12-2022

The Importance of Hand Motions for Communication and Interaction in Virtual Reality

Alex Adkins
adkins4@clemson.edu

The Importance of Hand Motions for Communication and Interaction in Virtual Reality

A Dissertation
Presented to
the Graduate School of
Clemson University

In Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Human-Centered Computing

by
Alex Adkins
December 2022

Accepted by:
Dr. Sophie Jörg, Committee Chair
Dr. Andrew Robb
Dr. Guo Freeman
Dr. Nathan McNeese

Abstract

Virtual reality (VR) is a growing method of communication and play. Recent advances have enabled hand-tracking technologies for consumer VR headsets, allowing virtual hands to mimic a user's real hand movements in real-time. A growing number of users now utilize hand-tracking when using VR to manipulate objects or to create gestures when interacting with others. As VR grows as a tool and communication platform, it is important to understand how the rising prevalence of hand-tracking technology might affect users' experiences.

The goal of this dissertation is to investigate, through a series of experiments, how using hand motions in VR influences our experience when we communicate with others or interact with the environment. In our daily lives hand motions play a major role in interpersonal communication. Our hands can help emphasize or clarify our speech, or even supplement words entirely. When interacting with the world, hands are our primary tool for manipulating objects and performing dexterous tasks. Bringing these capabilities into VR, a space that has so far been lacking in such detailed expression and interaction, may have unexpected effects.

In our first study, we explore the effects of common hand-tracking errors on communication. We use a virtual character who performs charades with detailed hand motions, alter the hand motions, and ask participants to guess the correct movie titles and to rate the personality of the virtual character. We find that the absence of finger motion significantly reduces comprehension and negatively affects people's perception of a virtual character and their social presence. Adding some hand motions, even random ones, does attenuate some of these effects. However, jittering hand motions should be avoided as they significantly decrease user comfort.

Our second study utilizes an in-depth VR Escape Room game to examine the effects of two input modalities (VR controllers vs. hand-tracking) and two grasping visualizations (continuously tracked hands vs. virtual hands that disappear when grasping) on ownership, realism, efficiency, enjoyment, and presence. Many VR users will be interacting in similar immersive game-like experiences, and so we must know if changes in interaction modality still matter when players are focused on a game rather than their hands. Amongst other results, we show that ownership, realism, enjoyment, and presence increased when using hand tracking for interactions compared to controllers in a game-like experience.

In the final study, we implement a collaborative virtual environment that utilizes communication and interaction to examine two input modalities (hand-tracking vs. VR controllers) and how they affect social presence, comprehension, team cohesion, and mental workload. Participants worked together, as embodied virtual Astronauts stranded on the moon, to rank items in order of importance to their survival. We found who participants that used hand-tracking together within the virtual scenario showed trends of higher social presence and lower task workload.

Overall, we show that using hand-tracking and hand motions in VR is beneficial to many metrics that are used to measure the quality of experiences in virtual environments. When using accurate hand motions, people feel more comfortable and embodied within their virtual avatars, or they feel more socially present. We recommend tracking and displaying hand motions in virtual environments if embodiment or communication are the most important criteria.

Acknowledgements

I am lucky to have many people to thank for help, guidance, and support provided throughout the creation of this dissertation. First of all, my advisor, Dr. Sophie Jörg, has been wonderfully patient and supportive during this entire multi-year journey. You have provided much needed guidance and I have learned so much from you over the years. You have been a joy to work for and with. I would not be who I am as a person today without your guidance, patience, and knowledge.

I must also thank various lab mates and collaborators. Ryan Canales, I enjoyed every moment of our banter and snark. This entire experience would have been worse off without you sitting at the desk behind me. I am looking forward to your and Lauren's wedding in April! Lorraine Lin, you set the theme for my time in the lab at Clemson and I learned from you of the importance of being comfortable with who you are. Aline Normoyle, thank you for the encouragement and guidance during my first year at Clemson, you reassured a fresh graduate student and I would recall your advice frequently. Jacob Justice, you brought a new perspective to the lab and I'm glad to have worked with you this past year. Grace Lim and Jackson Henry, you were both so helpful with my final projects and made the lab a very cheery place. Yuting Ye and Massimiliano Di Luca, you both were swell collaborators, I learned a lot from you two.

I also wish to thank my Committee members, Drs. Andrew Robb, Guo Freeman, Nathan McNeese and (previously) Larry Hodges. They were a source of constructive feedback and helpful guidance. Classes I took under these faculty were always a fount of knowledge that greatly aided in this entire endeavor. Additionally, Drs. Andrew Duchowski and Nicholas Widman provided guidance in my first teaching experience, cementing my desire to become an instructor.

On a more personal note I would like to thank various friends and family. My fiancée, Brian Risden, has been a supportive and comforting presence. Whenever I was having a challenging day, Brian was there encouraging me through it. My parents, Sandy Adkins and Stephen

Wright, have patiently listened to long soliloquies about my work and endeavors. They have supported me in my choices and have provided more help than I can ever express proper thanks for. Bet Walsh and Elizabeth Tompkins, thank you for always believing in me and knowing I could do it. Christina Fong, thank you for being a sympathetic and reliable exchanger of memes and rants. Emma Newland and Raquel Robinson, you both have been there forever and will be there forever. I love you!

I have had the honor of making many friends at Clemson during my time here. John Porter III, you have been an endless source of encouragement and inspiration, as well as many fun times and learning experiences. You inspired me to become a teacher, and I would not be on this life path without you. Ayush Bhargava and Divine Maloney, I so enjoyed my time here at Clemson and as a Human-Centered Computing student and much of that is thanks to you two. Rohith Venkatakrishnan, thank you for last-minute proofreading and helpful resources. Roshan Venkatakrishnan, I'm glad you've gotten into hands in VR; I'll have fun watching you finish up your degree! Thanks to you both for being fun and engaging and some of the best brothers anyone could hope for. Catherine Barwulor and Dane Acena, you have been wonderful friends and I wish you all the best in your future endeavors.

Best regards,

Alex Adkins (PhD)

Table of Contents

Title Page . . . . **i**

Abstract . . . . **ii**

List of Tables . . . . **vii**

List of Figures . . . . **viii**

1 Introduction **1**
- 1.1 Research Motivation 1
- 1.2 Research Questions 2
- 1.3 Overview of Studies 3

2 Related Works **5**
- 2.1 Gesture and Hand Motions 5
- 2.2 Communication 8
- 2.3 Virtual Reality 12

3 Study 1: How Important are Detailed Hand Motions for Communication for a Virtual Character **16**
- 3.1 Study Overview 18
- 3.2 Experiment 1: Alteration Experiment 23
- 3.3 Experiment 2: Intensity Experiment 34
- 3.4 Experiment 3: Virtual Reality Experiment 39
- 3.5 Conclusion, Limitations, and Future Work 45

4 Study 2: Evaluating Grasping Visualizations and Control Modes in a VR Game . **49**
- 4.1 Experimental Design 51
- 4.2 Results 58
- 4.3 Discussion 61
- 4.4 Conclusion, Limitations, and Future Work 64

5 Study 3: Survive on the Moon! How Interaction Controls Impact Collaborative VR **66**
- 5.1 Experimental Design 68
- 5.2 Results 75
- 5.3 Discussion 78

6 Conclusions and Future Works **80**

Appendices **84**

Bibliography **111**

List of Tables

3.1 Study 1: Summary of experiments. 18
3.2 Study 1: High-level summary of results. 46

4.1 Study 2: The four experiment conditions. 51
4.2 Study 2: Distribution of participants' gender and age throughout conditions. . . . 53

1 Study 1: Detailed results for the Alteration Experiment. 85
2 Study 1: Detailed results for the Intensity Experiment. 86
3 Study 1: Detailed results for the Virtual Reality experiment. 87
4 Study 1: Full details of the information gathering questionnaire. 88
5 Study 1: full detailed results of Alteration Experiment: Motion Alteration 89
6 Study 1: all Alteration Experiment: Motion Alteration Post-Hoc Comparisons . . 90
7 Alteration Experiment: Avatar Results 92
8 Alteration Experiment: Clip Length Results 94
9 Intensity Experiment: Jitter Intensity Results 96
10 Intensity Experiment: Jitter Motion Intensity Post-Hoc Comparisons 96
11 Intensity Experiment: Popping Intensity Results 98
12 Intensity Experiment: Smoothing Intensity Results 100
13 VR Compare Experiment Results 102
14 VR Compare vs. Alteration Experiment Results 102
15 VR Comfort: Motion Condition Results 105
16 VR Comfort: Motion Condition Post-Hoc Comparisons 107
17 VR Rank: Results 108
18 Study 2: Questionnaire and main results. 109
19 Study 3: Questionnaire and main results. 110

List of Figures

2.1 The Gesture Continuum 7

3.1 The virtual character acting out the short motion “three” under different alteration conditions. 16
3.2 The two avatars used in the experiments. 19
3.3 Alteration Experiment. Main effects of Motion Alteration for Comprehension. . . 28
3.4 Alteration Experiment. Main effects of Avatar and Clip Length for Comprehension. 29
3.5 Alteration Experiment. Interaction effect of Motion Alteration and Clip Length for Perceived Comprehension. 29
3.6 Alteration Experiment, Main effects of Motion Alteration on Perception of Character. 30
3.7 Alteration Experiment, Perception of Character Interaction Effects. 33
3.8 Alteration Experiment, Main effects of Motion Alteration and Avatar on Social Presence. 33
3.9 Intensity Experiment: Example of JitterHigh, PoppingHigh, and SmoothHigh intensities against the Original condition. 35
3.10 Intensity Experiment: Motion Intensity on Motion Comprehension. 36
3.11 Intensity Experiment: Jitter, main effects of Motion Intensity. 37
3.12 Interaction Experiment: Interaction effects. 37
3.13 VR Rank scene. 40
3.14 VR Compare: Main effect of Motion Alteration on Motion Comprehension (left). VR Rank: Average rankings of Motion Alterations in VRRank experiment (right). 41
3.15 Alteration Experiment and VR Compare: Significant differences between the two experiments. 42
3.16 VR Comfort: Main effect of Motion Alteration on Comfort and perceived Naturalness. 42

4.1 Panorama third-person view of the final game. 49
4.2 The Control Modes. 51
4.3 The Grasping Visualizations. 52
4.4 A (near) first-person view of a participant in the Controllers condition during the experiment. 54
4.5 Examples of puzzles. 55
4.6 Main effects for Ownership and Realism. 58
4.7 Main effects for Enjoyment. 59
4.8 Main effect of Control Mode for Presence. 60
4.9 Comparisons of our results with previous works. 62

5.1 An image of the VR scenario. 66
5.2 The virtual Astronaut hands. 69
5.3 The table and operating space that participants use to rank items. 71
5.4 Results for Social Presence 75
5.5 Results for Perceived Comprehension 76
5.6 Results for Team Cohesion 76
5.7 Results for Performance (left), Synergy (mid), and Duration (right). 77
5.8 Results for Task Workloads. 77

1 Alteration Experiment: Motion Alteration Graphs . . . 91
2 Alteration Experiment: Avatar Graphs . . . 93
3 Alteration Experiment: Clip Length Graphs . . . 95
4 Intensity Experiment: Jitter Graphs . . . 97
5 Intensity Experiment: Popping Graphs . . . 99
6 Intensity Experiment: Smoothing Graphs . . . 101
7 VR Compare: Motion Alteration Graphs . . . 103
8 VR Compare vs Alteration Graphs . . . 104
9 VR Comfort Graphs . . . 106
10 VR Rank Graphs . . . 108

Chapter 1

Introduction

1.1 Research Motivation

Virtual Reality (VR) is the virtual immersion into a simulated environment and experience. Advances in VR have allowed for growing access to virtual applications and environments. Today's publicly available VR technologies take the form of head-mounted displays (HMD) with internal sensors, lenses, and screens that create a stereoscopic view into a created virtual scene. Many modern head-mounted displays and their accompanying technologies enable user interaction within the virtual world with virtual items, menus, and even with other users. A primary goal of Virtual Reality is to build fully immersive and presence-inducing environments with seamless interaction. In recent years, more sophisticated HMDs have begun to track hand motions and gestures, which can be used as a way to interact in or with the virtual environment. Additionally, the number of VR users is growing, and many VR consumers have begun to use the technology as a social platform to interact with others in shared virtual spaces. These technologies are brand new and are still being integrated with one another. HMD developers have not yet settled on a standard interaction method, and varying costs and device availability will likely not have users all interacting with the same modality. Errors and technological mismatches have and will happen. Their effects and the full effects of interacting with others in an entirely virtual space are not yet known. As we come to use and rely on these technologies, we must understand what information might be lost or altered when communicating with others in VR.

Advancing technologies and interface development have allowed for more non-verbal communicative techniques to present themselves on the VR platform [133, 182] Users can now use gestures and hand motions when speaking with one another in VR applications such as VR-Chat or RecRoom. Gestures are a major component of human language and communication. They are the movements an individual will make with his or her hands and body as they speak. Gestures are inseparable from talking and are used to orchestrate speech [141], convey ideas and information, and possibly even help us think [77]. Detailed hand motions, which are an integral part of many gestures, play an important role in face-to-face communication to emphasize points, describe objects, clarify concepts, or replace words altogether [143].

Now, users can see each other within the same virtual spaces in VR and communicate with one another virtual-face to virtual-face. These interactions between users in shared virtual environments have only grown in relevance as users quarantine during the COVID-19 pandemic [22] and look to shared VR as an alternative to video-conferencing [11] or in-person events [176, 29].

Virtual simulations are only just beginning to capture real-life hand motions and to display corresponding virtual hands within the virtual environments. As fully-tracked hands are becoming more common within virtual spaces due to consumer availability of hand-tracking technologies, the VR community must understand the effects and consequences of applying hand motions to virtual hand representations in the pursuit of immersion.

1.2 Research Questions

There are unknowns in how important hand motions are for communication and interaction in virtual reality. We seek to investigate some of these unknowns within this dissertation by answering these questions:

- How important are accurate hand motions to non-verbal communication?
  - How important are accurate hand motions for interlocution comprehension?
  - Can hand motions change how we perceive the personality of another person?
  - Can hand motions affect how much we feel another person is socially present?
  - Will inaccurate hand motions induce discomfort in a viewer?
- How important are hand motions and controls when interacting in a virtual space?
  - Do different types of interaction modalities affect a user's perceived realism and ownership of their virtual hands?

  - Can a player's enjoyment of a game be impacted by how they are interacting with the virtual environment?
  - Will interaction controls affect how present a user feels within a VR game?

- How important are hand motions and controls when communicating and interacting with someone else in a virtual space?
  - How important are accurate hand motions for perceived comprehension in VR?
  - Can higher-fidelity hand motions increase how much a person feels another is socially present?
  - Will interaction modalities affect teamwork?

To answer these questions we investigate the effects of hand-tracking on communication and interaction in VR using state-of-the-art hand-tracking technologies. We design and implement a series of studies focused on the impact of hand motions on communication between virtual reality users, interactions with virtual objects, and combinations thereof.

1.3 Overview of Studies

In our first study, we investigate the consequences of errors in hand and finger motions on comprehension, character perception, social presence, and user comfort. We conduct three perceptual experiments where participants guess words and movie titles based on motion captured movements, as in the game charades. We introduce errors and alterations to the hand movements and apply techniques to synthesize or correct hand motions. We collect data from more than 1000 Amazon Mechanical Turk participants in two large experiments, and conduct a third experiment in VR. As results might differ depending on the realism of the virtual character, we investigate all effects on two virtual characters of different levels of realism. We furthermore use shorter clips and longer motions in our experiments.

Amongst other results, we show that the absence of finger motion significantly reduces comprehension and negatively affects people's perception of a virtual character and their social presence. Adding some hand motions, even random ones, does attenuate some of these effects when it comes to the perception of the virtual character or social presence, but it does not necessarily improve comprehension. Slightly inaccurate or erroneous hand motions are sufficient to achieve the same level of comprehension than with accurate hand motions. However, they might still affect the viewers' impression of a character. Finally, jittering hand motions should be avoided as they significantly decrease user comfort. These results have important implications for the animation of virtual characters, for the creation of effective virtual agents, and for the development of VR technologies. This study has been accepted to ACM *Transactions on Graphics (TOG)* peer-reviewed journal.

Most studies investigating hands in VR require participants to perform repetitive tasks. For the second study, we investigate if results of such studies translate into a real application and game-like experience. We designed a virtual escape room in which participants interact with various objects to gather clues and complete puzzles. In a between-subjects study, we examine the effects of two input modalities (controllers vs. hand-tracking) and two grasping visualizations (continuously tracked hands vs. virtual hands that disappear when grasping) on ownership, realism, efficiency, enjoyment, and presence.

Our results show that ownership, realism, enjoyment, and presence increased when using hand-tracking compared to controllers. Visualizing the tracked hands during grasps lead to higher ratings in one of our ownership questions and one of our enjoyment questions compared to having the virtual hands disappear during grasps as is common in many applications. We also confirm some of the main results of two studies that have a repetitive design in a more realistic gaming scenario that might be closer to a typical user experience. This work was presented at ACM *Symposium on Applied Perception (SAP)* in 2021, and subsequently published in ACM *Transactions on Applied Perception (TAP)* peer-reviewed journal.

For the final study, we create a scenario that combines communication (investigated in the first study) and interaction (examined in the second study) to investigate if and how different input modalities (hand-tracking vs. VR controllers) influence social presence, comprehension, team cohesion, and mental workload within a collaborative VR scenario. We find trends of input modalities on social presence and mental workload.

These studies highlight the importance of accurate hand motions in virtual reality when considering communication between individuals and interactions for users. We show that accurate hand motions are beneficial for character perception and comprehension, as well as increasing social presence and user comfort. Using hands to interact in VR also increases user ownership, realism, enjoyment, and presence, and potentially reduces a user's task workload. There are many benefits to using hand tracking and hand motions in VR, and it should strongly be considered as the new standard type of interaction in VR.

Chapter 2

Related Works

This dissertation examines the importance of hand motions for communication and interaction in virtual reality. To fully comprehend the scope of this dissertation we must examine these topics in detail. This section describes the current state of the relevant literature and prevailing theories. We describe literature on hand motions in the form of gestures, the motions our hands make when we speak; we then consider current theories of non-verbal communication and how hand motions and gestures contribute to these. Finally, we examine the current state of interactions in virtual reality and the literature gaps that surround these topics.

2.1 Gesture and Hand Motions

2.1.1 What are Gestures?

Gesture carries many definitions. This dissertation considers gesture in the context of bodily movement with a specific focus on hand motions. McNeill [141] defines gesture as "the intrinsic imagery of language." Goldin-Meadow [77], also focusing on the hands' contribution to gesture, defines gesture as "the way we move our hands while we speak." Both definitions highlight the prevalent idea that gestures are closely related to language and speech.

In his book "Why We Gesture" [141], McNeill hypothesizes that gestures orchestrate speech. Various theories follow this hypothesis and examine gesture and speech conjointly, though differ on whether to treat gesture and speech as one unit or two. Vygotski [192] codes gesture and speech together into a minimal combined unit, a perspective that McNeill also supports [141], whereas Kita's Information Processing Hypothesis [117] treats gesture and speech as independent cognitive streams. Either way, gesture and speech are inseparable, and language as we know it cannot exist without both. However, a gesture's purpose and execution may vary by context.

2.1.2 Gesture for the Individual

To an individual, gesturing is part of the thought process, not just expressions of thought but an act of thought itself [65]. McNeill [141] considers gestures as a component of a thought coming into being onto the physical plane [141] and Goldin-Meadow [77] hypothesizes that gesturing reduces a speaker's cognitive load. The rate of gesture increases when tasks become more difficult [80, 166], when speakers are describing a scene from memory [59, 200], and when individuals reason about problems [15, 54, 59]. Additionally, when speaking becomes difficult, or communication is unsuccessful, a speaker increases their number of gestures and effort towards producing them [139, 94, 71, 134]. In conversational settings, gestures can also substitute missing vocabulary [24] or help explain social structures [66].

The effect of speech and gesture works both ways. McNeill's [141] proposal that gesture orchestrates speech is supported by works by Ríme [161] and Alibali [16], which restrict speakers' movements and found decreased vividness in their descriptions and increased rate of perceptual explanations. Based upon these and further works, Goldin-Meadow [77] hypothesizes that gesture promotes spatial thinking and can affect what speakers are going to say.

Various studies have shown that life-long blind individuals gesture at the same rate as sighted individuals [102, 103, 101], which suggests that performing gestures is not a learned ability. Individuals gesture even when their conversational partner cannot see them and the gestures serve no communicative purpose [199]. However, gestures are usually performed for a listener, and will be adjusted accordingly.

2.1.3 Gestures for Communication

We have established that performing gestures is an ingrained ability that serves to facilitate thinking and speech, but the gestures themselves also often work to augment a speaker's point. Gestures are a major component of human communication [139, 140, 112, 76] and a key dimension of face-to-face dialogue [27]. Face-to-face dialogue has been proposed as the "fundamental site" of human language as it is our earliest and most prominent way of communicating [28, 25]. In face-to-face dialogue, conversational partners use gestures to "create and convey shared meaning" [27, 26]. These gestures take many forms: *iconic* where a person acts out an action as they describe it; *metaphoric* where a person treats an abstract concept as a physical object, for example, using a giving gesture to indicate generosity; *beat* where the hand moves with the rhythm of speech; and *deictic* which refers to pointing gestures [139]. Iconic, metaphoric, and deictic gestures can complement a verbal description and add to a viewer's understanding of an event, setting, or object [67] and some gestures can substitute words entirely [24].

Spontaneous gestures that accompany speech are classified as "gesticulations" under a gesture continuum [111], first defined by Kendon, formalized by McNeill [139], and outlined in Figure 2.1. This continuum orders gestures based on the necessity of speech, and the presence of language properties and regulated gestural signs.

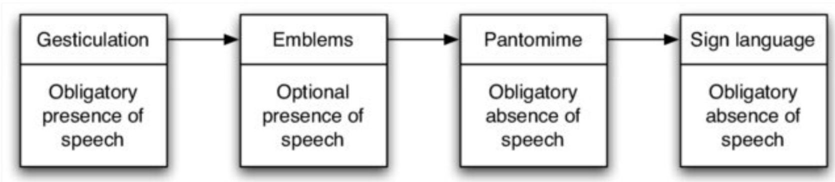


Figure 2.1: The Gesture Continuum

Language-like gestures fall between gesticulation and emblems, and usually serve to replace a word in a spoken statement; Emblems are known gestures with specific meanings (e.g. the thumb and forefinger OK sign); Pantomimes depict object or actions, as in the game charades, which is used to evaluate the quality of non-verbal communication in Chapter 4; Sign Languages are entire languages unto themselves composed almost entirely from gestures, though highly dependent on facial expressions. Each type of gesture provides some level of meaning in communication. In the case of Sign-Language, and Emblems and Pantomimes to some degree, the gestures themselves are the primary method of communication between individuals.

Kendon [112] further illustrates several kinds of contributions gestures make towards meaning and includes techniques of classifying gestural representation. To observers, an individual's gestures can imply meaning [112], change the observer's perception of the individual [195, 149], or even indicate deception [52, 61]. A lack of gestures may decrease or alter information comprehension, especially when communicating about shapes and objects [79, 55].

There are many works that show that gestures and hand motions are a vital aspect of human communication, but how do we define "communication" and why is it important?

2.2 Communication

2.2.1 Interpersonal Communication

In the previous section we consider communication as the interaction between individuals, where they are using gestures to emphasize or supplement their speech. While communication theory itself is a broad topic that considers a wide array of contexts, we continue to examine communication from the interpersonal perspective throughout this dissertation. This is called interpersonal communication, defined by Griffin [83] as "the ongoing process of using verbal and nonverbal messages with another person to create and alter the images in both of [the communicators'] minds."

An early theory of interpersonal communication is symbolic interactionism, developed by George Herbert Mead and his students [142, 41]. This theory proposes that meaning is not inherent in the world, that people create it and interpret it themselves, and then act on the basis of those internal meanings. As such, even the concept of the self and others are meanings created through an internal and subjective lens. When communicating with others a person may change their behavior to conform to expectations, and their impression of the other person is warped by their own perspective. A number of other interpersonal communication theories developed later, some using Mead's original premise as a base template. Berger's Planning theory of communication [33] explores how a person's mental plans influence interpersonal communication, whereas the older Goals-Plan-Action theory [62] accounts for an individual's personal goals that will shape the planning stage of communication. Another theory proposes that imagined interaction, int*ra*personal communication, is the foundation of all other types of communication, as communication processing happens within the individual [96].

A prevailing theme of these theories is that interpersonal communication is enacted and perceived by the people involved, and that any communication will be understood through an individual's own perspective and experience. Any changes made to the method of interpersonal communication, such as communicating through virtual reality instead of via face-to-face discourse, will by nature change the composition of the communication itself.

Claude Shannon, working for the Bell Telephone Company in the 1940s, expanded this concept into considering communication as a form of information processing [171], where a 'transmitter' sends a message through a channel to a 'receiver.' However, the information transmitted picks up 'noise' along the way, thus degrading the quality of communication. If we consider virtual reality as a channel or system of communication, discussed by Biocca and Levy as early as 1995 [38], then we must consider that using this channel will alter the communication that is traditionally transmitted face-to-face.

2.2.2 Media Richness and Communication in Virtual Reality

Media richness theory emerges as a major theory within the literature when examining computer mediated communication. Ishii *et al.* [100] consider modern technologies when expanding on Daft and Lengel's [57, 58] initial media richness theory, which proposes that a media's "richness," as determined by immediate feedback, multiple cues, language variety, and personal focus, should match the equivocality of the task. Ishii *et al.* [100] highlight the inconsistencies found within the literature when considering this recommendation, especially with newer technologies. For example, Aritz *et al.* [19] conducted a large study among many virtual teams and found that team members identified rich media channels as being more effective, even with less equivocal tasks, and Kahai and Cooper [180] found that richer media can have positive impacts on decision making quality. Additionally, Ledbetter *et al.* [124] found that enjoyment of a media was a better predictor of media usage frequency than task type.

Virtual reality, a medium with a high level of richness due to its high immersion level design, has been found to be both a rich channel of communication with minimal differences with face-to-face communication [10], as well as induce a high measure of enjoyment [193]. This higher level of media richness may indicate increased media usage according to recent media usage trends [100]. However, we must consider the negative effects of VR's high richness. Kolomaznik *et al.* [119] suggest that the potential for motion sickness and additional technological overhead leads to VR not being a remedy for engagement with teamwork. Some users are more prone to motion sickness and cybersickness than others [183], altering those individuals' perception of the VR experience by reducing their presence [198] and possibly hindering effective communication. As the technology develops we can hope to see improvements with these issues.

2.2.3 Non-verbal Communication

While gestures are covered extensively in Section 2.1, there is more to non-verbal communication than just hand motions, both conceptually and practically. Human non-verbal communication is also conveyed through facial expression, body language, proxemics, and more. Non-verbal cues are ubiquitous across communicative interactions and are meaningful across all face-to-face dialogues [30, 122, 73, 106, 51]. They have even been found to be more important than verbal conveyance, especially when considering deception [47, 46, 191, 154] or important relational communication [18]. How non-verbal communication is conveyed and perceived is dependent on many factors, including personality [74], age [69], gender [85], culture [135], and more. Despite differences in conveyance across populations, non-verbal communication has been found to be vital in conveying meaning and improving comprehension [40, 48, 43, 205].

Body and hand language influence people's judgements about personality and mood. Extroverts tend to gesture more often and with faster, larger hand movements [158] and viewers can detect angry hand gestures, regardless of the hand model's appearance [164]. Furthermore, editing the timing and poses of hand motions or gestures can affect the perception of a virtual character [107, 175, 195]. Such factors can even be incorporated into procedurally animated characters so that viewers clearly perceive the characters animated with different motions as having distinct personality types [149, 175]. These perceptions can influence how individuals perceive each other and thus how they communicate [147]. As technology develops, non-verbal communication is now happening in the virtual world.

2.2.4 Non-verbal Communication in Shared VR Spaces

With the growing popularity of consumer level headsets, shared VR spaces and multiplayer games have become common. Communication between individuals within VR happens primarily over text, voice chat, icons and cues [123, 128], or with non-verbal interactions such as gestures [39, 133], facial expressions [182, 97], or proxemics [42, 91, 44]. An old adage claims that 90% of communication is non-verbal, though research shows this number actually varies between 70-93% at any given time [143]. Research suggests that non-verbal body language and gestures are important in VR as well. Whole-body avatars have been shown to improve communication, presence, and usability when using a Head-Mounted Display (HMD) in VR and using them is somewhat comparable to face-to-face interaction [64, 147, 176, 194].

Because hand motions are often difficult to capture in real-time, the fingers are rarely considered in VR communication experiments [64], though hands heavily influence non-verbal communication. Of particular importance are gestures [76, 104, 140], but without hand tracking, gestures are constrained to what users can choose via interface or pantomime with controllers, often without meaningful control [182]. Only now are gestures entering into VR spaces, but few works examine how they might matter in group VR settings.

2.2.5 Communication for Teams

Teams and groups of people can come together for many reasons. Harris and Sherblom [89] list four common reasons: interpersonal attraction, personal need for affiliation, meaning or identity, commitment to group goals and activities, and assignment to the group by someone else. Whatever the reason a team is formed, team members will interact with and inter-depend on one another and possibly produce synergy, the ability to adapt with and often improve on individual members' performances [49]. These interactions and subsequent teamwork are possible due to communication between team members, which can take many forms and serve several functions. In a group setting verbal communication often functions to define and order group tasks, determine how tasks will be completed, and also to talk about the group itself. Non-verbal communication in groups can determine the quality of relationships between group members and will serve to add meaning and emphasis to interactions [89].

The quality of communication in a group setting can influence team cohesiveness. Team cohesiveness is the team members' desire to remain in the group and their willingness to work to accomplish the team's goals [113, 203]. Team cohesiveness is a popular metric in determining team quality and is a predictor of team performance and success [68, 178]. It has been used to investigate the effects of computer mediated communication on teamwork and group effectiveness.

2.2.6 Collaboration and Team Cohesion in Virtual Reality

As VR technology becomes more accessible, users have begun to utilize shared VR spaces to collaborate and coordinate with one another. While embodied VR is reportedly more like face-to-face interaction than other collaborative technologies [11], teamwork and collaboration in VR has rightfully been investigated as its own discipline.

Freeman *et al.* [72] explored collaboration in social VR with a series of in-depth interviews. They highlight the importance of embodiment in facilitating virtual collaboration and the value of realism and naturalness when collaborating with others. Kolomasnik et al. [119] found that VR interventions did not improve students' attitudes towards teamwork, though Wienrich et al. [202] did find that requiring collaboration for a task between VR users improved social presence and cooperation.

Team cohesion, a metric of team appeal and performance [95, 53, 197], has been used to investigate the potential of teamwork and collaboration in virtual reality. Torro *et al.* [186] found that team cohesion was limited in social virtual reality, but that developing VR technologies will likely soon mitigate these effects. Liszio et al. [131] found that interactive social entities increase team cohesion in a virtual environment. VR is also being proposed as a way to study team cohesion and mental health for long-term space missions [163].

2.3 Virtual Reality

2.3.1 Measuring the VR Experience

A major metric for evaluating experiences in VR is the sense of presence, which is the "sense of being" within a virtual space [172]. Presence can be broken down into three types: self presence, social presence, and physical presence [93, 125, 167]. Self presence, the feeling of embodying your virtual avatar, social presence, the feeling that others also exist within the virtual space with you, and physical presence, the sense of the environment being an actual space around you, all play a critical role in inducing presence within VR users [34, 125] and are often used as measures for VR research experiences as presence affects satisfaction, immersion, enjoyment, and more in virtual environments [45, 188].

Self presence can affect real life health and virtual social connections [31, 105]. Embodiment, the sense of being inside and controlling a body; body ownership, the sense of owning the body; agency, the sense of controlling your movements [115]; and further measures often serve as adjacent investigators to self presence [34]. Support for the Virtual Hand Illusion [208], where body ownership is induced for a fake virtual hand, suggests that the appearance of an avatar's hand does have an effect on perceived ownership and agency [17, 129], and further work shows that how a hand is steered in a virtual scene (controller vs. hand-tracking) also influences ownership [130].

Social presence [36] predicts learning satisfaction [98, 160], attraction [126], trust, enjoyment [92], and is often associated with positive communication outcomes [152]. A recent publication finds that social presence of others is affected by hand appearance in a collaborative virtual space [207], however there exists no work investigating how hand input may affect social presence in real-time communicative virtual reality. Further research into hand motions on presence is limited, but evidence shows higher reported self and social presence and enjoyment with intuitive controls over buttons and gamepads [14], though indirect button input was still preferred for some actions [155].

2.3.2 Hand Tracking for Virtual Spaces

Recent advances in consumer technologies have allowed hand motions to become a viable interaction method in VR. The Leap Motion device enabled hand tracking in consumer VR as early as 2012 [153], though its modular design was more popular with developers than the average user. Consumer headsets with built-in hand tracking, such as the Oculus Quest (2019), Oculus Quest 2 (2020) [1], and Valve Index (2019) [4], enable users to easily interact with their environments through hand tracking and representation. Several games have been adapted or crafted for users to experience these new technologies fully [21, 9], allowing what was once a rare interaction method to enter the mainstream. These headsets track and represent the users' hand motions in real-time using machine learning [87], and other hand-tracking techniques are still being explored [165, 75].

Existing hand-tracking methodologies involve marker-based optical tracking, sensored gloves, markerless optical tracking, and depth cameras [201, 109]. The Leap Motion uses depth cameras [8] and the Oculus Quest uses markerless optical tracking [7] but, while accessible, these methods are limited in their scope. Trained neural networks can account for occlusions [184, 127], but the cameras of these methods must be focused within a small area, and as such can only track hands within a limited region. Marker-based tracking and sensored gloves can allow for freer and more natural movement, but come with their own drawbacks. Marker-based tracking systems can be very accurate but require extensive setup and expensive cameras [118]. While sensored gloves do not have occlusions, they do not provide spatial positions, a major limitation for visualization and interaction within VR, and can be expensive and cumbersome [63].

2.3.3 Interactions in Virtual Reality

In current applications, controllers are still the consumer industry standard for interactions in virtual environments. Research suggests that more natural interaction techniques can increase user enjoyment and presence even if performance might be reduced [138, 146, 174], which could imply that using accurately tracked hand motions would be preferable to using controllers. A study by Lin *et al.* [130] directly examines the differences between controllers and high-fidelity hand tracking. They measured the effect of using controllers or tracked hands and of having different hand sizes on ownership and the virtual hand illusion. Among other results, they found that tracked hands lead to higher levels of ownership and perceived realism but to poorer perceived efficiency and longer task times.

Additional work has investigated how interactions with hands are implemented within VR, especially when grasping objects. Since the haptic feeling of grasping a real object is missing, several ways of giving users feedback when grasping have been suggested. Multiple studies have demonstrated that visual feedback has several advantages over no feedback [50, 120, 157, 190]. For example, users tend to prefer visual feedback over no feedback while grasping a virtual object [157, 190] and visual grasping feedback can improve efficiency [120]. Prachyabrued and Borst [157] investigated several visual feedback techniques for virtual grasping with a repetitive grasp and release task and found that preventing the virtual hand from entering the virtual object while grasping was preferred. Canales *et al.* [50] confirmed that preventing hand-object interpenetrations was preferred with a similar repetitive procedure, though grasping performance was reduced. Displaying the virtual hands when grasping lead to higher performance and perceived hand ownership than hiding the hands during a grasp.

2.3.4 The Effect of Interaction Control Methods on Communication

Few works consider both communication and interaction controls together in Virtual Reality. Aseeri & Interrante [20] develop a scenario where a dyad, one participant and one researcher, conducts a series of collaborative tasks such as ranking the importance of survival items and playing charades. They use three avatar representations: No Avatar, where both players are only represented by virtual controller models, Scanned Avatar, where the researcher had a 3D scanned model of themselves as their avatar and the participant had a gender and race matched avatar, and Real Avatar, where both players were viewed with a green screen background and then inserted into the virtual environment. This study examines the effects of avatar representations on communication satisfaction, social presence, interpersonal trust, and attention to behavioral cues. Aseeri & Interrante found that the Scanned and Real Avatars led to higher reported social presence and that a majority of participants preferred the Real Avatar.

Wu *et al.* [204] present a system that is both collaborative and includes interaction controls. This project compares a highly expressive system that includes high fidelity expressions, hand, and body movements created from using an advanced fusion algorithm to combine the inputs of many cameras, to a low expressive system that uses consumer-level body and expression tracking and only tracks the hand position via controller. Participant dyads each used one system and played charades. Wu *et al.* found that participants interacting with the highly expressive avatar felt higher social presence and attraction and performed better at the task.

Abdullah *et al.* [11] compare videoconferencing (e.g. Zoom) to shared VR spaces with full-body (including hands) tracking. They use various collaborative tasks. *e.g.* laying out an apartment floor-plan and planning a party to investigate behavioral differences: gaze, gestures, and conversational turns. This study found that videoconferencing required more effort to maintain the social connection, reduced conversational overlap, increased self-adaptor gestures (gestures not designed to communicate, often indicating increased anxiety [196]), and decreased deictic gestures (pointing). This study measures co-presence and clear communication but does not specifically target hands or investigates hand-tracking.

Smith & Neff [176] compare interactions in face to face, embodied VR, and non-embodied VR scenarios. A dyad of participants perform collaborative tasks such as negotiating an apartment's room uses and planning out furniture placement. The embodied VR condition uses marker-based optical tracking to capture and render the participants' body movements onto a virtual avatar, and the hands are tracked and controlled with VR controllers. The non-embodied VR condition only uses the VR controllers. This study found that embodied VR induced higher reported social presence and that the non-embodied condition led to loneliness and degraded communication.

While several studies that consider different interaction controls in shared virtual environments are listed above, none of the existing literature explicitly looks at the effects of hand motions and hand-tracking on communication in VR. Hand-tracking is growing as a way to interact with VR spaces. This work endeavors to investigate the importance of hand-tracking when communicating in VR, and examines the effects that different interaction controls might have when individuals are interacting in VR.

Chapter 3

Study 1: How Important are Detailed Hand Motions for Communication for a Virtual Character

(a) Original (b) Reduced (c) Jitter (d) Popping (e) Smooth (f) Passive (g) Random (h) Static

Figure 3.1: The virtual character acting out the short motion “three” under different alteration conditions. (a) Original corresponds to highly accurate motion capture; (b) Reduced only uses the information of the finger tips of the thumb, index, and pinky; (c) for Jitter random noise is added to the primary rotational axes; (d) Popping has the fingers occasionally freezing and then jumping back into place (e) in Smooth the motions are averaged over several frames; (f) in Passive finger motions only respond to gravity. In the shown frame, the fingers droop forward; (g) for Random we replace the hand motion with unrelated gestures; (h) Static uses a static hand pose, so the fingers do not move at all. We hide the face of the realistic avatar so that the lack of facial animation does not distract the viewer.

In this first study we investigate the role of detailed hand motions when conveying content. The results and conclusions of this work strengthen the notion that the presence of hand motions does matter when communicating with virtual characters, and that a complete lack of hand motions is detrimental. Higher fidelity hand motions seem to improve comprehension, perception of character, social presence, and user comfort, though any hand motions at all can mitigate at least some negative effects on social presence and user comfort.

Our study consists of three experiments. To explore several of the possible hand motion alterations, we use Amazon Mechanical Turk's ability to gather large numbers of participants in our first two experiments, in which participants watch videos of our stimuli. We then repeat a small subset of the conditions in a third experiment in VR. Our experiments are based on charades. Charades motions have correct answers that make for straightfoward evaluation; a multitude of incorrect answers will indicate a severe effect on communication fidelity. We hide the face of the Realistic virtual avatar with a Head-Mounted Display (HMD) so as not to distract participants with imperfect facial animations, and to avoid a confounding factor due to the facelessness of the Mannequin avatar. Our stimuli is pre-recorded which allows us to thoroughly post-process the detailed hand motions and guarantee the best possible quality for our baseline condition. This method furthermore gives us full control over our stimuli and ensures that each participant sees the same gestures.

In our first experiment, the **Alteration Experiment**, we alter the hand motions of a motion captured character to simulate errors or changes that would typically happen when creating, tracking, or post-processing hand motions (Figure 4.1). This experiment asks: How do hand animation and avatar appearance affect the viewer's ability to understand the character? How do they affect the viewer's impressions of the character? Are some motion errors more acceptable than others? Can we create acceptable hand motions without any data? In our second experiment, the **Intensity Experiment**, we ask: Which intensities of specific alterations, namely Jitter, Popping, and Smooth, are acceptable? What are acceptable thresholds for these errors? We vary their intensity from subtle to extreme to observe their effects. In the third experiment, the **VR Experiment**, we verify some of our results in a virtual environment and evaluate participants' comfort level when watching the character in VR.

Full detailed results are included in the Appendix.

Experiment	Design	Avatar	Clip Lens	Motion Alts	Total Conds	Ref	Valid
1. Alteration	Btwn	MQ/Real	Lng/Shrt	8	32	Sec. 3.2	871
2. Intensity	Btwn	MQ/Real	Lng/Shrt	10	40	Sec. 3.3	1198
3. VR Compare	Btwn	Real	Shrt	Orig/Stat	2	Sec. 3.4	31
VR Comfort	Wthn	Real	10s clips	14	14	Sec. 3.4	31
VR Rank	Wthn	Real	Lng	8	8	Sec. 3.4	31

Table 3.1: Study 1: Summary of experiments. Both the Alteration and Intensity experiments evaluate participants' comprehension, perception of the character, and social presence based on video clips. VR Compare investigates if those concepts are perceived in the same way in a virtual environment and repeats two conditions of the Alteration experiment in VR. VR Comfort and VR Rank examine participants' comfort level when watching the character in VR and establishes a preference ranking for all conditions.

3.1 Study Overview

Our study consists of three experiments: Alteration, Intensity, and VR (see Table 3.1). In our experiments, among other questions, participants are asked to guess acted movie titles or words displayed on a virtual character, in a similar way to the game "charades" where players pantomime words or phrases.

For both the Alteration and Intensity experiments, our independent variables are the Motion Condition (Motion Alteration or Motion Intensity), Avatar, and Clip Length. Our dependent variables are the participants' comprehension, perception of character, and social presence. Both experiments use a between-group design so that no participant would be asked to guess the same movie or word more than once. Therefore, each participant viewed a series of clips with consistent motion condition, avatar, and clip length. For both experiments, participants were recruited online through Amazon Mechanical Turk.

The VR experiment was conducted in-person at Clemson University and consists of three parts: VR Compare, VR Comfort, and VR Rank. VR Compare repeats the Alteration Experiment with a small subset of conditions. VR Comfort and VR Rank evaluate whether the hand motions influence how comfortable participants feel while watching the character: VR Comfort asks participants for comfort ratings and VR Rank asks them to rank motion conditions based on comfort in direct comparison. The independent variable in all three parts is the Motion Condition. The dependent variables of the first part are the same as in the Alteration and Intensity Experiments; in the second and third part they are the participants' comfort and ranking.

Figure 3.2: The two avatars used in the experiments. *Left*, a stylized wooden mannequin. *Right*, a realistic character wearing a head mounted display.

Our hypotheses are as follows:

- **H1, Comprehension**: Partially missing or inexact hand motion data reduces participants' comprehension of a character. Complete absence of motion data reduces it further.
- **H2, Perception of Character**: Changes to hand motions and character appearance will affect the perception of a character.
- **H3, Social Presence**: Less natural hand motions or a less realistic character will reduce social presence.
- **H4, Comfort**: Less natural hand motions will make people feel less comfortable.

3.1.1 Stimuli Creation

We captured a set of charades using an 18-camera Vicon optical motion capture system specifically set up to accurately capture the detailed hand motions of a standing performer. An actor wore 60 optical markers on his body and 24 markers on each hand. We then asked him to pantomime several movie titles. The actor was told that the virtual character's face would not be animated and that any facial expressions would be lost, so that he focused on his body motions when performing the charades. After verifying which movies could be guessed well based on videos of the capture, we labeled the markers of six movies and computed the skeletons for the body and each hand separately for highest possible accuracy. This process produced three separate joint skeletons – one each for the body, left hand and right hand – which were aligned using aim and point constraints and combined by reparenting each hand to the body's elbow joint. This approach ensured that the captured hand motions were not modified in any way and stayed as accurate as possible. We furthermore configured our virtual characters to match the captured skeleton rather than using retargeting, which can generate slight inaccuracies.

We use two character models to display the motions: a mannequin and a realistic avatar (see Figure 3.2). The realistic avatar wears an HMD to hide the non-animated face, so that the lack of facial animation does not distract viewers' attention from body language, and to equalize conveyed information between the two avatars as facial animation would otherwise be a confounding factor (which would invalidate any conclusion on the influence of the avatar). We experimented with several different face-hiding options to find one that would be as natural and inconspicuous as possible, including blurring, using a black rectangle, and hiding the face with various objects.

The HMD was chosen as it was the most unobtrusive to our pilot participants and did not seem to divert people's attention from the task.

With this procedure we created six long motions (movies) and 15 short motions (words), which are subsets of the long motions. The long motions last between 28 and 80 seconds (mean: 39.1s) and represent the movies Back to the Future, Eat Pray Love, The King's Speech, The Lion King, The Pianist, and The Three Musketeers. Each charade starts with several emblems. Five of our charade motions start with the emblem for movie (right fist describes vertical circles close to the head as if operating an antique video camera while one looks through the left hand that is shaped like a cylinder to represent the lens) and one charade starts with the sign for book (flat hands are opened similar to the pages of a book). Then the number of words is indicated by showing the corresponding number of fingers. Further signs can be used to clarify which word is being pantomimed (first, second, third, fourth) and which syllable of a word is being described. The short motions are between 1 and 17 seconds long (mean: 6.6s) and include individual words from the movie titles such as Eat, Lion, or Three. We include thirteen short motions in our analysis; the motions "four" and "two" were included as attention checks. The long motions give participants more time to notice errors and form an impression of the avatar. The short clips give participants less redundancy to guess the meaning of the motions and allow us to get insights on how the motion alterations might affect individual gestures and on how quickly participants might form an impression of the avatar.

We implemented each motion condition as a filter over the original motion in Unity. Finally, for each condition, we exported videos from Unity using the RockVR Video Capture Unity plugin and trimmed them with FFmpeg. The videos had no sound.

3.1.2 Measurements

Our goal is to measure the effect of changes in hand motions on people's comprehension, their perception of the virtual character, their social presence, and their comfort level. The full questionnaire can be seen in the Appendix, Table 4.

Motion comprehension is based on how well participants guessed the movie titles or words. Their answers were rated by two researchers on a scale of 0 - incorrect, 0.5 - partially correct, or 1 - correct. A third researcher solved any discrepancies. Guidelines for rating were established beforehand. If participants guessed a variation of the correct short word, we labeled their response as correct (e.g., "eat" and "eating," "pray" and "praying"). If they wrote down an answer with similar meaning to the correct one for the short motions or if they guessed parts of the long movie title, we labeled their response as partially correct (e.g, any words in "Eat, Pray, Love"). Answers that seemed straight forward based on the animations were also judged as correct (e.g., "monster" for "lion" or "shotgun" for "rifle" for the short clips, movies about famous composers for "The Pianist"). We averaged participants' scores into a final motion comprehension score between 0 and 1. In the Alteration and Intensity experiments, the averages ranged from 0 to 1, with a mean of 0.55. Out of 11882 answers given, 654 had discrepancies between the two initial researchers (5.5%). Inter-rater reliability was very high with an unweighted Cohen's κ of 0.90.

Our perceived comprehension measure is the mean of two 7pt-Likert scale questions that asked participants to judge how well they thought they understood the virtual character. These questions were adapted from Biocca *et al.*'s Networked Minds Measure of Social Presence Questionnaire [37].

To evaluate the perception of the virtual character, we use McDonnell *et al.*'s Perception of Virtual Character questionnaire [136] as well as the Ten Item Personality Inventory (TIPI) [78]. The TIPI questionnaire includes two measures for each of five personality traits; one question measures the personality trait positively, the other negatively. The questionnaire is based on the Big Five model of personality that has been common in psychology since the 1990s [56]. It measures extraversion, agreeableness, conscientiousness, emotional stability (reversed neuroticism), and openness to experience. For analysis we follow the procedure set by Gosling *et al.* and flip the negative measure, then take the average of the two values as the final measure for each personality trait.

Social presence was evaluated based on the questions by Nowak and Biocca [150]. Our social presence measure is the mean of the five social presence questions.

Finally, in VR Comfort, we asked participants to rate how comfortable they would feel interacting with this character for an extended period of time. In VR Rank, we asked them to rank how comfortable they would feel interacting with each character from most comfortable (1) to least (8).

3.2 Experiment 1: Alteration Experiment

The Alteration Experiment examines the role of hand animation accuracy and character appearance on participants' comprehension, perception of the character, and social presence. A between-group design was used, where each participant saw either 15 short clips or 6 long motions on one avatar (out of 2) with one motion alteration (out of 8), leading to a total of 32 different conditions: 8 (Motion Alteration) $\times$ 2 (Avatar) $\times$ 2 (Clip Length).

3.2.1 Motion Alterations

Our baseline motions are the original, unmodified motion captured data (Original) and the complete lack of hand motion (Static). Additionally, we created six motion alterations based on typical errors in the motion capture process or on methods to synthesize or post-process motion data. Based on our results, we realized that our alterations can be grouped into three categories: Full motion data displays the fully accurate motion data, Partial motion data represents data that has been altered from the captured data, and No motion data includes conditions where no information on the hand motions is used and hand motions are either lacking or synthesized from scratch. In the following, we summarize and detail the eight motion conditions.

- Full Motion Data
  - **Original**: unmodified motion captured data
- Partial Motion Data
  - **Reduced**: simplified motion capture
  - **Jitter**: random noise
  - **Popping**: periodic freezes
  - **Smooth**: moving average
- No Motion Data
  - **Passive**: passive hand motion
  - **Random**: unrelated motion capture data
  - **Static**: no movement

Original. Original corresponds to the detailed, unaltered motion captured motion. These motions were recorded with a high-fidelity motion capture system and manually post-processed. This quality can typically not be achieved with real-time, consumer level equipment (yet). It is our most accurate motion.

Reduced. The reduced condition simulates a reduced marker set from Hoyet *et al.* [99], assuming only 6 markers, 2 each on the thumb, index, and pinky fingers. We use the markers to get the fingertip positions for the index, pinky, and thumb. The fingertip positions for the middle and ring fingers are computed using linear interpolation. Based on the fingertip positions, we compute rotations for the finger joints using inverse kinematics. This type of motion happens when a hand tracking system is used that only records the fingertip positions [13].

Jitter. Jitter induces random rotation movement (jitter) along the primary rotational axes of the wrist, fingers, and thumb (Figure 3.9). This condition simulates the effects of noise from sensors, which can cause jumpiness and small fluctuations in the animation. For each frame, we compute a small random rotation pertubation by sampling an angle θ from a normal distribution: $\theta \sim \mathcal{N}(0, \sigma)$. J. Segen and S. Kumar [170] examined the ranges of jitter in hand tracking and proposed that typical jitter in orientation are less than 2 degrees, which also corresponds to our experience. We stay consistent with this result when setting the variance σ to 0.667 to create jitter. With this setting, the angle θ stays within -2 and 2 in 99.7% of cases. We also stay within the range used by Toothman and Neff [185] who add jitter to whole body motions to evaluate the impact of avatar tracking errors in virtual reality. They apply a rotational jitter between 0 and 0.5 degrees, then between 0 and 1 degree, and finally between 0 and 6 degrees. Jitter is also encountered in current consumer equipment and can increase in low light conditions [151].

Popping. The popping condition periodically freezes the joints of the wrist, fingers, and thumb and then pops them back to their current rotations. It simulates the effects of abrupt transitions in the motion such as those caused by temporary occlusions or loss of tracking. This type of error is common with head-mounted inside-out hand-tracking technologies when the hands leave the tracking space [70]. We induce popping with a freeze duration of 0.8 seconds at intervals between 7 and 9 seconds to prevent the popping from looking too regular. Pops are more visible if the hands are moving a lot (Figure 3.9). We ensured each clip had at least one pop.

Smooth. Most systems perform smoothing to counteract jitter from sensors. We implement this condition by applying an exponentially weighted average on the original animation curves of the wrist, fingers, and thumb, sampled at 30 frames per second. This smoothing technique blends the incoming frame f^{orig} with the previous computed frame f^{t-1} such that $f^t = f^{orig}\alpha + f^{t-1}(1-\alpha)$. Choosing a lower α weights the previous values over the new value, which produces a smoother curve at the expense of loss of detail. We set α to 0.2 to simulate a slight, not too obvious smoothing that would also be used in practice in such applications.

Passive. The Passive condition uses the method developed by Neff and Seidel [148] to implement digits that move solely under the effect of gravity. The result is a hand that seems uncontrolled and lax. The authors provide the results of simulation in a table, driven by wrist orientation, which we implement directly. The motivation for including this condition is that in cases where no information on the finger and thumb motion is available, it might look more realistic to add some motion than to have none at all.

Random Based on the same motivation as the Passive condition (some hand motion might be preferred to none), Random adds captured hand motions that might not fit the body motions: for each charade, we applied the hand motion from the next charade (order is alphabetical by title), starting at the middle of the charade to avoid similar beginnings. This technique creates somewhat random hand motions within the same style. The short clips were extracted from the resulting long motions.

Static The hand does not move. We set the wrist, fingers, and the thumb to a relaxed pose to make the effect more subtle. This condition occurs when an avatar's hands are shown but there is no detailed hand tracking, for example when using simple controllers.

3.2.2 Method

3.2.2.1 Participants

For the Alteration and Intensity experiments combined (they were run together), we recruited 1940 online participants using Amazon Mechanical Turk. A technical failure resulted in the loss of the motion comprehension data for 840 participants but preserved all other data. Participants found our experiment listed as a HIT (Human Intelligence Task) on the Mechanical Turk portal. We restricted participants to those who had an approval rate of over 95%, were located in the United States, and had not previously taken any other questionnaires distributed as part of the same project. Participants were compensated with $1 for a task that took about 10 minutes.

3.2.2.2 Cleaning Data

Research suggests that recruiting participants from Mechanical Turk does not lead to a significant degradation in data quality [177, 23] as long as some quality assurance is performed on the responses. Text responses were checked manually. Across both the Alteration and Intensity experiments, we excluded 39 participants due to nonsense written answers or errors playing back the video. An additional 36 participants were omitted due to non-consent or missing data. Ultimately, 1865 (96.13%) online participants remained for further processing. To ensure the quality of the survey responses, we computed Pearson correlation coefficients in a similar way as Smith and Neff [175] for each grouping of questions as they were shown to participants: two for the TIPI measure, one for McDonnell *et al.*'s questions on the perception of the character, and one for the questions on social presence. Participants whose answers greatly differed from the mean in a grouping compared with others in their same condition (same motion condition, same avatar, and same clip length, 19.3 participants on average) were flagged. Those with three flags or more were omitted from the analysis (resulting in 265 being omitted), leaving 1600 (82.47% of 1940 recruited) participants total for analysis. Similar quality assurance techniques have been used in various crowd-sourced studies [177, 175] with the assumption that not too many responses from an individual should deviate greatly compared to the other responses in that condition. We followed Smith and Neff's [175] example and, after extensive testing, used the same small threshold of 0.15 in order to preserve as many responses as possible. Spot checks revealed that this method correctly excluded participants whose ratings did not seem thought through, e.g., when the same rating was given to every question, or who did not answer our attention check motions correctly.

Out of the 1600 participants (927 of which had motion comprehension data), a total of 871 participants (505 with full data) were analyzed as part of the Alteration experiment, and 1198 participants (685 with full data) were analyzed as part of the Intensity experiment. There was an overlap of 469 participants (263 with full data) because the conditions Original, Jitter(Low), Popping(Low), and Smooth(Low) were considered in both experiments. The Alteration experiment had an average of 27.2 participants for each combination of conditions, 108.9 participants per motion alteration, and 425.5 per clip length and avatar.

3.2.2.3 Procedure

Participants were directed to a Qualtrics survey. Participants started by signing a consent form and providing demographic information. Participants watching the long clips were introduced to the rules of charades and told that they would be asked to guess a movie title. They were then asked to select the movies they were familiar with from a selection of 45 movie covers that included the titles for the charades used in the experiment. We had planned to use this information to eliminate participants that were not familiar with the movies they had to guess, but found that many participants were able to guess the movie titles even if they were not familiar with the actual movies. We therefore did not use this information. Participants who watched the short clips were told that they would be asked to guess a noun or verb.

For both clip lengths, participants watched the sequence of animation clips in randomized order and typed in their responses. Participants could only watch each clip once. After the video section, participants answered questions about their perceived comprehension, perception of the character, and social presence. The last question was open-ended and asked for comments and feedback.

3.2.3 Results and Discussion

If not otherwise mentioned, results were analyzed with an 8x2x2 repeated measures ANOVA with between-subjects factors Motion Alteration (8), Avatar, and Clip Length. As typical tests for normality do not provide reliable answers for large datasets, we inspected the distribution of the answers in the histograms. As the number of analyses run was large, p-values were adjusted for Type I error using False Discovery Rate (FDR) control over all values from the 15 measures [32].

If significance was found post-hoc testing used Tukey HSD comparisons. Only significant results are reported. Statistics for the Alteration experiment are provided in Table 1 of the Appendix. We follow the order in Table 1 when presenting and discussing our results, starting with main effects of Motion Alteration, Avatar, and Clip Length followed by any interaction effects for each examined concept.

Comprehension Our analysis revealed a main effect of Motion Alteration for Motion Comprehension and Perceived Comprehension; the No Motion Data conditions performed significantly worse than the Partial and Full Motion Data conditions, with the exception of a non-significant difference between Passive and Reduced for Perceived Comprehension, see Figure 3.3.

These results **support part of H1, that the complete absence of motion data reduces comprehension**. This effect could not be diminished by adding synthesized motions as in the Random and Passive conditions. However, **the first part of H1 was not supported, as errors**

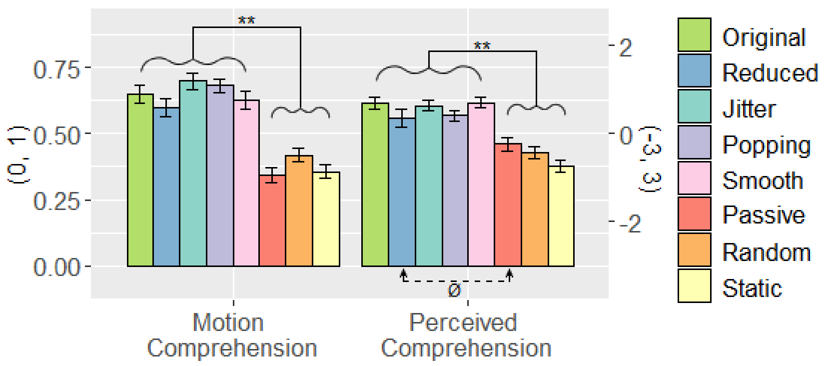


Figure 3.3: Alteration Experiment. Main effects of Motion Alteration for Motion Comprehension and Perceived Comprehension. For both measures the No Motion Data conditions (Passive, Random, and Static) performed significantly worse than the Partial Data (Reduced, Jitter, Popping, Smoothing) and Full Data (Original) conditions with the exception that the difference between the Reduced and Passive motion alterations is not significant (⊘) for Perceived Comprehension. Motion Comprehension values range from 0 to 1 and Perceived Comprehension is visualized in the graph as between -3 and 3 (no numbers were shown on the actual Likert scales). Error bars represent the standard error of the mean in all graphs. *** indicates a p-value of less than 0.001, ** a p-value <0.01, and * a p-value <0.05. If multiple conditions are grouped, the lowest p-value is used. These symbols are consistent throughout the paper.

or reduced information, at least up to the levels we tested, **did not affect comprehension** in our experiment. The hand motion data in our Partial Data conditions was sufficient to understand the meaning of our clips as correctly as when the accurate hand motion was depicted.

We also found main effects of Avatar for Motion Comprehension and Perceived Comprehension. As shown in Figure 3.4, participants were on average able to guess more words or movies with the Realistic avatar than with the Mannequin. Despite efforts to keep the avatars as similar as possible, including their degrees of freedom and primary colors, participants were not able to understand the Mannequin as well as the Realistic avatar. This result could be due to the fact that the shading of the hands lead to slightly less contrast for the Mannequin, or due to increased familiarity with the Realistic avatar. This result does show how important the design of the avatar is when accurate comprehension is key.

Furthermore, we found main effects of Clip Length for both Motion Comprehension and Perceived Comprehension, as the Long movies received lower ratings than the Short clips for both comprehension measures. The better results for the Short clips could be due to the fact that they were taken from the most comprehensible segments of the Long movies or maybe guessing words is an easier task then guessing movies. For the Long movies, participants might have guessed parts of the answer correctly but did not manage to infer the correct movie, which may have contributed

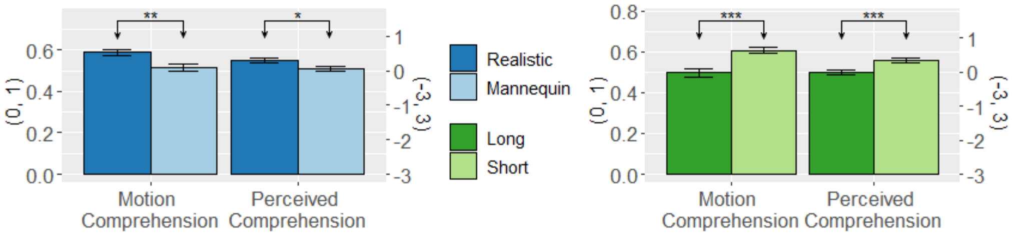


Figure 3.4: Alteration Experiment. Main effects of Avatar and Clip Length for Motion Comprehension and Perceived Comprehension. Main effects of Avatar for Motion Comprehension and Perceived Comprehension (left). Participants correctly understood the Realistic avatar significantly more often than the Mannequin. Main effects of Clip Length for both Motion Comprehension and Perceived Comprehension (right). The Long movies were and seemed significantly more difficult to comprehend than the Short words.

to the lower comprehension scores.

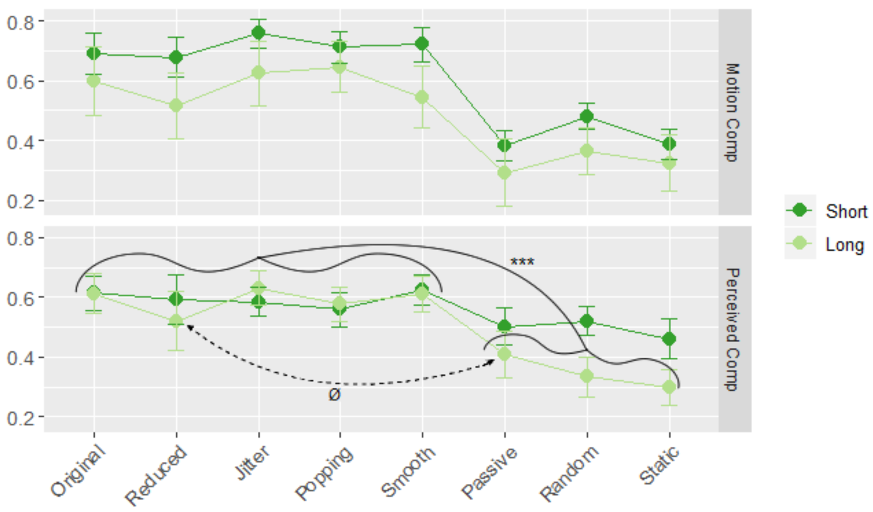


Figure 3.5: Alteration Experiment. Interaction effect of Motion Alteration and Clip Length for Perceived Comprehension. Interaction effect of Motion Alteration and Clip Length (bottom): The No Data alterations (Passive, Random, Static) as Long motions were rated significantly lower than many other conditions. The graphs for Motion Comprehension are shown at the top for comparison, but there is no interaction effect.

Finally, there was a significant interaction effect between Motion Alteration and Clip Length for Perceived Comprehension, see Figure 3.5. The effect occurs because for the Full and Partial Motion Data conditions, the Long and Short clips are perceived to be similarly comprehensible, whereas for the No Motion Data conditions the Long movies were perceived to be less comprehensible than the Short clips. Interestingly, this interaction effect is not present when it comes to actual Motion Comprehension. This result may imply that the user had enough time when viewing the Long clips to realize that not everything could be understood, leading to a lower perceived comprehension. Or this difference could be attributed to the differences in tasks and a different perception of task difficulty. One conclusion could be that to achieve a high level of perceived comprehension, accurately tracked hand motions are more important in longer interactions.

Perception of Character Main effects of Motion Alteration were present for nine of the twelve Perception of Character measures, see Figure 3.6. Agreeableness, Extraversion, and Emotional Stability were the exceptions. For each measure, some of the No Motion Data conditions were rated as significantly worse than some of the Full or Partial Data conditions. In most cases the Static condition received the least positive results. The only additional significant differences affect the Naturalness measure: Jitter was rated as significantly less natural than the Original condition and Random was perceived as significantly more natural than Static. The detailed significant differences for each measure are listed in Appendix Table 1.

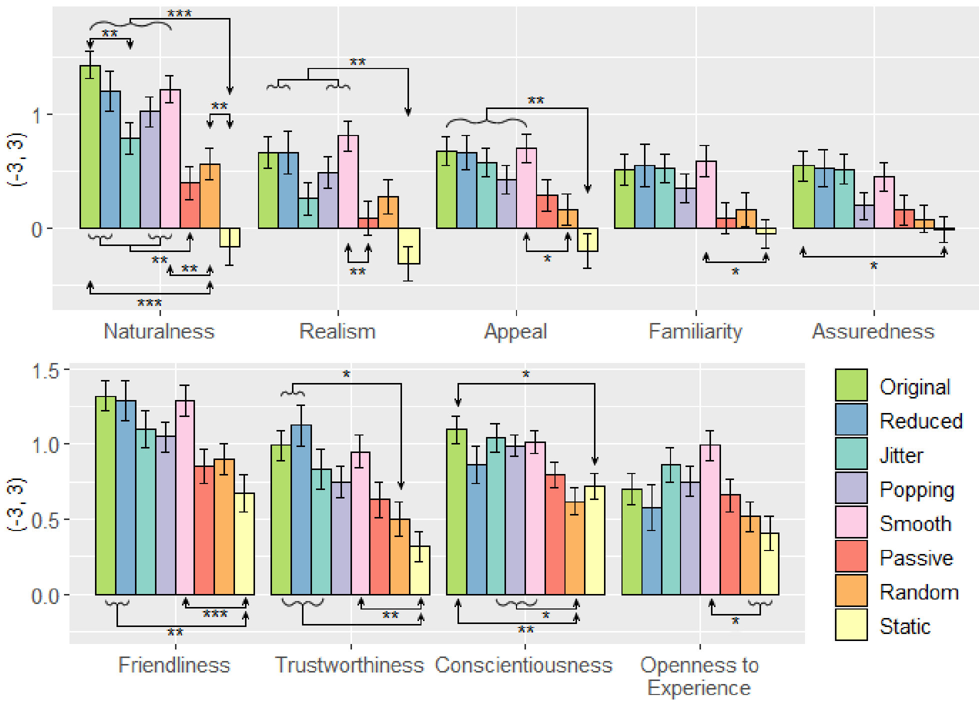


Figure 3.6: Alteration Experiment, Main effects of Motion Alteration on Perception of Character. Main effects of Motion Alteration for questions on the perception of the character. Some of the No Motion Data conditions were rated significantly more negatively than the Full or Partial Data conditions in many cases, with the Static condition rated worst most often. Questions were asked on a 7-pt Likert scale, and are represented here on a -3 to 3 scale.

To quantify these results, we counted how often each condition received a significantly higher value (+1) or a significantly lower value (-1) than any other condition for all measures related to the perception of the character. We found the following results: Original 12, Reduced 7, Jitter 3, Popping 5, Smooth 13, Passive -5, Random -9, Static -26, confirming our observations. According to these sums, the conditions can be divided into four groups: Original and Smooth were rated most favorably followed by Reduced, Jitter, and Popping. The next group consists of Passive and Random, seen as less positive than the Reduced, Jitter, and Popping conditions. Finally, in the Static condition the character was perceived least favorably by far.

These results **strongly support the first part of H2, that changes to the hand motions will affect the participants' perception of the character**. The significant effects are nearly all based on the No Motion Data conditions being rated less favorably. These results imply that hand motions are important when it comes to a positive impression of a virtual character. Surprisingly, the Partial Motion Data conditions did not significantly change participants' perception of the character when compared to the Full Motion Data condition, meaning that **errors in hand tracking did not significantly affect how people perceive a virtual character** at least up to the levels of error we tested in this first experiment. One exception is Jitter, but even Jitter only reduced the perceived Naturalness of the character, not other measures such as Familiarity.

A closer look at our results reveals further insights:

- When some type of hand motion is added (Passive and Random conditions), our virtual characters less often receive lower ratings than without any hand motion (Static). While these conditions still perform significantly worse than selected Full Motion Data or Partial Motion Data conditions for some measures, adding some motion and having correct wrist motions seems to be advantageous.
- While there were no significant differences between the Partial and the Full Motion Data conditions (except for the naturalness of Jitter), the Original and Smooth conditions were more often significantly different from the No Motion Data conditions than the other Partial Data conditions, so Original and Smooth were rated most positively overall.

Significant main effects of Avatar were found for Realism, Appeal, Familiarity, Assuredness, and Agreeableness. In all cases, the Mannequin avatar was ranked significantly lower than the Realistic avatar. These results **strongly support the second part of H2, that changes to character appearance will affect the participants' perception of the character**. It furthermore shows that the design of an avatar is a crucial element of any application where interaction with a virtual character is important.

Main effects of Clip Length were present for Naturalness, Realism, Appeal, Familiarity, and Openness to Experience. The Long movies were rated worse than the Short words in all cases. This result is in line with our results for Motion Comprehension and Perceived Comprehension, where longer movies also performed worse. These results indicate that negative effects are more noticeable when the motions are seen for longer times. Viewers might have more time to notice errors and imperfections. Finally, we found interaction effects between Avatar and Clip Length for several measures related to the perception of the character measures: Naturalness, Realism, Appeal, Familiarity, and Trustworthiness (Figure 3.7). In most cases, the Long movie clips with the Mannequin were rated significantly worse than all other conditions (see Appendix Table 1 for details).

Social Presence We found a main effect of Motion Alteration for Social Presence. Participants who watched the Static condition found that Social Presence was significantly lower than participants who watched any of the Full and Partial data conditions with the exception of Reduced, see Figure 3.8, left. Furthermore, there was a main effect of Avatar, with the Realistic avatar leading to significantly higher ratings than the Mannequin (Figure 3.8, right).

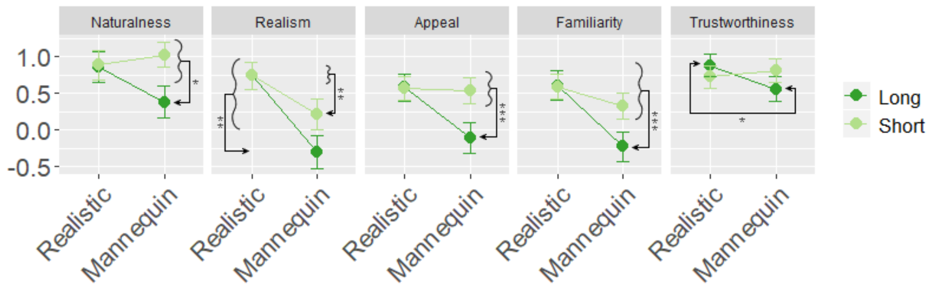


Figure 3.7: Alteration Experiment, Perception of Character Interaction Effects. Interaction effects of Clip Length and Avatar mostly due to the low ratings of the Long movie clips with the Mannequin.

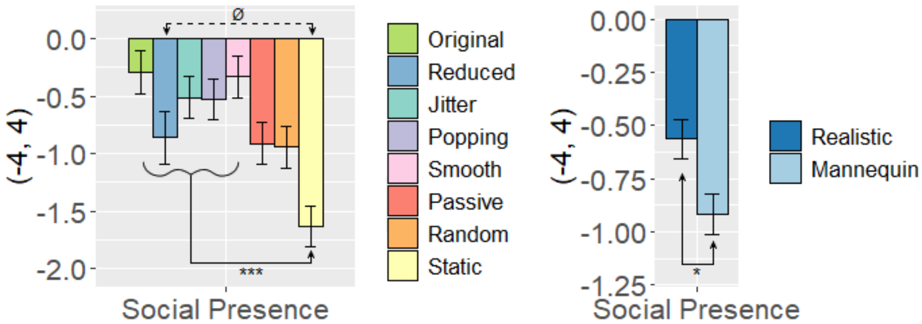


Figure 3.8: Alteration Experiment, Main effects of Motion Alteration and Avatar on Social Presence. Main effects of Motion Alteration and Avatar. The Static condition induced significantly lower Social Presence than the Full and Partial data conditions. The Mannequin avatar had lower ratings than the Realistic avatar. Social presence responses were asked on a 9-pt Likert scale, and are represented here on a -4 to 4 scale.

These results **support H3, that less natural hand motions or a less realistic character will reduce social presence**, as the Static condition and the Mannequin both had that effect. The other No Motion Data conditions, Passive and Random, did reduce the perceived social presence on average, but considerably less and without reaching significance, implying that some motion, even if partly incorrect, is still better than none. It is unclear why the difference between Reduced and Static was not significant, maybe the decrease in detail did impact social presence for the Reduced condition. Based on these results, we also recommend to use a more realistic avatar when high social presence is desired.

3.3 Experiment 2: Intensity Experiment

In our first experiment, we found that our Partial Motion conditions did not lead to many differences compared the our Original condition. They did not reduce Motion Comprehension or Perceived Comprehension at all. However, if the intensities of these errors are increased, at some point we expect them to influence comprehension as there is no meaningful data left. For example, smoothing a motion to an extreme point would result in a static, averaged hand pose, which corresponds to our Static condition. Increasing the intensity of Popping to an extreme level would result in one or a few random poses being held for a long time, and exaggerating the Jitter condition would result in an erratic, random-type motion. The levels of errors we added were very reasonable. Therefore, in our second experiment, we test higher levels of errors with the goal of finding thresholds up to which the errors would be acceptable.

3.3.1 Motion Intensities

The Intensity experiment uses the same design as the Alteration experiment, but changes the levels of intensity of specific motion alterations. We include the Original condition into our analysis as a baseline. This experiment tests three different intensities (low, medium, and high) for each of the motion alterations Jitter, Popping, and Smooth, see Figure 3.9. The low intensities are identical to the motion alterations from the Alteration experiment (e.g., PoppingLow in the Intensity experiment is Popping in the Alteration experiment). Medium intensity doubles the parameters and high intensity quadruples them. The Jitter alteration samples a normal distribution to obtain an offset to apply to the original rotation. This distribution has variance of 0.667 degrees for low, 1.32 for medium, and 2.67 for high. To increase the intensity for Popping, we decrease the time between the pops from 7-9 seconds (low), to 3-5 seconds (medium), and 1-3 seconds (high). Note that for our Short word dataset, we could not test all intensities of popping as some of the clips were too short. We ensured each clip had at least one pop. The Smooth alteration is implemented using an exponential moving average. To increase the intensity, we decrease the parameter α we use for blending. We use values of 0.2 for low smoothing, 0.1 for medium smoothing, and 0.05 for high smoothing.

3.3.2 Method

The Alteration and Intensity experiments were run in parallel on Mechanical Turk. The procedure in both experiments was identical as was the process to clean the data, described in Sec-

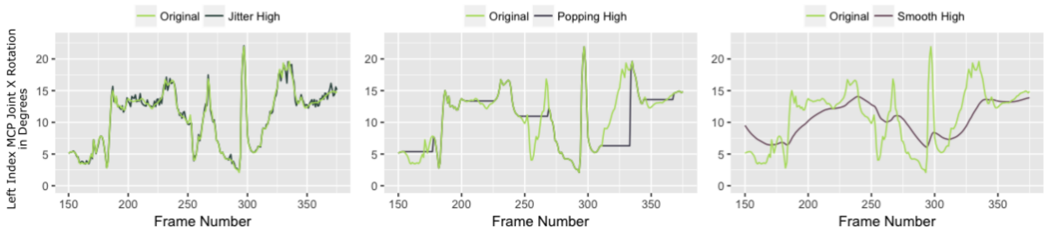


Figure 3.9: Intensity Experiment: Example of JitterHigh, PoppingHigh, and SmoothHigh intensities against the Original condition. This displays the curve for the first joint of the left hand index finger in the motion *The Pianist*.

tion 3.2.2.2. After post-processing, the Intensity experiment had 1198 participants, 685 had Motion Comprehension data. On average there were 30.0 people in each combination of conditions, 119.8 per motion intensity, and 599 per clip length and avatar.

3.3.3 Results and Discussion

We perform the analysis of our second experiment in a similar way to our first experiment: For each of the three alterations that we varied in intensity (Jitter, Popping, and Smooth), we ran a three-way 4x2x2 ANOVA with the between-subjects factors Motion Intensity (4; Original and Low, Med, High Intensities), Clip Length (2), and Avatar (2). P-values were adjusted for Type I error using False Discovery Rate control; FDR was run over the p-values of the 15 measures for each intensity. If significant main or interaction effects were found, a post-hoc Tukey HSD revealed the detailed significant differences between conditions. Detailed results are listed in Table 2 of the Appendix.

Comprehension There were no significant differences of Motion Intensity for Motion Comprehension or Perceived Comprehension for Jitter, Popping, or Smooth, see Figure 3.10.

This result comes as a surprise. When the changes to the original data become larger, one can see less and less of the original information. We expected comprehension measures to decrease as a result. We did not find this effect in our collected data. We conclude that relatively large errors can be applied before comprehension is affected in a significant way (at least in the way we measure it), which is good news for developers in that area. We can again **not support the first part of H1, that partially missing or inexact hand motion data reduces participants' comprehension** of a character.

We found main effects of Clip Length for Motion Comprehension for all three types of errors. The Short words were guessed correctly more often than the Long movies, which is in

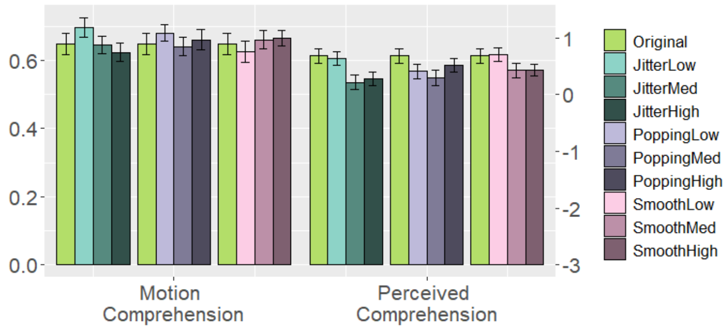


Figure 3.10: Intensity Experiment: Motion Intensity on Motion Comprehension. Motion Intensity on Motion Comprehension. Contrary to what we expected, there were no significant differences of Motion Intensity for Motion Comprehension or Perceived Comprehension for Jitter, Popping, or Smooth even higher intensities.

line with our results in the Alteration experiment. A main effect of Clip Length for Perceived Comprehension could not be confirmed in the Intensity experiment.

For Popping there was a main effect of Avatar for Perceived Comprehension; the Realistic avatar was perceived to be easier to understand than the Mannequin. This result supports our findings from the first experiment. There were no significant effects of Avatar for Motion Comprehension or for Perceived Comprehension for Jitter or Smooth, however, the Mannequin lead to lower scores on average in all five cases as well, suggesting a consistent trend.

Perception of Character We found main effects of Motion Intensity for several measures related to the perception of the character for Jitter, see Figure 3.11, namely Naturalness, Realism, Appeal, Assuredness, Conscientiousness, and Emotional Stability. In each case, JitterHigh was ranked as significantly worse than Original. For Naturalness, Assuredness, and Emotional Stability, the differences between JitterMedium and Original reached significance. JitterLow was only rated significantly worse than Original for Naturalness. Further details can be found in the Appendix in Table 2. The impact of jitter on personality is in line with results from Wang *et al.* [195] and Smith and Neff [175], who found that a resting hand pose conveys high emotional stability (jitter would be the most opposite to a resting hand pose) and disfluency in the arm motions reduces conscienciousness and emotional stability.

There were no main effects of Motion Intensity for Popping or Smooth, which again is surprising considering the large errors that are being introduced.

As in the Alteration experiment, a main effect of Avatar was present for multiple measures related to the perception of the character with the Mannequin avatar always yielding lower

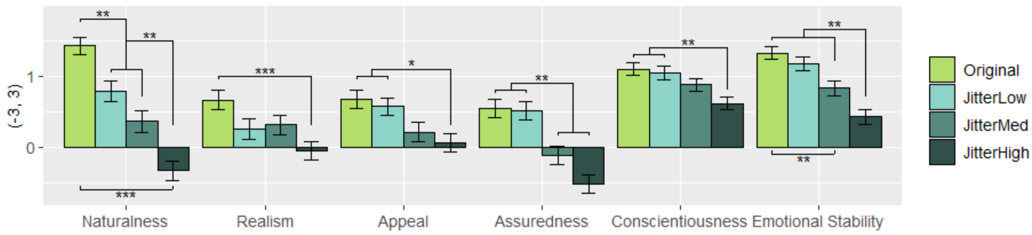


Figure 3.11: Intensity Experiment: Jitter, main effects of Motion Intensity. Main effects of Motion Intensity: JitterHigh is rated significantly lower than Original in six measurements, JitterMedium in three, and JitterLow in one.

scores than the Realistic avatar. This effect was present for all three types of errors for Realism, Familiarity, and Assuredness; for Jitter and Smooth it was additionally found for Appeal. These results are expected and **further support H2, that changes to character appearance will affect the participants' perception of the character.**

Finally, we found an interaction effect of Clip Length and Avatar for the Realism measure when analyzing the intensities of Smooth, mainly based on the fact that, when watching the Long movies, the Mannequin was perceived as significantly less realistic than in all other combinations of Avatar and Clip Length (see Figure 3.12, right), which is in line with our results from the Alteration Experiment.

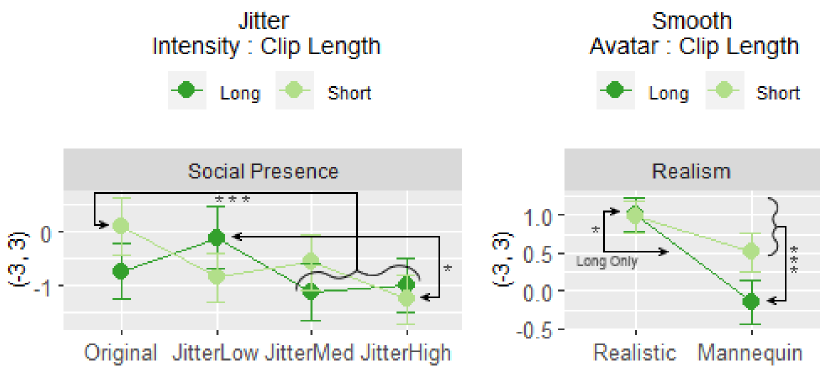


Figure 3.12: Interaction Experiment: Interaction effects. For Jitter there was an interaction effect of Motion Intensity and Clip Length (left) and for Smooth we found an interaction effect of Clip Length and Avatar for the Realism measurement (right).

Social Presence There was a main effect of Motion Intensity for Social Presence when analyzing the Jitter intensities. The post-hoc test revealed that Social Presence was reduced in the condition with the highest level of jitter compared to the Original condition. For Jitter, we furthermore found an interaction effect of Motion Intensity and Clip Length, visualized in Figure 3.12, left, mainly because participants rated social presence as significantly higher when watching the Short clips in the Original condition than for several of the other combinations with higher error intensities. For the Smooth intensities, a main effect of Avatar was based on a lower Social Presence rating for the Mannequin compared to the Realistic avatar, which is again in line with our results from the Alteration experiment. Together, these results **support H3, that less natural hand motions or a less realistic character will reduce social presence,** at least in some cases.

3.4 Experiment 3: Virtual Reality Experiment

3.4.1 Method

The main goal of our third experiment is to investigate if selected findings from our first experiment also apply for an avatar observed in a virtual environment or if the virtual character is perceived differently in VR. As users might have preferences that are not reflected in our measurements (comprehension, perception of the character, and social presence), we furthermore compare all of our conditions from the previous experiments in a within-subjects design and examine the viewers' comfort level with every condition and their preferences between conditions. The experiment has three parts, all three use the Realistic avatar.

The first part, **VR Compare**, recreates the Alteration experiment in virtual reality with the Original and Static conditions only. Participants wear an Oculus Rift head-mounted display (HMD) and are integrated into the same Unity scene used to generate the videos. We follow the procedure of the Alteration experiment, with the change that during the word guessing phase participants say their answers out loud so that they do not have to take off their HMD. The experimenter writes the answers down and starts the next clip. Participants are randomly assigned to see either the Original or the Static motion condition; all participants see all 15 of the Short word clips in their assigned condition. After viewing all of the motions, participants briefly remove their HMD to answer the same post-experiment questionnaire as in the Alteration and Intensity experiments, then put the HMD back on.

The second part, **VR Comfort**, asks participants to judge the viewing comfort and perceived naturalness of all motion conditions from the Alteration and Intensity experiments (14 conditions in total). In this experiment, we use random 10 second clips from each of the six charade motions. Between each clip, the experiment pauses to ask participants two questions: "How comfortable would you feel interacting with this character for an extended period of time?" (comfort) and "Please rate the naturalness of the character's motions" (naturalness). This experiment is self paced and participants choose their answers on a 7-point Likert scale using a gamepad controller. Each motion condition is shown twice (in random order), once with the Likert scales initialized with the lowest value selected and once with the scale initialized with the highest value. So each participant rates a total of 28 clips.

The third part, **VR Rank**, asks participants to rank each of the motion conditions from the Alteration Experiment from most comfortable to interact with (1) to least (8). In this experi-

Figure 3.13: VR Rank scene. Participants assigned rankings from 1 (most comfortable) to 8 (least comfortable) to each avatar. Each avatar was animated with a different condition. Participants were placed in the center of the avatars.

ment, participants are surrounded by eight clones in a slightly more than half circle as shown in Figure 3.13. Each clone is animated with a different motion condition. Placement is randomized. This configuration allows participants to make side-by-side comparisons. We decided not to show all 14 conditions based on pilot tests as the task becomes more complex and confusing with that many animated clones. This experiment is also self-paced, with no time limit. Participants assign a unique rank to each character using a gamepad controller.

We recruited 31 in-person participants through emails, flyers, and word of mouth. Upon arrival, participants fill out a consent form along with a demographics questionnaire. Next, participants put on the HMD. They start each experiment part in a virtual welcome room where they can become comfortable with the VR environment. They see a welcome screen with introductory text, which allows the experimenter to adjust the focus if necessary. During each experiment part participants can see the character in virtual reality as if they were standing in front of it. They have no virtual body of their own. Participants complete VR Compare, VR Comfort, and VR Rank. To wrap up, participants are asked open ended exit questions and compensated with a $5 gift card. For most participants, the experiment took 20-25 minutes to complete. The experiment was approved by our ethics committee.

3.4.2 Results

The significant results for VR Compare, VR Comfort, and VR Rank are also reported in Table 3 in the Appendix.

3.4.2.1 VR Compare

We used one-way ANOVAs to analyze VR Compare data with False Discovery Rate to correct for Type I errors. A main effect of Motion Alteration was found for Motion Comprehension. Participants who watched the Original condition were able to guess the words correctly significantly more often than participants viewing the Static condition (Figure 3.14, left). This result corresponds to our results in our first experiment and allows us to **confirm Hypothesis 1, that the absence of hand motion data reduces comprehension, for virtual environments** as well.

We also found a significant effect for Conscientiousness, with the character in the Static condition rated as significantly less conscientious than in the Original condition, which is an effect we also found in our first experiment.

There were no significant differences for our other measurements, which could be due to the smaller number of participants in this experiment. Results that are similar to the ones in previous experiments, such as the differences in Perceived Comprehension visualized in Figure 3.14 that did not reach significance, can be seen as supporting that explanation. However, it is also possible that the differences seen are less apparent in a virtual environment. The fact that the viewer can look around more freely in VR could contribute to these results.

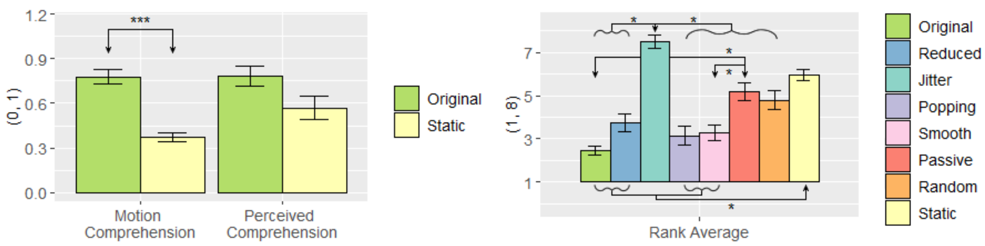


Figure 3.14: VR Compare: Main effect of Motion Alteration on Motion Comprehension (left). VR Rank: Average rankings of Motion Alterations in VRRank experiment (right). Participants were asked to rank how comfortable they would feel interacting with a character with each motion alteration from most comfortable (1) to least comfortable (8). The Original condition was rated best, the Jitter condition worst.

For further insights, we directly compared participants' reactions in VR Compare to those in the Alteration experiment with a two-way ANOVA with between-subjects factors Experiment (2) and Motion Alteration (2), including only participant results from the Alteration experiment under the same conditions as in VR Compare (Realistic avatar, Short word clips, Static and Original alterations). Here again, we used False Discovery Rate corrections. There were no significant effects for Motion Comprehension or Perceived Comprehension. We found several main effects of

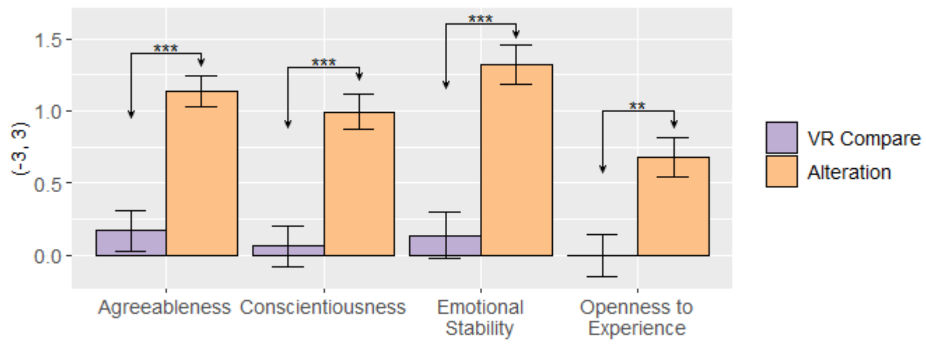


Figure 3.15: Alteration Experiment and VR Compare: Significant differences between the two experiments. The avatar was rated significantly more favorable when seen in videos than when seen in VR for four of our measures.

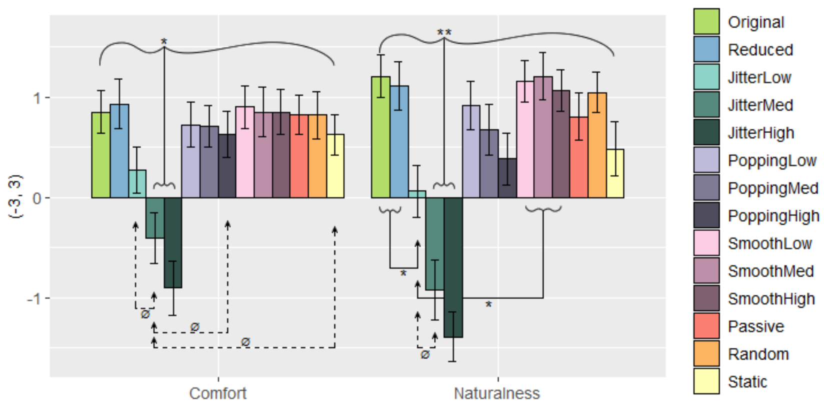


Figure 3.16: VR Comfort: Main effect of Motion Alteration on viewing Comfort and perceived Naturalness in VR. The JitterMed and JitterHigh alterations were rated significantly worse than all other conditions with few exceptions. Significant effects are visualized in the graphs; full details are available in Table 3 of the Appendix. Both questions were asked on a -3 to 3 Likert scale.

Experiment: for Agreeableness, Conscientiousness, Emotional Stability, and Openness to Experience (see Figure 3.15). In all cases the VR Comfort participants ranked these measures as lower on average than the participants from the Alteration experiment. So participants had a less favorable perception of the character when viewed in VR. Character design might be even more important in VR than it already is in videos.

There were no interaction effects, showing that the changes in hand motions might have a similar effect when in VR compared to when watching videos, or at least that differences were not strong enough to reach significance. The lack of interaction effects (which we expected) is not a proof that errors in hand motion have the same effect in VR than when watching videos. It is always possible that effects are present but that our experiment did not reach the power or design to reveal them. However, it is likely that any differences would be small.

A surprising result was the lack of a significant difference in social presence, which we would have expected between a video and a scene in virtual reality. We suspect that this result is due to the lack of reference. An experiment with a within-subjects design could verify this assumption.

3.4.2.2 VR Comfort

Using a one-way ANOVA and FDR to correct for Type I errors, we found main effects of Motion Condition for Comfort and Naturalness. In both cases, the JitterHigh and JitterMed conditions were rated significantly worse than nearly all other conditions. For Naturalness, even a small amount of jitter (JitterLow) had a significant negative effect compared to some other conditions (see Figure 3.16 and Table 3 for details). These results also **support Hypothesis 2**.

Comparing the results from Naturalness to the two previous experiments shows many similar tendencies as one would expect. Interestingly, the average ratings for the No Motion data conditions are much higher than they were in the Alteration experiment. This observation could be due to the fact that participants watched a specific condition for a much shorter timespan in the VR Comfort experiment (10 seconds in the VR Comfort experiment vs. 99 seconds (Short clips) or 235 seconds (Long movies) in the Alteration and Intensity experiments). They furthermore were not asked to understand the character, but only to watch and rate it. Finally, they had other conditions to compare the motions to, which can also influence the results.

The main effect of Motion Condition for Comfort **supports Hypothesis 4, that less natural hand motions will make people feel less comfortable**. However, while this effect is strong when adding medium or high levels of jitter, the other Partial or No motion data conditions did not result in a significantly lower perceived comfort. Considering the differences between the Short clips and Long movies in the previous experiment, we think that for some of these conditions (such as Static), this result could be due to the short viewing times of each condition. In retrospect, we should have shown each condition for a longer time period to give conclusive results with regard to comfort.

3.4.2.3 VR Rank

Averaging the given rankings across all participants results in the following ordering: Original (most comfortable), Popping, Smooth, Reduced, Random, Passive, Static, Jitter (least comfortable). A Friedman's ANOVA was used to analyze the ranked data between Motion Conditions (Figure 3.14, right) with Friedman Multiple Comparisons to determine individual differences. We found a significant main effect of Motion Alteration. Participants ranked the Jitter, Passive, and Static conditions as significantly less comfortable than several other conditions. Detailed significant differences can be found in Table 3 in the Appendix and are visualized in Figure 3.14. Not all of the differences are significant. For example, the differences between Original, Popping, and Smooth are statistically meaningless.

These results give further **support to Hypothesis 4, that less natural hand motions will make people feel less comfortable**. They show that Jitter should definitely be avoided, that some of the No Motion conditions reduce comfort, and that the differences between the Original motion and most of the Partial Motion conditions (Reduced, Popping, and Smooth) are not significant when it comes to comfort levels.

3.5 Conclusion, Limitations, and Future Work

In this project, we investigate the effects of errors in or the lack of hand motions on comprehension, perception of a virtual character, perceived social presence, and comfort when watching the character. We summarize our key findings in Table 3.2 and as follows:

1. Lack of hand motion data significantly reduces comprehension and social presence, as well as negatively affects the perception of the character, for example, appeal, friendliness, and conscientiousness are reduced.

2. Partial or erroneous hand data at the levels we tested is sufficient in many cases to avoid negative effects. Comprehension with partial hand data is not reduced compared to accurate hand motions even with the large errors we tested; the character is perceived similarly and social presence is similar to having accurate hand motions.

3. Adding unrelated motion to the digits and correct wrist motion does not improve comprehension but can reduce the negative effects of a fully static hand when it comes to social presence or the perception of the character. For example, social presence is not significantly reduced in the Passive and Random conditions compared to the full and partial data conditions whereas it is significantly reduced in the Static condition.

4. Jittery motions should be avoided. While the presence of jitter did not affect comprehension, our Jitter condition was preferred least and rated lowest for comfort and naturalness.

5. Our more realistic avatar performed better: comprehension was higher in some cases and many of the personality ratings were more positive. The negative effects of the mannequin were more pronounced when the viewers watched longer motions.

6. Comprehension of our realistic character for the tested conditions was similar (and not significantly different) in VR than when watching videos.

7. Watching our character in VR created a less favorable perception of the character as opposed to watching the same character on a screen.

Our experiments confirm the importance of detailed hand motions for communication, social presence, and for accurately conveying personality. Furthermore, we found several surprising results. We expected to see negative effects when showing static hands, but also when seeing hand motions with large errors. However, errors in hand motions did not reduce comprehension,

Hand motions	Comprehension	Perception of Character	Social Presence	User Comfort
Fully accurate	++	++	++	++
Partial, no jitter	++	+	++	++
Partial, with jitter	++	+	+	-
Unrelated to content	-	-	+	+
No motion	-	- -	-	+

Table 3.2: Study 1: High-level summary of results. This summary simplifies some of the details of our results. It is intended to give a quick idea of the consequences when choosing the accuracy of hand motions. The categories ++, +, -, and – are relative and do not reflect the importance of each category in a specific application. The Original condition represents the "Fully accurate" hand motions; "Partial, no jitter" includes the Reduced, Popping, and Smooth conditions; "Partial, with jitter" includes the Jitter conditions; Passive and Random are the conditions with hand motions "Unrelated to content"; and "No motion" is the Static condition.

even when the errors were very obvious and larger than what one would encounter in practice as was the case in our Intensity experiment. We assume that the redundancy in motions in large enough, so that viewers are able to extract or infer the information necessary for comprehension or forming impressions of a character even if the data is incomplete or noisy. The thresholds for reducing comprehension were higher than the values we tested. Jitter was perceived negatively at lower intensities in some cases, but still did not reduce comprehension even with a high intensity.

While adding random or passive motions to the hands in the absence of data did not help with comprehension, it did at least improve social presence and how the character was perceived to some degree. Both chosen methods (conditions Passive and Random) were rather simple. It would be interesting to see if other, more complex methods of creating hand motions when no data is available might help with motion comprehension or increase, for example, the perceived naturalness.

Finally, we found that character design is important (not surprisingly), and that it might be even more important when the character is seen for longer times and in virtual reality. This might be due to viewers noticing more details when given more time and when sharing a virtual environment with a character. While in our case the realistic character was perceived more positively than the mannequin when seen for longer times, this result might be different based on the exact design of the character, e.g., if we had used a highly appealing cartoony character our results might be different.

While we were able to answer many questions with our experiments, it also has limitations. In these experiments, the virtual character plays charades. We chose this type of motions as we specifically were looking for tasks with expressive motions where body motions and hand motions might be important and for a quantitative way to measure comprehension. We were furthermore trying to avoid any confounding effects that audio might have. However, gestures are used differently when playing charades where one might use more iconic and metaphoric gestures than during typical conversations where beat gestures are more common and detailed hand motions might therefore be less important. To quantify those differences, the gestures of all of the charades motions used in this study were labeled by two graduate students and classified as iconic, metaphoric, beat, or deictic based on the descriptions by McNeill [139, 141]. In the 3 minutes and one second of charades, we detected 32 iconic, 37 metaphoric, 8 deictic, and not a single beat gesture. As a comparison, the same process was applied to two motion databases from Jörg et al. [108], one called *Conversations database* that includes 8 minutes and 5 seconds of narrations and one called *Debates database* from the same actor as the charades and with 9 minutes and 34 seconds of debates. In the narrations, 42 gestures were coded as iconic, 27 as metaphoric, 9 as deictic, and 66 as beat gestures. In the debates, 28 gestures we categorized as iconic, 77 as metaphoric, 29 as deictic, and 74 as beat gestures. As expected, the charades show a higher frequency of iconic and metaphoric gestures and a lower frequency of beat gestures than the conversations or the debates. We also compare this distribution to findings from the literature: McNeill provides statistics of six cartoon narratives by English-language speaking university students. There are 261 iconic, 43 metaphoric, 28 deictic, and 268 beat gestures in an estimated 49 minutes of narration. Such a distribution would be expected in a narrative scenario and is closest to our conversations database.

While results might differ depending on the exact type of communication, accurate hand motions are likely to be less relevant in a conversational scenario with audio and more beat gestures. We conclude that, if errors in hand motions did not reduce motion comprehension when playing charades, they are unlikely to affect comprehension during a typical conversation with audio. Still, gestures are an integral part of conversations and the knowledge that the complete lack of hand motion reduces comprehension in some cases might be enough to attempt to add at least some hand motions all of the time.

A second design choice and limitation was to not animate the face of the realistic avatar to avoid confounding factors with the mannequin. Of course, the presence of detailed facial animations is likely to influence our results. We assume that differences between conditions would be less pronounced as facial animation might convey additional information and distract the viewer from the hands. Changes are likely to depend on the detailedness of the facial animation. In current VR social rooms, facial animation, if present, typically only includes motion of the jaw matching the audio. As that animation is very limited, we assume that our results apply well to current VR scenarios. However, future work will have to show the influence of accurate hand motions when detailed facial motions are present.

Future work would be able to further investigate these effects and answer further questions. Would people in a live scenario adapt and move differently or speak more clearly if errors occur in the motions? Do the results vary depending on the expressiveness of gestures, the personalities of the performers, the information conveyed, and the emotional content of the conversation? For a complete picture, many variables need to be taken into account in future work. The influence of the design of the avatar from stylized floating upper bodies with floating hands to realistic virtual characters should be investigated further. It would also be interesting to find out if hand motions can be learned that actually contribute to comprehension. Our experimental setup could serve as a test bed for such approaches.

Based on our findings, we have several recommendations for developers and animators to consider when creating virtual characters or interactions with avatars in VR. We recommend to capture at least partial hand motion whenever possible even if they contain some errors. Smaller errors and even most of the larger ones we tested did not affect comprehension or social presence or how the character was perceived. The main exception was jitter, which should be avoided or smoothed. However, even highly jittery motion contributes to comprehension. If no hand motions can be acquired, creating some substitute motions is still better than leaving the fingers immobile when it comes to social presence and how the character is perceived.

Chapter 4

Study 2: Evaluating Grasping Visualizations and Control Modes in a VR Game

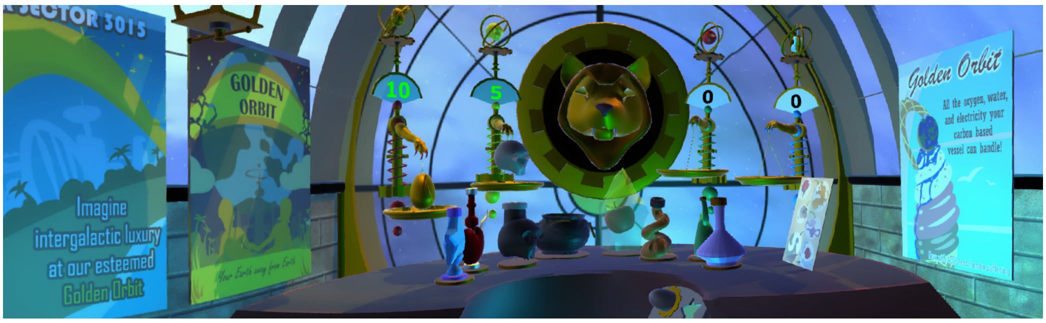


Figure 4.1: Panorama third-person view of the final game. This view from behind the chair where participants sit; the avatar is hidden in this image so that the environment can be seen.

The second study, published in TAP'21 [12], examines the role of hand motions when interacting with items in virtual environments in a game environment. As we study the influence of hand motions on communication between users, we should also consider how they may affect interactions between users and their environments. If we consider communication and interaction separately we may not develop a comprehensive view of how hand motions matter in VR. To that end, in this study we study the effects of two control modes (controllers vs. hand tracking) and two grasping visualizations (continuously tracked hands vs. virtual hands that disappear when grasping) on ownership, realism, efficiency, enjoyment, and presence. This results of this study indicate that implementing hand-tracking and hand motions when interacting with objects in VR is beneficial to these measures of immersion, sans efficiency.

Additionally, hand motions may matter more in some scenarios over others. Previous research has studied these or very similar effects [50, 130, 156] using more typical experimental designs with short tasks that are repeated in several conditions. These tasks do not reflect our experience in VR applications or games, and participants are often aware of the concepts studied and experience all conditions. In this study, our goal is to investigate these effects during an experience that might be closer to a typical VR experience where the user's attention is not focused on interaction conditions and instead on gameplay. To this aim, we designed a VR Escape Room game (see Figure 4.1). Can we still observe similar effects when the participants are not aware of the purpose of the experiment, when they are not able to compare different conditions, and when they might be distracted and not even pay attention to the interaction being used? Our design furthermore allows us to study effects that would be difficult to examine in a repetitive task such as the influence of control modes and grasping visualizations on enjoyment.

4.1 Experimental Design

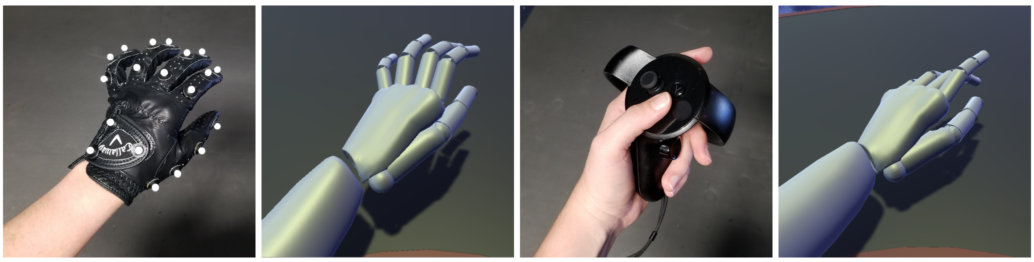
4.1.1 Conditions

Our study uses a 2x2 between-subjects experimental design comparing the independent variables of Control Modes (conditions: Controllers vs. tracked Gloves) and Grasping Visualizations (conditions: Tracked Hand vs. Disappearing Hand), see Table 4.1.

		Control Modes	
		Controllers	Gloves
Grasping Visualizations	Tracked Hand	ControllersTH	GlovesTH
	Disappearing Hand	ControllersDH	GlovesDH

Table 4.1: Study 2: The four experiment conditions. The Control Modes are represented at the top and the Grasping Visualizations are on the left.

In the **Control Modes** conditions, participants either use Oculus Touch **Controllers** to interact with the scene or have their hands tracked by wearing **Gloves** with 19 motion capture markers attached to each finger joint and the back of the hand (Figure 4.2). The markers are tracked at 120fps using 15 Optitrack Prime 17W cameras and labeled in real-time using Han *et al.*'s [87] optical marker based hand tracking algorithm . Participants can freely move their hands and the movements of the avatar's hands mimic their own. In the Controllers condition the fingers are directed by a thumb button, index finger trigger, and hand trigger (typically activated with the middle finger); fingers are extended if the buttons are untouched, partially extended if being touched, and pinched if the buttons are pressed. The avatar's arms are animated using inverse kinematics based on the position of the hands. Participants choose the glove that best tightly fits their hand out of six different sizes prior to entering the virtual environment. The hand size of the avatar is then adjusted accordingly.



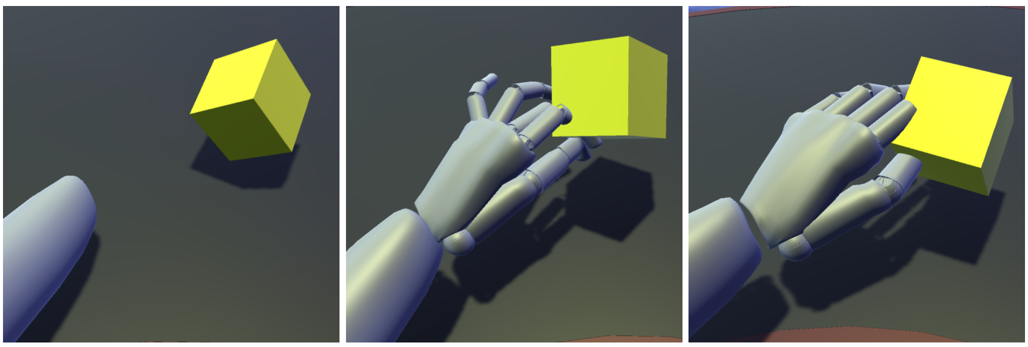
(A) (B) (C) (D)

Figure 4.2: The Control Modes. In the Gloves condition (A, B), the virtual hand follows the participants' hand motions. In the Controllers condition (C, D), a set of default poses are used.

We create two types of **Grasping Visualizations**: **Tracked Hand** and **Disappearing Hand** (Figure 4.3). In the Tracked Hand condition the virtual hands are always visible and follow the players' hands or the controller motions.

In the Disappearing Hand condition the virtual hands disappear once a participant grabs an item and reappear upon release. We chose the DH condition as it imitates grasping visualization in VR games such as *Job Simulator* or *I Expect You to Die*. It is simple to implement as the hand pose does not need to be adjusted based on the object, which might be why it is a popular approach. The Disappearing Hand is furthermore investigated in Canales *et al.*'s work [50] where it was rated significantly lower in questions related to ownership than some of the other tested conditions and preferred least on average out of all tested conditions.

Whether an item is grasped or not in the Gloves condition depends on the positions and velocities of the thumb and index fingers in relation to each other and on the number of contacts between a hand and an object. An item is detected as "grabbed" if the distance between the index finger and thumb is below 5mm or if the velocity between those two digits is greater than 15cm/s; additionally, the nearby item needs at least two points of contact with the hand. An item is released when the thumb and the index finger move apart at a velocity above a threshold of 30cm/s. The thresholds were adjusted through tests with multiple pilot participants.



(A) (B) (C)

Figure 4.3: The Grasping Visualizations. *A:* The Disappearing Hand visualization during a grasp in both control modes. *B:* The Tracked Hand condition when using Gloves. *C:* The Tracked Hand condition with Controllers.

4.1.2 Hypotheses

Our hypotheses on ownership, realism, and efficiency are based on Lin *et al.* [130] and Canales *et al.*'s work [50]. We anticipate higher presence and thus higher game enjoyment [181] with the Tracked Hand and the Gloves condition. So we formulate our hypotheses as follows:

1. **H1. Ownership:**
   (a) Greater ownership in the Gloves condition than in the Controllers condition ([130])
   (b) Greater ownership in the Tracked Hand condition than in the Disappearing Hand condition ([50])
2. **H2. Realism:** Greater realism in the Gloves condition than in the Controllers condition ([130])
3. **H3. Efficiency:** Greater efficiency in the Controllers condition than in the Gloves condition ([130])
4. **H4. Enjoyment:**
   (a) Greater enjoyment for the Tracked Hand than for the Disappearing Hand, as we assume presence increases for the Tracked Hand and increased presence leads to increased enjoyment
   (b) Greater enjoyment in the Gloves condition compared to the Controllers condition, as the Gloves are the more 'natural' control mode
5. **H5. Presence:**
   (a) Greater presence for the Tracked Hand than for the Disappearing Hand, as the Disappearing Hand may be slightly jarring and thus reduce presence
   (b) Greater presence in the Gloves condition compared to the Controllers condition, due to the gloves' more accurate tracking of the hand motions

		Gender			Age		
	Total	**F**	**M**	**O**	**Mean**	**Min**	**Max**
ControllersTH	15	7	7	1	24.67	20	35
ControllersDH	15	2	13		27.67	19	69
GlovesTH	15	5	10		24	19	29
GlovesDH	19	9	10		28.1	20	62

Table 4.2: Study 2: Distribution of participants' gender and age throughout conditions.

Figure 4.4: A (near) first-person view of a participant in the Controllers condition during the experiment.

4.1.3 Participants

A total of 72 participants were recruited for this IRB approved study, 17 for each Controllers condition and 19 for each Gloves condition as we expected technical issues; 62.5% identified as male, 36.1% as female and 1.4% as other. Ages of participants ranged from 19 to 69, with a mean age of 26. All participants were locally recruited through e-mail, reddit, and word of mouth, with a majority being university students. Participants were assigned conditions in sequential order, round-robin style with the two extra participants at the end. A total of eight participants were eliminated from results analysis: three participants were excluded as the motion capture system was not well calibrated, two for different unique technical errors, and three as they had difficulties understanding how to play the game in general. This left 64 participants for analysis, demographics are detailed in Table 4.2.

4.1.4 The Game

For this experiment, we designed an Escape Room type video game, modelled after the popular live-action activity, where a person or a group of people is locked in a room and has to get out by finding clues and solving puzzles. In our case, the player was locked to a chair in an escape pod in space and had to solve puzzles to find the key to the lock.

Advantages of that specific genre are that it uses a first-person player perspective and that the player would not walk or run around. Participants stayed seated during the duration of the game and all necessary puzzle-solving objects were provided within our tracking space. The puzzles allowed us to create a variety of interactions and to design a fun experience where players would use their virtual hands. Early pilots showed that placing all clues in front of the players at the same time was confusing. Therefore, only the objects related to the current puzzle were placed on a table in front of the participant. When a puzzle was solved, the table surface was lowered, then rose with the next puzzle's objects in place. A total of seven puzzles were implemented, with four primary complex puzzles. A quick playthrough of the game can be seen in the video, impressions of the game are shown in Figures 4.1, 4.4, and 4.5.

Figure 4.5: Examples of puzzles. *Left:* Participants must match the cage combination lock to the cards. *Right:* Mimic that "bites" on any hand that tries to retrieve the statue.

For one of the puzzles we implemented a mimic - a box with teeth that suddenly closes when one tries to retrieve the object in it - as a threat condition. The mimic was used as an indication for the strength of the virtual hand illusion in a similar way that has been done in other studies [129, 130, 50, 17, 132, 208].

Objects became highlighted when picked or when the players' hands were in touching distance. In addition, objects that could interact with held objects also became highlighted when both objects touched. There was no gravity: if an object was let go of in mid-air, it stayed there until it was picked up again.

As a neutral actor, we used a robot from Unity's 4.0 Mecanim Animation Tutorial [189], which we modified in Maya 2017 and Unity 5.6.1 to allow resizable hands. The participant could look down and see their virtual body. The avatar hand provided all degrees of freedom for movement of the twenty finger joints, but did not perform subtler movements such as skin stretching and palm flexing. The game models were created in Maya 2017, textures were designed in Adobe Photoshop CC 2017 and PaintTool SAI, and game functionality was implemented in Unity 5.6.1.

4.1.5 Procedure

At the start of the experiment, participants are asked to sign a consent form and answer a cybersickness pre-experiment motion sickness questionnaire. Participants who answer "yes" to more than two of four cybersickness questions are eliminated from the study; none were.

Before putting on the Oculus Rift headset, participants are guided on how to adjust the spacing between the lenses to match their interocular distance and, if necessary, how to put on the headset with glasses. Participants are assisted with tightening and adjusting the headset for a satisfactory fit.

Prior to entering the VR environment, participants in the Gloves condition are instructed to pick up items by pinching with their thumb, index, and middle fingers. Participants using the Controllers condition are instructed to pick up items by grabbing with the primary thumb button and index finger trigger, resulting in a similar motion to the pinching action of the Gloves condition. Participants in all conditions determine their hand size by trying on the tracking gloves; the size of their virtual hands in the game environment is then adjusted to match their real-world hand size for increased presence. The avatar height and arm length are also adjusted to match those of each participant.

Participants are introduced to the concept of the experiment, an Escape Room video game in VR, described in Section 4.1.4, at the start of the study. During the course of the game, participants who take more than a set amount of time to solve a puzzle (dependent on the puzzle and determined in pilot tests; max: 255s, min: 27s, mean: 137s) are prompted with situational clues such as, "There is something below you that can be interacted with" or "That stove could use some fuel" and the key puzzle items also flash briefly.

Finally, participants are given time to explore and practice grasping, moving, and placing items in a training phase. Participants can color-match simple shapes and blocks to grow comfortable with the interaction methods and the virtual environment.

Game completion takes on average 7 minutes and 33 seconds, not including the average 92 seconds it takes for participant calibration and training. Once participants finish the game, they are offered congratulations and directed to complete a post-experiment questionnaire (Table 18) on a nearby desktop computer. When the questionnaire is completed, participants are asked whether they noticed the threat condition (the toothy mimic that tries to bite their hand), what they thought of it, and what they thought of the game. After these questions, participants choose whether to sign a release of information form for the data gathered during the experiment (all participants signed) and then receive their incentive card.

4.1.6 Measures

We investigate the influence of our four interaction conditions on the players' feeling of ownership of the virtual hands, the perceived realism, the perceived efficiency of the interactions, the players' enjoyment, and the players' feeling of presence. The effect of different interaction types on ownership, realism, and efficiency have been investigated in two recent studies [130, 50], and we compare our results to theirs. We furthermore examine the effect of our interaction conditions on presence and enjoyment, which are typical measurements for game experiences. Our questions are listed in Table 18. The questions on ownership, realism, and efficiency were adapted from previous studies. We use the Pens Presence [162] questionnaire as a measure of game presence, as PENS is statistically validated and generally comparable to other popular questionnaires IEQ and EEngQ [60]. Seven items slightly altered from the Intrinsic Motivation Inventory [6] are used to measure game enjoyment.

4.2 Results

Results of our experiment were analyzed with a two-way independent ANOVA. Levene's Test was used to assess the homogenity of variance across groups. A significant difference of variance ($F(3, 60) = 3.47, p = 0.022$) was found in one measure, E2 on the IMI Enjoyment Questionnaire (Table 18). All measures were significantly non-normal. Therefore, we attempted a robust ANOVA that included trimming the means, but it did not yield any differences in significant results compared to the two-way ANOVA, and thus was not included in the results. All questionnaire results are summarized in Table 18.

Ownership We found a significant main effect of Control Mode for questions O3 ($F(1, 60) = 7.95, p < 0.01$) and O4 ($F(1, 60) = 7.47, p < 0.01$). As expected, participants reported higher levels of ownership when using gloves compared to using controllers. A significant main effect of Grasping Visualization was found for question O2 ($F(1, 60) = 7.72, p < 0.01$). Ownership was perceived to be greater when participants used the Tracked Hand visualization for grasping (see Figure 4.6).

Of the 52 (out of 64) participants who responded when asked about the threatening mimic, 27 (51%) reported that it was frightening in some way. Of those 27, 7 (26%) participants used controllers and 20 (74%) used gloves. Of the 25 (49%) who reported it as non-frightening or unnoticed, 14 (56%) used controllers and 11 (44%) used gloves. Pearson's chi-squared test showed a significant association between the type of Control Mode and whether participants reported the mimic as scary ($\chi^2 = 4.88, p < .05$). The odds of participants reporting the mimic as frightening were 3.5 (CI: 0.99, 13.9) times higher in the Gloves condition than in the Controllers condition.

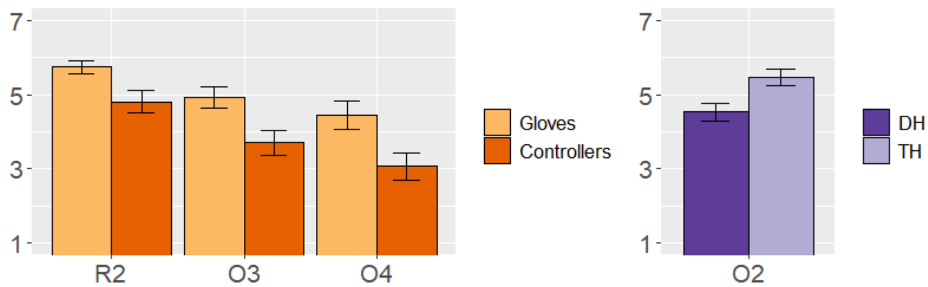


Figure 4.6: Main effects for Ownership and Realism. *Left:* Realism and Ownership were rated higher in the Gloves conditions than in the Controllers conditions. *Right:* Ownership was rated as greater in the Tracked Hand conditions than in the Disappearing Hand conditions.

Participants in either condition made comments such as "I didn't want to lose my hand" or

"I hesitated until I remembered it was VR." Twenty-two participants visibly jumped or exclaimed when the mimic chomped.

Realism A significant main effect of Control Mode was present for question R2 ($F(1, 60) = 7.66, p < 0.01$), see Figure 4.6. The motion of the hand was perceived to be more realistic in the Gloves condition than in the Controllers condition.

Efficiency We found no significant effects on perceived efficiency. However, analysis of game completion time showed a significant main effect of Grasping Visualization ($F(1, 55) = 4.14, p < 0.05$), with participants in the Disappearing Hand condition taking longer to complete the game than those in the Tracked Hand condition.

Enjoyment Significant main effects of Control Mode were present for three of the five game enjoyment questions: E2 ($F(1, 60) = 6.54, p < 0.05$), E3 ($F(1, 60) = 5.95, p < 0.05$), and E4 ($F(1, 60) = 4.72, p < 0.05$), with enjoyment being rated as higher by participants who used the gloves. Additionally, an effect of Grasping Visualization was found for E2 ($F(1, 60) = 4.28, p < 0.05$), with participants reporting enjoying the Tracked Hand visualization more. A significant interaction effect was found for question E1, but a Tukey HSD post-hoc test did not show any significant results. Main effects of enjoyment can be seen in Figure 4.7.

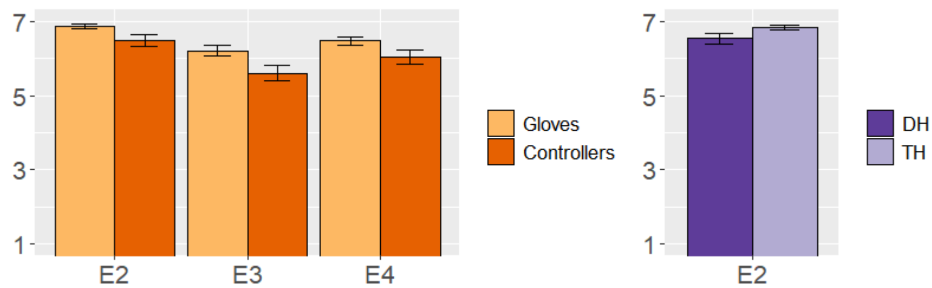


Figure 4.7: Main effects for Enjoyment. *Left:* Enjoyment was rated greater in the Gloves conditions than when using controllers. *Right:* Enjoyment was rated higher in the Tracked Hand conditions than in the Disappearing Hand conditions.

Presence Measuring presence yielded significant main effects for Control Mode for four of the nine presence questions: P5 $(F(1,60) = 7.13, p < 0.01)$, P7 $(F(1,60) = 5.87, p < 0.05)$, P8 $(F(1,60) = 4.42, p < 0.05)$, and P9 $(F(1,60) = 5.81, p < 0.05)$. For all effects the Glove condition induced higher perceived presence than the controllers (see Figure 4.8). An additional interaction effect was found for question P3; however, a Tukey HSD post-hoc test did not show any significant differences between conditions.

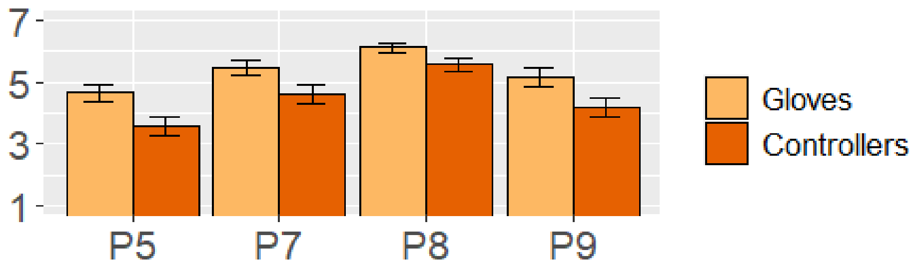


Figure 4.8: Main effect of Control Mode for Presence. For all significant effects of Control Mode on presence, the Gloves condition resulted in higher ratings than the Controllers condition.

4.3 Discussion

In this section, we discuss if our hypotheses are supported, compare our results to previous studies, discuss the advantages and disadvantages of using a game-like experience to study interactions in VR, and give tips for preparing such studies.

Ownership and Realism Our results confirm our hypotheses H1 (a) based on significant differences for questions O3 and O4 as well as the reactions to the threat. In all cases, ownership was perceived to be higher in the Gloves condition where the motions of the virtual fingers corresponded to the players' motions than in the Controllers condition that only displayed base poses. H1 (b) is only supported through one question (O2), so our evidence is only weak in this case. Participants in the Tracked Hand condition experienced higher ownership than those in the Disappearing Hand condition.

Realism metric R2 showed that participants perceived the movement of the virtual hands to be more realistic in the Gloves condition than in the Controllers condition, confirming hypothesis H2.

These overall results correspond to the findings from Lin *et al.* [130] and Canales *et al.* [50]. However, the results for each individual question are not always the same. Lin *et al.* averaged the answers to their ownership questions in their analysis and found a significant effect. We ran that analysis and also find a significant effect in that case. However, they did not find a significant effect for O3 when considered individually, which we do. Lin *et al.* also find a significant effect for question R1 (Gloves rated as more realistic than Controllers), which we do not (they did not ask R2). Canales *et al.* find a significant difference for O1 and two ownership questions that we did not ask, but not for O3 or O4. Findings from the different studies are shown in direct comparison in Figure 4.9

Efficiency We did not find significant differences between the Controllers condition and the Gloves condition when it comes to the perceived efficiency or the actual game completion. Thus we cannot confirm our hypothesis H3. Lin *et al.* find that the controllers were perceived to be more efficient than the gloves. In a simpler grasping task and in direct comparison, differences in efficiency might be more noticeable than in a relatively slow-paced game such as this one that is focused on solving puzzles.

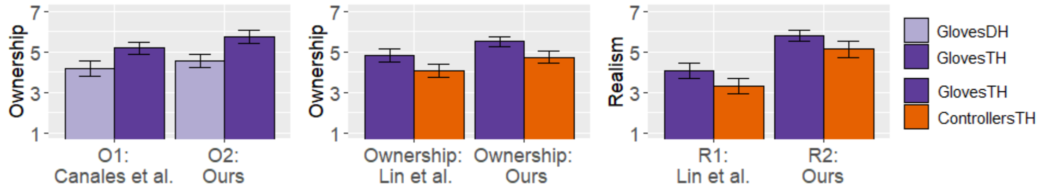


Figure 4.9: Comparisons of our results with previous works. *Left:* Both our work and Canales *et al.* [50] had significant main effects of Grasping Visualization on ownership, where the Tracked Hand condition (called Inner Hand in Canales *et al.*'s work) resulted in higher perceived ownership. *Center:* Taking the mean of all ownership averages shows a significant main effect of Control Mode in both Lin *et al.* [130] and our work. Using the Gloves caused higher perceived ownership. *Right:* Realism was perceived as significantly higher when using the Gloves as opposed to the Controllers in both Lin *et al.* [130] and our work. These graphs show that similar results were found in these experiments with different experimental setups and procedures.

Enjoyment and Presence We can confirm Hypothesis H4 (b), that enjoyment was higher in the Gloves condition compared to the Controllers condition, based on the significant differences in the answers of E2, E3, and E4. Enjoyment was rated very high in general, which shows that we successfully created an enjoyable game experience. Hypothesis H4 (a), that enjoyment will be greater for the Tracked Hand than for the Disappearing Hand, was only supported by a significant effect of E2, so the evidence in this case is too weak to draw confident conclusions.

We find evidence to support Hypothesis H5 (b) but not H5 (a). Presence questions P5, P7, P8, and P9 all showed that using gloves to interact in VR lead to a greater feeling of presence when compared to using controllers.

Experiments with Game-like Experiences As a goal of VR research is to understand our perception to create better VR experiences, our findings and hypotheses should be confirmed in scenarios that are similar to actual user experiences outside of lab settings in addition to experiments with repetitive tasks (not instead). However, the design of such studies also presents many challenges: The development of a suitable game can be very laborious, the variance between participants' reactions can be increased through further confounding factors such as how skilled participants are at playing specific games, and the number of participants needed is typically larger (based on the estimated variance and the fact that such studies might require between-subjects designs). Furthermore, effects might become diluted in some types of games. For example, it is more difficult to measure efficiency and performance in a game that focuses on slow-speed puzzles than in a first-person shooter where speed is a key to success. Being able to compare different conditions without distractions in a repetitive design might lead to the participants' "recalibration of the scale" and show more subtle differences. However, these differences might then not be important in a more immersive application. Despite the challenges, we consider experiments using more realistic applications as a necessary and important addition to studies with procedures using repetitive tasks because they can provide more true-to-life observations of immersive virtual experiences.

When planning such an experiment, we recommend to adjust the game type to the concepts being studied. Different types of games might need to be used to evaluate different concepts, and ideally, the same concepts would be tested in several scenarios. Ideally, a series of applications of different types would be accessible for experiments in the community, so that hypotheses can be tested in a variety of genres.

4.4 Conclusion, Limitations, and Future Work

In this paper, we present a study that investigates the effect of two control modes (Gloves vs. Controllers) and two grasping visualizations (Tracked Hands vs. Disappearing Hands) on ownership, realism, efficiency, enjoyment, and presence when playing an Escape Room game in which players interact with objects to solve puzzles. Our results show that ownership, realism, enjoyment, and presence significantly increased when using hand tracking (Gloves) as an input modality compared to controllers. We also found limited evidence that a Tracked Hand visualization increases ownership and enjoyment compared to a virtual hand that disappears during grasps.

We therefore recommend to take hand-tracking into account as an input modality instead of controllers when creating VR applications, and to continue to improve this technology and increase its accuracy for consumers. Our results were found using a motion capture system that was specifically developed to track hand motions in real-time. Further studies would need to demonstrate if our findings would be the same with current commercially available hand tracking devices.

A limitation of this work is that the user's hands are represented by a robotic model low in realism. While this model is in line with the model used in previous studies and allows for better comparison, the results might look different with a more realistic hand model. Additionally, our grasping representations are not realistic as the participants' fingers intersect with the object when grasping if they do not disappear. Interestingly, none of the participants commented on the hands moving through the objects. Future work could investigate the effect of hand model and grasping representation realism in game-like experiences. It would also be interesting to investigate whether visualizing the hands with controllers in the Controllers condition would affect results. Finally, we only tried one game and can not generalize our results to other games or genres. Exploring our results with experiments that use other game genres of varying levels of immersion, use players of different experience, or use existing games with modifications would further the generalizability of our findings.

While our results can not be generalized to other games, one has to also be cautious when generalizing studies with a repetitive design. We often can not confirm with certainty that such results will still be the same with an altered task or a different participant sample, who might, for example, have more experience with virtual reality. Most research progress in our field (and in any other field) is not made through individual studies but through many studies. Findings need to be replicated and validated in different contexts. While we do not replicate other studies — we would need to accurately follow the exact same protocol to do so — verifying how specific conditions are perceived in different situations can reinforce and strengthen findings, which is one of the main contributions of this paper.

Chapter 5

Study 3: Survive on the Moon! How Interaction Controls Impact Collaborative VR

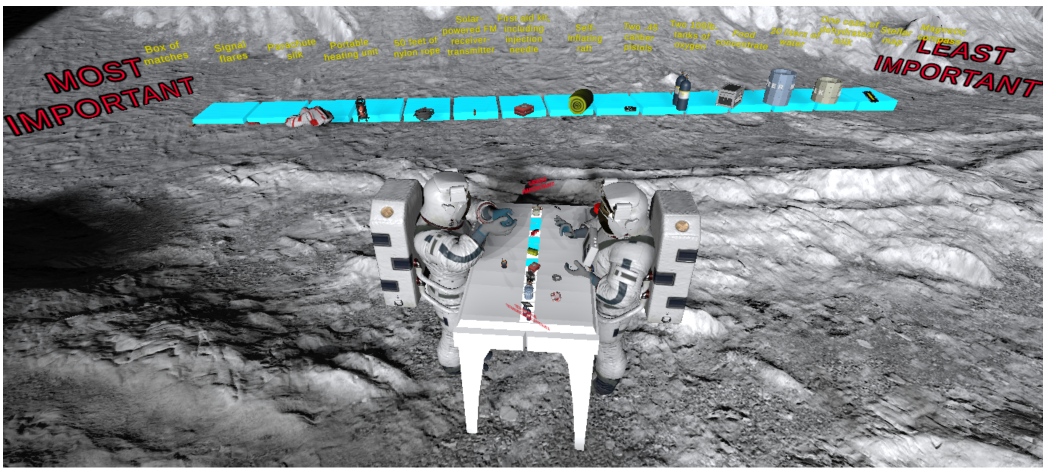


Figure 5.1: An image of the VR scenario. Two participants are negotiating which items are important to their survival in NASA's Survival on the Moon task, using the VR Controller condition.

This final study synthesizes themes of hand motions on communication from the first study and interaction from the second into a single project. This study explores whether the beneficial effects of hand motions on communication and presence hold true when studied in a one-to-one real-time collaborative environment. Will high-fidelity tracking still improve social presence and perceived comprehension for users interacting with each other, or are the simple gestures created by common VR controllers enough? We use Interaction Control conditions (similar to the Control Mode conditions from Study 2) to investigate this question.

In this between-subjects study, dyads of participants work first alone, then together, within a virtual space to solve NASA's *Survival on the Moon* [84] teamwork evaluation exercise. Participants first individually, then collaboratively, rank the importance of 15 items to their survival within a given scenario. They can communicate both verbally and non-verbally.

We implement two Interaction Control conditions for the dyads to interact with their environments: Hand-Tracking and VR Controllers. Participants enter the virtual scene using the Oculus Quest 2 head-mounted display; the Hand-Tracking level of control utilizes the Oculus Quest 2's hand-tracking technology and the VR Controller level uses the Oculus Quest 2's Touch Controllers. All participants view and interact with an identical virtual scene and use the same Interaction Control condition as their partner.

To determine effects of Interaction Control in a shared environment, we measure participants' levels of social presence and perceived comprehension of their partner. We also investigate the impact on team cohesion and we calculate a cognitive synergy score [144] determined by the individual and the dyad's final ranking of the items. Finally, we utilize NASA's Task Load Index Questionnaire as a measure of mental load, which may be affected by Interaction Control.

While we do not find any significant main effects of Interaction Control on the primary measures, we do find near-significant trends for Social Presence and Task Workload in favor of the Hand-Tracking condition. We find a significant effect on Task Duration; participants who use Hand-Tracking take significantly longer than those who use VR Controllers to individually rank the items.

5.1 Experimental Design

This study uses a between-subjects design to evaluate the effects of the Interaction Control conditions Hand-Tracking and VR Controller on social presence, perceived comprehension, team cohesion, and group synergy between dyads. We furthermore investigate task workload. Participants work collaboratively on a negotiation task within a virtual space while using one of the Interaction Control conditions with an Oculus Quest 2 VR head-mounted display (HMD).

5.1.1 Interaction Control Conditions & Virtual Hands

Interaction Control Conditions	
Hand-Tracking	VR Controller

All participants use the Oculus Quest 2 HMD; VR Controller participants use the Oculus Quest 2 Touch controllers while the Hand-Tracking participants use their own hands to interact within the VR scene. Both Interaction Control conditions use the native Oculus Quest 2 tracking technologies, implemented in Unity 2020.3.34f1 with Oculus' Interaction SDK version 40.0 [3] and using Oculus' Legacy OVRPlugin [5]. Each player saw their own hands as represented by the Oculus Interaction SDK package, virtual hands that either closely mimic their own hand movements for the Hand-Tracking condition, or are selected from a series of common hand poses for the VR Controller condition. The poses in the VR Controller condition are determined by what buttons are being pressed or touched on the controller. Sensors exist for the thumb button, index finger trigger, and middle/ring finger trigger, creating a total of 8 poses for the virtual hands. Participants can grab items in the Hand-Tracking condition by pinching the item with their thumb and index finger. VR Controller participants press the thumb button and the index finger trigger when hovering over an item to grasp it; the virtual hand makes a similar pinching pose as in the Hand-Tracking condition.

The participants' virtual astronaut hands are each rendered twice, once locally for their own viewing, and again on the server for their partner to see. The local hands are rendered by the Oculus Interaction SDK which uses input from the Oculus Quest 2 HMD to set the hand joints' positions and rotations. These orientations are sent to the server to set the server's hands' joint positions and rotations. Both sets of virtual hands are scaled to match the participants' own hand size. Unfortunately, due to inverse kinematic constraints, the local hands would pull away from the avatar arms if the player reached far, as the avatar's torso does not move. All items were placed close to the players to mitigate this behavior.

Figure 5.2: The virtual Astronaut hands. The "relaxed" (top) and "grasping" (bottom) virtual hands as controlled by Hand-Tracking (left) and VR Controllers (right).

5.1.2 Hypotheses

Our hypotheses are as follows:

- **H1. Social Presence**: Hand-Tracking will increase perceived Social Presence compared to the VR Controller condition.

Study 1 of this dissertation (Chapter 3) finds a lack of hand motions degrades perceived social presence. We expect this effect to persist when communicating in real-time with the different levels of hand motion fidelity produced by the Interaction Controls.

- **H2. Perceived Comprehension:** Hand-Tracking will increase Perceived Comprehension compared to the VR Controller condition.

Results from Study 1 (Chapter 3) indicate that a lack of hand motions highly reduces comprehension and perceived comprehension. We expect this effect to persist when communicating in real-time with the varying levels of hand motion fidelity produced by the Interaction Control conditions.

- **H3. Team Cohesion**: Hand-Tracking will increase Team Cohesion compared to the VR Controller condition.

Xue and Mbarika found that face-to-face teams had higher cohesion than virtual teams [206]. We expect that the higher media richness of the Hand-Tracking condition that is closer to a face-to-face experience will improve team cohesion.

- **H4. Group Synergy**: Interaction Control will not influence group synergy.

Previous works have not found that performance improves between mediums [11, 179] for intellective tasks as defined by McGrath [137]. Kim *et al.* [116] similarly find that differing multimedia modes do not affect group synergy. We do not anticipate a difference in our study.

- **H5. Task Workload**: Hand-Tracking will increase task workload when compared to the VR Controller conditions.

Hameed *et al.* [86] found that mental workload increased for a reach-pick-place task when using hand tracking vs. controllers. We expect to find similar results due to the more complex grasping and movement of the Hand-Tracking condition.

5.1.3 Participants

We recruited 42 participants for this IRB approved study; 20 participants (10 dyads) used the Hand-Tracking condition, and 22 (11 dyads) used the Controllers condition. 36% reported themselves as female and 64% as male. Participants' ages ranged from 18 to 44, with a mean age of 25. Participants were recruited locally via flyers, hand-outs, e-mail, Reddit, and word of mouth, with most being university students. Conditions alternated and were assigned based on the participants' identification number, which were assigned sequentially. An uncounted dyad was excluded from analysis as one of the pair could not fit their glasses into the headset, preventing them from participating.

5.1.4 The Astronaut Avatar

The virtual avatars within the shared virtual spaces are identical for each participant. The full-body avatar is represented as an astronaut and has a non-expressive face rendered as an opaque reflective helmet. Previous works show higher usability when utilizing a full-body avatar [194], so inverse kinematics steered the astronaut avatar's arms based on the hands' positioning. Additionally, the astronaut's helmet rotates within a limit according to the participant's headset rotation. The avatars are standing, but only their torsos and up are visible above the virtual table unless participants lean far out of the operating space. These settings are present in both the single and multi-player task phases.

Previous work has found that more realistic hand representations create a higher sense of body ownership [129], but that a mismatch between the gender of the participant and their virtual hand can cause reduced presence [168], and that a race mismatch can alter body ownership and behavior [114]. The anonymous astronaut avatar allows us to use the same avatar for all participants without creating any mismatches, and it matches the task setting (described in Section 5.1.5).

5.1.5 The Scenario: NASA's Survival on the Moon Task

Participants are placed into a virtual scene designed around NASA's Survival on the Moon task (Figure 5.1). The scenario is that the participants have crash-landed on the moon 200 miles from their rendezvous point with a mother ship. Their survival depends on reaching the mother ship. Participants must rank 15 items scavenged from their ship (e.g., a box of matches, tanks of oxygen, flare gun, stellar map, etc.) in order of importance to their survival of the 200-mile trip. The virtual environment is set on the moon, with miniature virtual models of the items to be ranked placed on a table in front of them, visible in Figure 5.3. The mini items' descriptions appear when the item is being grasped. Real-size models of the items are visible in the near distance and always have their descriptions hovering above them (see Figure 5.1). Upon the table are 15 spaces to place the items ordered from most to least important. The miniature items snap into place when set upon one of the spaces and the space changes color to indicate it is occupied. The real-size items appear on correlating spaces nearby when the miniature items are placed.

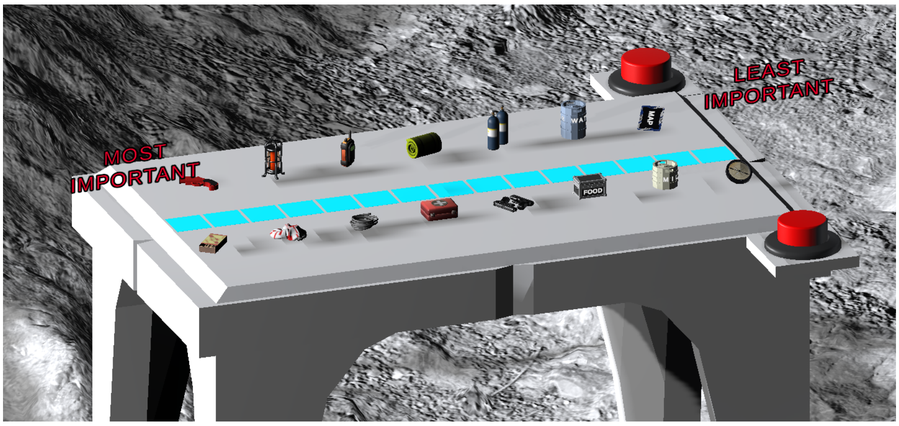


Figure 5.3: The table and operating space that participants use to rank items. An image of the multi-player operating space where participants negotiate the importance of the items. In the single-player phase the items appear on the same side of the table as the participant. Here, in the multi-player phase, the items are distributed to either side of the table. Participants are randomly placed on either side of the table and they remain on that side for both phases.

Each participant first ranks the items individually by placing them onto the ranked spaces within individual scenes without seeing or communicating with their partner. Once both participants have completed ranking the items individually, they are placed within a shared scene together. They can see each other's astronaut avatar and speak with one another through the HMD's microphone and speakers. They then collaboratively negotiate and discuss which scavenged items are the most important to their survival. They rank the items accordingly by picking up the items and placing them into ranked item spaces. Both participants can interact with all scavenged items, though only one hand may grasp one item at a time. If items leave the operating space above the table they reappear in their original positions so that an item may not be lost. There are no time restrictions on either phase of the task.

5.1.6 Procedure

Two participants are welcomed concurrently by two researchers into separate laboratory spaces. They are asked to complete and sign consent and cybersickness forms. If there are no disqualifiers among either participant they fill out a demographics questionnaire on a nearby desktop computer. The researcher measures the participants' interpupillary distance (IPD) using the GlassesOn mobile app [2], then presents the scenario description and instructions on a piece of paper. While participants review the scenario the researchers adjust the HMD to their IPD. Participants are directed to rank the items based on their textual names, as in the original NASA scenario, instead of their visual appearance. Once participants have reviewed and confirmed their understanding of the scenario, the researchers instruct them on how to grasp and manipulate the virtual items, as described in subsection 5.1.1. Finally, the participants are shown to their respective VR station and are aided in putting on and adjusting the HMD VR headset. Participants remain seated for the duration of the scenario. They are guided on how to open the Moon application and how to input their participant number. Once they have done so they are now present within the virtual moon environment and may begin ranking the virtual items upon the virtual table in front of them.

Once participants have ranked the items individually, they press a virtual button and are placed into the multi-player environment. This environment is identical to the single-player environment, except that the virtual items do not appear until both players have entered the scene. Once both participants have moved into the multi-player phase, their microphones activate and they can now negotiate and rank the items' importance together.

Once the VR phase of the experiment is complete, participants are asked to fill out the post-experiment questionnaire (described in subsection 5.1.7 and listed in Table 19) on the nearby desktop computers. Finally, each participant is invited to a debrief together in an adjacent laboratory space. They are told their individual and group scores as compared to expert rankings and awarded a $5 incentive card, and thanked for participating. Ideally participants do not know each other, but an additional 5pt Likert question in the post-experiment questionnaire asks how well they knew their partner before the experiment.

5.1.7 Measures

5.1.7.1 Social Presence & Perceived Comprehension

Social Presence, the sense of another being present within the same space, is measured with Nowak and Biocca's [150] measure of social presence, originally sourced from Short *et al*, [173] as in Study 1 (Chapter 3). Social presence is a common measure in social VR spaces and predicts many other outcomes.

Biocca *et al*'s Networked Minds Social Presence Inventory [36, 35], includes a measure of perceived comprehension that we use for this study. Accurate comprehension implies effective communication and will likely be affected by Interaction Controls, as seen in Study 1 (Chapter 3). Both of these measures use 7-pt Likert scales.

5.1.7.2 Team Cohesion

Team Cohesion is widely defined but often serves as a measure of team performance and dynamics [53, 187]. Michalisin et al. [145] define a 5-pt Likert team cohesion assessment (see Table 19) that focuses on good working relationships, high contribution levels, and shared commitment to completing the group task, all metrics that are strongly associated with cohesion [169, 110, 197].

5.1.7.3 Performance & Group Synergy

NASA's Survival on the Moon task (Section 5.1.5) provides rankings performed by experts, with rationale for each item. These official rankings act as an answer ke and are provided to participants during the post-experiment debrief. Participants' rankings, individual and group, are scored compared to NASA's rankings. A lower score indicates better task performance with a range from 0 to 112.

Participants' scores are compared to produce weak and strong group synergy scores, which are measures of the group's performance compared to the individual's [121]. Group synergy is based on the gain in performance between individuals and the group and has been previously investigated with NASA's Survival on the Moon task [144]. Weak synergy is stronger if the collective performance is better than the average individual performance, and strong synergy increases if collective performance is better than the best individual in the group.

5.1.7.4 Task Workload

Participants' workload is measured via NASA's Task Load Index (TLX) [90]. NASA's TLX questionnaire is commonly used in VR research. Harris *et al.* [88] recently developed and validated an updated version of the index explicitly designed for Virtual Reality. All questions are asked on a 1-10 rating or as a choice between two factors. Previous work has indicated that VR [82, 81] and hand-tracking [86] increase mental task load, but this has not been investigated in a collaborative environment.

5.2 Results

Results of the experiment were analyzed using the Wilcoxon Mann-Whitney Rank Sum Test. Levene's Test found no significant difference in the homogeneity of variance across the conditions. All measures were found to be significantly non-normal using Shapiro-Wilks normality test. Details are summarized in Table 19.

We found no significant main effects for any questionnaire-based measures, though trends ($p < 0.15$) were found in Social Presence (Figure 5.4), and Task Workload (Figure 5.8) measures. A significant main effect was found for individual task duration (Figure 5.7).

5.2.1 Social Presence

No significant effects were found for Social Presence (Figure 5.4), but a trend did exist for SP3 (To what extent was this like you were in the same room with the virtual character?): $W = 285.5, p = 0.0796, r = -0.271$. Perceived social presence was higher for Hand-Tracking ($median = 6$) than for VR Controllers ($median = 5$).

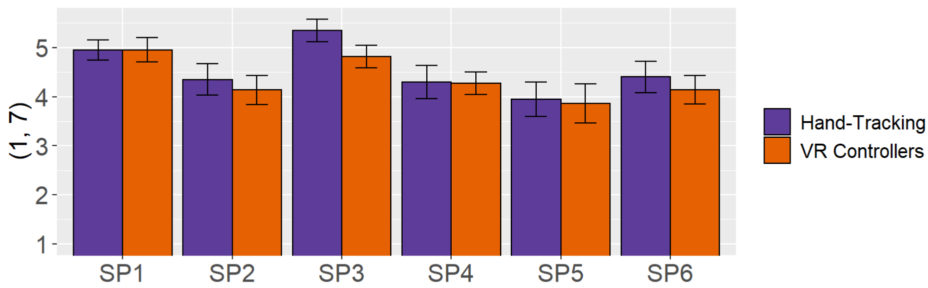


Figure 5.4: Results for Social Presence

5.2.2 Perceived Comprehension

We found no significant effects or trends for Perceived Comprehension (Figure 5.5).

5.2.3 Team Cohesion

We found no significant effects for Team Cohesion (Figure 5.6). However, dyads knowing their partner (asked on the post-experiment questionnaire: Before this experiment, I knew my teammate very well) did significantly predict responses for TC2 (I wish I were on a different team. (R)) when analyzed with a linear regression model $R^2 = 0.0766, F_{(1,40)} = 4.4, p = 0.0434$.

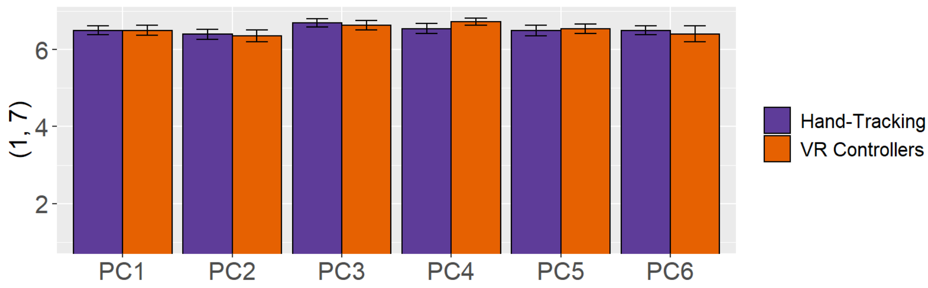


Figure 5.5: Results for Perceived Comprehension

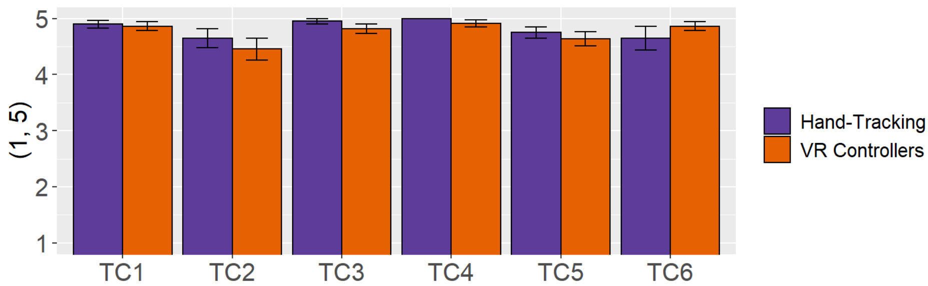


Figure 5.6: Results for Team Cohesion

5.2.4 Performance, Synergy, and Task Duration

Performance is measured as the accuracy of the item rankings when compared to NASA experts' rankings. The absolute differences of each item's ranking, compared to its official NASA ranking, are summed as a performance score. A lower score indicates better performance. Scores were recorded as individual scores ($mean = 41.05, std = 11.75$) and group scores ($mean = 36.86, std = 7.6$); there were no trends or significant effects of the Interaction Control. Groups nearly performed significantly better than individuals when analyzed independently $W = 1094, p = 0.05782, r = -0.207$.

Group synergy was calculated as two factors: weak synergy and strong synergy, as in Meslec and Curşeu's [144] investigation of group synergy when using the Survival on the Moon Task 5.1.5. Weak synergy is computed as the difference between the group's score and the mean of the pair's individual scores. Strong synergy is the difference between the group score and the best performing individual's. Lower synergy calculations indicate better synergy due to lower performance scores being the better scores. We did not find any trends or significant effects.

Task duration was recorded for both individual ($mean = 319.83, std = 143.25$) and group phases ($mean = 361.89, std = 193.33$). We found a significant main effect of Interaction

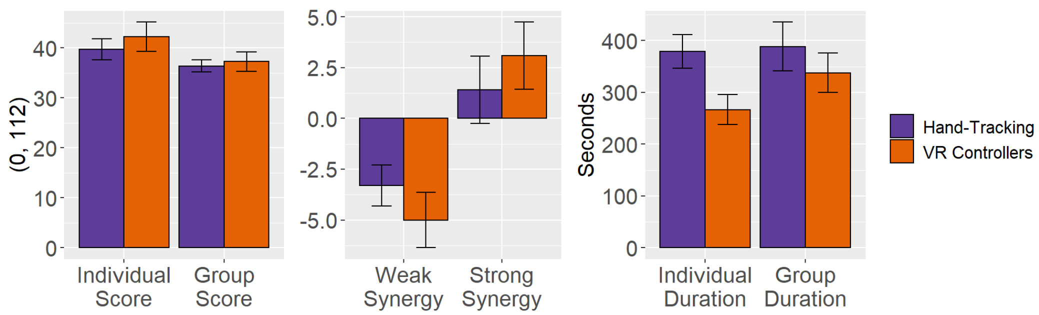


Figure 5.7: Results for Performance (left), Synergy (mid), and Duration (right). Lower values indicate better performance and synergy. Duration is in seconds.

Control on individual Task Duration $W = 257, p = 0.024, r = -0.161$, with Hand-Tracking taking longer ($median = 324.4$s) than VR Controllers ($median = 277.9$s).

5.2.5 Task Workload

The adjusted task workloads were calculated by multiplying the task demand (TLX 1-6, see Table 19) by their weights, which are the summed cumulative score derived from the 15 comparisons between demand dimensions. An adjusted Overall Workload was determined by summing the Task Workloads and dividing by 15 to account for the comparisons.

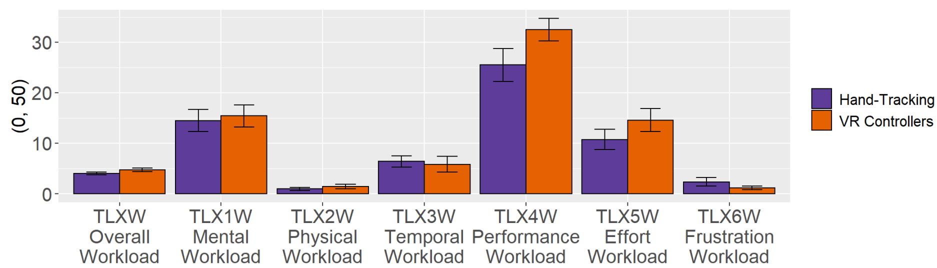


Figure 5.8: Results for Task Workloads. These metrics are adjusted (W) from the initial TLX measures.

No significant effects were found for the task workload metrics. However, trends were found for the adjusted Overall workload $W = 153, p = 0.0939, r = -0.259$ and the Performance workload $W = 155.5, p = 0.105, r = -0.25$ (TLX4. How successful were you in accomplishing what you were asked to do in the overall experience?). For both metrics workload was found to be higher for the VR Controllers (Overall $median = 4.567$; Performance $median = 30$) than for the Hand-Tracking (Overall $median = 4$; Performance $median = 25.5$).

5.3 Discussion

While we found no significant effects of Interaction Control for our main measures, trends ($p < 0.15$) can be seen for Social Presence and Task Workload. These trends indicate higher Social Presence for Hand-Tracking and higher Task Workload when using VR Controllers.

We hypothesized in **H1** on **Social Presence** that Hand-Tracking would increase perceived Social Presence; these trends match this hypothesis and are supported by Study 1 (Chapter 3). So while we cannot definitively confirm H1 for this study, there is strong support for Hand-Tracking increasing Social Presence in shared and embodied virtual experiences. This effect on Social Presence could be due to the more accurate hand motions of the Hand-Tracking players' virtual avatars while participants were communicating with one another.

Both **H2** on **Perceived Comprehension** and **H3** on **Team Cohesion** hypothesized a positive effect of Hand-Tracking on these measures. Participants were able to audibly speak with each other over the HMD microphone and speakers; this active auditory communication mechanism likely offset the need for higher fidelity Interaction Controls when considering comprehension and cohesion. Participants did not need to rely on non-verbal communication for this negotiating decision-making task while being able to utilize the information-rich audio channel.

H4 did not predict any effect of Interaction Control on **Group Synergy**. Group Synergy was based on participants' individual and group performances; as the *Survival on the Moon* task is a decision-making negotiation task, performance was not likely to depend on Interaction Control. We do find one significant effect of Interaction Control on Individual Task Duration. While we did not see this effect in Study 2 (Chapter 4), previous work by Lin et al. [130] also found that task duration was significantly longer for participants that used the Glove condition rather than the Controllers. The more complex grabbing interaction of pinching fingers for Hand-Tracking as opposed to the button-pressing of VR Controllers may contribute to this effect.

We predicted in **H5** on **Task Workload** that using Hand-Tracking would increase mental workload, as found by Hameed *et al.* [86], but our trends do not support this and we cannot provide any support for **H5**. Instead, the Task Workload trends indicate that VR Controllers increase Overall and Performance workload in a collaborative VR space. This effect could be attributed to the ease of grasping and naturalness of use of the Hand-Tracking condition. Further work in this space could explore these contradictory results.

For all hypotheses, it is possible that the expected effects were smaller than anticipated so that the number of participants was not sufficient to expose them. We may also consider Richard *et al.*'s [159] findings that increasing the number of participants in a between-subjects virtual embodiment study did not replicate previous results; a power-analysis indicated that between-subjects virtual embodiment studies may require as many as four times as many participants than a within-subjects design to show similar effects. Either way, as none of our effects are significant, we cannot confirm any of our hypotheses. We can conjecture about our lack of results.

Due to the nature of the task, participants seemed to focus more on the virtual items and their auditory conversations than the appearance and movements of their partner. Implementing a task that requires more critical deictic and iconic gestures may be better suited to answering this research question. Alternatively, the emulation of hand movements by modern VR controllers (the Oculus Quest Touch controllers, in this case) may be adequate for effective communication and team cohesion.

We intend to continue this work by removing the auditory communication channel and running a new set of participants under this and the Interaction Control conditions. It may be that the inclusion of a high capacity communication channel like audio mitigated the impact of gestures. We hypothesize that there will be an interaction effect between the presence of an audio channel and the Interaction Control conditions. With the absence of an information-rich auditory communication channel, the level of detail of hand motions and gestures will become more important to discussion and negotiation of item rankings.

Chapter 6

Conclusions and Future Works

This dissertation describes a series of studies that are amongst the first to investigate the influences of hand motions and interaction modalities on interacting and communicating within VR spaces. Recent advances in VR technologies have enabled the use of hand-tracking on the consumer level. VR developers and users have been quick to take advantage of these new features, but little research yet exists on the use hand-tracking in shared VR spaces. To evaluate how users' perception of themselves and others may be impacted by this new technology, we designed and developed three human-subject studies. The first study examines how a lack of hand motion data can affect communication, the second investigates how the interaction modality used impacts a user even in more immersive and game-like environments, and the third combines these concepts to look at how interaction modality influences communication between users in a shared VR environment. We ultimately find that detailed hand motions play a significant role in non-verbal communication and interactions that occur within virtual spaces. We show that a complete lack of hand motion data has severe detrimental effects when viewing a virtual character and that using controllers instead of hand-tracking decreases a user's immersion within a VR scene. We also find positive trends of using hand-tracking as an interaction modality when communicating with others in shared VR spaces.

Our first study explicitly examines the effects of hand motion fidelity on non-verbal communication. We recruited more than 1600 participants to investigate how important accurate hand motions are to comprehension, perception of a virtual character, and social presence. We tasked participants with watching videos of a character playing charades with various motion errors and alterations applied to his hand and finger motions. We find that small and even larger errors or alterations to hand and finger motions do not cause great detrimental effects but that a total lack of the original motion data has repercussions on comprehension, social presence, perception of the character, and user comfort. Jittering fingers were especially disliked by participants. The results obtained from this study indicate the importance of rendering hand motions without jitter, and the potential ramifications of other common errors that appear when tracking hands.

The second study utilizes an in-depth Escape Room VR game to explore the effects of now-common VR interaction modalities: hand-tracking and controllers. Using a between-subjects study design wherein we manipulated the control modes (gloves vs. controllers) and grasping visualizations (disappearing vs. tracked hand), we studied how these manipulations affect the user experience associated with performing interactions in a virtual escape room. Previous works have investigated control modes [130] and grasping visualizations [50], but none have done so together, and in a game-like environment. Many VR research studies investigate effects with repetitive and straight-forward experimental designs, but most VR users will not be interacting with virtual environments in this way. We sought to confirm whether effects observed in repetitive studies persisted within a more game-like VR scenario, more akin to a general VR user's experience. Results show that ownership, realism, enjoyment, and presence are all improved by using hand-tracking to solve puzzles in the immersive game environment. These results are comparable to other works and suggest that effects found in repetitive task experiments can be replicated in certain game-like experiments.

From our first study we learned that a lack of hand motion data is detrimental to comprehension, character perception, and social presence, but that even partial motion loss can have a negative effect. In the second study we find that using hand-tracking as an interaction modality improves and increases the immersion of a virtual reality experience. We devise our third study based on these results.

In our last study we investigate how the input modalities of hand-tracking and VR controllers might affect perceptions between users in shared virtual spaces. We know that input modalities affect user experiences associated with interactions and so hypothesize that they may also affect communicative user experiences. We utilize motion data degradation, which we know reduces comprehension and social presence, to examine how the limited motion data provided by VR controllers may affect real-time communication between users in a collaborative virtual reality experience. Dyadic teams of users were tasked with collaboratively ranking items in order of importance by manipulating them with one of the input modalities. We measured aspects such as social presence, team cohesion and user comprehension to understand what implications our manipulations had on the overall collaborative experience. We find trends of input modality on social presence as in the first study, but no effects on perceived comprehension or team cohesion. We also find a trend of increased workload for VR controllers, which contradicts previous works [86] and will require further studies to be fully understood.

With no significant effects, the results from our last study are not conclusive. Hypotheses of improved perceived comprehension derived from the results of our first study were not supported. It is likely that adding an information rich communication channel like audio mitigates the importance of detailed hand motions for communication. However, adding a non-verbal condition to a future version of this final study would possibly yield an interaction effect of audio and interaction modality. Without the benefit of speech, participating dyads would need to rely heavily on non-verbal methods of communication. The higher fidelity hand motions enabled by hand-tracking would likely contribute to better perceived comprehension when compared to the limited finger motions generated by controllers. Other measures, like team cohesion and social presence, may also yield observable effects with a lack of audio to aid in communication.

Another future avenue of investigation lies in the contradictory results on workload when compared to previous works. We found a trend of lower perceived worload when using hand-tracking, opposite to what Hameed *et al.* [86] found. The difference in findings could be attributed to differences in technology between studies (Oculus Quest vs. Oculus Quest 2), differences in task design and execution, or perhaps simply being in a shared virtual space had an influence on perceived workload. Further works could investigate the discrepancy in results.

There also exists a wide avenue of potential research when examining both hand motions and communication in VR. How may a user's perception of another be dependent on the presence and accuracy of that other user's hand motions? Are user avatars perceived as more friendly or persuasive when they have detailed hand motions compared to others who do not? Is deception harder or easier to detect when hand motions are absent in shared VR spaces? We can also consider how differing types of interaction modalities between users may matter. In a collaborative task, if one user can finely manipulate objects within a scene and the other cannot, will frustration or workload increase?

As VR progresses as a viable remote collaboration tool, accounting for the answers to these questions in HMD and application design may very well make or break a user's VR experience. New users of a technology can be easily turned off by frustrating experiences, so ensuring that communication and interaction within virtual spaces is as intuitive and seamless as possible will be vital in ensuring Virtual Reality's future as a useful and popular collaborative tool.

Appendices

Alteration: Comprehension

Question	Effect	F-Test and p-value	Post-hoc
Motion Comprehension	Motion Alteration	$F_{(7, 473)} = 27.94, p < 0.001$	No Data conditions < all Full and Partial data conditions
	Avatar	$F_{(1, 473)} = 9.47, p < 0.01$	Mannequin < Realistic
	Clip Length	$F_{(1, 473)} = 30.59, p < 0.001$	Long < Short
Perceived Comprehension	Motion Alteration	$F_{(7, 839)} = 18.61, p < 0.001$	Random and Static < all Full and Partial data conditions
			Passive < all Full and Partial data conditions except Reduced
	Clip Length	$F_{(1, 839)} = 12.17, p < 0.01$	Long < Short
	Avatar	$F_{(1, 839)} = 5.51, p < 0.05$	Mannequin < Realistic
	Motion Alt : Clip Len	$F_{(7, 839)} = 3.71, p < 0.01$	StLong < Long(O, Re, J, Po, Sm), Short(O, Re, J, Po, Sm, Pa, Ra, St)
			PaLong < Long(O, J, Po, Sm), Short(O, Re, J, Po, Sm)
			StShort < Long(J, Sm), Short(O, Sm)
			RaLong < Long(O, Re, J, Po, Sm), Short(O, Re, J, Po, Sm, Pa, Ra)

Alteration: Perception of Character

Question	Effect	F-Test and p-value	Post-hoc
Naturalness	Motion Alteration	$F_{(7, 839)} = 14.51, p < 0.001$	Jitter < Original; Passive < Original, Reduced, Popping, Smooth
			Random < Original, Smooth
			Static < Random, all Full and Partial data conditions
	Clip Length	$F_{(1, 839)} = 11.18, p < 0.01$	Long < Short
	Avatar : Clip Len	$F_{(1, 839)} = 10.22, p < 0.01$	Mannequin:Long < all other conditions
Realism	Motion Alteration	$F_{(7, 839)} = 6.13, p < 0.001$	Passive < Smooth; Static < Orig, Reduced, Popping, Smooth
	Avatar	$F_{(1, 839)} = 55.65, p < 0.001$	Mannequin < Realistic
	Clip Length	$F_{(1, 839)} = 6.39, p < 0.05$	Long < Short
	Avatar : Clip Len	$F_{(1, 839)} = 7.71, p < 0.05$	Mannequin:Long < all other conditions
			Mannequin:Short < Realistic:Long, Realistic:Short
Appeal	Motion Alteration	$F_{(7, 839)} = 5.74, p < 0.001$	Random < Smooth; Static < all Full and Partial data conditions
	Avatar	$F_{(1, 839)} = 13.45, p < 0.01$	Mannequin < Realistic
	Clip Length	$F_{(1, 839)} = 10.52, p < 0.01$	Long < Short
	Avatar : Clip Len	$F_{(1, 839)} = 13.41, p < 0.01$	Mannequin:Long < all other conditions
Familiarity	Motion Alteration	$F_{(7, 839)} = 3.06, p < 0.05$	Static < Smooth
	Avatar	$F_{(1, 839)} = 30.70, p < 0.001$	Mannequin < Realistic
	Clip Length	$F_{(1, 839)} = 7.09, p < 0.05$	Long < Short
	Avatar : Clip Len	$F_{(1, 839)} = 9.40, p < 0.05$	Mannequin:Long < all other conditions
Assuredness	Motion Alteration	$F_{(7, 839)} = 3.00, p < 0.05$	Static < Original
	Avatar	$F_{(1, 839)} = 10.07, p < 0.01$	Mannequin < Realistic
Friendliness	Motion Alteration	$F_{(7, 839)} = 4.18, p < 0.01$	Static < Original, Reduced, Smooth
Trustworthiness	Motion Alteration	$F_{(7, 839)} = 5.20, p < 0.001$	Random < Original, Reduced
			Static < Original, Reduced, Jitter, Smooth
	Avatar : Clip Len	$F_{(1, 839)} = 6.69, p < 0.05$	Mannequin:Long < Realistic:Long
Agreeableness	Avatar	$F_{(1, 839)} = 13.60, p < 0.01$	Mannequin < Realistic
Conscientiousness	Motion Alteration	$F_{(7, 839)} = 3.93, p < 0.01$	Random < Original, Jitter, Popping, Smooth
			Static < Original
Openness to Experience	Motion Alteration	$F_{(7, 839)} = 3.12, p < 0.05$	Random < Smooth; Static < Smooth
	Clip Length	$F_{(1, 839)} = 13.55, p < 0.01$	Long < Short

Alteration: Social Presence

Question	Effect	F-Test and p-value	Post-hoc
Social Presence	Motion Alteration	$F_{(7, 839)} = 5.67, p < 0.001$	Static < Full and Partial conditions except Reduced
	Avatar	$F_{(1, 839)} = 6.26, p < 0.05$	Mannequin < Realistic

Table 1: Study 1: Detailed results for the Alteration Experiment.

All significant results with $p < 0.05$ are listed. Non-significant effects are not included.

Jitter: Comprehension

Question	Effect	F-Test and p-value	Post-hoc
Motion Comprehension	Clip Length	$F_{(1, 255)} = 15.02, p < 0.01$	Long < Short

Jitter: Perception of Character

Naturalness	Motion Intensity	$F_{(3, 458)} = 29.53, p < 0.001$	JitterLow < Original JitterMed < Original JitterHigh < Original, JitterLow, JitterMed
Realism	Motion Intensity Avatar	$F_{(3, 458)} = 4.72, p < 0.05$ $F_{(1, 458)} = 25.78, p < 0.001$	JitterHigh < Original Mannequin < Realistic
Appeal	Motion Intensity Avatar	$F_{(3, 458)} = 5.45, p < 0.05$ $F_{(1, 458)} = 12.09, p < 0.01$	JitterHigh < Original, JitterLow Mannequin < Realistic
Familiarity	Avatar	$F_{(1, 458)} = 23.80, p < 0.001$	Mannequin < Realistic
Assuredness	Motion Intensity Avatar	$F_{(3, 458)} = 16.27, p < 0.001$ $F_{(1, 458)} = 14.07, p < 0.01$	JitterMed < Original, JitterLow JitterHigh < Original, JitterLow Mannequin < Realistic
Conscientiousness	Motion Intensity	$F_{(3, 458)} = 6.44, p < 0.01$	JitterHigh < Original, JitterLow
Emotional Stability	Motion Intensity	$F_{(3, 458)} = 16.10, p < 0.001$	JitterMed < Original JitterHigh < Original, JitterLow, JitterMed

Jitter: Social Presence

Social Presence	Motion Intensity Motion Int : Clip Len	$F_{(3, 458)} = 4.22, p < 0.05$ $F_{(3, 458)} = 4.20, p < 0.05$	JitterHigh < Original JMed:Long < Orig:Short JHigh:Short < Orig:Short, JLow:Long JHigh:Long < Orig:Short

Popping: Comprehension

Motion Comprehension	Clip Length	$F_{(1, 247)} = 16.22, p < 0.01$	Long < Short
Perceived Comprehension	Avatar	$F_{(1, 456)} = 11.21, p < 0.05$	Mannequin < Realistic

Popping: Perception of Character

Realism	Avatar	$F_{(1, 456)} = 18.35, p < 0.01$	Mannequin < Realistic
Familiarity	Avatar	$F_{(1, 456)} = 11.80, p < 0.05$	Mannequin < Realistic
Assuredness	Avatar	$F_{(1, 456)} = 11.16, p < 0.05$	Mannequin < Realistic

Smooth: Comprehension

Motion Comprehension	Clip Length	$F_{(1, 263)} = 19.06, p < 0.001$	Long < Short

Smooth: Perception of Character

Realism	Avatar Avatar : Clip Len	$F_{(1, 466)} = 39.17, p < 0.001$ $F_{(1, 466)} = 8.91, p < 0.05$	Mannequin < Realistic Mannequin:Short < Long:Realistic Mannequin:Long < all other conditions
Appeal	Avatar	$F_{(1, 466)} = 9.52, p < 0.05$	Mannequin < Realistic
Familiarity	Avatar	$F_{(1, 466)} = 14.70, p < 0.01$	Mannequin < Realistic
Assuredness	Avatar	$F_{(1, 466)} = 14.08, p < 0.01$	Mannequin < Realistic

Smooth: Social Presence

Social Presence	Avatar	$F_{(1, 466)} = 12.11, p < 0.05$	Mannequin < Realistic

Table 2: Study 1: Detailed results for the Intensity Experiment.

Non-significant effects are not included.

VR COMPARE

Question	Effect	F-Test and p-value	Post-hoc
MOTION COMPREHENSION	MOTION CONDITION	$F_{(1, 31)} = 58.52$, $p < 0.001$	Static < Original
CONSCIENTIOUSNESS	MOTION CONDITION	$F_{(1, 31)} = 5.06$, $p < 0.05$	Static < Original

VR COMPARE - ALTERATION EXPERIMENT

AGREEABLENESS	EXPERIMENT	$F_{(1, 85)} = 28.37$, $p < 0.001$	VR Compare < Alteration Experiment
CONSCIENTIOUSNESS	EXPERIMENT	$F_{(1, 85)} = 12.76$, $p < 0.001$	VR Compare < Alteration Experiment
EMOTIONAL STABILITY	EXPERIMENT	$F_{(1, 85)} = 30.12$, $p < 0.001$	VR Compare < Alteration Experiment
OPENNESS TO EXPERIENCE	EXPERIMENT	$F_{(1, 85)} = 9.77$, $p < 0.01$	VR Compare < Alteration Experiment

VR COMFORT

COMFORT	MOTION CONDITION	$F_{(13, 448)} = 6.68$, $p < 0.001$	JMed < All other conditions except JLow, JHigh, PoHigh JHigh < All other conditions except JMed
NATURALNESS OF MOTION	MOTION CONDITION	$F_{(13, 448)} = 12.76$, $p < 0.001$	JLow < Original, Reduced, JHigh, Smooth(Low, Med, High) JMed < All other conditions except JLow, JHigh JHigh < All other conditions except JMed

VR RANK

Question	Effect	χ^2	Alteration 1	Alteration 2	Differences
RANKING COMFORT OF INTERACTION	MOTION CONDITION	$\chi^2(7) = 95.77$, $p < 0.001$	Original	Jitter	141
			Original	Passive	71
			Original	Static	93
			Reduced	Jitter	123
			Reduced	Static	75
			Jitter	Popping	135
			Jitter	Smooth	146
			Jitter	Passive	70
			Jitter	Random	93
			Popping	Passive	65
			Popping	Static	87
			Smooth	Passive	76
			Smooth	Static	98

Table 3: Study 1: Detailed results for the Virtual Reality experiment.

Non-significant effects are not included.

Factor	Question(s)	Format	Scale	Origin
Motion Comprehension	What word or noun is being acted out? What movie title is being acted out?	Text	n/a	-
Perceived Comprehension	I understood what the other meant. The other's thoughts were clear to me.	7pt Likert	Strongly Agree - Strongly Disagree	Biocca *et al.* [37]
Naturalness⋆	I see the virtual character as:	7pt Likert	Extremely Unnatural - Very Natural	-
Realism			Extremely Abstract - Extremely Realistic	McDonnell *et al.* [136]
Appeal			Extremely Unappealing - Extremely Appealing	
Familiarity			Extremely Unfamiliar - Extremely Familiar	
Assuredness			Extremely Eerie - Extremely Reassuring	
Friendliness			Extremely Unfriendly - Extremely Friendly	
Trustworthiness			Extremely Untrustworthy - Extremely Trustworthy	
Extraversion	I see the virtual character as: Extraverted, enthuestiac Reserved, quiet (reversed)	7pt Likert	Disagree Strongly - Agree Strongly	Gosling *et al.* [78]
Agreeablness	Critical, quarrelsome (reversed) Sympathetic, warm			
Conscientiousness	Dependable, self-disciplined Disorganized, careless (reversed)			
Emotional Stability	Anxious, easily upset (reversed) Calm, emotionally stable			
Openness to Experience	Open to new experiences, complex Conventional, uncreative (reversed)			
Social Presence	To what extent was this like a face-to-face meeting?	9pt Likert	A lot like face-to-face - Not like face-to-face at all	Nowak and Biocca [150]
	To what extent was this like you were in the same room with the virtual character?		A lot like being in the same room - Not like being in the same room	
	To what extent did the virtual character seem "real"?		Very real - Not real at all	
	How likely is it that you would choose to use this system of interaction for a meeting in which you wanted to persuade others of something?		Very likely - Not likely at all	
	To what extent did you feel you could get to know someone that you met only through this system?		Very well - Not at all	
Comfort	How comfortable would you feel interacting with this character for an extended period of time? ∗	7pt Likert	Not at all comfortable - Very comfortable	-
	Rank how comfortable you would feel interacting with this character from most comfortable (1) to least (8). ○	Assigning Ranks	1 - 8	-

Table 4: Study 1: Full details of the information gathering questionnaire.

All questions were asked in Alteration, Intensity, and VR Compare Experiments.
⋆ : Question also asked in VR Comfort; ∗ : Question only asked in VR Comfort; ○ : Question only asked in VR Rank.

Alteration Experiment: Motion Alteration Results

Measure	*Motion Comprehension*	*Perceived Comprehension*	*Naturalness*	*Realism*	*Appeal*
Values	$F_{(7, 473)} = 27.94$, $p < 0.001$	$F_{(7, 839)} = 18.61$, $p < 0.001$	$F_{(7, 839)} = 14.51$, $p < 0.001$	$F_{(7, 839)} = 6.13$, $p < 0.001$	$F_{(7, 839)} = 5.74$, $p < 0.001$
Means & STDs	Original = (0.65, 0.25) Reduced = (0.6, 0.26) Jitter = (0.7, 0.23) Popping = (0.68, 0.2) Smooth = (0.63, 0.26) Passive = (0.34, 0.21) Random = (0.42, 0.18) Static = (0.36, 0.2)	Original = (0.68, 1.39) Reduced = (0.35, 1.53) Jitter = (0.63, 1.21) Popping = (0.41, 1.37) Smooth = (0.7, 1.26) Passive = (-0.24, 1.55) Random = (-0.42, 1.41) Static = (-0.75, 1.47)	Original = (1.43, 1.29) Reduced = (1.2, 1.41) Jitter = (0.79, 1.48) Popping = (1.02, 1.38) Smooth = (1.21, 1.26) Passive = (0.39, 1.57) Random = (0.57, 1.46) Static = (-0.16, 1.71)	Original = (0.66, 1.5) Reduced = (0.66, 1.55) Jitter = (0.26, 1.56) Popping = (0.49, 1.51) Smooth = (0.81, 1.37) Passive = (0.09, 1.63) Random = (0.27, 1.55) Static = (-0.31, 1.58)	Original = (0.67, 1.34) Reduced = (0.66, 1.19) Jitter = (0.58, 1.29) Popping = (0.42, 1.34) Smooth = (0.7, 1.36) Passive = (0.29, 1.46) Random = (0.16, 1.44) Static = (-0.2, 1.55)

Measure	*Familiarity*	*Assuredness*	*Friendliness*	*Trustworthiness*	*Extraversion*
Values	$F_{(7, 839)} = 3.06$, $p < 0.01$	$F_{(7, 839)} = 3.00$, $p < 0.05$	$F_{(7, 839)} = 4.18$, $p < 0.01$	$F_{(7, 839)} = 5.20$, $p < 0.001$	$F_{(7, 839)} = 2.77$, $p = 0.05$
Means & STDs	Original = (0.51, 1.48) Reduced = (0.55, 1.5) Jitter = (0.52, 1.34) Popping = (0.35, 1.37) Smooth = (0.58, 1.46) Passive = (0.09, 1.47) Random = (0.16, 1.57) Static = (-0.05, 1.33)	Original = (0.55, 1.4) Reduced = (0.52, 1.31) Jitter = (0.51, 1.4) Popping = (0.2, 1.33) Smooth = (0.45, 1.38) Passive = (0.16, 1.41) Random = (0.08, 1.27) Static = (-0.01, 1.19)	Original = (1.32, 1.11) Reduced = (1.29, 1.07) Jitter = (1.1, 1.31) Popping = (1.05, 1.13) Smooth = (1.29, 1.15) Passive = (0.85, 1.24) Random = (0.9, 1.13) Static = (0.67, 1.33)	Original = (0.99, 1.06) Reduced = (1.12, 1.08) Jitter = (0.83, 1.38) Popping = (0.75, 1.19) Smooth = (0.95, 1.17) Passive = (0.63, 1.29) Random = (0.5, 1.2) Static = (0.32, 1.03)	Overall = (0.92, 1.16)

Measure	*Agreeableness*	*Conscientiousness*	*Emotional Stability*	*Openness to Experience*	*Social Presence*
Values	$F_{(7, 839)} = 1.44$, $p = 0.43$	$F_{(7, 839)} = 3.93$, $p < 0.01$	$F_{(7, 839)} = 1.18$, $p = 0.95$	$F_{(7, 839)} = 3.12$, $p < 0.05$	$F_{(7, 839)} = 5.67$, $p < 0.001$
Means & STDs	Overall = (0.99, 0.87)	Original = (1.1, 0.95) Reduced = (0.86, 1.01) Jitter = (1.04, 1.01) Popping = (0.99, 0.81) Smooth = (1.02, 0.83) Passive = (0.79, 0.94) Random = (0.62, 0.93) Static = (0.72, 0.89)	Overall = (1.2, 0.96)	Original = (0.7, 1.15) Reduced = (0.58, 1.19) Jitter = (0.86, 1.18) Popping = (0.75, 1.1) Smooth = (0.99, 1.1) Passive = (0.66, 1.2) Random = (0.52, 1.06) Static = (0.4, 1.18)	Original = (-0.3, 2.02) Reduced = (-0.86, 1.85) Jitter = (-0.51, 1.9) Popping = (-0.53, 2) Smooth = (-0.33, 1.97) Passive = (-0.92, 1.94) Random = (-0.95, 1.99) Static = (-1.63, 1.84)

Table 5: Study 1: Full detailed results of Alteration Experiment: Motion Alteration. Motion Comprehension is on a scale of (0, 1); Social Presence is on a scale of (-4, 4). All other measures are on a scale of (-3, 3).

Alteration Experiment: Motion Alteration Post-Hoc Comparisons

Measure	**O**riginal							**Re**duced							**J**itter							**Po**pping						
	Re	J	Po	Sm	Ps	Ra	St	O	J	Po	Sm	Pa	Ra	St	O	Re	Po	Sm	Pa	Ra	St	O	Re	J	Sm	Pa	Ra	St
Motion Comprehension	0.9	0.92	0.99	1.0	***	***	***	0.9	0.18	0.37	1.0	***	***	***	0.92	0.18	1.0	0.58	***	***	***	0.99	0.37	1.0	0.83	***	***	***
Perceived Comprehension	0.77	1.0	0.79	1.0	***	***	***	0.77	0.88	1.0	0.7	0.11	**	***	1.0	0.88	0.91	1.0	***	***	***	0.79	1.0	0.91	0.71	**	***	***
Naturalness	0.97	*	0.34	0.95	***	***	***	0.97	0.58	0.99	1.0	**	0.08	***	*	0.58	0.92	0.32	0.43	0.94	***	0.34	0.99	0.92	0.96	*	0.23	***
Realism	1.0	0.44	0.99	1.0	0.07	0.5	***	1.0	0.65	0.99	1.0	0.2	0.7	***	0.44	0.65	0.93	0.09	0.99	1.0	0.08	0.99	0.99	0.93	0.71	0.43	0.95	***
Appeal	1.0	1.0	0.85	1.0	0.4	0.09	***	1.0	1.0	0.95	1.0	0.64	0.25	***	1.0	1.0	0.99	1.0	0.76	0.29	***	0.85	0.95	0.99	0.75	1.0	0.81	*
Familiarity	1.0	1.0	0.99	1.0	0.3	0.55	0.06	1.0	1.0	0.98	1.0	0.39	0.62	0.12	1.0	1.0	0.98	1.0	0.28	0.52	0.05	0.99	0.98	0.98	0.9	0.84	0.97	0.39
Assuredness	1.0	1.0	0.45	1.0	0.34	0.14	*	1.0	1.0	0.75	1.0	0.64	0.39	0.17	1.0	1.0	0.6	1.0	0.47	0.22	0.07	0.45	0.75	0.6	0.82	1.0	1.0	0.94
Friendliness	1.0	0.84	0.64	1.0	0.06	0.13	***	1.0	0.97	0.88	1.0	0.25	0.41	*	0.84	0.97	1.0	0.93	0.77	0.92	0.13	0.64	0.88	1.0	0.77	0.91	0.98	0.23
Trustworthiness	1.0	0.97	0.75	1.0	0.29	*	***	1.0	0.76	0.43	0.98	0.13	*	***	0.97	0.76	1.0	1.0	0.91	0.42	*	0.75	0.43	1.0	0.89	0.99	0.76	0.1
Conscientiousness	0.72	1.0	0.99	1.0	0.2	***	*	0.72	0.9	0.99	0.96	1.0	0.69	0.98	1.0	0.9	1.0	1.0	0.44	*	0.15	0.99	0.99	1.0	1.0	0.73	*	0.35
Openness to Experience	1.0	0.96	1.0	0.5	1.0	0.93	0.51	1.0	0.73	0.97	0.25	1.0	1.0	0.98	0.96	0.73	1.0	0.99	0.87	0.3	0.05	1.0	0.97	1.0	0.72	1.0	0.75	0.27
Social Presence	0.56	0.99	0.98	1.0	0.23	0.18	***	0.56	0.94	0.95	0.63	1.0	1.0	0.17	0.99	0.94	1.0	1.0	0.76	0.69	***	0.98	0.95	1.0	0.99	0.79	0.72	***

Measure	**Sm**ooth							**Pa**ssive							**Ra**ndom							**St**atic						
	O	Re	J	Po	Pa	Ra	St	O	Re	J	Po	Sm	Ra	St	O	Re	J	Po	Sm	Pa	St	O	Re	J	Po	Sm	Pa	Ra
Motion Comprehension	1.0	1.0	0.58	0.83	***	***	***	***	***	***	***	***	0.59	1.0	***	***	***	***	***	0.59	0.78	***	***	***	***	***	1.0	0.78
Perceived Comprehension	1.0	0.7	1.0	0.71	***	***	***	***	0.11	***	**	***	0.97	0.1	***	**	***	***	***	0.97	0.63	***	***	***	***	***	0.1	0.63
Naturalness	0.95	1.0	0.32	0.96	***	*	***	***	**	0.43	*	***	0.99	0.07	***	0.08	0.94	0.23	*	0.99	***	***	***	***	***	***	0.07	***
Realism	1.0	1.0	0.09	0.71	**	0.12	***	0.07	0.2	0.99	0.43	**	0.98	0.48	0.5	0.7	1.0	0.95	0.12	0.98	0.06	***	***	0.08	***	***	0.48	0.06
Appeal	1.0	1.0	1.0	0.75	0.28	*	***	0.4	0.64	0.76	1.0	0.28	1.0	0.12	0.09	0.25	0.29	0.81	*	1.0	0.5	***	***	***	*	***	0.12	0.5
Familiarity	1.0	1.0	1.0	0.9	0.12	0.29	*	0.3	0.39	0.28	0.84	0.12	1.0	1.0	0.55	0.62	0.52	0.97	0.29	1.0	0.96	0.06	0.12	0.05	0.39	*	1.0	0.96
Assuredness	1.0	1.0	1.0	0.82	0.71	0.41	0.16	0.34	0.64	0.47	1.0	0.71	1.0	0.98	0.14	0.39	0.22	1.0	0.41	1.0	1.0	*	0.17	0.07	0.94	0.16	0.98	1.0
Friendliness	1.0	1.0	0.93	0.77	0.09	0.21	***	0.06	0.25	0.77	0.91	0.09	1.0	0.95	0.13	0.41	0.92	0.98	0.21	1.0	0.83	***	*	0.13	0.23	***	0.95	0.83
Trustworthiness	1.0	0.98	1.0	0.89	0.44	0.08	***	0.29	0.13	0.91	0.99	0.44	0.99	0.49	*	*	0.42	0.76	0.08	0.99	0.94	***	***	*	0.1	***	0.49	0.94
Conscientiousness	1.0	0.96	1.0	1.0	0.58	*	0.23	0.2	1.0	0.44	0.73	0.58	0.84	1.0	***	0.69	*	*	*	0.84	0.99	*	0.98	0.15	0.35	0.23	1.0	0.99
Openness to Experience	0.5	0.25	0.99	0.72	0.32	*	***	1.0	1.0	0.87	1.0	0.32	0.98	0.7	0.93	1.0	0.3	0.75	*	0.98	1.0	0.51	0.98	0.05	0.27	***	0.7	1.0
Social Presence	1.0	0.63	1.0	0.99	0.29	0.23	***	0.23	1.0	0.76	0.79	0.29	1.0	0.1	0.18	1.0	0.69	0.72	0.23	1.0	0.13	***	0.17	***	***	***	0.1	0.13

Table 6: Study 1: All Alteration Experiment: Motion Alteration Post-Hoc Comparisons. * indicates a post-hoc p-value of <0.05, ** of <0.01, and *** values are <0.001. Conditions are abbreviated: Original = O, Reduced = Re, Jitter = J, Popping = Po, Smooth = Sm, Passive = Pa, Random = Ra, Static = St.

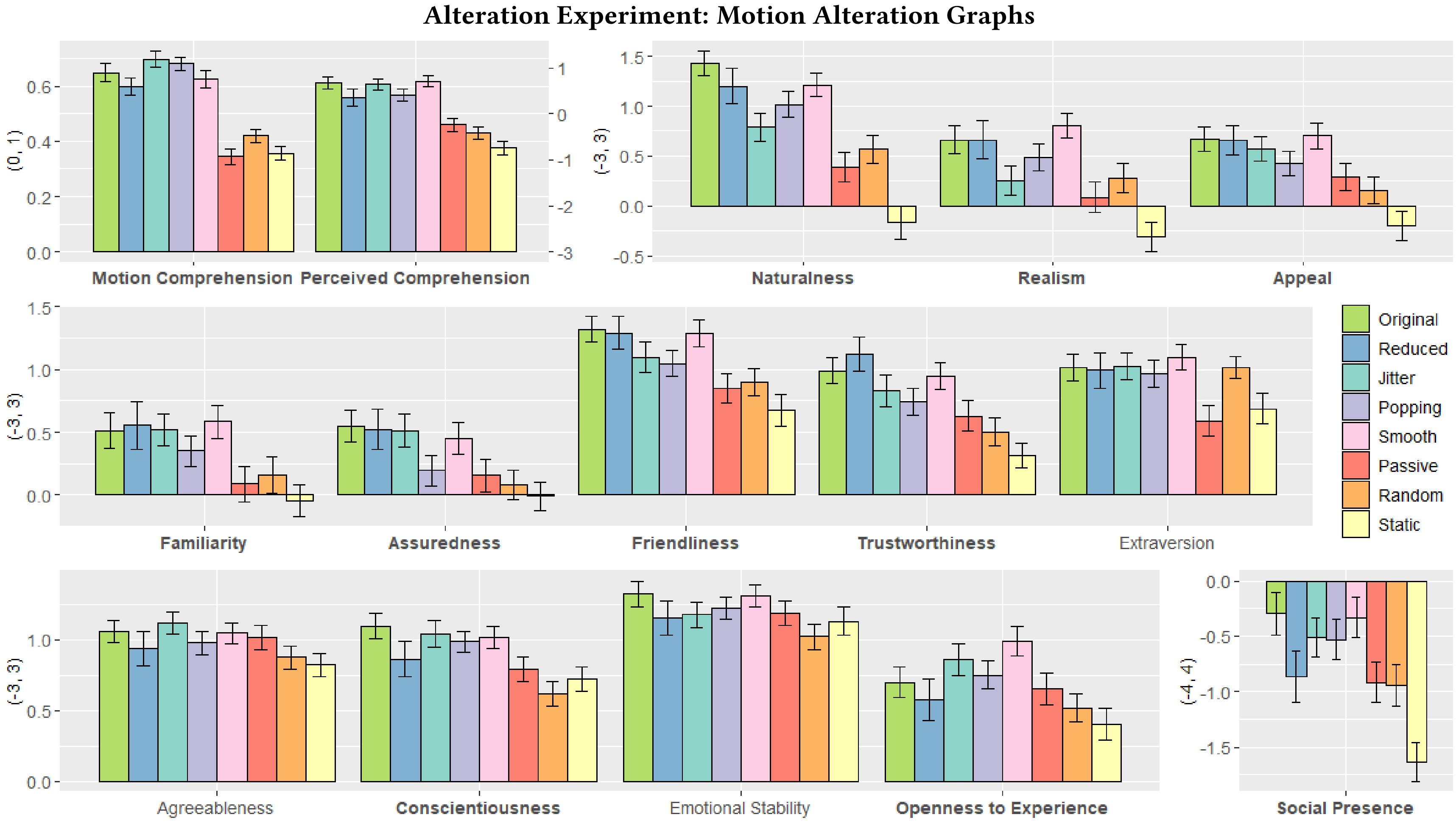


Figure 1: Study 1: All graphs for the Motion Alteration condition of the Alteration Experiment. Bolded measures had significant differences between Motion Alterations. Motion Comprehension is on a scale of (0, 1); Social Presence is on a scale of (-4, 4). All other measures are on a scale of (-3, 3).

Alteration Experiment: Avatar Results

Measure	*Motion Comprehension*	*Perceived Comprehension*	*Naturalness*	*Realism*	*Appeal*
Values	$F_{(1, 473)} = 9.47$, $p < 0.01$	$F_{(1, 839)} = 5.51$, $p < 0.05$	$F_{(1, 839)} = 1.45$, $p = 0.32$	$F_{(1, 839)} = 55.65$, $p < 0.001$	$F_{(1, 839)} = 13.45$, $p < 0.01$
Means & STDs	Realistic = (0.6, 0.26) Mannequin = (0.54, 0.26)	Realistic = (0.29, 1.5) Mannequin = (0.05, 1.46)	Overall = (0.79, 1.53)	Realistic = (0.74, 1.42) Mannequin = (-0.04, 1.6)	Realistic = (0.58, 1.32) Mannequin = (0.22, 1.48)
Post-Hoc	$p < 0.01$	$p < 0.05$		$p < 0.001$	$p < 0.001$

Measure	*Familiarity*	*Assuredness*	*Friendliness*	*Trustworthiness*	*Extraversion*
Values	$F_{(1, 839)} = 30.7$, $p < 0.001$	$F_{(1, 839)} = 10.07$, $p < 0.05$	$F_{(1, 839)} = 3.77$, $p = 0.09$	$F_{(1, 839)} = 1.48$, $p = 0.31$	$F_{(1, 839)} = 0.00$, $p = 0.99$
Means & STDs	Realistic = (0.6, 1.39) Mannequin = (0.06, 1.46)	Realistic = (0.44, 1.31) Mannequin = (0.15, 1.37)	Overall = (1.05, 1.21)	Overall = (0.75, 1.21)	Overall = (0.92, 1.16)
Post-Hoc	$p < 0.001$	$p < 0.01$			

Measure	*Agreeableness*	*Conscientiousness*	*Emotional Stability*	*Openness to Experience*	*Social Presence*
Values	$F_{(1, 839)} = 13.60$, $p < 0.01$	$F_{(1, 839)} = 0.37$, $p = 0.69$	$F_{(1, 839)} = 0.20$, $p = 0.96$	$F_{(1, 839)} = 1.24$, $p = 0.34$	$F_{(1, 839)} = 6.26$, $p < 0.05$
Means & STDs	Realistic = (1.09, 0.85) Mannequin = (0.88, 0.89)	Overall = (0.9, 0.93)	Overall = (1.2, 0.96)	Overall = (0.69, 1.15)	Realistic = (-0.56, 1.98) Mannequin = (-0.92, 1.97)
Post-Hoc	$p < 0.001$				$p < 0.05$

Table 7: Study 1: Full detailed results of Alteration Experiment: Avatar. Motion Comprehension is on a scale of (0, 1); Social Presence is on a scale of (-4, 4). All other measures are on a scale of (-3, 3).

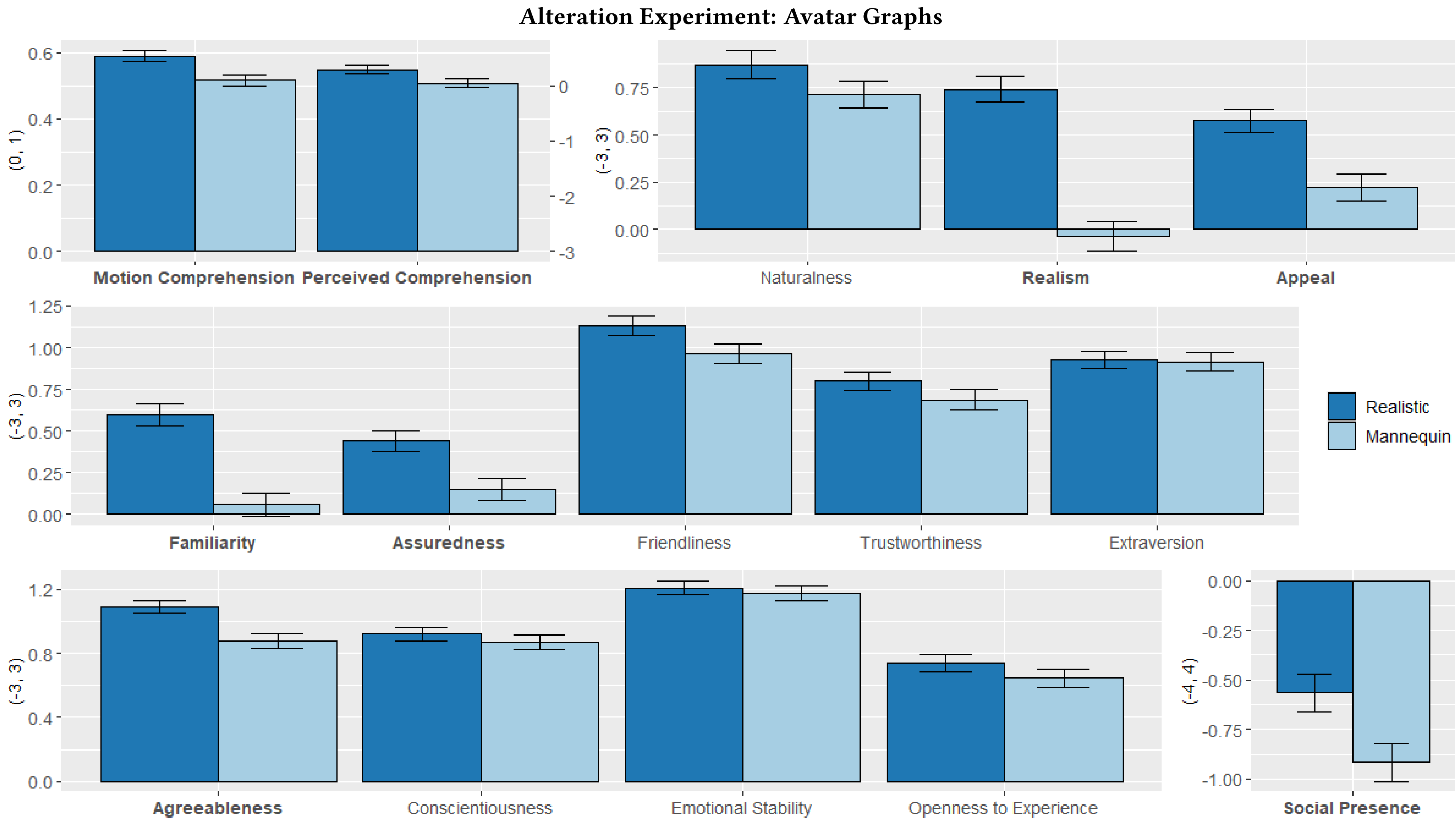


Figure 2: Study 1: All graphs for the Avatar condition of the Alteration Experiment. Bolded measures had significant differences between Avatars. Motion Comprehension is on a scale of (0, 1); Social Presence is on a scale of (-4, 4). All other measures are on a scale of (-3, 3).

Alteration Experiment: Clip Length Results

Measure	*Motion Comprehension*	*Perceived Comprehension*	*Naturalness*	*Realism*	*Appeal*
Values	$F_{(1, 473)} = 30.59, p < 0.001$	$F_{(1, 839)} = 12.17, p < 0.01$	$F_{(1, 839)} = 11.18, p < 0.01$	$F_{(1, 839)} = 6.39, p < 0.05$	$F_{(1, 839)} = 10.52, p < 0.01$
Means & STDs	Short = (0.61, 0.21) Long = (0.5, 0.3)	Short = (0.34, 1.33) Long = (-0.01, 1.62)	Short = (0.96, 1.44) Long = (0.62, 1.59)	Short = (0.49, 1.51) Long = (0.22, 1.61)	Short = (0.55, 1.31) Long = (0.24, 1.49)
Post-Hoc	$p < 0.001$	$p < 0.001$	$p < 0.01$	$p < 0.01$	$p < 0.01$

Measure	*Familiarity*	*Assuredness*	*Friendliness*	*Trustworthiness*	*Extraversion*
Values	$F_{(1, 839)} = 7.09, p < 0.05$	$F_{(1, 839)} = 1.92, p = 0.29$	$F_{(1, 839)} = 4.45, p = 0.08$	$F_{(1, 839)} = 0.22, p = 0.64$	$F_{(1, 839)} = 1.02, p = 0.73$
Means & STDs	Short = (0.46, 1.35) Long = (0.19, 1.54)	Overall = (0.3, 1.35)	Overall = (1.05, 1.21)	Overall = (0.75, 1.21)	Overall = (0.92, 1.16)
Post-Hoc	$p < 0.01$		$p < 0.05$		

Measure	*Agreeableness*	*Conscientiousness*	*Emotional Stability*	*Openness to Experience*	*Social Presence*
Values	$F_{(1, 839)} = 3.62, p = 0.20$	$F_{(1, 839)} = 0.29, p = 0.69$	$F_{(1, 839)} = 0.19, p = 0.96$	$F_{(1, 839)} = 13.55, p <0.01$	$F_{(1, 839)} = 3.29, p = 0.10$
Means & STDs	Overall = (0.99, 0.87)	Overall = (0.9, 0.93)	Overall = (1.2, 0.96)	Short = (0.83, 1.1) Long = (0.54, 1.19)	Overall = (-0.74, 1.98)
Post-Hoc				$p < 0.001$	

Table 8: Study 1: Full detailed results of Alteration Experiment: Clip Length. Motion Comprehension is on a scale of (0, 1); Social Presence is on a scale of (-4, 4). All other measures are on a scale of (-3, 3).

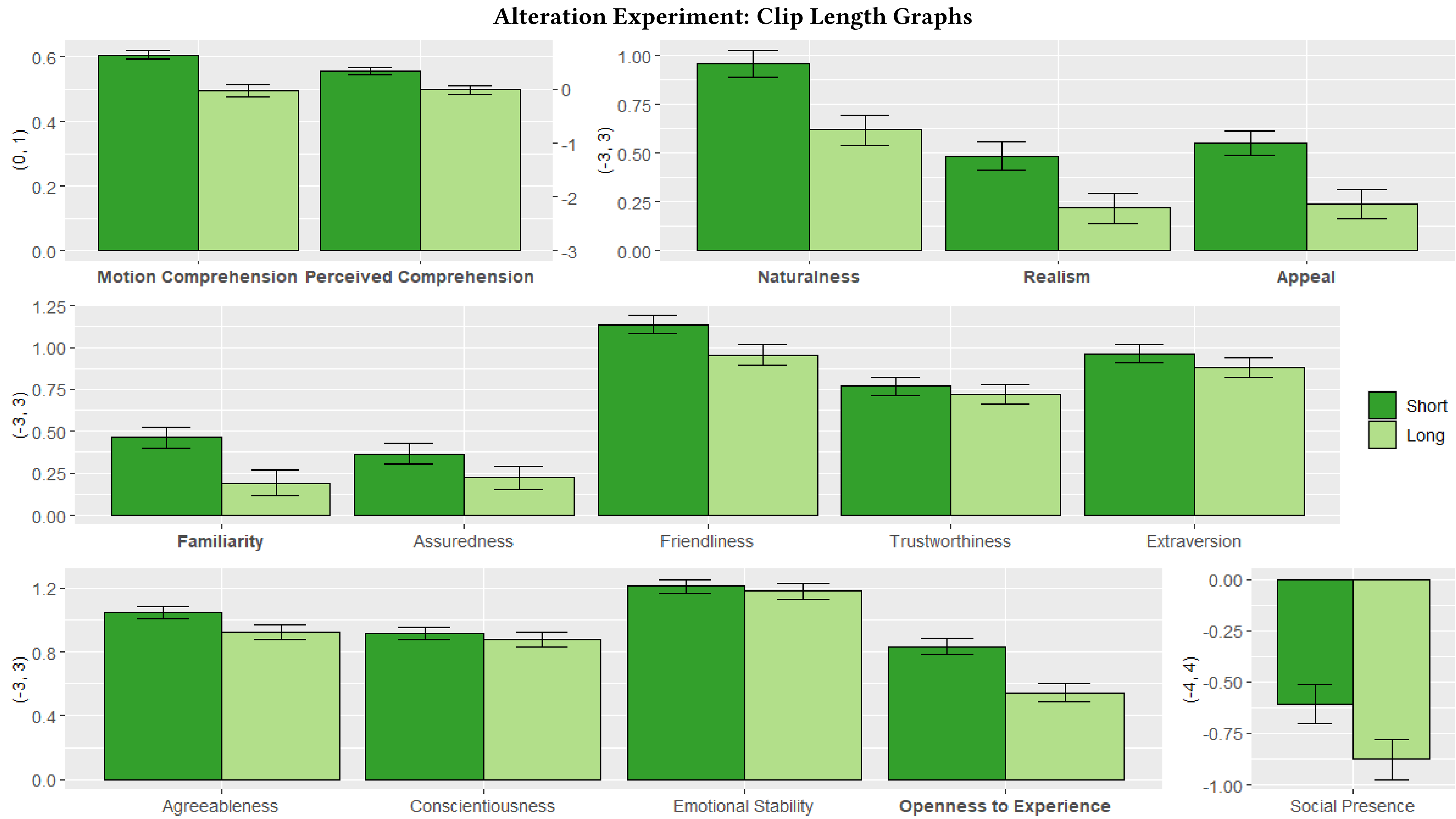


Figure 3: Study 1: All graphs for the Clip Length condition of the Alteration Experiment. Bolded measures had significant differences between Clip Lengths. Motion Comprehension is on a scale of (0, 1); Social Presence is on a scale of (-4, 4). All other measures are on a scale of (-3, 3).

Intensity Experiment: Jitter Intensity Results

Measure	*Motion Comprehension*	*Perceived Comprehension*	*Naturalness*	*Realism*	*Appeal*
Values	$F_{(3, 255)} = 1.32$, p = 0.48	$F_{(3, 458)} = 3.67$, p = 0.08	$F_{(3, 458)} = 29.53$, $p < 0.001$	$F_{(3, 458)} = 4.74$, $p < 0.05$	$F_{(3, 458)} = 5.45$, $p < 0.01$
Means & STDs	Overall = (0.65, 0.23)	Overall = (0.44, 1.37)	Original = (1.43, 1.29) JitterLow = (0.79, 1.48) JitterMed = (0.36, 1.64) JitterHigh = (-0.33, 1.59)	Original = (0.66, 1.5) JitterLow = (0.26, 1.56) JitterMed = (0.31, 1.55) JitterHigh = (-0.05, 1.47)	Original = (0.67, 1.34) JitterLow = (0.58, 1.29) JitterMed = (0.21, 1.47) JitterHigh = (0.05, 1.47)

Measure	*Familiarity*	*Assuredness*	*Friendliness*	*Trustworthiness*	*Extraversion*
Values	$F_{(3, 458)} = 3.03$, p = 0.10	$F_{(3, 458)} = 16.27$, $p < 0.001$	$F_{(3, 458)} = 3.13$, p = 0.07	$F_{(3, 458)} = 3.29$, p = 0.14	$F_{(3, 458)} = 0.58$, p = 0.74
Means & STDs	Overall = (0.34, 1.42)	Original = (0.55, 1.4) JitterLow = (0.51, 1.4) JitterMed = (-0.12, 1.43) JitterHigh = (-0.52, 1.49)	Overall = (1.07, 1.19)	Overall = (0.78, 1.16)	Overall = (0.95, 1.13)

Measure	*Agreeableness*	*Conscientiousness*	*Emotional Stability*	*Openness to Experience*	*Social Presence*
Values	$F_{(3, 458)} = 1.70$, p = 0.39	$F_{(3, 458)} = 6.44$, $p < 0.01$	$F_{(3, 458)} = 16.10$, $p < 0.001$	$F_{(3, 458)} = 1.52$, p = 0.36	$F_{(3, 458)} = 4.23$, $p < 0.05$
Means & STDs	Overall = (1, 0.88)	Original = (1.1, 0.95) JitterLow = (1.04, 1.01) JitterMed = (0.88, 0.92) JitterHigh = (0.62, 0.95)	Original = (1.33, 0.97) JitterLow = (1.18, 1) JitterMed = (0.83, 1.17) JitterHigh = (0.43, 1.23)	Overall = (0.75, 1.14)	Original = (-0.3, 2.02) JitterLow = (-0.51, 1.9) JitterMed = (-0.83, 1.98) JitterHigh = (-1.12, 1.92)

Table 9: Study 1: Full detailed results of Intensity Experiment: Jitter Intensity . Motion Comprehension is on a scale of (0, 1); Social Presence is on a scale of (-4, 4). All other measures are on a scale of (-3, 3).

Intensity Experiment, Jitter: Motion Intensity Post-Hoc Comparisons

Measure	**O**riginal			Jitter**L**ow			Jitter**M**ed			Jitter**H**igh		
	JL	JM	JH	O	JM	JH	O	JL	JH	O	JL	JM
Naturalness	**	***	***	**	0.14	***	***	0.14	**	***	***	**
Realism	0.16	0.27	***	0.16	0.99	0.35	0.27	0.99	0.21	***	0.35	0.21
Appeal	0.96	0.05	**	0.96	0.18	*	0.05	0.18	0.82	**	*	0.82
Assuring	1.0	**	***	1.0	**	***	**	**	0.13	***	***	0.13
Conscientiousness	0.98	0.31	***	0.98	0.56	**	0.31	0.56	0.13	***	**	0.13
Emotional Stability	0.74	**	***	0.74	0.08	***	**	0.08	*	***	***	*
Social Presence	0.83	0.15	**	0.83	0.6	0.07	0.15	0.6	0.65	**	0.07	0.65

Table 10: Study 1: All Intensity Experiment: Jitter Motion Intensity Post-Hoc Comparisons. * indicates a post-hoc p-value of <0.05, ** of <0.01, and *** values are <0.001. Conditions are abbreviated: Original = O, JitterLow = JL, JitterMed = JM, JitterHigh = JH.

Intensity Experiment: Jitter Graphs

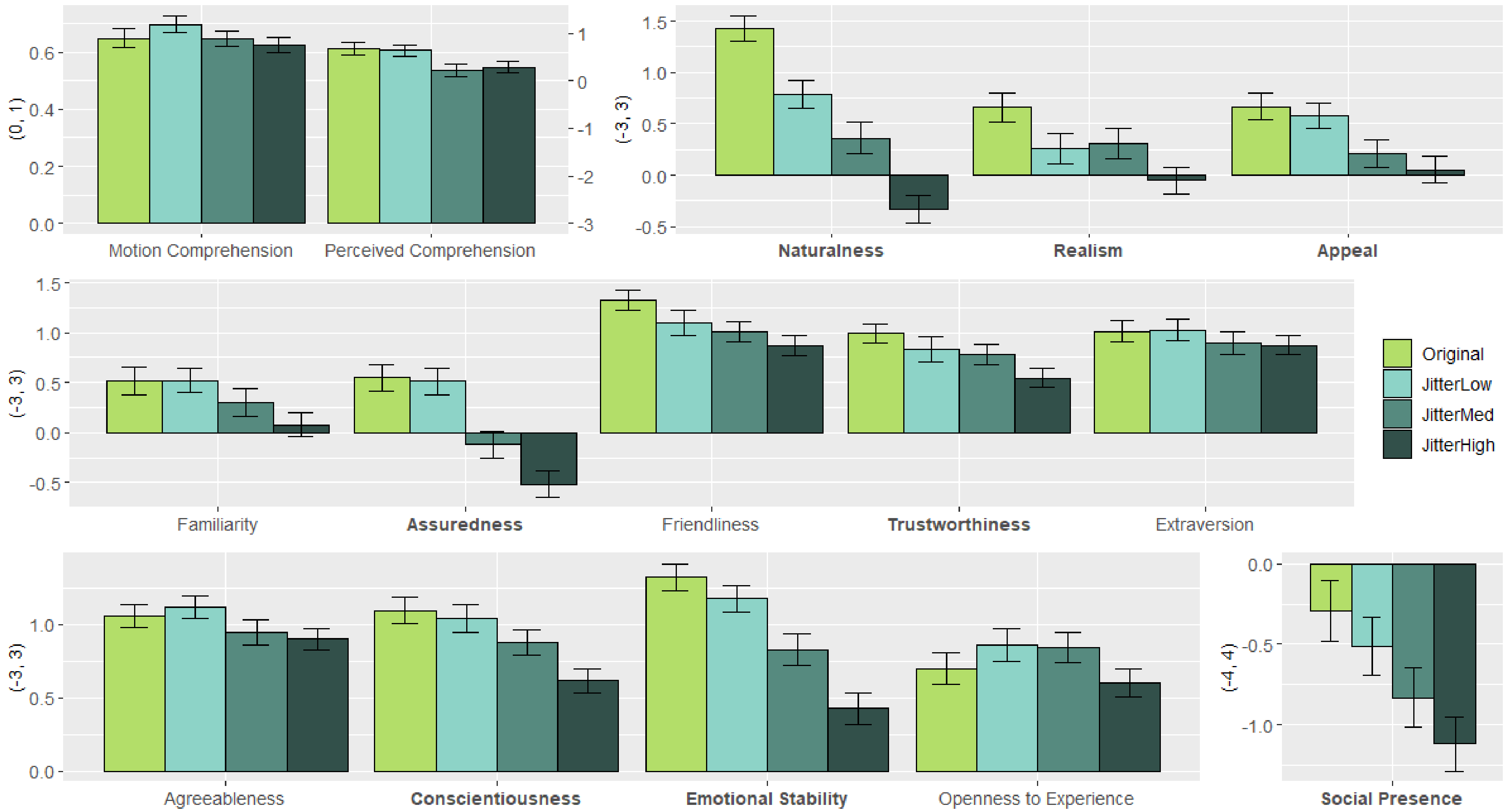


Figure 4: Study 1: All Intensity Experiment: Jitter Graphs. All graphs for the Jitter Intensity Levels of the Intensity Experiment. Bolded measures had significant differences between levels. Motion Comprehension is on a scale of (0, 1); Social Presence is on a scale of (-4, 4). All other measures are on a scale of (-3, 3).

Intensity Experiment: Popping Intensity Results

Measure	*Motion Comprehension*	*Perceived Comprehension*	*Naturalness*	*Realism*	*Appeal*
Values	$F_{(3, 247)} = 0.64$, p = 0.88	$F_{(3, 456)} = 1.49$, p = 0.51	$F_{(3, 456)} = 2.42$, p = 0.23	$F_{(3, 456)} = 0.35$, p = 0.90	$F_{(3, 456)} = 0.86$, p = 0.54
Means & STDs	Overall = (0.66, 0.22)	Overall = (0.47, 1.38)	Overall = (1.13, 1.34)	Overall = (0.59, 1.47)	Overall = (0.58, 1.32)
Measure	*Familiarity*	*Assuredness*	*Friendliness*	*Trustworthiness*	*Extraversion*
Values	$F_{(3, 456)} = 0.63$, p = 0.73	$F_{(3, 456)} = 1.43$, p = 0.74	$F_{(3, 456)} = 1.12$, p = 0.48	$F_{(3, 456)} = 1.00$, p = 0.59	$F_{(3, 456)} = 1.37$, p = 0.49
Means & STDs	Overall = (0.46, 1.41)	Overall = (0.35, 1.33)	Overall = (1.16, 1.17)	Overall = (0.83, 1.15)	Overall = (1.08, 1.05)
Measure	*Agreeableness*	*Conscientiousness*	*Emotional Stability*	*Openness to Experience*	*Social Presence*
Values	$F_{(3, 456)} = 0.19$, p = 0.97	$F_{(3, 456)} = 2.02$, p = 0.39	$F_{(3, 456)} = 0.56$, p = 0.75	$F_{(3, 456)} = 1.12$, p = 0.43	$F_{(3, 456)} = 0.46$, p = 0.71
Means & STDs	Overall = (1, 0.88)	Overall = (0.95, 0.9)	Overall = (1.24, 0.95)	Overall = (0.8, 1.13)	Overall = (-0.46, 1.95)

Table 11: Study 1: Full detailed results of Intensity Experiment: Popping Intensity. Motion Comprehension is on a scale of (0, 1); Social Presence is on a scale of (-4, 4). All other measures are on a scale of (-3, 3).

Intensity Experiment: Popping Graphs

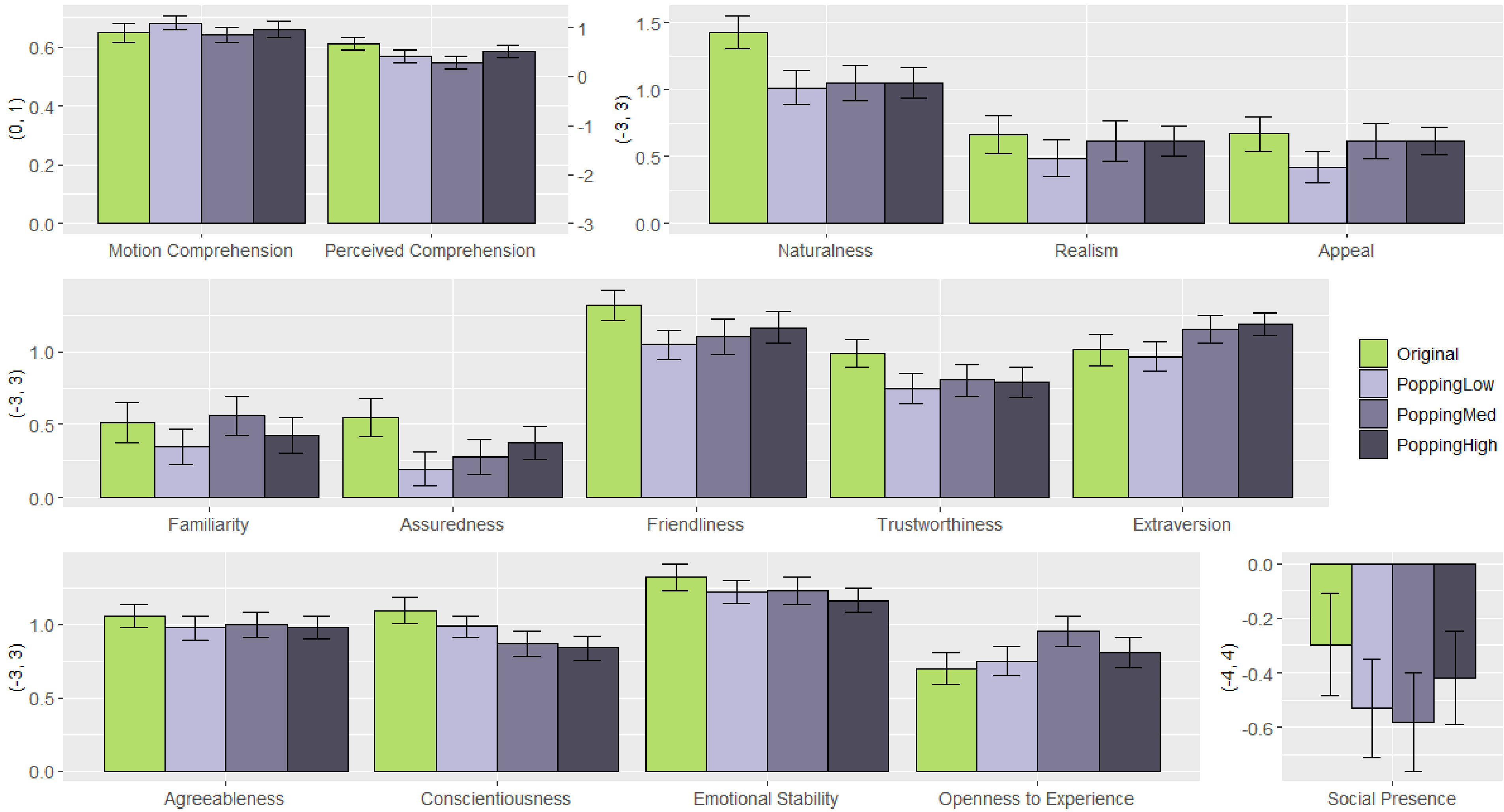


Figure 5: Study 1: All graphs for the Popping Intensity Levels of the Intensity Experiment. Bolded measures had significant differences between levels. Motion Comprehension is on a scale of (0, 1); Social Presence is on a scale of (-4, 4). All other measures are on a scale of (-3, 3).

Intensity Experiment: Smoothing Intensity Results

Measure	*Motion Comprehension*	*Perceived Comprehension*	*Naturalness*	*Realism*	*Appeal*
Values	$F_{(3, 263)}$ = 0.37, p = 0.90	$F_{(3, 466)}$ = 1.54, p = 0.41	$F_{(3, 466)}$ = 3.44, p = 0.11	$F_{(3, 466)}$ = 2.02, p = 0.19	$F_{(3, 466)}$ = 2.52, p = 0.13
Means & STDs	Overall = (0.65, 0.23)	Overall = (0.55, 1.33)	Overall = (1.15, 1.34)	Overall = (0.61, 1.44)	Overall = (0.55, 1.4)

Measure	*Familiarity*	*Assuredness*	*Friendliness*	*Trustworthiness*	*Extraversion*
Values	$F_{(3, 466)}$ = 0.34, p = 0.88	$F_{(3, 466)}$ = 0.69, p = 0.70	$F_{(3, 466)}$ = 1.04, p = 0.69	$F_{(3, 466)}$ = 1.08, p = 0.46	$F_{(3, 466)}$ = 0.25, p = 0.86
Means & STDs	Overall = (0.51, 1.43)	Overall = (0.41, 1.38)	Overall = (1.21, 1.2)	Overall = (0.87, 1.17)	Overall = (1.06, 1.11)

Measure	*Agreeableness*	*Conscientiousness*	*Emotional Stability*	*Openness to Experience*	*Social Presence*
Values	$F_{(3, 466)}$ = 0.05, p = 0.99	$F_{(3, 466)}$ = 0.47, p = 0.82	$F_{(3, 466)}$ = 0.22, p = 0.88	$F_{(3, 466)}$ = 1.48, p = 0.37	$F_{(3, 466)}$ = 1.79, p = 0.35
Means & STDs	Overall = (1.07, 0.84)	Overall = (1.02, 0.91)	Overall = (1.33, 0.92)	Overall = (0.85, 1.13)	Overall = (-0.52, 1.98)

Table 12: Study 1: Full detailed results of Intensity Experiment: Smoothing Intensity. Motion Comprehension is on a scale of (0, 1); Social Presence is on a scale of (-4, 4). All other measures are on a scale of (-3, 3).

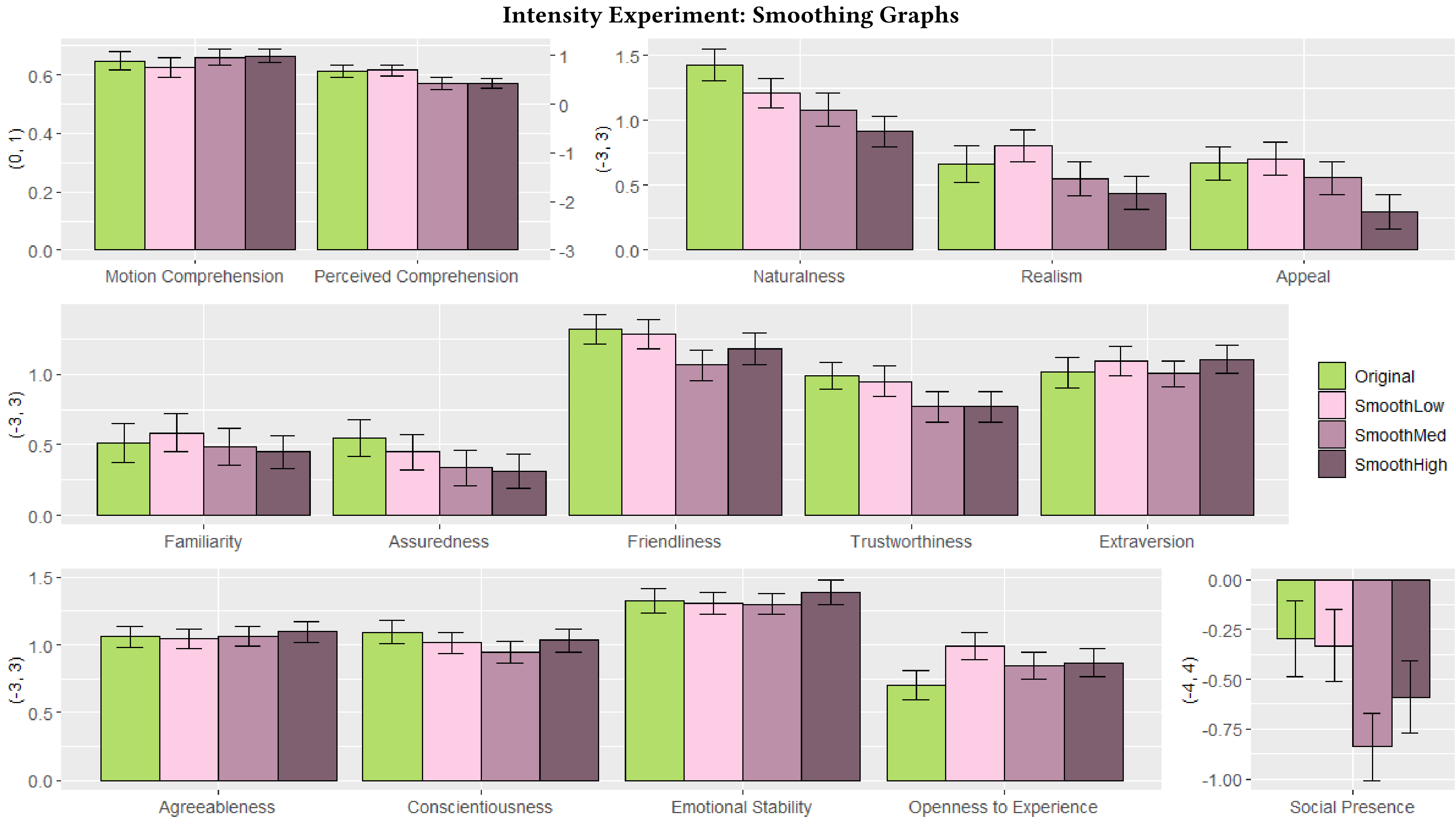


Figure 6: Study 1: All graphs for the Smoothing Intensity Levels of the Intensity Experiment. Bolded measures had significant differences between levels. Motion Comprehension is on a scale of (0, 1); Social Presence is on a scale of (-4, 4). All other measures are on a scale of (-3, 3).

VR Compare Experiment Results

Measure	*Motion Comprehension*	*Perceived Comprehension*	*Naturalness*	*Realism*	*Appeal*
Values	$F_{(1, 31)} = 58.52$, $p < 0.001$	$F_{(1, 31)} = 4.15$, $p = 0.05$	$F_{(1, 31)} = 0.59$, $p = 0.45$	$F_{(1, 31)} = 3.95$, $p = 0.06$	$F_{(1, 31)} = 0.00$, $p = 0.99$
Means & STDs	Original = (0.78, 0.18) Static = (0.38, 0.11)	Overall = (0.58, 1.31)	Overall = (0.82, 1.31)	Overall = (0.91, 1.18)	Overall = (1.06, 1.2)
Post-Hoc	$p < 0.001$				
Measure	*Familiarity*	*Assuredness*	*Friendliness*	*Trustworthiness*	*Extraversion*
Values	$F_{(1, 31)} = 1.67$, $p = 0.21$	$F_{(1, 31)} = 0.05$, $p = 0.82$	$F_{(1, 31)} = 0.01$, $p = 0.93$	$F_{(1, 31)} = 1.59$, $p = 0.22$	$F_{(1, 31)} = 0.17$, $p = 0.69$
Means & STDs	Overall = (1, 1.09)	Overall = (0.3, 1.26)	Overall = (1.39, 1.12)	Overall = (0.67, 1.02)	Overall = (0.38, 0.79)
Post-Hoc					
Measure	*Agreeableness*	*Conscientiousness*	*Emotional Stability*	*Openness to Experience*	*Social Presence*
Values	$F_{(1, 31)} = 0.02$, $p = 0.89$	$F_{(1, 31)} = 5.06$, $p < 0.05$	$F_{(1, 31)} = 0.99$, $p = 0.33$	$F_{(1, 31)} = 0.04$, $p = 0.84$	$F_{(1, 31)} = 0.16$, $p = 0.69$
Means & STDs	Overall = (0.17, 0.82)	Original = (0.38, 0.81) Static = (-0.24, 0.75)	Overall = (0.14, 0.94)	Overall = (0, 0.84)	Overall = (-0.38, 1.99)
Post-Hoc		$p < 0.05$			

Table 13: Study 1: VR Compare Experiment Results details. Motion Comprehension is on a scale of (0, 1); Social Presence is on a scale of (-4, 4). All other measures are on a scale of (-3, 3).

VR Compare vs. Alteration Experiment Results

Measure	*Motion Comprehension*	*Perceived Comprehension*	*Naturalness*	*Realism*	*Appeal*
Values	$F_{(1, 62)} = 0.35$, $p = 0.83$	$F_{(1, 85)} = 2.35$, $p = 0.19$	$F_{(1, 85)} = 0.16$, $p = 0.69$	$F_{(1, 85)} = 1.78$, $p = 0.28$	$F_{(1, 85)} = 5.87$, $p = 0.05$
Means & STDs	Overall = (0.59, 0.26)	Overall = (0.37, 1.48)	Overall = (0.8, 1.62)	Overall = (0.72, 1.42)	Overall = (0.64, 1.43)
Post-Hoc					
Measure	*Familiarity*	*Assuredness*	*Friendliness*	*Trustworthiness*	*Extraversion*
Values	$F_{(1, 85)} = 3.36$, $p = 0.13$	$F_{(1, 85)} = 0.35$, $p = 0.55$	$F_{(1, 85)} = 2.35$, $p = 0.31$	$F_{(1, 85)} = 0.17$, $p = 0.91$	$F_{(1, 85)} = 3.71$, $p = 0.17$
Means & STDs	Overall = (0.71, 1.29)	Overall = (0.44, 1.32)	Overall = (1.13, 1.33)	Overall = (0.61, 1.01)	Overall = (0.66, 1)
Post-Hoc					
Measure	*Agreeableness*	*Conscientiousness*	*Emotional Stability*	*Openness to Experience*	*Social Presence*
Values	$F_{(1, 85)} = 28.37$, $p < 0.001$	$F_{(1, 85)} = 22.51$, $p < 0.001$	$F_{(1, 85)} = 30.11$, $p < 0.001$	$F_{(1, 85)} = 9.77$, $p < 0.01$	$F_{(1, 85)} = 0.14$, $p = 0.71$
Means & STDs	VR Compare = (0.17, 0.82) Alteration = (1.13, 0.8)	VR Compare = (0.06, 0.83) Alteration = (0.99, 0.9)	VR Compare = (0.14, 0.94) Alteration = (1.32, 1)	VR Compare = (0, 0.84) Alteration = (0.68, 1.04)	Overall = (-0.44, 2.2)
Post-Hoc	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.01$	

Table 14: Study 1: Full details of VR Compare vs. Alteration Experiment. Motion Comprehension is on a scale of (0, 1); Social Presence is on a scale of (-4, 4). All other measures are on a scale of (-3, 3).

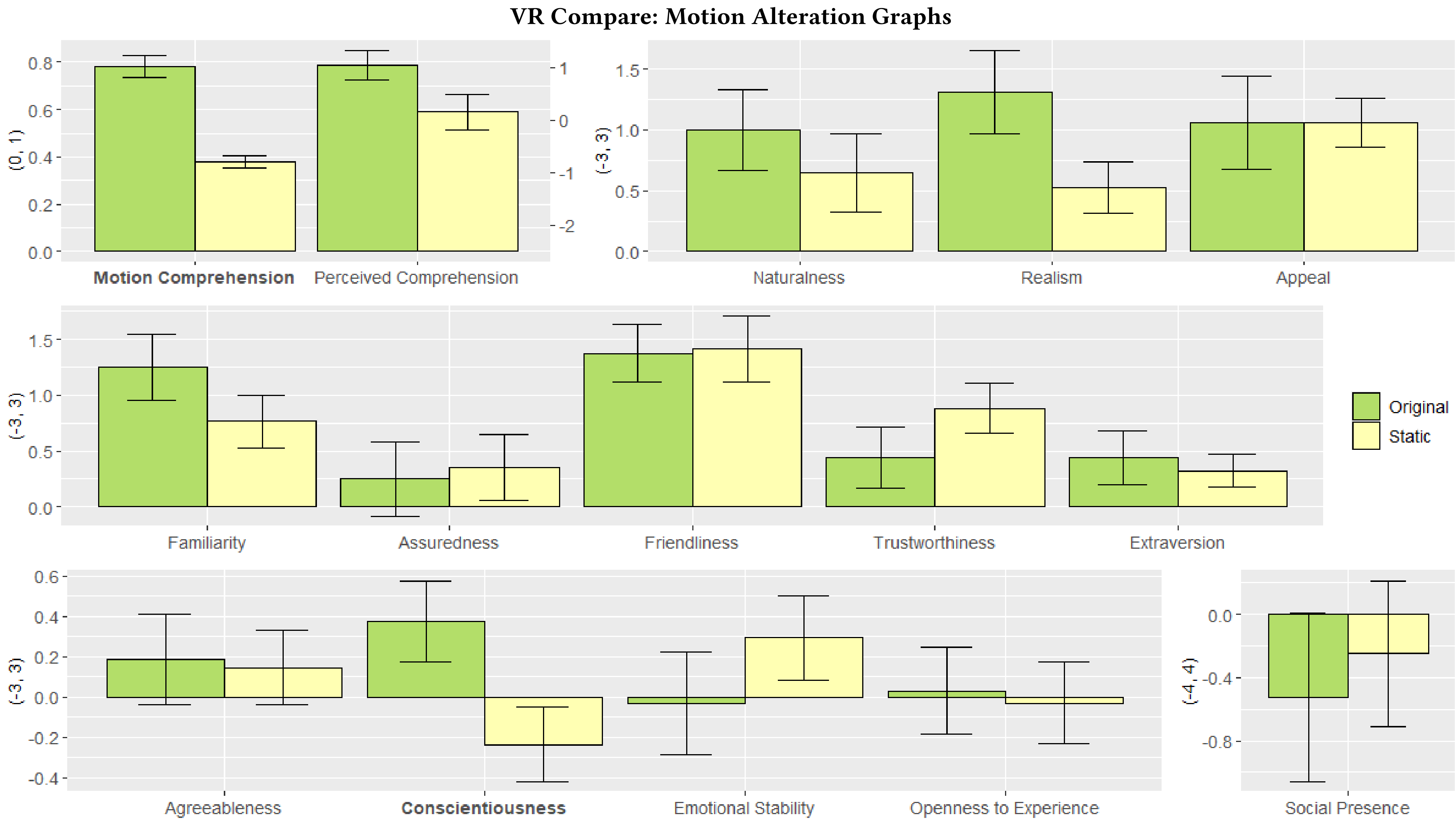


Figure 7: Study 1: All graphs for the Motion Alterations Original and Static for VR Compare. Bolded measures had significant differences between levels. Motion Comprehension is on a scale of (0, 1); Social Presence is on a scale of (-4, 4). All other measures are on a scale of (-3, 3).

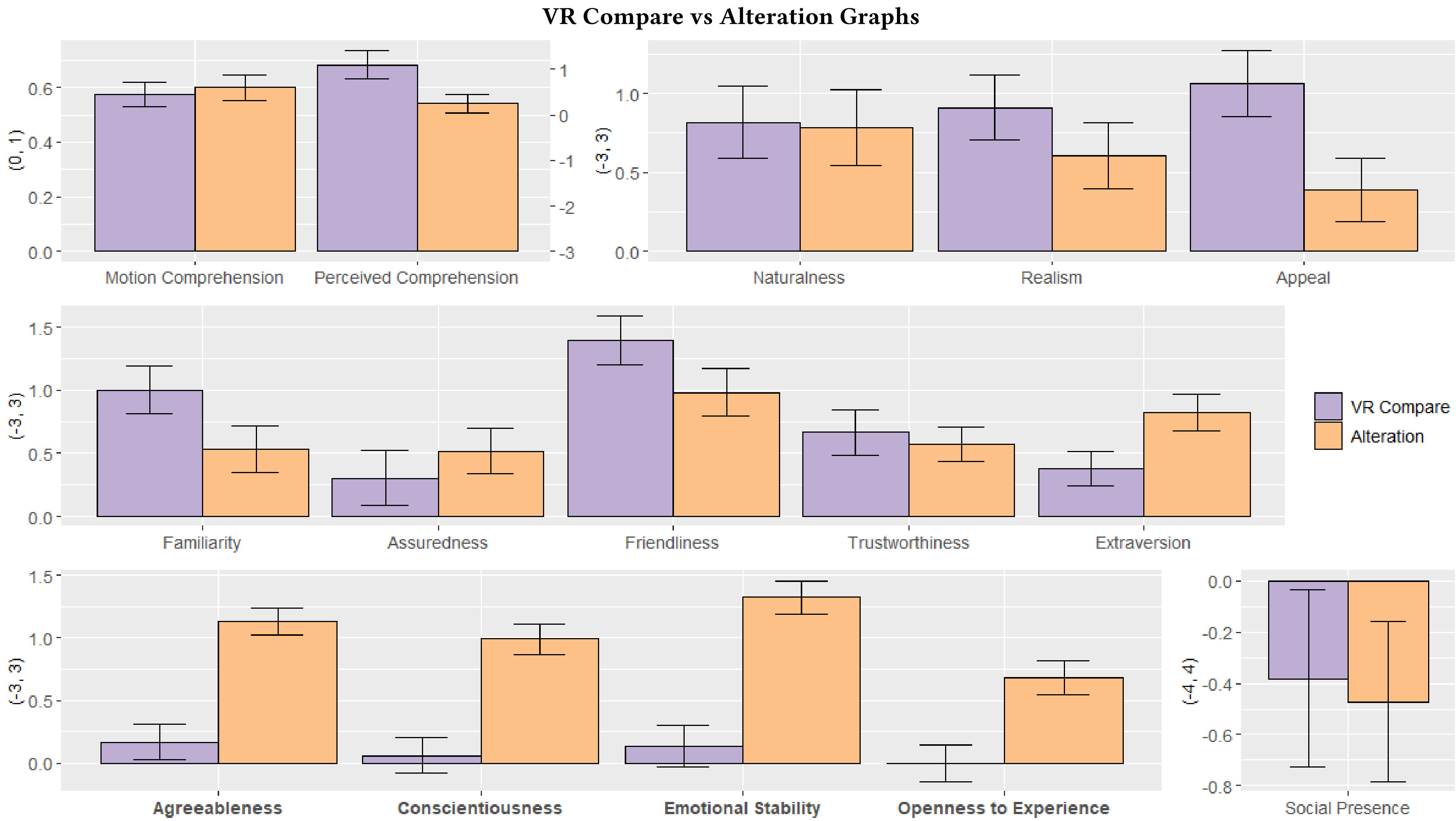


Figure 8: Study 1: All graphs of VR Compare vs. Alteration experiment. Bolded measures had significant differences between levels. Motion Comprehension is on a scale of (0, 1); Social Presence is on a scale of (-4, 4). All other measures are on a scale of (-3, 3).

VR Comfort: Motion Condition Results

Measure	*Comfort*	*Naturalness*
Values	$F_{(13, 448)} = 6.68$, $p < 0.001$	$F_{(13, 448)} = 12.76$, $p < 0.001$
Means & STDs	Original = (0.82, 1.18)	Original = (1.18, 1.16)
	Reduced = (0.94, 1.36)	Reduced = (1.14, 1.29)
	JitterLow = (0.14, 1.34)	JitterLow = (-0.08, 1.49)
	JitterMed = (-0.56, 1.49)	JitterMed = (-1.05, 1.67)
	JitterHigh = (-1.03, 1.54)	JitterHigh = (-1.48, 1.37)
	PoppingLow = (0.64, 1.28)	PoppingLow = (0.82, 1.37)
	PoppingMed = (0.68, 1.15)	PoppingMed = (0.67, 1.36)
	PoppingHigh = (0.52, 1.35)	PoppingHigh = (0.24, 1.52)
	SmoothLow = (0.89, 1.17)	SmoothLow = (1.12, 1.21)
	SmoothMed = (0.8, 1.37)	SmoothMed = (1.23, 1.28)
	SmoothHigh = (0.85, 1.22)	SmoothHigh = (1.08, 1.14)
	Random = (0.71, 1.4)	Random = (0.92, 1.21)
	Passive = (0.79, 1.08)	Passive = (0.79, 1.29)
	Static = (0.56, 1.17)	Static = (0.45, 1.51)

Table 15: Study 1: Detailed results of VR Comfort: Motion Condition. Comfort and Naturalness are both on a scale of (-3, 3).

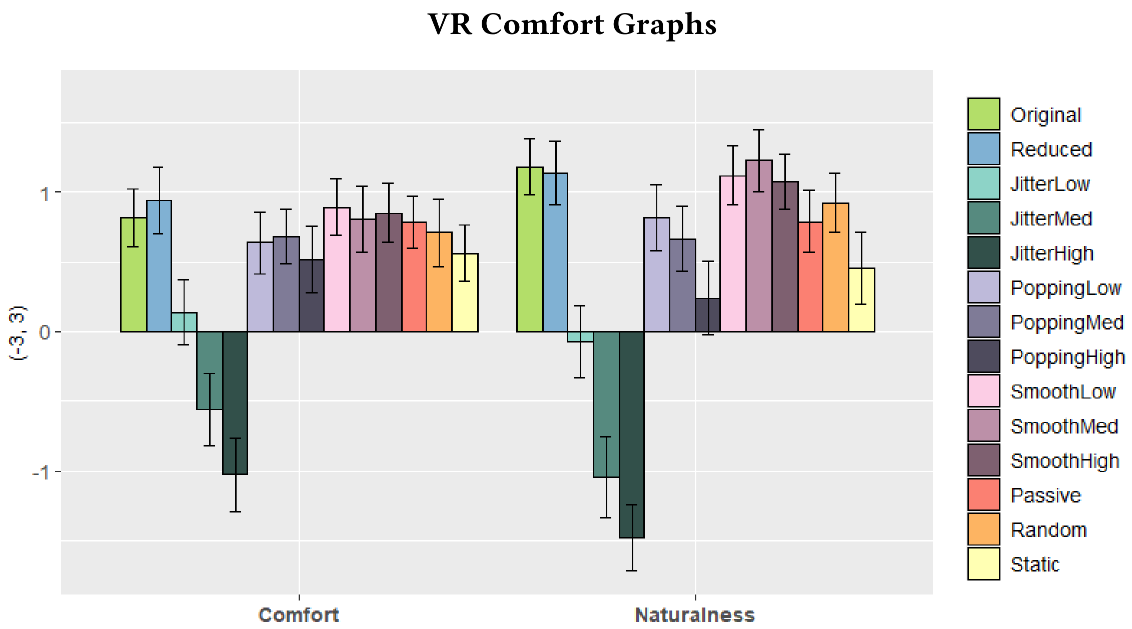


Figure 9: Study 1: VR Comfort Graphs. Measures of Comfort and Naturalness for all Motion Condition conditions. Both measures had significant differences between levels. Comfort and Naturalness are on a scale of (-3, 3).

VR Comfort: Motion Condition Post-Hoc Comparisons

Measure	Original													Reduced												
	Re	JL	JM	JH	PoL	PoM	PoH	SmL	SmM	SmH	Pa	Ra	St	O	JL	JM	JH	PoL	PoM	PoH	SmL	SmM	SmH	Pa	Ra	St
Comfort	1.0	0.68	**	***	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.4	***	***	1.0	1.0	0.99	1.0	1.0	1.0	1.0	1.0	1.0
Naturalness	1.0	*	***	***	1.0	0.96	0.22	1.0	1.0	1.0	1.0	1.0	0.64	1.0	*	***	***	1.0	0.98	0.29	1.0	1.0	1.0	1.0	1.0	0.74

Measure	JitterLow													JitterMed												
	O	Re	JM	JH	PoL	PoM	PoH	SmL	SmM	SmH	Pa	Ra	St	O	Re	JL	JH	PoL	PoM	PoH	SmL	SmM	SmH	Pa	Ra	St
Comfort	0.68	0.4	0.65	*	0.95	0.91	1.0	0.51	0.71	0.61	0.74	0.88	0.99	**	***	0.65	0.97	*	**	0.05	***	**	**	**	**	*
Naturalness	*	*	0.18	**	0.29	0.61	1.0	*	**	*	0.35	0.14	0.95	***	***	0.18	0.99	***	***	**	***	***	***	***	***	***

Measure	JitterHigh													PoppingLow												
	O	Re	JL	JM	PoL	PoM	PoH	SmL	SmM	SmH	Pa	Ra	St	O	Re	JL	JM	JH	PoM	PoH	SmL	SmM	SmH	Pa	Ra	St
Comfort	***	***	*	0.97	***	***	***	***	***	***	***	***	***	1.0	1.0	0.95	*	***	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Naturalness	***	***	**	0.99	***	***	***	***	***	***	***	***	***	1.0	1.0	0.29	***	***	1.0	0.91	1.0	0.99	1.0	1.0	1.0	1.0

Measure	PoppingMed													PoppingHigh												
	O	Re	JL	JM	JH	PoL	PoH	SmL	SmM	SmH	Pa	Ra	St	O	Re	JL	JM	JH	PoL	PoM	SmL	SmM	SmH	Pa	Ra	St
Comfort	1.0	1.0	0.91	**	***	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.99	1.0	0.05	***	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Naturalness	0.96	0.98	0.61	***	***	1.0	0.99	0.99	0.92	0.99	1.0	1.0	1.0	0.22	0.29	1.0	**	***	0.91	0.99	0.32	0.16	0.41	0.94	0.74	1.0

Measure	SmoothLow													SmoothMed												
	O	Re	JL	JM	JH	PoL	PoM	PoH	SmM	SmH	Pa	Ra	St	O	Re	JL	JM	JH	PoL	PoM	PoH	SmL	SmH	Pa	Ra	St
Comfort	1.0	1.0	0.51	***	***	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.71	**	***	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Naturalness	1.0	1.0	*	***	***	1.0	0.99	0.32	1.0	1.0	1.0	1.0	0.77	1.0	1.0	**	***	***	0.99	0.92	0.16	1.0	1.0	0.99	1.0	0.54

Measure	SmoothHigh													Passive												
	O	Re	JL	JM	JH	PoL	PoM	PoH	SmL	SmM	Pa	Ra	St	O	Re	JL	JM	JH	PoL	PoM	PoH	SmL	SmM	SmH	Ra	St
Comfort	1.0	1.0	0.61	**	***	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.74	**	***	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Naturalness	1.0	1.0	*	***	***	1.0	0.99	0.41	1.0	1.0	1.0	1.0	0.85	1.0	1.0	0.35	***	***	1.0	1.0	0.94	1.0	0.99	1.0	1.0	1.0

Measure	Random													Static												
	O	Re	JL	JM	JH	PoL	PoM	PoH	SmL	SmM	SmH	Pa	St	O	Re	JL	JM	JH	PoL	PoM	PoH	SmL	SmM	SmH	Pa	Ra
Comfort	1.0	1.0	0.88	**	***	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.99	*	***	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Naturalness	1.0	1.0	0.14	***	***	1.0	1.0	0.74	1.0	1.0	1.0	1.0	0.98	0.64	0.74	0.95	***	***	1.0	1.0	1.0	0.77	0.54	0.85	1.0	0.98

Table 16: Study 1: All VR Comfort: Motion Condition Post-Hoc Comparisons. * indicates a post-hoc p-value of <0.05, ** of <0.01, and *** values are <0.001. Conditions are abbreviated: Original = O, Reduced = Re, JitterLow = JL, JitterMed = JM, JitterHigh = JH, PoppingLow = PoL, PoppingMed = PoM, PoppingHigh = PoH, SmoothLow = SmL, SmoothMed = SmM, SmoothHigh = SmH, Passive = Pa, Random = Ra, Static = St.

VR Rank: Results

$\chi^2(7) = 95.77$, $p < 0.005$; Critical Difference = 62.16

Original							**Reduced**							**Jitter**							**Popping**						
Re	J	Po	Sm	Ps	Ra	St	O	J	Po	Sm	Pa	Ra	St	O	Re	Po	Sm	Pa	Ra	St	O	Re	J	Sm	Pa	Ra	St
18	**141**	6	5	**71**	48	**93**	18	**123**	12	23	53	30	**75**	**141**	**123**	**135**	**146**	**70**	**93**	48	6	12	**135**	11	**65**	42	**87**

Smooth							**Passive**							**Random**							**Static**						
O	Re	J	Po	Pa	Ra	St	O	Re	J	Po	Sm	Ra	St	O	Re	J	Po	Sm	Pa	St	O	Re	J	Po	Sm	Pa	Ra
5	23	**146**	11	**76**	53	**98**	**71**	53	**70**	**65**	**76**	23	22	48	30	**93**	42	53	23	45	**93**	**75**	48	**87**	**98**	22	45

Table 17: Study 1: Detailed results of VR Rank. Bolded measures indicate significant differences between conditions. Conditions are abbreviated: Original = O, Reduced = Re, Jitter = J, Popping = Po, Smooth = Sm, Passive = Pa, Random = Ra, Static = St.

VR Rank Graphs

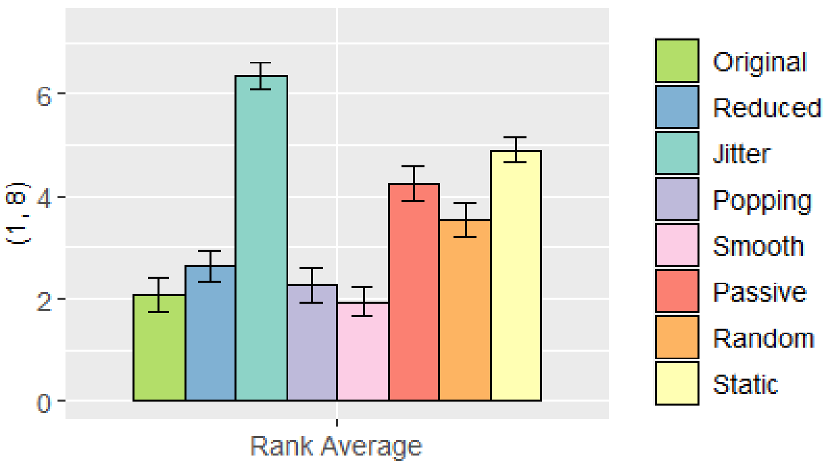


Figure 10: Study 1: VR Rank Graphs. Ranking measures from Most Comfortable (1) to Least Comfortable (8).

Measure	Questions	Results	Mean, Std.	Origin
Ownership	O1. I felt as if the virtual hands were part of my body. ⋆†		(5.19, 1.08)	Lin *et al.* [130] Canales *et al.* [50]
	O2. It sometimes seemed my own hands were located on the screen.⋆	TH > DH	TH (5.47, 1.25) DH (4.53, 1.44)	
	O3. It sometimes seemed like my own hands came into contact with the virtual object. ⋆†	Glv > Ctrl	Glv (4.91, 1.64) Ctrl (3.7, 1.8)	
	O4. I thought that the virtual hands could be harmed by the virtual danger. †	Glv > Ctrl	Glv (4.44, 2.16) Ctrl (3.07, 2)	
Realism	R1. I thought the virtual hands looked realistic. ⋆		(4.31, 1.73)	Lin *et al.* [130]
	R2. I thought the movement of the virtual hands looked realistic.	Glv > Ctrl	Glv (5.74, 1.05) Ctrl (4.8, 1.67)	
Efficiency	F1. I felt like I could very efficiently use my virtual hands to interact with the environment. ⋆		(5.53, 1.3)	Lin *et al.* [130]
Enjoyment Not true at all - Very true	E1. While I was playing this game, I was thinking about how much I enjoyed it.		(5.92, 1.15)	IMI Enjoyment [6]
	E2. This game did not hold my attention at all. (R)	Glv > Ctrl TH > DH	TH (6.83, 0.38) DH (6.53, 0.83) Glv (6.85, 0.36) Ctrl (6.47, 0.86)	
	E3. I would describe this game as very interesting.	Glv > Ctrl	Glv (6.21, 0.81) Ctrl (5.6, 1.19)	
	E4. I enjoyed playing this game very much.	Glv > Ctrl	Glv (6.47, 0.66) Ctrl (6.03, 1)	
	E5. This game was fun to play.		(6.28, 0.93)	
Presence Do Not Agree - Strongly Agree	P1. When playing the game, I feel transported to another time and place.		(5.56, 1.1)	PENS Presence [162]
	P2. Exploring the game world feels like taking an actual trip to a new place.		(5.25, 1.33)	
	P3. When moving through the game world I feel as if I am actually there.		(5.31, 1.42)	
	P4. I am not impacted emotionally by events in the game. (R)		(3.61, 1.84)	
	P5. The game was emotionally engaging.	Glv > Ctrl	Glv (4.65, 1.69) Ctrl (3.57, 1.65)	
	P6. I experience feelings as deeply in the game as I have in real life.		(3.27, 1.97)	
	P7. When playing the game I feel as if I was part of the story.	Glv > Ctrl	Glv (5.47, 1.31) Ctrl (4.6, 1.63)	
	P8. When I accomplished something in the game I experienced genuine pride.	Glv > Ctrl	Glv (6.12, 0.95) Ctrl (5.57, 1.19)	
	P9. I had reactions to events and characters in the game as if they were real.	Glv > Ctrl	Glv (5.18, 1.7) Ctrl (4.2, 1.73)	

Table 18: Study 2: Questionnaire and main results.

All measures were on a 7-pt Likert Scale, with values from "Strongly agree" to "Strongly disagree" with the exception of the Enjoyment and Presence questions. Values from questions marked with (R) were reversed before analysis.
⋆ : Question also measured in Lin *et al.* [130]; † : Question also measured in Canales *et al.* [50]

Measure	**Questions**	**Scale**	**Mean, Std.**
Social Presence - 7pt Likert - Nowak & Biocca [150]	SP1. To what extent did you feel able to assess your partner's reactions to what you said?	Not able to assess reactions - Able to assess reactions	(4.95, 1.06)
	SP2. To what extent was this like a face-to-face meeting?	Not like face-to-face at all - A lot like face-to-face	(4.24, 1.39)
	SP3. To what extent was this like you were in the same room with the virtual character?	Not like being in the same room at all - A lot like being in the same room	(5.07, 1.09)
	SP4. To what extent did the virtual character seem "real?"	Not real at all - Very real	(4.29, 1.29)
	SP5. How likely is it that you would choose to use this system of interaction for a meeting in which you wanted to persuade others of something?	Not likely at all - Very likely	(3.9, 1.71)
	SP6. To what extent did you feel you could get to know someone that you met only through this system?	Not at all - Very well	(4.26, 1.38)
Perceived Comprehension - 7pt Likert - Biocca & Harms [35]	PC1. I was able to communicate my intentions clearly to my partner.	Strongly disagree - Strongly agree	(6.5, 0.55)
	PC2. My thoughts were clear to my partner.		(6.38, 0.66)
	PC3. I was able to understand what my partner meant.		(6.67, 0.53)
	PC4. My partner was able to communicate their intents clearly to me.		(6.64, 0.53)
	PC5. My partner's thoughts were clear to me.		(6.52, 0.59)
	PC6. My partner was able to understand what I meant.		(6.45, 0.77)
Team Cohesion - 5pt Likert - Michalisin *et al.* [145]	TC1. I enjoyed working with my teammates.	Strongly disagree - Strongly agree	(4.88, 0.33)
	TC2. I wish I were on a different team. (R)		(4.55, 0.83)
	TC3. The team worked well together.		(4.88, 0.33)
	TC4. Everyone contributed to the discussion.		(4.95, 0.22)
	TC5. The team wasted a lot of time. (R)		(4.69, 0.52)
	TC6. I trust that my teammates will do their fair share of the work.		(4.76, 0.69)
NASA's Task Load Index - 10pt Likert - Harris *et al.* [88]	TLX1. How mentally demanding was the overall experience?	Low - High	(3.83, 2.04)
	TLX2. How physically demanding was the overall experience?		(1.55, 0.97)
	TLX3. How hurried or rushed was the pace of the overall experience?		(2.26, 1.7)
	TLX4. How successful were you in accomplishing what you were asked to do in the overall experience?		(8.88, 1.63)
	TLX5. How hard did you have to work to accomplish your level of performance in the overall experience?		(3.88, 2.34)
	TLX6. How insecure, discouraged, irritated, stressed and annoyed were you with the overall experience?		(1.48, 1.33)

Table 19: Study 3: Questionnaire and main results.

Values from questions marked with (R) were reversed before analysis.

Bibliography

[1] Getting started with Hand Tracking on Meta Quest 2 and Meta Quest — Meta Store. https://store.facebook.com/help/quest/articles/headsets-and-accessories/controllers-and-hand-tracking/hand-tracking-quest-2/. Accessed on 2022-07-08.

[2] GlassesOn — Pupils & Lenses - Apps on Google Play. https://play.google.com/store/apps/details?id=com.sixoversix.copyglass&hl=en_US&gl=US. Accessed on 2022-07-11.

[3] Oculus Integration — Integration — Unity Asset Store. https://assetstore.unity.com/packages/tools/integration/oculus-integration-82022. Accessed on 2022-06-29.

[4] Valve Index Controllers. https://www.valvesoftware.com/en/index/controllers. Accessed on 2021-06-23.

[5] VrApi Input API — Oculus Developers. https://developer.oculus.com/documentation/native/android/mobile-vrapi-input-api/. Accessed on 2022-07-12.

[6] Intrinsic Motivation Inventory (IMI) – selfdeterminationtheory.org, 1987. Accessed on 2021-06-25.

[7] Introducing Hand Tracking on Oculus Quest—Bringing Your Real Hands into VR. https://www.oculus.com/blog/introducing-hand-tracking-on-oculus-quest-bringing-your-real-hands-into-vr/, September 2019. Accessed on 2021-06-14.

[8] How Hand Tracking Works — Ultraleap. https://www.ultraleap.com/company/news/blog/how-hand-tracking-works/, September 2020. Accessed on 2021-06-14.

[9] Oculus Picks: 5 Hand Tracking Experiences on Quest. https://www.oculus.com/blog/oculus-picks-5-hand-tracking-experiences-on-quest/, May 2020. Accessed on 2021-06-14.

[10] Ali Abbas, Minji Choi, JoonOh Seo, Seung Hyun Cha, and Heng Li. Effectiveness of Immersive Virtual Reality-based Communication for Construction Projects. *KSCE Journal of Civil Engineering*, 23(12):4972–4983, December 2019.

[11] Ahsan Abdullah, Jan Kolkmeier, Vivian Lo, and Michael Neff. Videoconference and Embodied VR: Communication Patterns Across Task and Medium. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2):453:1–453:29, October 2021.

[12] Alex Adkins, Lorraine Lin, Aline Normoyle, Ryan Canales, Yuting Ye, and Sophie Jörg. Evaluating Grasping Visualizations and Control Modes in a VR Game. *ACM Transactions on Applied Perception*, 18(4):19:1–19:14, October 2021.

[13] Advanced Realtime Tracking. Fingertracking, 2022.

[14] Monthir Ali and Rogelio E. Cardona-Rivera. Comparing Gamepad and Naturally-Mapped Controller Effects on Perceived Virtual Reality Experiences. In *ACM Symposium on Applied Perception 2020*, SAP '20, pages 1–10, 2020.

[15] Martha W. Alibali, Sotaro Kita, and Amanda J. Young. Gesture and the process of speech production: We think, therefore we gesture. *Language and Cognitive Processes*, 15(6):593–613, December 2000.

[16] Martha Wagner Alibali, Sotaro Kita, Lisa J Bigelow, Chai M Wolfman, and Susan M Klein. Gesture plays a role in thinking for speaking. In *Oralité et Gestualité: Interactions et Comportements Multimodaux Dans La Communication [Orality and Gestuality: Multimodal Interaction and Behavior in Communication]. Proceedings of the Meeting of ORAGE*, pages 407–410, 2001.

[17] Ferran Argelaguet, Ludovic Hoyet, Michaël Trico, and Anatole Lécuyer. The role of interaction in virtual embodiment: Effects of the virtual hand representation. In *2016 IEEE Virtual Reality (VR)*, pages 3–10. IEEE, 2016.

[18] Michael Argyle, Florisse Alkema, and Robin Gilmour. The communication of friendly and hostile attitudes by verbal and non-verbal signals. *European Journal of Social Psychology*, 1:385–402, 1971.

[19] Jolanta Aritz, Robyn Walker, and Peter W. Cardon. Media Use in Virtual Teams of Varying Levels of Coordination. *Business and Professional Communication Quarterly*, 81(2):222–243, June 2018.

[20] Sahar Aseeri and Victoria Interrante. The Influence of Avatar Representation on Interpersonal Communication in Virtual Social Environments. *IEEE Transactions on Visualization and Computer Graphics*, 27(5):2608–2617, May 2021.

[21] Harry Baker. Top 10 Best Oculus Quest Hand Tracking Games & Apps, May 2021. Accessed on 2021-06-14.

[22] Christopher Ball, Kuo-Ting Huang, and Jess Francis. Virtual reality adoption during the COVID-19 pandemic: A uses and gratifications perspective. *Telematics and Informatics*, 65:101728, December 2021.

[23] Christoph Bartneck, Andreas Duenser, Elena Moltchanova, and Karolina Zawieska. Comparing the Similarity of Responses Received from Studies in Amazon's Mechanical Turk to Studies Conducted Online and with Direct Recruitment. *PLOS ONE*, 10(4):e0121595, April 2015.

[24] Janet Bavelas, Christine Kenwood, Trudy Johnson, and Bruce Phillips. An experimental study of when and how speakers use gestures to communicate. *Gesture*, 2(1):1–17, January 2002.

[25] Janet B Bavelas and Jennifer Gerwing. Conversational hand gestures and facial displays in face-to-face dialogue. *Social communication*, pages 283–308, 2007.

[26] Janet Beavin Bavelas and Nicole Chovil. Visible Acts of Meaning: An Integrated Message Model of Language in Face-to-Face Dialogue. *Journal of Language and Social Psychology*, 19(2):163–194, June 2000.

[27] Janet Beavin Bavelas and Nicole Chovil. Nonverbal and Verbal Communication: Hand Gestures and Facial Displays as Part of Language Use in Face-to-face Dialogue. In *The Sage Handbook of Nonverbal Communication*, pages 97–115. Sage Publications, Inc, Thousand Oaks, CA, US, 2006.

[28] Janet Beavin Bavelas, Sarah Hutchinson, Christine Kenwood, and Deborah Hunt Matheson. Using Face-to-face Dialogue as a Standard for Other Communication Systems. *Canadian Journal of Communication*, 22(1), January 1997.

[29] Alejandro Beacco, Ramon Oliva, Carlos Cabreira, Jaime Gallego, and Mel Slater. Disturbance and Plausibility in a Virtual Rock Concert: A Pilot Study. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pages 538–545, March 2021.

[30] Sherry L Beaumont and Shannon L Wagner. Adolescent-parent verbal conflict: The roles of conversational styles and disgust emotions. *Journal of language and social psychology*, 23(3):338–368, 2004.

[31] Elizabeth Behm-Morawitz. Mirrored selves: The influence of self-presence in a virtual world on health, appearance, and well-being. *Computers in Human Behavior*, 29(1):119–128, January 2013.

[32] Yoav Benjamini and Yosef Hochberg. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1):289–300, 1995.

[33] Charles R. Berger. Planning theory of communication. In *Engaging Theories in Interpersonal Communication: Multiple Perspectives*, pages 89–101. Sage Publications, Inc, Thousand Oaks, CA, US, 2008.

[34] Frank Biocca. The Cyborg's Dilemma: Progressive Embodiment in Virtual Environments. *Journal of Computer-Mediated Communication*, 3(2), September 1997.

[35] Frank Biocca and Chad Harms. Guide to the Networked Minds Social Presence Inventory v. 1.2. http://cogprints.org/6743/, 2003.

[36] Frank Biocca, Chad Harms, and Judee Burgoon. Towards A More Robust Theory and Measure of Social Presence: Review and Suggested Criteria. *Presence*, 12:456–480, October 2003.

[37] Frank Biocca, Chad Harms, and Jenn Gregg. The Networked Minds Measure of Social Presence: Pilot Test of the Factor Structure and Concurrent Validity. page 10, 2001.

[38] Frank Biocca and Mark R. Levy. Virtual reality as a communication system. In *Communication in the Age of Virtual Reality*, LEA's Communication Series, pages 15–31. Lawrence Erlbaum Associates, Inc, Hillsdale, NJ, US, 1995.

[39] Frank Biocca and Mark R. Levy. *Communication in the Age of Virtual Reality.* Routledge, February 2013.

[40] Ray L Birdwhistell. *Kinesics and Context: Essays on Body Motion Communication.* University of Pennsylvania press, 2010.

[41] Herbert Blumer. *Symbolic Interactionism: Perspective and Method.* University of California Press, 1986.

[42] Andrea Bönsch, Sina Radke, Heiko Overath, Laura M. Asché, Jonathan Wendt, Tom Vierjahn, Ute Habel, and Torsten W. Kuhlen. Social VR: How Personal Space is Affected by Virtual Agents' Emotions. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 199–206, March 2018.

[43] R Thomas Boone and Ross Buck. Emotional Expressivity and Trustworthiness: The Role of Nonverbal Behavior in the Evolution of Cooperation. *Journal of Nonverbal Behavior*, page 20, 2003.

[44] Lauren E. Buck, John J. Rieser, Gayathri Narasimham, and Bobby Bodenheimer. Interpersonal Affordances and Social Dynamics in Collaborative Immersive Virtual Environments: Passing Together Through Apertures. *IEEE Transactions on Visualization and Computer Graphics*, 25(5):2123–2133, May 2019.

[45] Saniye Tugba Bulu. Place presence, social presence, co-presence, and satisfaction in virtual worlds. *Computers & Education*, 58(1):154–161, January 2012.

[46] Buller Burgoon, D Buller, and W G. Woodall. Nonverbal communication: The unspoken dialogue, 1996.

[47] Judee K Burgoon, Laura K Guerrero, and Valerie Manusov. Nonverbal signals. *Handbook of interpersonal communication*, pages 239–280, 2011.

[48] Judee K. Burgoon and Beth A. Le Poire. Nonverbal cues and interpersonal judgments: Participant and observer perceptions of intimacy, dominance, composure, and formality. *Communication Monographs*, 66(2):105–124, June 1999.

[49] Burke C. Shawn, Pierce Linda G., and Salas Eduardo. *Understanding Adaptability : A Prerequisite for Effective Performance within Complex Environments.* Number v. 6 in Advances in Human Performance and Cognitive Engineering Research. JAI Press Inc, 2006.

[50] Ryan Canales, Aline Normoyle, Yu Sun, Yuting Ye, Massimiliano Di Luca, and Sophie Jörg. Virtual grasping feedback and virtual hand ownership. In *ACM Symposium on Applied Perception 2019*, pages 1–9, 2019.

[51] Joseph N Cappella and Darren M Schreiber. The interaction management function of nonverbal cues: Theory and research about mutual behavioral influence in face-to-face settings. 2006.

[52] Letizia Caso, Fridanna Maricchiolo, Marino Bonaiuto, Aldert Vrij, and Samantha Mann. The Impactof Deception and Suspicious on Different Hand Movements. *Journal of Nonverbal Behavior*, 30(1):1–19, March 2006.

[53] François Chiocchio and Hélène Essiembre. Cohesion and Performance: A Meta-Analytic Review of Disparities Between Project Teams, Production Teams, and Service Teams. *Small Group Research*, 40(4):382–420, August 2009.

[54] Akiba A. Cohen and Randall P. Harrison. Intentionality in the use of hand illustrators in face-to-face communication situations. *Journal of Personality and Social Psychology*, 28:276–279, 1973.

[55] Susan Wagner Cook and Michael K. Tanenhaus. Embodied communication: Speakers' gestures affect listeners' actions. *Cognition*, 113(1):98–104, October 2009.

[56] Paul T. Costa and Robert R. McCrae. Four ways five factors are basic. *Personality and Individual Differences*, 13(6):653–665, June 1992.

[57] Richard L. Daft and Robert H. Lengel. Information Richness. A New Approach to Managerial Behavior and Organization Design. Technical report, Texas A and M Univ College Station College of Business Administration, May 1983.

[58] Richard L. Daft and Robert H. Lengel. Organizational Information Requirements, Media Richness and Structural Design. *Management Science*, 32(5):554–571, May 1986.

[59] Jan-Peter de Ruiter. *Gesture and Speech Production.* Number 6 in MPI Series in Psycholinguistics. Univ, Nijmegen, 1998.

[60] Alena Denisova, A Imran Nordin, and Paul Cairns. The convergence of player experience questionnaires. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play*, pages 33–37. ACM, 2016.

[61] Bella M. DePaulo, James J. Lindsay, Brian E. Malone, Laura Muhlenbruck, Kelly Charlton, and Harris Cooper. Cues to deception. *Psychological Bulletin*, 129(1):74–118, 2003.

[62] James Price Dillard, L Baxter, and DO Braithwaite. Goals-plans-action theory of message production. *Engaging theories in interpersonal communication: Multiple perspectives*, pages 65–76, 2008.

[63] Laura Dipietro, Angelo M. Sabatini, and Paolo Dario. A Survey of Glove-Based Systems and Their Applications. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 38(4):461–482, July 2008.

[64] Trevor J. Dodds, Betty J. Mohler, and Heinrich H. Bülthoff. Talk to the Virtual Hands: Self-Animated Avatars Improve Communication in Head-Mounted Display Virtual Environments. *PLoS ONE*, 6(10):e25759, October 2011.

[65] Hubert Dreyfus. Being-in-the-world. A commentary on Heidegger's being and time. Division I, Cambridge, Massachusets, 1994.

[66] N. J. Enfield. The Body as a Cognitive Artifact in Kinship Representations: Hand Gesture Diagrams by Speakers of Lao. *Current Anthropology*, 46(1):51–81, February 2005.

[67] Randi Alexandra Engle. *Toward a Theory of Multimodal Communication: Combining Speech, Gestures, Diagrams, and Demonstrations in Instructional Explanations.* PhD thesis, Stanford University, United States – California, 2000.

[68] Charles R. Evans and Kenneth L. Dion. Group cohesion and performance: A meta-analysis. *Small Group Research*, 22(2):175–186, 1991.

[69] Robert S Feldman and James M Tyler. Factoring in age: Nonverbal communication across the life span. 2006.

[70] Ylva Ferstl, Rachel McDonnell, and Michael Neff. Evaluating study design and strategies for mitigating the impact of hand tracking loss. In *ACM Symposium on Applied Perception 2021*, SAP '21, Virtual Event, France, 2021.

[71] Pierre Feyereisen. Manual Activity During Speaking in Aphasic Subjects*. *International Journal of Psychology*, 18(1-4):545–556, 1983.

[72] Guo Freeman, Dane Acena, Nathan J. McNeese, and Kelsea Schulenberg. Working Together Apart through Embodiment: Engaging in Everyday Collaborative Activities in Social Virtual Reality. *Proceedings of the ACM on Human-Computer Interaction*, 6(GROUP):1–25, January 2022.

[73] Donna Frick-Horbury. The effects of hand gestures on verbal recall as a function of high-and low-verbal-skill levels. *The Journal of general psychology*, 129(2):137–147, 2002.

[74] Robert Gifford. Personality and nonverbal behavior: A complex conundrum. 2006.

[75] Oliver Glauser, Shihao Wu, Daniele Panozzo, Otmar Hilliges, and Olga Sorkine-Hornung. A stretch-sensing soft glove for interactive hand pose estimation. In *ACM SIGGRAPH 2019 Emerging Technologies*, pages 1–2, Los Angeles California, July 2019. ACM.

[76] Susan Goldin-Meadow. The role of gesture in communication and thinking. *Trends in Cognitive Sciences*, 3(11):419–429, November 1999.

[77] Susan Goldin-Meadow. *Hearing Gesture: How Our Hands Help Us Think.* Harvard University Press, October 2005.

[78] Samuel D Gosling, Peter J Rentfrow, and William B Swann. A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, 37(6):504–528, December 2003.

[79] Jean Ann Graham and Michael Argyle. A cross-cultural study of the communication of extra-verbal meaning by gestures(1). *International Journal of Psychology*, 10(1):57–67, 1975.

[80] Theresa A. Graham. The Role of Gesture in Children's Learning to Count. *Journal of Experimental Child Psychology*, 74(4):333–355, December 1999.

[81] Scott W Greenwald, Wiley Corning, Markus Funk, and Pattie Maes. Comparing learning in virtual reality with learning on a 2D screen using electrostatics activities. *Journal of Universal Computer Science*, 24(2):220–245, 2018.

[82] Scott W. Greenwald, Zhangyuan Wang, Markus Funk, and Pattie Maes. Investigating Social Presence and Communication with Embodied Avatars in Room-Scale Virtual Reality. In Dennis Beck, Colin Allison, Leonel Morgado, Johanna Pirker, Foaad Khosmood, Jonathon Richter, and Christian Gütl, editors, *Immersive Learning Research Network*, Communications in Computer and Information Science, pages 75–90, Cham, 2017. Springer International Publishing.

[83] Em Griffin. *A First Look at Communication Theory, 6th Ed.* A First Look at Communication Theory, 6th Ed. McGraw-Hill, New York, NY, US, 2006.

[84] Jay Hall and W. H. Watson. The Effects of a Normative Intervention on Group Decision-Making Performance. *Human Relations*, 23(4):299–317, August 1970.

[85] Judith A Hall. Women's and men's nonverbal communication: Similarities, differences, stereotypes, and origins. 2006.

[86] Asim Hameed, Andrew Perkis, and Sebastian Möller. Evaluating Hand-tracking Interaction for Performing Motor-tasks in VR Learning Environments. In *2021 13th International Conference on Quality of Multimedia Experience (QoMEX)*, pages 219–224, June 2021.

[87] Shangchen Han, Beibei Liu, Robert Wang, Yuting Ye, Christopher D. Twigg, and Kenrick Kin. Online optical marker-based hand tracking with deep labels. *ACM Transactions on Graphics*, 37(4):166:1–166:10, July 2018.

[88] David Harris, Mark Wilson, and Samuel Vine. Development and validation of a simulation workload measure: The simulation task load index (SIM-TLX). *Virtual Reality*, 24(4):557–566, December 2020.

[89] Thomas E Harris and John C Sherblom. *Small Group and Team Communication.* Waveland Press, 2018.

[90] Sandra G. Hart. NASA Task Load Index (TLX), January 1986.

[91] Béatrice S. Hasler and Doron A. Friedman. Sociocultural Conventions in Avatar-Mediated Nonverbal Communication: A Cross-Cultural Analysis of Virtual Proxemics. *Journal of Intercultural Communication Research*, 41(3):238–259, November 2012.

[92] Khaled Hassanein and Milena Head. Manipulating perceived social presence through the web interface and its impact on attitude towards online shopping. *International Journal of Human-Computer Studies*, 65(8):689–708, August 2007.

[93] Carrie Heeter. Being There: The Subjective Experience of Presence. *Presence: Teleoperators and Virtual Environments*, 1(2):262–271, January 1992.

[94] Marieke Hoetjes, Emiel Krahmer, and Marc Swerts. On what happens in gesture when communication is unsuccessful. *Speech Communication*, 72:160–175, September 2015.

[95] Michael A Hogg. The social psychology of group cohesiveness: From attraction to social identity, 1992.

[96] James M Honeycutt. Imagined interaction theory. *Engaging theories in interpersonal communication: Multiple perspectives*, pages 77–87, 2008.

[97] Ruud Hortensius, Felix Hekele, and Emily S. Cross. The Perception of Emotion in Artificial Agents. *IEEE Transactions on Cognitive and Developmental Systems*, 10(4):852–864, December 2018.

[98] Carol Hostetter and Monique Busch. Measuring up Online: The Relationship between Social Presence and Student Learning Satisfaction. *Journal of Scholarship of Teaching and Learning*, 6(2):1–12, October 2006.

[99] Ludovic Hoyet, Kenneth Ryall, Rachel McDonnell, and Carol O'Sullivan. Sleight of hand: Perception of finger motion from reduced marker sets. In *Proceedings of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games - I3D '12*, page 79, Costa Mesa, California, 2012. ACM Press.

[100] Kumi Ishii, Mary Madison Lyons, and Sabrina A. Carr. Revisiting media richness theory for today and future. *Human Behavior and Emerging Technologies*, 1(2):124–131, 2019.

[101] Jana M. Iverson and Susan Goldin-Meadow. What's communication got to do with it? Gesture in children blind from birth. *Developmental Psychology*, 33:453–467, 1997.

[102] Jana M. Iverson and Susan Goldin-Meadow. Why people gesture when they speak. *Nature*, 396(6708):228–228, November 1998.

[103] Jana M. Iverson and Susan Goldin-Meadow. The resilience of gesture in talk: Gesture in blind speakers and listeners. *Developmental Science*, 4(4):416–422, 2001.

[104] Naomi Jacobs and Alan Garnham. The role of conversational hand gestures in a narrative task. *Journal of Memory and Language*, 56(2):291–303, February 2007.

[105] Seung-A Annie Jin. Avatars Mirroring the Actual Self versus Projecting the Ideal Self: The Effects of Self-Priming on Interactivity and Immersion in an Exergame, Wii Fit. *CyberPsychology & Behavior*, 12(6):761–765, December 2009.

[106] Elizabeth Jones, Cynthia Gallois, Victor Callan, and Michelle Barker. Strategies of accommodation: Development of a coding system for conversational interaction. *Journal of Language and Social Psychology*, 18(2):123–151, 1999.

[107] Sophie Jörg, Jessica Hodgins, and Carol O'Sullivan. The perception of finger motions. In *Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization - APGV '10*, page 129, Los Angeles, California, 2010. ACM Press.

[108] Sophie Jörg, Jessica Hodgins, and Alla Safonova. Data-driven finger motion synthesis for gesturing characters. *ACM Trans. Graph.*, 31(6):189:1–189:7, November 2012.

[109] Sophie Jörg, Yuting Ye, Franziska Mueller, Michael Neff, and Victor Zordan. Virtual hands in VR: Motion capture, synthesis, and perception. In *SIGGRAPH Asia 2020 Courses*, pages 1–32, Virtual Event, November 2020. ACM.

[110] Steven J. Karau and Kipling D. Williams. The effects of group cohesiveness on social loafing and social compensation. *Group Dynamics: Theory, Research, and Practice*, 1(2):156–168, 1997.

[111] Adam Kendon. How gestures can become like words. In *Cross-Cultural Perspectives in Nonverbal Communication*, pages 131–141. Hogrefe & Huber Publishers, Ashland, OH, US, 1988.

[112] Adam Kendon. *Gesture: Visible Action as Utterance*. Cambridge University Press, September 2004.

[113] Joann Keyton and Jeff Springston. Redefining cohesiveness in groups. *Small Group Research*, 21(2):234–254, 1990.

[114] Konstantina Kilteni, Ilias Bergstrom, and Mel Slater. Drumming in Immersive Virtual Reality: The Body Shapes the Way We Play. *IEEE Transactions on Visualization and Computer Graphics*, 19(4):597–605, April 2013.

[115] Konstantina Kilteni, Raphaela Groten, and Mel Slater. The Sense of Embodiment in Virtual Reality. *Presence: Teleoperators and Virtual Environments*, 21(4):373–387, November 2012.

[116] Daesang Kim, Dong-Joong Kim, and Woo-Hyung Whang. Cognitive Synergy in Multimedia Learning. *International Education Studies*, 6(4):p76, March 2013.

[117] Sotaro Kita. How representational gestures help speaking. *Language and gesture*, 2000.

[118] Midori Kitagawa and Brian Windsor. *MoCap for Artists: Workflow and Techniques for Motion Capture*. CRC Press, October 2020.

[119] Miela Kolomaznik, Miriam Sullivan, and Kate VyVyan. Can Virtual Reality Engage Students with Teamwork? *International Journal of Innovation in Science and Mathematics Education*, 25(4), August 2017.

[120] Meng Chun Lam, Haslina Arshad, Anton Satria Prabuwono, Siok Yee Tan, and Seyed Mostafa Mousavi Kahaki. Interaction techniques in desktop virtual environment: The study of visual feedback and precise manipulation method. *Multimedia Tools and Applications*, 77(13):16367–16398, 2018.

[121] James R. Larson. *In Search of Synergy in Small Group Performance*. Psychology Press, 2010.

[122] Beth A Le Poire. Attraction toward and nonverbal stigmatization of gay males and persons with AIDS evidence of symbolic over instrumental attirudinal structures. *Human Communication Research*, 21(2):241–279, 1994.

[123] Alex Leavitt, Brian C. Keegan, and Joshua Clark. Ping to Win? Non-Verbal Communication and Team Performance in Competitive Online Multiplayer Games. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 4337–4350, New York, NY, USA, May 2016. Association for Computing Machinery.

[124] Andrew M. Ledbetter, Samuel Hardman Taylor, and Joseph P. Mazer. Enjoyment fosters media use frequency and determines its relational outcomes: Toward a synthesis of uses and gratifications theory and media multiplexity theory. *Computers in Human Behavior*, 54:149–157, January 2016.

[125] Kwan Min Lee. Presence, explicated. *Communication theory*, 14(1):27–50, 2004.

[126] Kwan Min Lee, Younbo Jung, Jaywoo Kim, and Sang Ryong Kim. Are physically embodied social agents better than disembodied social agents?: The effects of physical embodiment, tactile interaction, and people's loneliness in human–robot interaction. *International Journal of Human-Computer Studies*, 64(10):962–973, October 2006.

[127] Rui Li, Zhenyu Liu, and Jianrong Tan. A survey on 3D hand pose estimation: Cameras, methods, and datasets. *Pattern Recognition*, 93:251–272, September 2019.

[128] Meng-Yun Liao, Ching-Ying Sung, Hao-Chuan Wang, Wen-Chieh Lin, and Fu-Yin Cherng. Embodying Historical Learners' Messages as Learning Companions in a VR Classroom. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI EA '19, pages 1–6. ACM, May 2019.

[129] Lorraine Lin and Sophie Jörg. Need a Hand? How Appearance Affects the Virtual Hand Illusion. In *Proceedings of the ACM Symposium on Applied Perception*, SAP '16, pages 69–76, 2016.

[130] Lorraine Lin, Aline Normovle, Alexandra Adkins, Yu Sun, Andrew Robb, Yuting Ye, Massimiliano Di Luca, and Sophie Jörg. The effect of hand size and interaction modality on the virtual hand illusion. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 510–518. IEEE, 2019.

[131] Stefan Liszio, Katharina Emmerich, and Maic Masuch. The influence of social entities in virtual reality games on player experience and immersion. In *Proceedings of the 12th International Conference on the Foundations of Digital Games*, pages 1–10, Hyannis Massachusetts, August 2017. ACM.

[132] Ke Ma and Bernhard Hommel. The virtual-hand illusion: Effects of impact and threat on perceived ownership and affective resonance. *Frontiers in Psychology*, 4, 2013.

[133] Divine Maloney, Guo Freeman, and Donghee Yvette Wohn. "Talking without a Voice": Understanding Non-verbal Communication in Social Virtual Reality. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW2):1–25, October 2020.

[134] Luis R. Marcos. Hand Movements and Nondominant Fluency in Bilinguals. *Perceptual and Motor Skills*, 48(1):207–214, February 1979.

[135] David Matsumoto. Culture and nonverbal behavior. *The SAGE handbook of nonverbal communication*, pages 219–235, 2006.

[136] Rachel McDonnell, Martin Breidt, and Heinrich H. Bülthoff. Render me real? Investigating the effect of render style on the perception of animated virtual humans. *ACM Transactions on Graphics*, 31(4):91:1–91:11, July 2012.

[137] Joseph Edward McGrath. *Groups: Interaction and Performance*, volume 14. Prentice-Hall Englewood Cliffs, NJ, 1984.

[138] Ryan P McMahan, Alexander Joel D Alon, Shaimaa Lazem, Robert J Beaton, David Machaj, Michael Schaefer, Mara G Silva, Anamary Leal, Robert Hagan, and Doug A Bowman. Evaluating natural interaction techniques in video games. In *2010 IEEE Symposium on 3D User Interfaces (3DUI)*, pages 11–14. IEEE, 2010.

[139] David McNeill. *Hand and Mind : What Gestures Reveal About Thought.* University of Chicago Press, 1992.

[140] David McNeill. *Gesture and Thought.* University of Chicago Press, September 2008.

[141] David McNeill. *Why We Gesture: The Surprising Role of Hand Movements in Communication.* Cambridge University Press, 2016.

[142] George Herbert Mead and Cornelius Schubert. *Mind, Self and Society*, volume 111. University of Chicago press Chicago, 1934.

[143] Albert Mehrabian. *Nonverbal Communication.* Routledge, July 2017.

[144] Nicoleta Meslec and Petru Lucian Curşeu. Too Close or Too Far Hurts: Cognitive Distance and Group Cognitive Synergy. *Small Group Research*, 44(5):471–497, October 2013.

[145] Michael D. Michalisin, Steven J. Karau, and Charnchai Tangpong. The effects of performance and team cohesion on attribution: A longitudinal simulation. *Journal of Business Research*, 57(10):1108–1115, October 2004.

[146] M. Moehring and B. Froehlich. Effective manipulation of virtual objects within arm's reach. In *2011 IEEE Virtual Reality Conference*, pages 131–138, March 2011.

[147] Fares Moustafa and Anthony Steed. A longitudinal study of small group interaction in social virtual reality. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*, pages 1–10, Tokyo Japan, November 2018. ACM.

[148] Michael Neff and Hans-Peter Seidel. Modeling Relaxed Hand Shape for Character Animation. In Francisco J. Perales and Robert B. Fisher, editors, *Articulated Motion and Deformable Objects*, Lecture Notes in Computer Science, pages 262–270, Berlin, Heidelberg, 2006. Springer.

[149] Michael Neff, Yingying Wang, Rob Abbott, and Marilyn Walker. Evaluating the Effect of Gesture and Language on Personality Perception in Conversational Agents. In Jan Allbeck, Norman Badler, Timothy Bickmore, Catherine Pelachaud, and Alla Safonova, editors, *Intelligent Virtual Agents*, Lecture Notes in Computer Science, pages 222–235, Berlin, Heidelberg, 2010. Springer.

[150] 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, October 2003.

[151] Oculus VR. Introducing High Frequency Hand Tracking. https://developer.oculus.com/blog/high-frequency-hand-tracking/, 2021-04-28, 2021. Accessed on 2021-04-28.

[152] Catherine S. Oh, Jeremy N. Bailenson, and Gregory F. Welch. A Systematic Review of Social Presence: Definition, Antecedents, and Implications. *Frontiers in Robotics and AI*, 5, 2018.

[153] Matthew Panzarino. Leap Motion Launches Software Developer Program, Sends Test Units. https://thenextweb.com/news/leap-motion-launches-software-developer-program-and-starts-sending-test-units-of-its-3d-controller, 2012. Accessed on 2021-06-14.

[154] W. Barnett Pearce and Kimberly A. Pearce. Extending the Theory of the Coordinated Management of Meaning (CMM) Through a Community Dialogue Process. *Communication Theory*, 10(4):405–423, November 2000.

[155] John Porter III, Matthew Boyer, and Andrew Robb. Guidelines on Successfully Porting Non-Immersive Games to Virtual Reality: A Case Study in Minecraft. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*, CHI PLAY '18, pages 405–415, New York, NY, USA, 2018. ACM.

[156] Mores Prachyabrued and Christoph W. Borst. Visual interpenetration tradeoffs in whole-hand virtual grasping. In *2012 IEEE Symposium on 3D User Interfaces (3DUI)*, pages 39–42, March 2012.

[157] Mores Prachyabrued and Christoph W. Borst. Visual feedback for virtual grasping. In *2014 IEEE Symposium on 3D User Interfaces (3DUI)*, pages 19–26, 2014.

[158] Alex Rayon, Timothy Gonzalez, and David Novick. Analysis of gesture frequency and amplitude as a function of personality in virtual agents. In *Proceedings of the Workshop on Multimodal Analyses Enabling Artificial Agents in Human-Machine Interaction*, MA3HMI '16, pages 3–9, New York, NY, USA, November 2016. Association for Computing Machinery.

[159] Grégoire Richard, Thomas Pietrzak, Ferran Argelaguet, Anatole Lécuyer, and Géry Casiez. Within or Between? Comparing Experimental Designs for Virtual Embodiment Studies. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 186–195, March 2022.

[160] Jennifer C. Richardson, Yukiko Maeda, Jing Lv, and Secil Caskurlu. Social presence in relation to students' satisfaction and learning in the online environment: A meta-analysis. *Computers in Human Behavior*, 71:402–417, June 2017.

[161] Bernard Ríme, Loris Schiaratura, Michel Hupet, and Anne Ghysselinckx. Effects of relative immobilization on the speaker's nonverbal behavior and on the dialogue imagery level. *Motivation and Emotion*, 8(4):311–325, December 1984.

[162] Richard M. Ryan, C. Scott Rigby, and Andrew Przybylski. The Motivational Pull of Video Games: A Self-Determination Theory Approach. *Motivation and Emotion*, 30(4):344–360, December 2006.

[163] Nick Salamon, Jonathan M. Grimm, John M. Horack, and Elizabeth K. Newton. Application of virtual reality for crew mental health in extended-duration space missions. *Acta Astronautica*, 146:117–122, May 2018.

[164] Ali-Akbar Samadani, Brandon J DeHart, Kirsten Robinson, Dana Kulić, Eric Kubica, and Rob Gorbet. A study of human performance in recognizing expressive hand movements. In *2011 RO-MAN*, pages 93–100, July 2011.

[165] Francesco Santoni, Alessio De Angelis, Antonio Moschitta, and Paolo Carbone. MagIK: A Hand-Tracking Magnetic Positioning System Based on a Kinematic Model of the Hand. *IEEE Transactions on Instrumentation and Measurement*, 70:1–13, 2021.

[166] Geoffrey B. Saxe and Rochelle Kaplan. Gesture in Early Counting: A Developmental Analysis. *Perceptual and Motor Skills*, 53(3):851–854, December 1981.

[167] Martijn J. Schuemie, Peter van der Straaten, Merel Krijn, and Charles A.P.G. van der Mast. Research on Presence in Virtual Reality: A Survey. *CyberPsychology & Behavior*, 4(2):183–201, April 2001.

[168] Valentin Schwind, Pascal Knierim, Cagri Tasci, Patrick Franczak, Nico Haas, and Niels Henze. "These are not my hands!": Effect of Gender on the Perception of Avatar Hands in Virtual Reality. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pages 1577–1582, New York, NY, USA, May 2017. Association for Computing Machinery.

[169] Stanley E Seashore. Group Cohesiveness in the Industrial Work Group. page 10, 1954.

[170] J. Segen and S. Kumar. Human-computer interaction using gesture recognition and 3D hand tracking. In *Proceedings 1998 International Conference on Image Processing. ICIP98 (Cat. No.98CB36269)*, pages 188–192 vol.3, October 1998.

[171] Claude Elwood Shannon. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423, 1948.

[172] Thomas B. Sheridan. Musings on Telepresence and Virtual Presence. *Presence: Teleoperators and Virtual Environments*, 1(1):120–126, January 1992.

[173] John Short, Ederyn Williams, and Bruce Christie. *The Social Psychology of Telecommunications*. Toronto; London; New York: Wiley, 1976.

[174] Paul Skalski, Ron Tamborini, Ashleigh Shelton, Michael Buncher, and Pete Lindmark. Mapping the road to fun: Natural video game controllers, presence, and game enjoyment. *New Media & Society*, 13(2):224–242, 2011.

[175] Harrison Jesse Smith and Michael Neff. Understanding the impact of animated gesture performance on personality perceptions. *ACM Transactions on Graphics*, 36(4):1–12, July 2017.

[176] Harrison Jesse Smith and Michael Neff. Communication Behavior in Embodied Virtual Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–12, Montreal QC Canada, April 2018. ACM.

[177] Jon Sprouse. A validation of Amazon Mechanical Turk for the collection of acceptability judgments in linguistic theory. *Behavior Research Methods*, 43(1):155–167, March 2011.

[178] Rex Stockton, Renate I. Rohde, and James Haughey. The effects of structured group exercises on cohesion, engagement, avoidance, and conflict. *Small Group Research*, 23(2):155–168, 1992.

[179] Susan G Straus and Joseph E McGrath. Does the medium matter? The interaction of task type and technology on group performance and member reactions. *Journal of applied psychology*, 79(1):87, 1994.

[180] Surinder Singh Kahai and Randolph B. Cooper. Exploring the Core Concepts of Media Richness Theory: The Impact of Cue Multiplicity and Feedback Immediacy on Decision Quality. *Journal of Management Information Systems*, 20(1):263–299, July 2003.

[181] Ron Tamborini and Paul Skalski. The role of presence in the experience of electronic games. *Playing video games: Motives, responses, and consequences*, 1:225–240, January 2006.

[182] Theresa Jean Tanenbaum, Nazely Hartoonian, and Jeffrey Bryan. "How do I make this thing smile?": An Inventory of Expressive Nonverbal Communication in Commercial Social Virtual Reality Platforms. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–13, Honolulu HI USA, April 2020. ACM.

[183] Nana Tian, Phil Lopes, and Ronan Boulic. A review of cybersickness in head-mounted displays: Raising attention to individual susceptibility. *Virtual Reality*, 26(4):1409–1441, December 2022.

[184] Jonathan Tompson, Murphy Stein, Yann Lecun, and Ken Perlin. Real-Time Continuous Pose Recovery of Human Hands Using Convolutional Networks. *ACM Transactions on Graphics*, 33(5):169:1–169:10, September 2014.

[185] Nicholas Toothman and Michael Neff. The Impact of Avatar Tracking Errors on User Experience in VR. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 756–766, March 2019.

[186] Osku Torro, Jani Holopainen, Henri Jalo, Henri Pirkkalainen, and Antti Lähtevänoja. How to get things done in social virtual reality – A study of team cohesion in social virtual reality–enabled teams. In *Hawaii International Conference on System Sciences*, 2022.

[187] Katherine M. Tsui, Munjal Desai, Holly A. Yanco, and Chris Uhlik. Exploring use cases for telepresence robots. In *2011 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 11–18, March 2011.

[188] Iis P. Tussyadiah, Dan Wang, Timothy H. Jung, and M. Claudia tom Dieck. Virtual reality, presence, and attitude change: Empirical evidence from tourism. *Tourism Management*, 66:140–154, June 2018.

[189] Unity. Unity 4.0 - Mecanim Animation Tutorial. https://www.youtube.com/watch?v=Xx21y9eJq1U, November 2012.

[190] Spyros Vosinakis and Panayiotis Koutsabasis. Evaluation of visual feedback techniques for virtual grasping with bare hands using Leap Motion and Oculus Rift. *Virtual Reality*, 22(1):47–62, March 2018.

[191] Aldert Vrij. Nonverbal communication and deception. 2006.

[192] Lev S. Vygotsky. *Thought and Language, Revised and Expanded Edition.* MIT Press, July 2012.

[193] Adam Wagler and Michael D. Hanus. Comparing Virtual Reality Tourism to Real-Life Experience: Effects of Presence and Engagement on Attitude and Enjoyment. *Communication Research Reports*, 35(5):456–464, October 2018.

[194] Tzu-Yang Wang, Yuji Sato, Mai Otsuki, Hideaki Kuzuoka, and Yusuke Suzuki. Effect of Full Body Avatar in Augmented Reality Remote Collaboration. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 1221–1222, March 2019.

[195] Yingying Wang, Jean E. Fox Tree, Marilyn Walker, and Michael Neff. Assessing the Impact of Hand Motion on Virtual Character Personality. *ACM Transactions on Applied Perception*, 13(2):9:1–9:23, March 2016.

[196] Peter H. Waxer. Nonverbal cues for anxiety: An examination of emotional leakage. *Journal of Abnormal Psychology*, 86(3):306–314, 1977.

[197] Barbara A. Wech, Kevin W. Mossholder, Robert P. Steel, and Nathan Bennett. Does Work Group Cohesiveness Affect Individuals' Performance and Organizational Commitment?: A Cross-Level Examination. *Small Group Research*, 29(4):472–494, August 1998.

[198] Séamas Weech, Sophie Kenny, and Michael Barnett-Cowan. Presence and Cybersickness in Virtual Reality Are Negatively Related: A Review. *Frontiers in Psychology*, 10, 2019.

[199] Carolyn Y. Wei. Not Crazy, Just Talking On The Phone: Gestures And Mobile Phone Conversations. In *2006 IEEE International Professional Communication Conference*, pages 299–307, October 2006.

[200] Richard Wesp, Jennifer Hesse, Donna Keutmann, and Karen Wheaton. Gestures maintain spatial imagery. *The American Journal of Psychology*, 114(4):591–600, 2001.

[201] Nkenge Wheatland, Yingying Wang, Huaguang Song, Michael Neff, Victor Zordan, and Sophie Jörg. State of the Art in Hand and Finger Modeling and Animation. *Computer Graphics Forum*, 34(2):735–760, 2015.

[202] C. Wienrich, K. Schindler, N. Döllinqer, S. Kock, and O. Traupe. Social Presence and Cooperation in Large-Scale Multi-User Virtual Reality - The Relevance of Social Interdependence for Location-Based Environments. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 207–214, March 2018.

[203] Hal Wiiteman. Group member satisfaction: A conflict-related account. *Small Group Research*, 22(1):24–58, 1991.

[204] Yuanjie Wu, Yu Wang, Sungchul Jung, Simon Hoermann, and Robert Lindeman. Using a Fully Expressive Avatar to Collaborate in Virtual Reality: Evaluation of Task Performance, Presence, and Attraction. *Frontiers in Virtual Reality*, 2, April 2021.

[205] Katherine R. Xin. Asian American Managers: An Impression Gap?: An Investigation of Impression Management and Supervisor-Subordinate Relationships. *The Journal of Applied Behavioral Science*, 40(2):160–181, June 2004.

[206] Yajiong Xue, Chetan S Sankar, and Victor WA Mbarika. Multimedia and virtual teams: Results of an experimental research. In *American Society for Engineering Education Southeast Section Conference, Macon, GA*, 2003.

[207] Boram Yoon, Hyung-il Kim, Seo Young Oh, and Woontack Woo. Evaluating Remote Virtual Hands Models on Social Presence in Hand-based 3D Remote Collaboration. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 520–532, November 2020.

[208] Ye Yuan and Anthony Steed. Is the rubber hand illusion induced by immersive virtual reality? In *2010 IEEE Virtual Reality Conference (VR)*, pages 95–102, March 2010.