.
I sometimes pretended to pay attention to the other individual.
The other individual paid close attention to me
I paid close attention to the other individual.
My partner was easily distracted when other things were going on around us.
I was easily distracted when other things were going on around me
When I was happy, the other was happy.
When the other was happy, I was happy.

My interaction partner seemed to find our interaction stimulating.
My interaction partner communicated coldness rather than warmth.
My interaction partner seemed detached during our interaction.
My interaction partner was unwilling to share personal information with me.
My interaction partner created a sense of closeness between us.
My interaction partner was interested in talking to me.
I wanted to maintain a sense of distance between us.
I was interested in talking to my interaction partner
I perceive that I am in the presence of another person in the room with me.
I feel that the person is watching me and is aware of my presence.
The thought that the person is not a real person crossed my mind often.
The person appears to be sentient (conscious and alive) to me.
I perceive the person as being only a computerized image, not as a real person.

A.3 Immersive Tendencies Questionnaire

Do you easily become deeply involved in movies or tv dramas?
Do you ever become so involved in a television program or book that people have problems getting your attention?
How mentally alert do you feel at the present time?
Do you ever become so involved in a movie that you are not aware of things happening around you?
How frequently do you find yourself closely identifying with the characters in a story line?
Do you ever become so involved in a video game that it is as if you are inside the game rather than moving a joystick and watching the screen?
How physically fit do you feel today?
How good are you at blocking out external distractions when you are involved in something?
When watching sports, do you ever become so involved in the game that you react as if you were one of the players?
Do you ever become so involved in a daydream that you are not aware of things happening around you?
Do you ever have dreams that are so real that you feel disoriented when you awake?
When playing sports, do you become so involved in the game that you lose track of time?
How well do you concentrate on enjoyable activities?
How often do you play arcade or video games? (OFTEN should be taken to mean every day or every two days, on average.)

Appendix B : Results

Description of results for all users with respect to each response variable.

B.1 Presence scores

ANOVA for P scores					
	<i>SUMMARY</i>	<i>Count</i>	<i>Sum</i>	<i>Average</i>	<i>Variance</i>
User 1		4	466	116.5	2033.66667
User 2		4	507	126.75	2508.25
User 3		4	445	111.25	2918.25
User 4		4	455	113.75	3296.91667
User 5		4	508	127	3734.66667
User 6		4	526	131.5	3887
User 7		4	517	129.25	3540.91667
User 8		4	522	130.5	3363
User 9		4	515	128.75	2992.91667
User 10		4	516	129	3916.66667
User 11		4	515	128.75	3666.25
User 12		4	525	131.25	3588.25
User 13		4	530	132.5	3292.33333
User 14		4	534	133.5	3165.66667
User 15		4	524	131	2445.33333
User 16		4	529	132.25	4157.58333
User 17		4	529	132.25	3007.58333
User 18		4	530	132.5	2803.66667
User 19		4	521	130.25	3224.91667
User 20		4	524	131	3404.66667
User 21		4	542	135.5	3049.66667
User 22		4	533	133.25	3070.91667
User 23		4	508	127	3734.66667
User 24		4	585	146.25	1770.25
Modality: Keyboard, Mouse		24	1740	72.5	75.6521739
Modality: Kinect for face and upper torso, Keyboard		24	2221	92.5416667	139.302536
Modality: Full face and body motion		24	3850	160.416667	57.2101449
Modality: Full face and body motion with Sensory Fusion		24	4595	191.458333	99.5634058

Table B.1: User Presence scores

B.2 Co-Presence scores

ANOVA for CO-P					
<i>SUMMARY</i>	<i>Count</i>	<i>Sum</i>	<i>Average</i>	<i>Variance</i>	
User 1	4	401	100.25	2828.25	
User 2	4	402	100.5	2745	
User 3	4	406	101.5	2883.6667	
User 4	4	388	97	2578	
User 5	4	405	101.25	2906.25	
User 6	4	397	99.25	2944.9167	
User 7	4	389	97.25	2702.9167	
User 8	4	400	100	2854.6667	
User 9	4	405	101.25	2802.25	
User 10	4	406	101.5	2692.3333	
User 11	4	414	103.5	2805.6667	
User 12	4	388	97	3223.3333	
User 13	4	408	102	2798.6667	
User 14	4	414	103.5	2821.6667	
User 15	4	383	95.75	2722.25	
User 16	4	413	103.25	2763.5833	
User 17	4	421	105.25	2687.5833	
User 18	4	389	97.25	2630.9167	
User 19	4	417	104.25	2906.9167	
User 20	4	423	105.75	2204.25	
User 21	4	368	92	2538	
User 22	4	420	105	2672.6667	
User 23	4	443	110.75	2860.9167	
User 24	4	426	106.5	2703	
Modality: Keyboard, Mouse	24	1200	50	25.826087	
Modality: Kinect for face and upper torso, Keyboard	24	1616	67.333333	17.101449	
Modality: Full face and body motion	24	2962	123.41667	15.644928	
Modality: Full face and body motion with Sensory Fusion	24	3948	164.5	24.608696	

Table B.2: User Co-Presence scores

B.3 Self reported Co-Presence scores

ANOVA for Self-reported copresence:

<i>SUMMARY</i>	<i>Count</i>	<i>Sum</i>	<i>Average</i>	<i>Variance</i>
User 1	4	88	22	205.333333
User 2	4	88	22	205.333333
User 3	4	88	22	205.333333
User 4	4	88	22	205.333333
User 5	4	88	22	205.333333
User 6	4	88	22	205.333333
User 7	4	83	20.75	180.916667
User 8	4	87	21.75	178.916667
User 9	4	85	21.25	191.583333
User 10	4	89	22.25	191.583333
User 11	4	91	22.75	203.583333
User 12	4	88	22	205.333333
User 13	4	89	22.25	204.916667
User 14	4	90	22.5	217
User 15	4	85	21.25	196.25
User 16	4	90	22.5	193
User 17	4	90	22.5	171
User 18	4	86	21.5	189.666667
User 19	4	91	22.75	212.25
User 20	4	88	22	205.333333
User 21	4	82	20.5	182.333333
User 22	4	89	22.25	166.25
User 23	4	94	23.5	246.333333
User 24	4	91	22.75	187.583333
self-reported copresence:	24	198	8.25	0.36956522
self-reported copresence:	24	298	12.4166667	0.86231884
self-reported copresence:	24	712	29.6666667	4.14492754
self-reported copresence:	24	908	37.8333333	0.66666667

Table B.3: User Self reported Co-Presence scores

B.4 Empathy scores

ANOVA for Empathy Scores

<i>SUMMARY</i>	Count	Sum	Average	Variance
User 1	4	65	16.25	44.9166667
User 2	4	65	16.25	44.9166667
User 3	4	65	16.25	44.9166667
User 4	4	65	16.25	44.9166667
User 5	4	65	16.25	44.9166667
User 6	4	64	16	48.6666667
User 7	4	63	15.75	47.5833333
User 8	4	66	16.5	51
User 9	4	66	16.5	41.6666667
User 10	4	67	16.75	42.9166667
User 11	4	67	16.75	47.5833333
User 12	4	61	15.25	60.9166667
User 13	4	67	16.75	42.9166667
User 14	4	66	16.5	41.6666667
User 15	4	62	15.5	47
User 16	4	67	16.75	47.5833333
User 17	4	69	17.25	40.9166667
User 18	4	62	15.5	47
User 19	4	69	17.25	40.9166667
User 20	4	70	17.5	31
User 21	4	60	15	54.6666667
User 22	4	69	17.25	40.9166667
User 23	4	71	17.75	36.25
User 24	4	70	17.5	51.6666667
Modality: Keyboard, Mouse	24	266	11.0833333	0.34057971
Modality: Kinect for face and upper torso, Keyboard	24	274	11.4166667	2.6884058
Modality: Full face and body motion	24	436	18.1666667	1.01449275
Modality: Full face and body motion with Sensory Fusion	24	605	25.2083333	0.34601449

Table B.4: User Empathy scores

B.5 Mutual Awareness scores

ANOVA for Mutual Awareness scores

<i>SUMMARY</i>	Count	Sum	Average	Variance
User 1	4	95	23.75	121.583333
User 2	4	95	23.75	121.583333
User 3	4	95	23.75	121.583333
User 4	4	95	23.75	121.583333
User 5	4	95	23.75	121.583333
User 6	4	94	23.5	129
User 7	4	93	23.25	134.25
User 8	4	97	24.25	118.916667
User 9	4	97	24.25	122.916667
User 10	4	97	24.25	109.583333
User 11	4	99	24.75	116.916667
User 12	4	91	22.75	154.25
User 13	4	98	24.5	113.666667
User 14	4	99	24.75	124.916667
User 15	4	90	22.5	117.666667
User 16	4	99	24.75	116.916667
User 17	4	100	25	126
User 18	4	93	23.25	109.583333
User 19	4	98	24.5	129
User 20	4	101	25.25	87.5833333
User 21	4	85	21.25	110.25
User 22	4	100	25	126
User 23	4	107	26.75	139.583333
User 24	4	101	25.25	115.583333
Modality: Keyboard, Mouse	24	328	13.6666667	5.36231884
Modality: Kinect for face and upper torso, Keyboard	24	387	16.125	0.28804348
Modality: Full face and body motion	24	731	30.4583333	1.21557971
Modality: Full face and body motion with Sensory Fusion	24	868	36.1666667	3.53623188

Table B.5: User Mutual Awareness scores

B.6 Attention Allocation scores

ANOVA for Attention Allocation scores					
<i>SUMMARY</i>	<i>Count</i>	<i>Sum</i>	<i>Average</i>	<i>Variance</i>	
User 1	4	53	13.25	23.5833333	
User 2	4	53	13.25	23.5833333	
User 3	4	53	13.25	23.5833333	
User 4	4	53	13.25	23.5833333	
User 5	4	53	13.25	23.5833333	
User 6	4	51	12.75	28.9166667	
User 7	4	53	13.25	23.5833333	
User 8	4	54	13.5	23	
User 9	4	54	13.5	23	
User 10	4	53	13.25	23.5833333	
User 11	4	54	13.5	23	
User 12	4	50	12.5	30.3333333	
User 13	4	53	13.25	23.5833333	
User 14	4	54	13.5	23	
User 15	4	50	12.5	30.3333333	
User 16	4	54	13.5	23	
User 17	4	56	14	18.6666667	
User 18	4	50	12.5	23	
User 19	4	56	14	24.6666667	
User 20	4	56	14	18.6666667	
User 21	4	49	12.25	26.9166667	
User 22	4	56	14	18.6666667	
User 23	4	57	14.25	21.5833333	
User 24	4	56	14	22	
Modality: Keyboard, Mouse	24	169	7.04166667	0.47644928	
Modality: Kinect for face and upper torso, Keyboard	24	295	12.2916667	1.43297101	
Modality: Full face and body motion	24	384	16	0.08695652	
Modality: Full face and body motion with Sensory Fusion	24	433	18.0416667	0.21557971	

Table B.6: User Attention Allocation scores

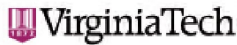
Appendix C : List of body expressions

Participants were given the task sheets for each experimental task and were expected to enact out and guess the expressions within a fixed amount of time for each experimental task, without stating or explicitly alluding to the caption on the list. The expectation was that the number of body expressions successfully guessed and enacted from their designated lists, would increase in tasks with higher interaction fidelity.




Appendix D : IRB Approval Letter

The user study required prior approval from the Institutional Review Board at Virginia Tech.

	<p>Office of Research Compliance Institutional Review Board North End Center, Suite 4120, Virginia Tech 300 Turner Street NW Blacksburg, Virginia 24061 540/231-4606 Fax 540/231-0959 email irb@vt.edu website http://www.irb.vt.edu</p>
---	--

MEMORANDUM

DATE: July 13, 2017 

TO: Ivica Bukvic, Siddharth Narayanan

FROM: Virginia Tech Institutional Review Board (FWA00000572, expires January 29, 2021)

PROTOCOL TITLE: Cinemacraft 2.0: Real time performance capture with Sensory Fusion

IRB NUMBER: 17-669

Effective July 12, 2017, the Virginia Tech Institution Review Board (IRB) Chair, David M Moore, approved the New Application request for the above-mentioned research protocol.

This approval provides permission to begin the human subject activities outlined in the IRB-approved protocol and supporting documents.

Plans to deviate from the approved protocol and/or supporting documents must be submitted to the IRB as an amendment request and approved by the IRB prior to the implementation of any changes, regardless of how minor, except where necessary to eliminate apparent immediate hazards to the subjects. Report within 5 business days to the IRB any injuries or other unanticipated or adverse events involving risks or harms to human research subjects or others.

All investigators (listed above) are required to comply with the researcher requirements outlined at: <http://www.irb.vt.edu/pages/responsibilities.htm>

(Please review responsibilities before the commencement of your research.)

PROTOCOL INFORMATION:

Approved As:	Expedited, under 45 CFR 46.110 category(ies) 7
Protocol Approval Date:	July 12, 2017
Protocol Expiration Date:	July 11, 2018
Continuing Review Due Date*:	June 27, 2018

*Date a Continuing Review application is due to the IRB office if human subject activities covered under this protocol, including data analysis, are to continue beyond the Protocol Expiration Date.

FEDERALLY FUNDED RESEARCH REQUIREMENTS:

Per federal regulations, 45 CFR 46.103(f), the IRB is required to compare all federally funded grant proposals/work statements to the IRB protocol(s) which cover the human research activities included in the proposal / work statement before funds are released. Note that this requirement does not apply to Exempt and Interim IRB protocols, or grants for which VT is not the primary awardee.

The table on the following page indicates whether grant proposals are related to this IRB protocol, and which of the listed proposals, if any, have been compared to this IRB protocol, if required.

Invent the Future

VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY
An equal opportunity, affirmative action institution