A STUDY ON THE EFFECTIVENESS OF USING TELEPRESENCE AND MULTIPLE CAMERAS IN REMOTE PHYSICAL THERAPY

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF COMMUNICATION
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

HANSEUL JUN

JUNE 2022

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Jeremy Bailenson, Primary Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Gabriella Harari

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Byron Reeves

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Gordon Wetzstein

Approved for the Stanford University Committee on Graduate Studies.

Stacey F. Bent, Vice Provost for Graduate Education

This signature page was generated electronically upon submission of this dissertation in electronic format. An original signed hard copy of the signature page is on file in University Archives.

Abstract

The present research investigates the effectiveness of using a telepresence system compared to a video conferencing system and the effectiveness of using two cameras compared to one camera for remote physical therapy. The telepresence system that was used is Telegie, which allows users to see a place in 3D through a VR headset. This telepresence system provides additional spatial information to its users. Similarly, using two cameras with a video conferencing system allows users to see a place from multiple angles and provides additional spatial information. These two approaches of providing users additional spatial information were examined and compared in the context of remote physical therapy. In this dissertation, a telepresence system will be introduced and the design criteria (real-time, multi-user, simplicity, behavioral realism, spatial realism, openness) that led to the current implementation of the telepresence system will be explained citing existing telepresence systems. These will be followed by the implementation details of the telepresence system, the design of the study, and finally, the analysis of the study.

In the study, the participants included 11 physical therapists from the University of North Carolina at Chapel Hill and 76 patients from Stanford University. The study was 2x2 factorial design: video conferencing compared to telepresence, and 1-camera compared to 2-camera. Each patient was assigned to one of the four conditions (e.g., 2-camera video conferencing). Results showed that none of the eight hypotheses predicting the main effects on many outcome measures for both telepresence and using two cameras were supported via t-tests. In additional analyses, with the individual differences both the therapist and patient controlled in a linear mixed model as random effects, using two cameras showed a marginally significant positive

effect on physical therapy evaluations from the therapists. The interaction between using telepresence and using two cameras showed a marginally significant negative effect on the evaluations from therapists. Other findings include the positive effects of using telepresence when the video clarity level is controlled and observation of the patients' spatial ability as a strong predictor of therapists' evaluations on sessions.

The findings of this dissertation indicate that video fidelity of remote communication systems matters and therefore suggest telepresence systems should provide sufficient video clarity matching the purposes of remote communication, which was remote physical therapy in this case. At the same time the findings, especially with the analyses with video clarity level controlled, suggest that with improved video clarity, telepresence may provide a better user experience compared to video conferencing. While it was not hypothesized as a primary mechanism, the spatial ability of patients was found as a strong predictor of physical therapy session evaluation, which should be further examined in the future.

Acknowledgments

I had wonderful years during my graduate studies. I deeply thank everyone including my advisors, friends, and family for all the support who are the reason this dissertation exists.

First of all, I would like to thank Jeremy Bailenson, who has been a better advisor than I could have imagined. Since the interview over a video call, I have only received amazing help and never even had a moment regretting joining this program. Thank you so much for all the guidance and being a role model for many aspects of life.

I would like to also thank the dissertation committee members, Byron Reeves, Gabriella Harari, and Gordon Wetzstein for their help making this dissertation come true. I remember the days I felt inspired taking your courses and it is my honor to have th opportunity to propose and discuss this dissertation project with you all.

Fortunately, I was able to collaborate with colleagues from the University of North Carolina at Chapel Hill. I greatly appreciate the help I have received from Henry Fuchs and Michael Lewek for setting the bar higher and pushing me forward toward conducting research that I would not have been able to imagine doing alone. Also, I thank Shaik for spending all the hours together with me going through the countless obstacles that were in front of this dissertation project. I really appreciate your tenacity.

All the work I have done in the lab was possible due to great help from the lab managers of VHIL, Tobin Asher, Elise Ogle, Talia Weiss, Crystal Chan, and Brian Beams. I deeply thank you all for answering my unclear questions and helping my ideas come true. Also, I would like to thank Janine Zacharia for the ongoing help I have received that starts from the welcoming party to the recruitment for this

dissertation study.

I would like to thank labmates and fellow students for the enormous amount of emotional support that I have received. I thank Fernanda Herrera for showing me how to do all the daily work as a graduate student and Catherine Oh for the extra emotional support. I thank Geraldine Fauville, Anna Carolina Queiroz, Marijn Mado, Eugy Han, and Cyan DeVeaux for being wonderful labmates and friends making the lab such a colorful place, and Mark for all the hours of discussions and being my partner researcher in crime. For the fellow students of the communication department, I thank you all for providing me such a warm, fun, and supporting environment. While I am not going to try to name everyone due to my anxiety of not listing someone's name here, the third floor of McClatchy Hall will always have a special place in my heart. For my cohort, Hannah Mieczkowski, Sabrina Huang, Sanna Ali, and Yingdan Lu, I would like to leave a special thank you for letting me have all such wonderful and fun memories, and also for Dan Muise, for being a patient listener of my incomprehensible jokes.

To Lily, thank you for your being such a wonderful gift of my life. It would have been impossible going through these years without your love and support. You are the best thing that happened to me.

Finally, I thank my parents, Youngsook Ko and Youngchul Jun for the unconditional love they have provided me since the beginning of my life. I find it unbelievable that I am submitting this dissertation as someone who was once a child of two graduate students. I thank my Aunt and Uncle, Jennifer Min and SK Min for showing me so much love, providing me the luxury of staying next to family in this place I thought would be so far from home.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Through the advent of new media, more and more types of information have become available to people for remote communication. For mass communication, radios, black and white televisions, and color televisions enabled the usage of audio, black and white video, then color video (Ayres, 2021). For interpersonal communication, video conferencing (VC) followed telephony (Kraut & Fish, 1995), which has become a widespread medium recently.

As the Media Ecology theory (McLuhan, 1964; Postman, 1974) points out, when a new medium supersedes an existing medium, the new medium rarely fully replaces the existing medium. Rather, it partially replaces the usage of the existing medium within the use cases where the new medium can outperform the existing medium. For example, while VC is quickly gaining popularity as a new medium for remote interpersonal communication, it is unlikely VC will ever fully replace telephony. VC is being used for circumstances that would benefit from having video available as it. For example, when a person is asking how to fix their bicycle, with the video information from VC, the person on the other side telling how to fix the bicycle can not only more accurately understand the situation but also read the face of the person trying to fix the bicycle and slow down when there is frustration. However, telephony is still preferred in cases where people would like to have a conversation with less effort. To introduce use cases of VC that can be explained from this perspective, Kydd and Ferry (1994) found corporate meetings with equivocality can benefit from using VC compared to emails. Researchers interviewed managers asking which are the

conditions needed for a successful VC meeting. Managers found preparation for the meetings and having participants already knowing each other before the meetings as the conditions. They also answered saying that the number of participants should not exceed fifteen. Bolle, Larsen, Hagen, and Gilbert (2009) found that multidisciplinary teams at hospitals performed better on VC than telephony when discussing issues about patients. The researchers of this study found that having visual information of the patients together with VC is especially better.

For remote communication, a strong candidate as the next step is the actual support of the 3D space in which we live. Currently, when we watch television or when we use VC systems, viewers can only see others from the perspective of the camera that captures the others. Telepresence (TP) describes experiences that provide *the feeling of being there* (Steuer, 1992). In TP systems, users are not restricted by the perspective of the camera. Instead, TP systems provide their users a 3D scene of other places the users can freely move around. For example, with a VR headset, the users can move around naturally relying on their headset to show them what they would see if they were actually at the other places and were moving their head in the same way.

In Section 3, an implementation of a TP system will be introduced, then, from Section 4 the effectiveness of TP for remote communication will be empirically examined in the context of remote physical therapy alongside the number of cameras the systems use (i.e., one camera vs. two cameras) as an additional experimental variable. Remote physical therapy has been chosen as the task for the study as it would very likely benefit from additional spatial information, fitting as a use case that TP will replace VC in the perspective of the Media Ecology theory. For example, when a physical therapy patient performs a squat (lowering hips from a standing position and standing back up), the therapist needs to see how far are the legs placed between each other, and at the same time check the angles of the knees. Having

3D information will likely help in this scenario. For example, Mishra, Skubic, and Abbott (2015) and Saraee et al. (2017) built systems to help remote physical therapy by providing additional 3D information based on the same belief that remote physical therapy will benefit from additional 3D information. Mishra et al. (2015) developed a video conferencing system that uses Kinect devices to capture the motions of a physical therapist patient and provides the accuracy of movements computed in 3D to the therapists. Saraee et al. (2017) also built a system for remote physical therapy using Kinect devices. Their system is a remote monitoring and evaluation platform for physical therapy and also uses the 3D information from the Kinect devices to evaluate patient movements. Both systems were introduced without a user study, which makes our comparison of systems based on evaluation from people more timely.

In this dissertation, the term TP will only cover systems that (1) utilize sensors that can capture depth (e.g., RGBD cameras, Lidar), (2) render the captured information in 3D, and (3) utilize displays that support depth perception via binocular disparity (e.g., AR/VR headsets, stereo displays). Originally, when TP was coined by Minsky (1980) as a term, the definition included high-quality haptic feedback in addition to audiovisual feedback. While haptic feedback is still valuable for TP systems, given that technology and hardware for haptic feedback are not ready for mass adoption yet, only audiovisual feedback will be discussed.

As this dissertation aims to examine TP as a potential successor to VC, the following empirical study will compare TP to VC. And additionally, the effectiveness of using another camera to capture space from another angle will be examined as another technique to provide additional spatial information. Adding another camera is adopted as an easy technique to provide spatial information. As adding another camera shares a similar goal with TP and is not harder to adopt than using TP, using TP should be better than adding another camera to be considered worth adopting. In other words, the additional camera technique exists as a bar for measuring the

worth of using TP.

Telegie[1] was used as the TP system for the following empirical study. The system runs on RGBD cameras (e.g., Azure Kinect) and VR headsets (e.g., Meta Quest 2), which are commodity hardware that is not difficult for the general population to obtain. Telegie has been chosen as it is the best among the TP systems that run on commodity hardware. Google Meet[2] was used as the VC system to compare. While there was a gap in terms of video resolution between Telegie and Google Meet, I decided to use Google Meet without degradation as an effort to maintain the comparison to actual media used in the real world.

1.1 Application of Virtual Reality for Body Movements

There are many VR applications built for physical therapy, though not many of them use VR for a TP experience. Most of them rather apply "solo" VR to provide instructions to the physical therapy patients without having a therapist in place. A subset of these systems allows remote therapists to monitor physical therapy patients. Many VR applications being introduced in this section will utilize other hardware they consider VR, but not headsets. In these cases, the researchers only aim to utilize the 3D aspect of VR, but not its immersiveness. For example, there can be a system that utilizes hand controllers to capture users' motions. This system allows users to move game characters with their hand motions and is structured to incentivize the users to move in certain ways that help the rehabilitation of the users. Many systems were built this way as many researchers found the 3D aspect of VR more crucial than the immersive nature of VR for physical therapy. VR applications for other body

[1] `https://telegie.com`
[2] `https://meet.google.com/`

movements will be introduced after the applications for physical therapy, given those systems help inform the state of the art in the field.

1.1.1 Physical Therapy Applications

Popescu, Burdea, Bouzit, and Hentz (2000) introduced an orthopedic rehabilitation system that utilizes a haptic control interface. In the paper, they provided sensors for hands and outlined future plans for extending the application to knees and elbows. In the application, the patient can see a virtual hand inside a monitor and use their haptic controller to control the virtual hand. They can practice squeezing a rubber ball or could practice more sophisticated movements such as playing with a peg-board. Therapists were able to monitor the activities of patients from remote. The researchers have reported the stability measure of their system but did not include user evaluation of their system.

With three adolescents with hemiplegic cerebral palsy, Golomb et al. (2009) conducted a clinical pilot study for 6 to 11 months using a VR telerehabilitation application for their treatment. For using the system, patients wore sensing gloves that could measure their hand movements. Wearing these gloves, patients were able to practice hand movements following gamified instructions from computer monitors. The main lesson from the study the researchers state is that "remote electronic monitoring is not enough; humans must be heavily involved in remote monitoring. Human contact and human understanding are key to the success of telerehabilitation" (p. 27).

One approach for VR applications to provide physical therapy instructions is through gamification. Lange et al. (2012) introduced JewelMine, a game built for rehabilitation, utilizing the Kinect to measure patient movements. In this game, users go down through a virtual mine inside a monitor, where they need to spread their arms at certain angles to grab virtual jewels inside the game to gain points. In this process, the game mechanics cause movements that are helpful for rehabilitation.

Applying this VR gamification approach for rehabilitation to the older population, Rendon et al. (2012) applied VR (i.e., Nintendo Wii Fit) and gamification on improving the balancing of older adults. In their study with 40 participants between 60 and 95 years of age, they divided the participants into two groups: one group used their application for improving balance for six weeks, and the other did not. They measured participants' ability to balance themselves before and after the study period. In their comparison between the participant groups, the researchers found participants who had used their application did significantly better in the 8-foot Up & Go test and the Activities-specific Balance Confidence Scale.

Applying VR to many use cases, Bertrand et al. (2013) introduced a set of VR applications they have built for training clinicians, neurorehabilitation of stroke patients, training nurses, and training technical college students. In their application for training clinicians, there is a floor of a hospital with virtual healthcare workers and virtual patients. The users of this system go through five scenes where they can learn about hand hygiene. In the neurorehabilitation game Duck Duck Punch, stroke patients perform reaching tasks to pop up targets on a virtual carnival shooting gallery. In the training system for nurses, there are multiple virtual patients including a patient who is rapidly deteriorating. The user of the system should quickly detect who is the deteriorating patient based on their observations inside the training system. In the training system for technical college students, users practice precision measurement with trackers attached to both of their hands.

VR has not only been used for providing instructions but also for building instruction programs. Camporesi, Kallmann, and Han (2013) built a VR system that allows therapists to build new therapy programs intuitively by direct demonstration and automatic delivery of these programs to the patients. Using the Kinect, their system can capture movements from the therapist and record them as a therapy program. Therapists can also monitor the activity of the patients. This work has been fur-

ther improved in subsequent research (Kallmann, Camporesi, & Han, 2015). In this system, patients can see the difference between their own movements and therapists' movements recorded for the therapy program by visualization of joint angle errors.

In their effort to apply VR to treating Parkinson's Disease patients, Feng et al. (2019) compared VR-based therapy to conventional physical therapy in their 12-week study with 28 Parkinson's Disease patients. Fourteen of the participants in the control group went through a traditional rehabilitation exercises protocol, while the 14 in the experimental group went through gamified VR applications of touching balls appearing in different positions on the screen, using the upper body to prevent their body from falling into the virtual waver, and walking through a maze. In this study, the researchers found that the experimental group outperformed the control group in terms of balance and gait after the 12 weeks compared to the group who received conventional physical therapy. The difference between the groups before the treatment was not significant, but the differences before and after the treatment and differences between the improvements were both statistically significant.

Postolache et al. (2020) aimed to combine VR serious games with wearable sensors that can improve the engagement of patients and evaluate their performances. They have introduced Cans Down challenge and Coffee Pong challenges as the serious games, which require virtually throwing balls and swinging rackets. These serious games utilized sensors that are shaped as gloves and a headband.

1.1.2 Applications for Other Body Movements

There are applications that are not built for physical therapy but are relevant to this dissertation as the design and evaluation process are very similar given they are looking at body movements in other domains.

Yang and Kim (2002) implemented a VR motion training system named Just Follow Me. The system uses an intuitive ghost metaphor, which the users of the VR

headsets can see and follow. The ghost metaphor is initially superimposed on the user, then demonstrates motions. Users can learn the motions by trying to fit into the ghost. The researchers have utilized 3D mouse devices to demonstrate teaching certain hand motions. In their paper, they have provided an example of utilizing their system for learning calligraphy.

Teaching Tai Chi using VR has also been studied. J. Bailenson et al. (2008) compared VR to a video as learning environments for Tai Chi. In the VR condition of this paper, participants were wearing polarized glasses and watched polarized projection on a wall that they saw as 3D. In the video condition, the participants saw a typical video projected on the same wall. In the two studies of this paper, participants were asked to learn Tai Chi moves using VR and using videos. In both studies, the participants have self-reported that they have learned better when using VR and attributed their ability to see themselves as the major reason, which was possible in the VR condition of the study. A subset of participants was allowed to review how they followed the moves by manipulating a 3D recording of their avatar on a computer screen, but this did not lead to a significant effect on their task performance.

In this study, an atypical VR display has been utilized: point light displays. Eaves, Breslin, Van Schaik, Robinson, and Spears (2011) studied the effectiveness of VR feedback using a point light display in learning dances. They divided 30 participants into three groups with one group receiving feedback on 12 joints, another group receiving feedback on 4 joints, and the other group receiving no feedback. The feedback provided in real-time using a point light display showed where the joints should be while the participants were learning 5 dance moves. The researchers found that the participants who received feedback on 4 joints performed better than the participants who received feedback on 12 joints or none of them in terms of following the dance moves with smaller errors.

There was also an application using Kinect to detect ballet movements. Trajkova

and Ferati (2015) introduced Super Mirror, a system that uses a Kinect to detect movements of ballet learners and shows the detected movements next to the ideal motions on a monitor in front of the ballet learners. When compared to actual teachers and asked preferences, Super Mirror was more highly evaluated by high school students in lower grades than upper grades. Between the motions, Super Mirror was preferred the most for Plie (knee bend in ballet), a motion the Kinect can capture easily.

There are many VR applications for body movements. The applications mostly utilize the 3D aspect of VR to give instructions for the movements, gamify certain movements to cause users to follow the movements, and allow instructors, or therapists in the case of physical therapy to monitor the users. Between these applications, the lack of a TP application is noticeable. There are systems that allow instructors to prepare certain programs and monitor users from remote, and also there are systems that utilize VC for real-time communication between the instructors and users. However, no study utilizes a TP system for remotely teaching movements in an immersive 3D environment, especially for physical therapy.

Chapter 2

Theoretical Foundation

In this chapter, theories that provide the foundation for this dissertation, in particular to motivate the hypotheses and research question of the study will be introduced. Media Richness theory (Daft & Lengel, 1986) provides a framework for understanding the partial replacement of telephony by VC, and is expected to explain the next partial replacement of VC by TP. Social presence as a construct—originally from the Social Presence Theory (Short, Williams, & Christie, 1976)—provides a clear and simple goal for remote communication systems: elicit higher social presence. The Social Influence model (Blascovich, 2002) finds more human-like virtual humans to elicit more social influence, which supports TP as a medium channeling more social influence than VC. In the following sections, more details of these theories with relevant experimental results will be discussed.

Media richness theory (Daft & Lengel, 1986) introduces richness of media as its main construct, providing a thought framework for understanding which media should be used for which task. The theory defines a medium as richer if it supports more communication channels. For example, the theory finds VC richer than telephony as VC supports the video channel on top of the audio channel, which telephony also provides. Ordering media with their levels of richness, the theory maps tasks to certain levels of media richness, claiming there is an optimal level of richness for conducting each task.

Kahai and Cooper (2003) provides an example of an empirical study using this Media Richness theory as its framework. In their study with 94 participants divided

into 31 groups, each group was assigned to one of the four conditions with different levels of media richness: face-to-face, electronic meeting, electronic conferencing, and electronic mail communication system. The participants' groups were asked to come up with plans for problems related to substance abuse and student housing. From this study, many hypotheses confirming the Media Richness theory were confirmed. For example, richer media resulted in greater socio-emotional communication. Also, richer media resulted in greater message clarity and increased participants' perceptions that they can identify others' deception.

While the media richness theory originally argues that providing a higher level of richness than needed harms task performance, empirical research often did not support this argument (e.g., Dennis & Kinney, 1998; Suh, 1999). In these papers, the researchers have found higher media richness to result in higher task performance without having an optimal midpoint in terms of media richness. Dennis and Kinney (1998) examined tasks with different levels of equivocality asking participants to conduct them with media with different levels of richness. While the Media Richness theory predicts a medium with low richness to outperform a medium with high richness for a task with low equivocality, the medium with high richness outperformed the medium with low richness for tasks with low equivocality and high equivocality. In their study with four different media with different richness levels, Suh (1999) also found media with higher richness to outperform media with lower richness in both intellective and negotiation tasks. The researcher finds the lack of an interaction effect between the richness levels and types of tasks as a non-support of the Media Richness theory.

Similar to the video channel being what VC adds to telephony, I propose spatial arrangement as the communication channel TP adds to VC. Spatial arrangement, in brief, is information on the location of other people in the conversation, and is elaborated in Appendix A.

Social presence as a construct (Short et al., 1976) is another useful tool in making predictions on TP as a new immersive medium and on the effects of having additional spatial information. Due to the various positive outcomes of having higher levels of social presence in computer-mediated communication (Oh, Herrera, & Bailenson, 2019), aiming for a higher level of social presence has been considered a rule-of-thumb goal when designing remote communication systems. Slater and Wilbur (1997) defined presence as a subjective feeling of being there and immersion as a characteristic of technology which is a necessary ingredient for technology to provide presence. Using this terminology, given that AR/VR headsets are widely known as immersive devices, and that TP natively supports these immersive devices, TP is likely to produce positive outcomes based on the higher levels of social presence, especially for the therapists, as TP is likely to provide higher social presence than VC.

The Social Influence Model from Blascovich (2002) argues that four factors decide whether virtual humans can influence real people: anthropomorphic realism, behavioral realism, photographic realism, and agency. According to the theory, with higher realisms and agency, virtual humans are more likely to influence real people seeing and hearing them. Between these four factors, additional spatial information that TP can provide is expected to increase both anthropomorphic realism and behavioral realism, leading to a prediction that TP may outperform VC in terms of social influence, which is needed when, for example, giving remote physical therapy instructions. As VC systems are technically more mature than TP systems, VC systems are expected to provide higher photographic realism.

In an empirical study, Zibrek and McDonnell (2019) tested the effect of photorealism on social presence, place illusion, and embodiment. In this within-subjects experiment, 27 participants met both realistic and simple-style characters. Participants reported significantly higher social presence and place illusion when they met the realistic characters compared to the simple-style characters. In another study,

Zibrek, Martin, and McDonnell (2019) examined whether photorealism in VR affects the perception of virtual humans with emotional expressions. The researchers found that higher photorealism increased the emotional response of participants in empathetic scenarios. Again in this study, as it did in their previous study, higher photorealism led to an increase in social presence and place illusion. These results resonate with the Social Influence model which predicts higher social influence from a virtual human with higher photographic realism.

2.1 Psychological Mechanisms

As TP and VC are media that are originally defined by their technical characteristics, to examine their psychological aspects, it is necessary to connect their technical characteristics to psychological mechanisms. In this section, the following four psychological mechanisms regarding the differences between TP systems to VC systems will be discussed: agency, stereoscopy, fidelity, and comfort.

Regarding agency, TP provides more agency than VC as it allows control. The users of TP can move around freely inside the virtual environment they are within. They can change their perspective naturally through head movements using VR headsets. In VC, they can only see the other side from the perspective of the camera capturing the other side. Bystrom and Barfield (1999) found that the level of control and head tracking had a significant positive effect on task performance and that support of head tracking improved spatial realism of a virtual environment. Kim and Sundar (2013) studied the effect of having a more realistic controller when playing a violent video game and found having better control via the more realistic controller led to higher levels of aggression, presence, and arousal. Markowitz, Laha, Perone, Pea, and Bailenson (2018) found more movement in VR leading to better learning outcomes. In their study, participants experienced a VR field trip that was to learn about ocean

acidification. The researchers have found a positive correlation between the distance the participants moved inside VR with how much they have learned about ocean acidification. citeAherrera2021virtual examined the effect of avatar representation and head movement on prosocial behavior. With their study with 937 participants, the researchers have found that participants with higher head movement, which can be shown as a representation of agency, sign a supporting petition more often after conducting a VR perspective-taking task on homelessness. They also found that the participants signed the petition more often when they could choose the skin tone of their virtual hands. These studies demonstrate how providing higher agency for users, especially in VR, leads to positive outcomes including higher task performance, higher presence, better learning outcomes, and more prosocial behavior.

In terms of stereoscopy, TP provides stereoscopic vision via VR headsets, while VC does not. Baños et al. (2008) found no significant effect from stereoscopy in their experiment with relaxing or joyful environments in terms of the change of emotions. Van Schooten, Van Dijk, Zudilova-Seinstra, Suinesiaputra, and Reiber (2010) examined the effect of stereoscopy and motion cues on 3D interpretation task performance with 32 participants. They found that motion cues were more important than stereoscopy for the task performance and that stereoscopy did not add value when motion cues were already present. citeAtakatalo2011user tested the effect of stereoscopic vision using a stereoscopic monitor. They have compared the original version of a game to a version of it converted for the stereoscopic monitor. The researchers found higher presence levels from the participants when they played the stereoscopic version of the game. Ling, Brinkman, Nefs, Qu, and Heynderickx (2012) examined the effect of stereoscopy with VR headsets in a virtual environment for public speaking and found stereoscopy to significantly improve level of presence but not involvement or realism. Souchet, Philippe, Lévêque, Ober, and Leroy (2021) conducted a study with 42 participants playing a serious game simulating a job in-

terview. The participants were assigned to 3 groups that used PC, monocular VR headset, and stereoscopic VR headset. They did not find any statistically significant evidence that stereoscopy affects discomfort levels or learning outcome except from a direct paired comparison between the group who used PC and the group who used stereoscopic VR headset. Participants who used stereoscopic VR headset reported higher visual discomfort compared to the participants who used PC for the serious game. The studies manipulating stereoscopy levels find the effect of stereoscopy not clear. There was evidence that stereoscopy leads to higher social presence but the studies did not find stereoscopy to lead to higher task performance.

In terms of fidelity, currently, VC systems provide higher levels of visual fidelity than TP systems, mainly due to their maturity as systems. In other words, current generation VC systems operate with higher pixel resolutions than TP systems. Theoretically, fidelity can be connected to the Social Influence model as it overlaps with photographic realism as a construct (J. N. Bailenson et al., 2005). A remote communication system with higher fidelity can render a virtual human with higher photographic realism, which the social influence finds as a factor that can lead to social influence. Smets and Overbeeke (1995) examined the effect of image resolution on task performance on a spatial puzzle task. This study with a 3x3 factorial design assigned participants in three temporal resolution levels (active, passive, still) and three spatial resolution levels (full resolution of PAL 625 system, 36x30 mosaic, 18x15 mosaic). They found spatial resolution was not important with higher temporal resolution, while spatial resolution was significantly important in terms of task performance when participants were viewing static images with lower temporal resolution. With 40 participants, Lok, Naik, Whitton, and Brooks (2003) tested the effect of visual fidelity of self-avatars. In the design task using real or virtual wooden blocks, the researchers found that the visual fidelity did not lead to significant difference in reported presences levels. Bracken and Skalski (2009) examined the effect

of image quality on presence when playing video games. They found a significant positive effect from image quality to the reported levels of presence.

In terms of comfort, VC provides a higher level of comfort than TP since TP requires usage of VR headsets. VR headsets have been considered as a source of discomfort. For example, there have been studies on the discomfort level due to their weight (Yan, Chen, Xie, Song, & Liu, 2018) and their influence on temperature and humidity (Wang, He, & Chen, 2020). In their paper, Wang et al. (2020) reported 7.8 ℃ and 3.5% increase of temperature and relative humidity when participants wore headsets for 45 minutes. Murtza, Monroe, and Youmans (2017), in their manuscript for finding categories to prioritize in evaluating VR headsets, found headset comfort as the second leading factor in terms of how often it was mentioned as a factor to consider when they asked VR headset hardware designers. The factor mentioned the most often was the level of immersion between the nine categories the researchers evaluated.

Chapter 3

System

In this chapter, the TP system that will be used for the following study will be introduced. There will be an overview of previously introduced TP systems, a proposal of a list of design criteria, a discussion of the design space for TP systems with the design criteria, then the implementation details of the TP system—Telegie. The design of the system was based on an approach with questions, options, and criteria (MacLean, Young, Bellotti, & Moran, 1991). In other words, the design criteria that will be proposed have been applied in deciding how Telegie should be implemented. The implementation is based on the work of Jun, Bailenson, Fuchs, and Wetzstein (2018), the pipeline that connects an RGBD camera (i.e., Kinect 2) to an AR headset (i.e., HoloLens).

3.1 Recent History of Telepresence Systems

Among the many attempts to deliver the TP experience, I will concentrate on the ones that utilize RGBD cameras and AR/VR headsets. The TP systems that try to provide the TP experience mainly through wider screens or by letting users control robots will not be introduced here.

In 2011, Maimone and Fuchs (2011) implemented a TP system using RGBD cameras and autostereoscopic monitors. This system renders the spaces captured by RGBD cameras on the autostereoscopic monitors. In a subsequent iteration of the system with the adoption of AR headsets, Maimone et al. (2013) replaced the au-

tostereoscopic displays with AR headsets, using projectors to supplement the brightness of AR headsets as displays. The projectors complemented the narrow field of view and low brightness of AR headsets.

In 2012, Steed et al. (2012) introduced Beaming, a TP system that supports an asymmetric setting of a single person beaming into a group of people. This asymmetric TP system has a VR system, called the transporter, for one particular user. The person using the transporter can visit a remote location through this VR system. It is partially multi-user as there may be many users at the destination who can see the visitor through monitors, projectors, and AR displays.

In 2013, Beck, Kunert, Kulik, and Froehlich (2013) presented a group-to-group TP system with projectors and depth cameras. Their system supports connection between groups from two different locations with everyone in the groups reconstructed into meshes. By wearing shutter glasses, users can stereoscopically see the reconstructed meshes of the group on the other side.

In 2015, Roberts et al. (2015) introduced the withyou system. Users of their system are inside cubic immersive displays based on projectors and surrounded by multiple cameras. Within the cubes, the users can see the reconstructed version of others for remote communication. Their system supports stereoscopic rendering with the use of stereo glasses and also the usage of more than two cubic displays.

Also in 2015, Kowalski, Naruniec, and Daniluk (2015) introduced LiveScan3D, an open-source 3D data acquisition system using multiple Kinect 2 devices. The purpose of the researchers was to provide an open-source reconstruction system based on multiple cameras that is easy to use. In 2017, the same scholars released an application for HoloLens devices to receive and render point clouds from LiveScan3D[1]. This extension with an AR headset as a viewer of reconstructed scenes makes the system a TP system. Source code for LiveScan3D and its extension are both available.

[1] `https://github.com/MarekKowalski/LiveScan3D-Hololens`

In 2016, Room2Room (Pejsa, Kantor, Benko, Ofek, & Wilson, 2016) and Holoportation (Orts-Escolano et al., 2016) were introduced. Room2Room allows remote communication between people in two different rooms. It captures a room using an RGBD camera and displays it in the other room using a projector. Using an RGBD camera and having depth information, instead of projecting the person from the other room on a flat wall, the system projects on top of existing objects (e.g., sofas). In their study where participants were asked to perform a collaborative assembly task, the researchers found their system superior to video chat in completion time, presence, and communication efficiency. In addition to the manuscript they provided describing the system, they published RoomAlive (Jones et al., 2014), the underlying system they used to build Room2Room. The source code of RoomAlive is available[2]. Holoportation captures users with multiple surrounding RGBD cameras to reconstruct them into high-quality textured meshes. Through meshes displayed on AR headsets, the system allows dyadic remote communication. This system provides the highest visual quality out of all the systems summarized.

HoloBeam[3] is an AR TP system from ValoremReply that used a single external RGBD camera installed in front of each user as the capture device that allows users to see each other through an AR headset. With this system, they demonstrated remote communication with three people from different places[4] and a business meeting scenario utilizing a monitor for people without AR headsets[5].

A remote communication system from Spatial[6] connected people through AR headsets, VR headsets, and monitors. Their system allows their users to meet in a space with their own avatar controlled by the device they are using. They provide different methods to control the avatars for different devices. For users with an AR

[2] https://github.com/microsoft/RoomAliveToolkit
[3] https://www.microsoft.com/en-us/p/holobeam-tech/9nblggh555zf
[4] https://www.valoremreply.com/post/holobeam-ces2018
[5] https://www.valoremreply.com/post/holobeamignite2018
[6] https://spatial.io/

headset, without an external camera, their system provides face and hand animation.

Kolkmeier et al. (2018) introduced OpenIMPRESS, a software toolkit for mixed reality remote collaboration systems. In their paper, they also introduce an asymmetric TP system built with their toolkit. With this system, a person can wear a VR headset with a hand tracking system attached to communicate with a person at a remote location who is wearing an AR headset. Through the AR headset, the person can see the head position and hands of the VR user. Source code of this toolkit is available[7].

Rhee, Thompson, Medeiros, Dos Anjos, and Chalmers (2020) introduced Augmented Virtual Teleportation (AVT). Their asymmetric TP system connects a VR headset user to an AR headset user using an omnidirectional camera. The VR user can see the place of the AR user through the camera and the AR user can see the avatar of the visiting VR user. Jones, Zhang, Wong, and Rintel (2020) introduced Virtual Robot Overlay for Online Meetings (VROOM), a TP system that combines AR and robotics. With this system, a user can control a robot at a remote place through a VR headset and hand controllers, and people at the remote place wearing an AR headset can see the avatar of the VR user.

Starline[8] is a system from Google that allows the users of the systems to have a high quality video conferencing call that is spatially realistic. By sitting in front of a wide stereoscopic display, users of the system can see another remote user also sitting in front of a stereoscopic display rendered in a way similar to both users sitting in front of each other in the real world.

Fender and Holz (2022) introduced a VR system that can playback events that happened while a VR user is immersed in another task. This system, which can be seen as an extension of Velt (a framework for multi RGBD camera systems; Fender & Müller, 2018), allowed the users of the system to see the past event in a manner

[7] `https://github.com/OpenIMPRESS/OpenIMPRESS`
[8] `https://blog.google/technology/research/project-starline/`

that preserves causality—in which order the events have happened.

TP systems made impressive progress from the point where capturing and rendering in 3D itself was a challenge to a point where there are systems built for specific use cases in a decade. This was partially due to the technical advances of the capture devices and computers, and partially due to the accumulation of software engineering hours including the application of machine learning algorithms. As a result, there is a sufficient number of TP systems to justify having a framework for understanding them.

3.2 System Design Criteria

Based on assimilating the best aspects of the previous work for today's consumer technological ecosystem, we introduce the following design criteria for TP systems.

Real-time (RT) is an essential criterion for synchronous remote communication systems. Lower latency fosters interpersonal synchrony, which leads to better communication results (Chartrand & Bargh, 1999; Wiltermuth & Heath, 2009; Mogan, Fischer, & Bulbulia, 2017).

Multi-user (MU) support of more than two people is important for AR/VR-based TP systems as it is these situations that truly justify using headsets, as opposed to simple VC systems. In a dyadic conversation, though there is a slight mismatch based on the camera location, VC systems can provide the approximate experience of mutual gaze. For example, even though the eyes are not directly looking at one another, people can at least detect when the other person is looking at them. But mutual gaze completely unravels when there are more than two people on a VC system. As Sellen (1995) stated, "[In VC], one participant may believe that he or she is making eye contact, but this is not perceived by the other participant." A TP system with headsets can preserve spatial relations and allow mutual gaze in groups

as headsets can render in 3D, avoiding projection to a monitor that makes preserving, for example, gaze directions impossible.

Simplicity (S) is critical for actual use. A TP system may have stellar quality. However, if using the system is too difficult, it is unlikely that people will use the system, as there are other options for remote communication. The benefit of using the system should surpass the cost.

Behavioral realism (BR) is from the Social Influence model (Blascovich, 2002). For virtual humans (including avatars that are controlled by people) to have social influence on other people, they need to display high levels of behavioral realism. Users should be represented as displaying a rich set of body motions, as opposed to only tracking and rendering a handful of degrees of freedom of movement.

Spatial realism (SR) is whether a system connects users preserving the spatial aspect of gestures and physical surroundings of the users. This relates to the discussion regarding the eye gaze issue with VC systems. For a TP system to address such issues, the system should provide spatial realism. Nguyen and Canny (2005) described their VC system that provides a certain level of spatial realism as "spatially faithful" and Valli, Hakkarainen, and Siltanen (2021) also used the term spatially faithful in their review paper on TP systems.

Openness (O) is more for research purposes than for the users of the systems. A TP system that possesses high quality in all other criteria may still have less impact and value to other researchers if the system is not open. While this criterion is hard to achieve for proprietary systems, for a system to foster the advance of TP technology, it should have a detailed description of the system and allow installation of it at other places. Source code should be made available if this field is to progress.

Between the criteria, it should be noticed that real-time and behavioral realism are not independent from each other as a system that is not real-time is unlikely to provide a high level of behavioral realism.

System	Criteria						Rendering Device(s)
	RT	MU	S	BR	SR	O	
Maimone and Fuchs (2011)	○			○	○		Stereoscopic Monitor
Maimone et al. (2013)	○			○	○		AR Headset, Projector
Beaming (Steed et al., 2012)	○			○	○		VR Headset, Projector, Monitor
Beck et al. (2013)	○			○	○		Stereoscopic Projector
Withyou (Roberts et al., 2015)	○	○		○	○		Stereoscopic Projector
LiveScan3D (Kowalski et al., 2015)			○		○	○	AR Headset
Room2Room (Pejsa et al., 2016)	○			○		○	Projector
Holoportation (Orts-Escolano et al., 2016)	○			○	○		AR Headset
HoloBeam	○	○	○	○	○		AR Headset
Spatial	○	○	○		○		AR Headset, VR Headset, Monitor
OpenIMPRESS (Kolkmeier et al., 2018)	○			○	○	○	AR Headset, VR Headset
AVT (Rhee et al., 2020)	○		○	○	○		AR Headset, VR Headset
Starline	○			○	○		Stereoscopic Monitor
VROOM (Jones et al., 2020)	○			○	○		AR Headset, VR Headset
Fender and Holz (2022)				○	○		VR Headset

Table 3.1: Summary of existing TP systems. RT, MU, S, BR, SR, and O stand for real-time, multi-user, simplicity, behavioral realism, spatial-realism, and openness, respectively.

Table 3.1 summarizes existing TP systems based on AR/VR technology. These pioneering systems all have strengths and weaknesses and outperform the system I propose in many aspects. The criteria are designed to focus on a particular set of affordances, which have advantages in the current technological space.

3.3 Design Space

In this section, criteria introduced in the previous section will be applied to the design of a TP system. Between the options that are technically available, the criteria above were used to decide which option to use (MacLean et al., 1991). From the whole design space for TP systems, I narrowed it down to a set of design decisions for TP systems.

3.3.1 Number of External Cameras

For a TP system to work, there should be a way for the system to capture its users to render them in front of others. While operating without any external cameras would be ideal in terms of simplicity, cameras attached on AR or VR headsets are too poorly

positioned for removing all external cameras. The attached cameras are all right next to the heads of their users that makes them much worse at capturing bodies of the users than external cameras. For this reason, I chose to have an external RGBD camera, but require only one external RGBD camera per user to keep the installation of the system as easy as possible. As it will be demonstrated in the following study, using multiple cameras is supported to provide better visual quality, but only as an option, not as a requirement.

3.3.2 Spatial Symmetry

In the physical reality, if one sees a person from a certain distance, the other person sees the one from the same distance, and I call this property of our real world as *spatial symmetry*. In virtual environments, including the ones of TP systems, this symmetry can be broken. For example, a user of a TP system can be a meter away from another person from the user's perspective while the other person is two meters away from the user from the other person's perspective. While supporting spatial asymmetry may provide a more personalized TP experience, the lack of spatial symmetry would make estimating the perspectives of others harder. Based on the criterion of simplicity, I decided to maintain spatial symmetry inside this TP system to not introduce additional cognitive burden to the users of the system.

3.3.3 Position and Rotation Constraints

Creating a TP system that maintains spatial symmetry is equivalent to creating a single virtual environment where the users inside the virtual environment can see each other. From a technical perspective, a TP system with spatial symmetry places its users with a single acyclic directional graph—scene—with its users having a position and rotation, 6 degrees of freedom . Notice that a TP system without spatial symmetry may have a scene per user instead of one scene for all. Scaling can be added to

each user to make a user larger than others for example. However, this can introduce asymmetric social influences that would cause additional cognitive burden, so I decided to not consider scaling at least for now. Leveraging that the external cameras are stationarily positioned in most use cases, the system uses where the camera is as the anchor position for the user in front of the camera.

Below, I propose a set of constraints that matches the design criteria and determines the 6 degrees of freedom (position and rotation) of users in scenes. In the following discussion of position (x, y, z) and rotation (yaw, pitch, roll) values, y-axis will be in the opposite direction of gravity, z-axis will be the direction of the camera projected to have no y component, and yaw will be the rotation along the y-axis.

3.3.3.1 Matching Floors

When people meet each other in the real world, usually they are standing on the same floor. A person floating in the air or sunken underground might be entertaining; however, it would often become uncanny or at least unnecessary. Therefore, we decided to match floors of users. As this means not allowing rising, sinking, or tilting floors, matching floors determines y values and the rotation values except yaw. Figure 3.1 depicts matching floors and not matching floors.

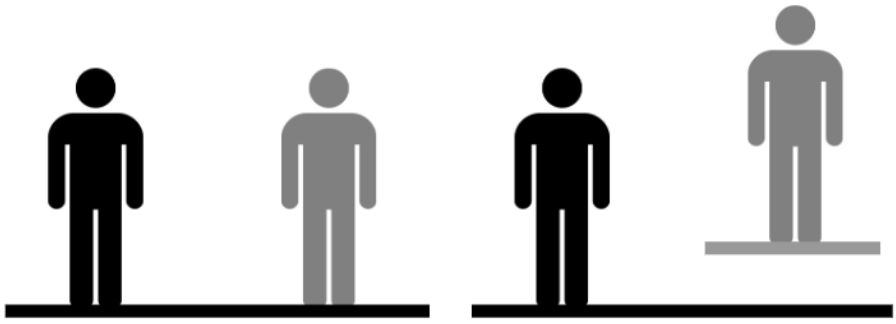


Figure 3.1: People on the same floor level (left) and people on different floor levels (right).

3.3.3.2 Angles between People

When people are positioned in a circle, the angles between people from the center of the circle contain contextual information. While unequal angles may become useful in certain situations (e.g., a conversation in a hierarchical organization), I found equal angles to be more desirable as the default setting of a TP system. When there are N users in the system, this makes the angle between each neighboring pair $2\pi/N$, determining yaw values. Figure 3.2 depicts equal and unequal angles between people.

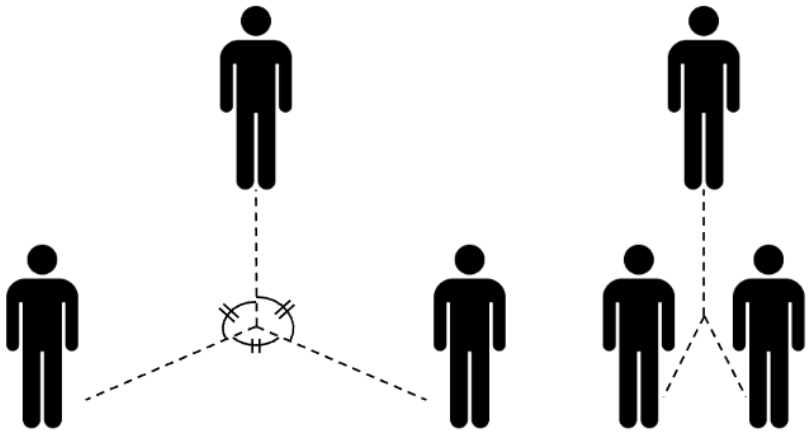


Figure 3.2: People with equal angles between each other (left) and people with unequal angles between each other (right).

3.3.3.3 Distances between People

Aside from the angles between the users, one can easily predict the distance between each user from the center of the circle to become another determining factor of the user experience. While angles and distances possess similar importance, deciding the settings for distances cannot follow the same manner of equal angles since there is no option equivalent to equal angles. With recent AR studies with virtual humans finding similarity between behavior towards real people and virtual humans (Lee, Bruder, Höllerer, & Welch, 2018; Miller et al., 2019), personal space literature (Hayduk, 1983) sheds light on this issue finding social context as a major factor deciding which dis-

tance is the most preferable. For example, people prefer to maintain larger distances towards strangers than towards their friends and, of course, than towards their significant others. Based on this contextual nature of distances, I decided to provide a user interface to set the distances. This determines x and z values.

3.3.4 Background Removal

Whether people would want others to see their background depends on their purpose of communication and social context. For example, one may want to introduce an object or the background itself. In this case, and this case, the person would want to share the background. In another case, there may be nothing in the background the person wants to share and may prefer not having the background shared. Therefore, in the implementation, background removal is provided as an option.

In terms of computational cost, background removal itself is additional computation, but at the same time largely reduces the burden for networking and rendering by removing many depth pixels. In total, background removal usually reduces the computational burden.

3.3.5 Visualization Technique

The simplest representation for an incoming stream of an RGBD camera would be a point cloud. While there are more sophisticated visualization techniques, such as skeleton-based avatars and reconstructed meshes, more sophisticated visualization techniques will be explored in future versions of the system. While visualization is an important aspect of TP, the search for an advanced visualization technique for virtual humans can be handled as an independent problem. Also, unless a visualization technique is much better than point clouds, the adoption of the visualization technique may not improve the user experience of the TP system. For example, Gamelin et al. (2021) found point clouds to outperform skeleton-based avatars for collaboration

purposes.

3.3.6 Shared Space Interpretation

A traditional interpretation of TP is letting people be in front of other people in remote places. I recommend an alternative interpretation of TP that is especially useful for understanding multi-user TP systems. With many users, the traditional interpretation leads to a quadratic growth of complexity as every user needs to be in front of everyone else, leaving $N(N-1)$ cases of someone being at another place when there are N users. By understanding a TP system as the creation of an additional virtual environment where people enter to see each other, the growth of complexity becomes linear as N users lead to N cases of someone entering the additional environment. For VR, this additional virtual environment would be the end. For AR, this additional virtual environment gets overlaid in front of the users. I call this the shared space interpretation of TP systems.

The root of the name for this interpretation—shared space—can be found in previous literature (Billinghurst, Weghorst, & Furness, 1998; Billinghurst, Poupyrev, Kato, & May, 2000). The difference between this interpretation proposed here and previous literature would be the reluctance on individuality—allowing individuals to customize the display to their needs (Szalavári, Schmalstieg, Fuhrmann, & Gervautz, 1998). Previous literature suggests allowing individuals to customize their space, but given this interpretation is aiming for simplicity, I do not recommend individual customization in TP systems.

3.4 System Implementation

The Telegie system consists of two applications: transmitter and viewer. The transmitter sends RGBD streams from cameras to viewers, and the viewer renders the

incoming RGBD streams to the users. The transmitter is supported for Windows 10 computers connected to Azure Kinect and iPhone devices. The viewer was built as a web application[9] that can run on any device with a web browser including VR headsets. The majority of the codebase is written in C++ and is shared between the transmitter and the viewer. Emscripten[10] has been used to compile the functionalities into WebAssembly for their use in the viewer.

A TP call between two people—equivalent to a video conference call—happens in three steps. First, one user creates a room using their transmitter and obtains a room ID. Then the other user joins the same room with their transmitter using the obtained room ID. Finally, both users enter the room using their viewers and start seeing each other.

Broadcasting can happen in two steps. The broadcasting user can create a room using a transmitter and obtain a room ID, then other users can enter the room through the website that shows the list of rooms.

It is also possible for a user to use multiple transmitters to increase the visual quality. After the first transmitter, additional transmitters can be added by joining the same room created by the first transmitter and after calibration, can provide additional information for the viewers. This process requires the additional transmitters to be manually calibrated based on their relative positions and rotations to the first transmitter. The feature has been added for the following study.

3.4.1 Transmitter-Viewer Pipeline

The transmitter-viewer pipeline of Telegie delivers color, depth, floor, and audio information from a transmitter to a viewer. Color, depth, and floor information form a video message for every camera frame. For color information, color pixels get encoded

[9] `https://telegie.com`
[10] `https://emscripten.org/`

in VP8 using libvpx[11]. For depth information, depth pixels go through optional background removal, mapping to the color camera's coordinate system, then Temporal RVL compression (Jun & Bailenson, 2020). Floor information gets extracted from depth pixels. Audio information is encoded in the Opus codec[12] and gets sent separately from video messages. Figure 3.3 describes the construction of video messages.

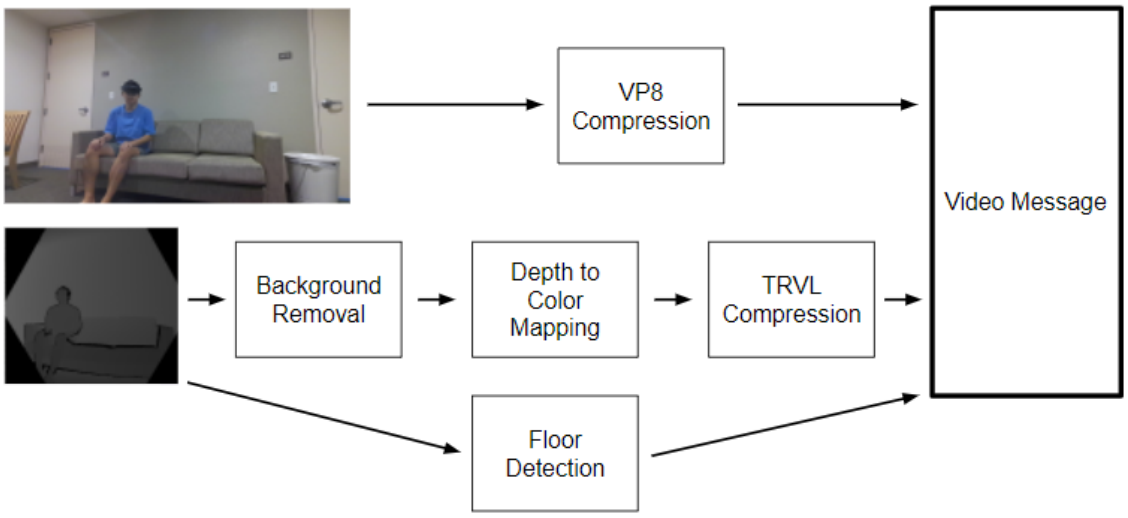


Figure 3.3: Construction of a video message from the color and depth pixels of an RGBD camera frame.

3.4.2 Networking Between Transmitter and Viewer

For users to see each other, network packets including video messages should be sent from transmitters to viewers. To support viewers running on web browsers and connections outside of local networks, across routers and firewalls, we utilized libdatachannel[13]—an implementation of WebRTC data channels. Unreliable data channels were used for real-time communication with lower latency. Unfortunately, this introduces packet loss, with which has to be dealt.

To handle packet loss, our system adopted a fountain code—Wirehair[14]. From a

[11] `https://chromium.googlesource.com/webm/libvpx`
[12] `https://opus-codec.org/`
[13] `https://github.com/paullouisageneau/libdatachannel`
[14] `https://github.com/catid/wirehair`

set of packets, a fountain code can provide limitless packets for packet loss recovery. From the receiving side, the original set of packets can be recovered after receiving sufficient fountain code packets. After splitting video messages into packets and encoding them in Wirehair, Telegie transmitters send packets with 50% of redundancy.

3.4.3 Visualization

Telegie viewers visualize every depth pixel into a quad and use color pixels to map color on the quads. We chose quads as the geometry for visualization as depth pixels are arranged in 2D grids. By default, our system operates with a color resolution of 1280x720 and a depth resolution of 640x360. In the following study, we used the half resolution of them, 720x360 and 320x180, to avoid networking issues such as occasional frame drops to complicate our analyses.

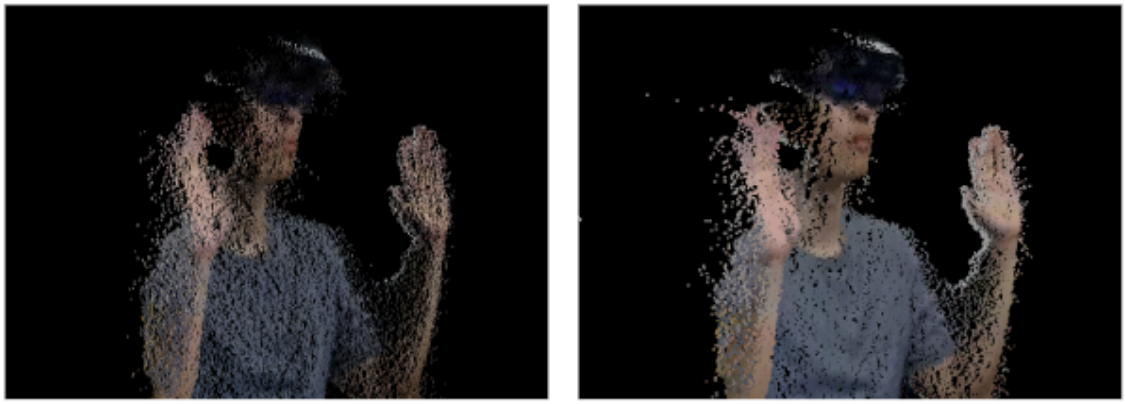


Figure 3.4: Comparison between a side view of quads facing the center of the camera which originally captured them (left) and another side view with quads rotated towards the user seeing the quads (right).

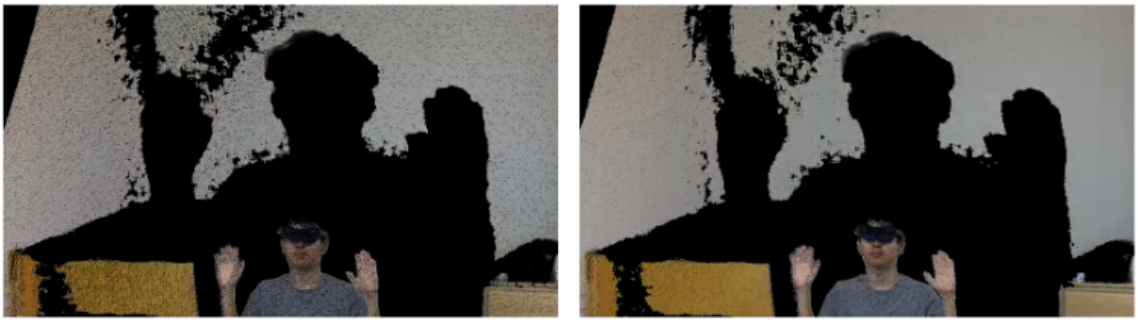


Figure 3.5: Comparison between without (left) and with (right) quad enlargement by the factor of 1.2.

The quads corresponding to depth pixels get rotated toward the users to enhance their visibility. Without additional rotations, there can be wide gaps between the quads. While these gaps can be seen as natural, reducing these gaps improves the visibility especially when the user is seeing the quads from the side (see Figure 3.4). Quads are also scaled by the factor of 1.2 to further increase their visibility (See Figure 3.5).

Chapter 4

Study

The study used a remote communication system for conducting remote physical therapy sessions between therapists and patients. Remote physical therapy was chosen as the task since it requires spatial information that TP and using two cameras can provide, and since physical therapists can participate in the study and evaluate as experts on the task. There were two independent variables with two conditions each: media type (VC vs. TP) and the number of cameras (one camera vs. two cameras).

With media type as a variable, the effectiveness of a TP system (i.e., Telegie) was compared to a VC system (i.e., Google Meet) in the context of general interpersonal communication and as a tool for remote physical therapy. With the number of cameras as a variable, using one camera for a remote communication system—showing one angle of the patient to the therapist—was compared to having two cameras—providing two angles to the therapist—in the context of general interpersonal communication and as a tool for remote physical therapy.

The two independent variables are pragmatically interesting. However, they are not by themselves psychological mechanisms that have been studied by communication researchers. Therefore, the independent variables will be connected to the psychological mechanisms including the ones previously introduced in Section 2.1. The variables will be mapped in a one-to-many manner to the five psychological mechanisms. Four—agency, stereoscopy, fidelity, and comfort—were previously introduced and cognitive load will be introduced below.

As described in Section 2.1, TP is expected to provide higher levels of agency and

stereoscopy by letting the users move around more naturally and providing proper depth perception. VC is expected to have higher levels of fidelity and comfort as the mature VC systems have higher video resolution than TP systems and the use of VR headsets for TP causes discomfort.

The number of cameras as a variable maps onto the constructs of agency and cognitive load. In terms of agency, the two-camera condition provides more agency than the one-camera condition as the additional camera provides an additional option for users to choose in terms of which direction to look at the other side. In terms of cognitive load, having two cameras elicit higher cognitive load than having one camera due to the additional perspective leading to additional cognitive load. Van Cauwenberge, Schaap, and Van Roy (2014) conducted a study on second-screen viewing and learning and found that when people watched news through a second-screen, the factual recall and comprehension were worse than single-screen viewing. This can be explained by the cognitive load theory (Chandler & Sweller, 1991). Another related study comes from Ophir, Nass, and Wagner (2009), who studied the effect of chronic media multitasking. The researchers have found heavy multitaskers to be ironically worse at task-switching, suggesting the existence of negative effects from having more screens.

The mapping between the independent variables and their conditions to the five psychological mechanisms and their levels is summarized in Table 4.1. Based on previous literature (Cummings & Bailenson, 2016), the effects sizes of agency and stereoscopy are larger than fidelity and comfort. Therefore, it is expected for TP to have more positive outcomes than VC. In terms of the number of cameras, as agency is expected to have larger effects than cognitive load in the following study, it is expected for the 2-camera condition to have more positive outcomes than the 1-camera condition. This is especially true given that therapists will be able to concentrate on one of the perspectives from the cameras in the 2-camera condition.

Between the variables, given media type maps to both agency and stereoscopy, while the number of cameras only maps to agency, even to a smaller extent than media type, it is expected for the effects sizes of media type to be larger than the number of cameras.

	Media Type	Number of Cameras	Direction of Effect
Agency	TP > VC	2-camera > 1-camera	Positive ($r = 0.41$)
Stereoscopy	TP > VC	N/A	Positive ($r = 0.32$)
Fidelity	VC > TP	N/A	Positive ($r = 0.15$)
Comfort	VC > TP	N/A	Positive
Cognitive Load	N/A	2-camera > 1-camera	Negative

Table 4.1: Conditions of independent variables and the levels of psychological mechanisms. Cells describe which condition maps to the higher levels of the conditions and the expected directions of effect from the mechanisms to the dependent variables on interpersonal communication and remote physical therapy. The r values in the directions of effect are from Cummings and Bailenson (2016).

4.1 Method

4.1.1 Participants

There were two types of participants: therapists and patients. For therapists, 11 physical therapy students were recruited from the University North Carolina at Chapel Hill. The average age of the therapists was 24.45 years old (SD = 1.51). Eight of them were female and three of them were male. Every therapist experienced all four experimental conditions of the 2x2 factorial design being scheduled to meet eight patients.

Seventy-six participants were recruited from Stanford University as patients. The participants as patients were not actual patients but patients in the sense they received instructions from the therapists. The average age of the patients was 25.71 years old (SD = 6.68) and there were 47 female and 29 male patients. Nineteen patients were

assigned for the VC1 (VC with one camera) condition, 20 patients for the TP1 (TP with one camera) condition, 19 patients for the VC2 (VC with 2 cameras) condition, and 18 patients for the TP2 (telepresence with 2 cameras) condition. The recruitment and experiment processes were approved by the Stanford IRB under protocol IRB-63483.

4.1.2 Materials and Apparatus

Therapists and patients met each other for the physical therapy sessions through monitors or VR headsets (i.e., Meta Quest 2). In VC conditions, therapists were seeing patients on a TV screen, and in TP conditions, therapists were seeing patients through a VR headset. Patients were seeing the therapist through a tablet (i.e., iPad). In all conditions, therapists were captured by a webcam for the patients to see. Patients were also captured by webcams in the VC conditions. In the TP conditions, patients were captured by both webcams and RGBD cameras (i.e., Azure Kinect). There was a microphone placed in front of both the therapists and patients.

4.1.3 Design and Procedure

This study was 2x2 factorial with the medium type (VC vs. TP) and the number of cameras (one camera vs. two cameras) as its independent variables. Due to the scarcity of therapists as participants, it was a within-participant study for therapists and a between-participant study for patients. Each therapist was scheduled for eight experimental sessions, two sessions per each of the four experimental conditions. The order of the conditions for the first four sessions was assigned based on the 4x4 Latin square. The latter four of the eight sessions were the former four conditions repeated in the reverse order. Table 4.2 shows the results of this application of the Latin square design. After the fourth therapist, the conditions were repeated by the first therapist. The conditions were mainly for therapists and the patients always saw the therapists

captured with one camera through a tablet placed in front of the patients.

	Therapist 1	Therapist 2	Therapist 3	Therapist 4
Session 1	VC1	VC2	TP1	TP2
Session 2	VC2	TP1	TP2	VC1
Session 3	TP2	VC1	VC2	TP1
Session 4	TP1	TP2	VC1	VC2
Session 5	TP1	TP2	VC1	VC2
Session 6	TP2	VC1	VC2	TP1
Session 7	VC2	TP1	TP2	VC1
Session 8	VC1	VC2	TP1	TP2

Table 4.2: The order of experimental conditions assigned to the sessions. After the fourth therapist, the conditions were repeated by the first therapist.

Before the experiment, all participants answered questions on demographics, prior VR experience, and were tested on their spatial ability. The full pre-questionnaire is in Appendix B. Patients also answered whether they had prior physical therapy experience. As the familiarity to the conditions may largely differ (therapists likely had experience using VC1 but not with VC2 or the TP conditions), therapists spent 10 minutes getting used to the experimental conditions before participating in the sessions. Patients did not go through this step as they always saw the therapists through a VC system with one camera.

During the experimental sessions, the lab for the therapists had only one webcam in front of the therapists. (See Figure 4.1.). From the patient-side, the lab had two webcams for VC and two RGBD cameras. One webcam and an RGBD camera faced the front side of the patients and another webcam and an RGBD camera faced the right side of the patients. (See Figure 4.2.) There was also a microphone placed in front of both the therapists and patients. During the sessions, all webcams and microphones were recorded.

Figure 4.1: The environment for the therapists had a webcam placed in front of the therapists and a monitor for the therapists to see the patients in VC conditions.

Figure 4.2: Photos of the front (left) and side (right) cameras of the environment for the patient. There are both a webcam and a RGBD camera installed both from the front and right side of the patients.

For the VC1 condition, the camera stream from the side camera of the patient was hidden to the therapist. For the VC2 condition, all camera streams were visible to the participants. For the TP conditions, the therapist was wearing a VR headset, thus was not able to see the TV screen in front of them, effectively hiding the VC camera streams from the therapists. For the TP1 condition, the therapist was able to see the patient through the RGBD camera placed in front of the patient. For the TP2 condition, the therapist was able to see the patient not only through the front RGBD camera, but also the side RGBD camera through their VR headset. Figure 4.3

demonstrates how patients looked like to the therapists wearing VR headsets in the TP1 condition.

Figure 4.3: Captures of Telegie from a user's perspective wearing a VR headset in the TP1 condition.

During each session, with the assigned condition, the therapist gave instructions on six exercises for fifteen minutes. The exercises were lunge, elastic band bilateral horizontal abduction, plank, exercises ball bridge upper back, side lying external rotation, and squat. From here, instead of using their full names, the exercises are called lunge, band, plank, ball, rotation, and squat. Figure 4.4 includes captures of a person performing the six exercises. Diverse exercises were chosen to examine VC and TP from various different aspects. For example, lunge requires the therapist to make sure the patient performs leg movements with proper joint angles. Plank requires the therapist to see whether the patient's back is straight. The therapists instructed two sets of each exercise for the patients. After each session, both the therapists and patients answered a post-questionnaire. See Appendix C for more details on the post-questionnaire.

When therapists were done with all eight sessions they were assigned, they were interviewed by the experimenter who ran the study at their site. The therapists answered questions comparing experimental conditions and additional follow-up questions which were asked based on their answers. See Appendix D for the interview questions comparing the conditions.

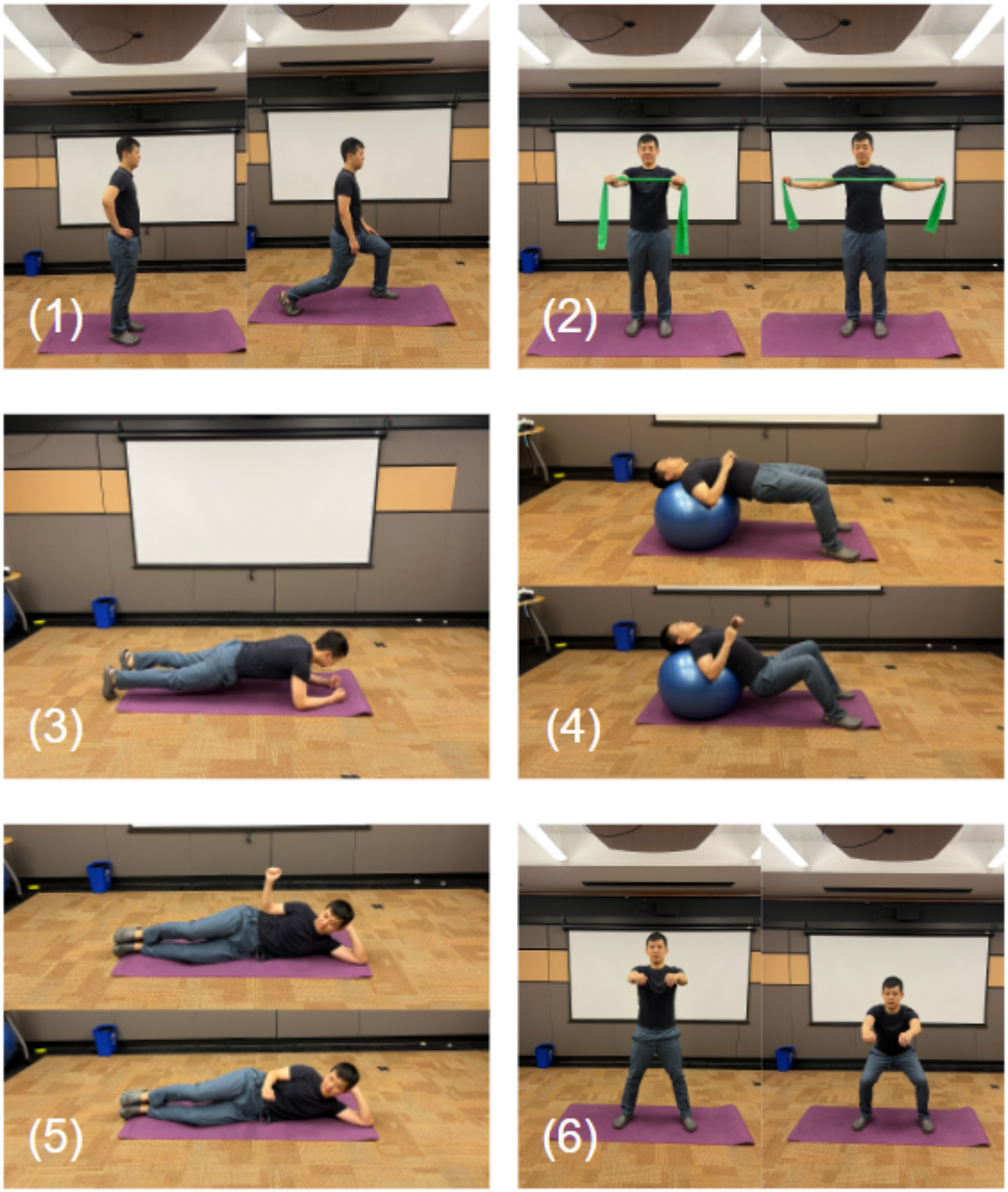


Figure 4.4: Every session consists of six exercises: (1) lunge, (2) elastic band bilateral horizontal abduction, (3) plank, (4) ball bridge upper back, (5) side lying external rotation, and (6) squat.

4.1.4 Measures

Prior VR Experience Participants were asked whether they have prior VR experience. The prior VR experience was asked as it may affect the TP conditions as TP conditions involve usage of VR. We only asked whether they have prior experience, not the amount of prior experience. Two out of the 11 therapists had prior VR experience and 49 out of the 76 patients had prior VR experience.

Prior Physical Therapy Experience Patients were asked whether they have prior physical therapy experience as familiarity with physical therapy was expected to affect their performance. We only asked whether they have prior experience, not the amount of prior experience. Forty-five out of the 76 patients had prior physical therapy experience.

Spatial Ability We measured the spatial ability of the individuals through a mental rotation test (Shepard & Metzler, 1971; Peters et al., 1995). Five questions were asking whether two figures are the same except for their orientations. Spatial ability was predicted to relate to the task performance of therapists and patients. The average score of therapists was 4.55 out of 5 (SD = 0.69), and the average score of patients was 4.50 (SD = 0.82).

Social Presence Participants were asked the level of social presence they have experienced after each experimental session by a 5-point Likert scale questionnaire from Herrera, Oh, and Bailenson (2020). The reliability of the five questions was high for both the therapists (Cronbach's $\alpha = 0.78$) and patients (Cronbach's $\alpha = 0.80$). The mean value was 3.63 (SD = 0.62) from the therapists and 3.51 (SD = 0.70) from the patients.

Communication Satisfaction Participants were asked how satisfying their communication was after each experimental session by a 5-point Likert scale questionnaire from Oh et al. (2019). The reliability of the four questions was high for both the therapists (Cronbach's $\alpha = 0.95$) and patients (Cronbach's $\alpha = 0.89$). The mean value was 3.36 (SD = 0.97) from the therapists and 3.37 (SD = 0.87) from the patients.

Interpersonal Liking Participants were asked how much they like their partner after each experimental session by a 5-point Likert scale questionnaire from Oh et al. (2019). The reliability of the three questions was high for both the therapists (Cronbach's $\alpha = 0.95$) and patients (Cronbach's $\alpha = 0.82$). The mean value was 3.73 (SD = 0.78) from the therapists and 3.61 (SD = 0.84) from the patients.

Inclusion of Other in the Self Participants were asked the level of Inclusion of Other in the Self (IOS; Aron, Aron, & Smollan, 1992) by a 7-point pictogram-based question after each experimental session. The mean value was 3.08 (SD = 1.30) from the therapists and 2.46 (SD = 1.10) from the patients.

Video Clarity Participants were asked how clear the video stream was after each experimental session in a 5-point Likert scale question. The video clarity levels were asked as the TP system does not have the same video quality as the VC system, mainly due to the VC system being a mature commercial system and the TP system being a relatively experimental system. This measured the subjective video clarity level. The mean value was 3.07 (SD = 1.21) from the therapists and 3.83 (SD = 0.81) from the patients.

Perceived Patient Motivation Therapists were asked how motivated the patient was in the experimental session in a 5-point Likert scale question. The perceived patient motivation levels were asked in an effort to capture the difference between the

patients that cannot be fully captured by measuring their spatial ability. The mean value was 3.83 (SD = 0.84).

Physical Therapy Questionnaire for Therapists Therapists were asked about their physical therapy experience after each experimental session. Leveraging that the therapists are domain experts who can provide relatively objective measures of the quality of the sessions, they were asked how well the patients learned in terms of accuracy and quickness on 5-point Likert scales per each exercise, which means 6 pairs of responses for each session. While it was planned to examine accuracy and quickness separately, due to their high correlation (Cronbach's $\alpha = 0.92$), they have been merged as the physical therapy evaluation from therapists. The mean value of physical therapy evaluation from therapists was 4.43 (SD = 0.54).

Physical Therapy Questionnaire for Patients Patients were asked about their physical therapy experience after each experimental session. The 16 questions are originally from J. Bailenson et al. (2008). The responses were coded responses with higher levels representing responses with more positive valence and their average score was analyzed as the physical therapy evaluation from patients. The 16 questions were highly correlated to each other (Cronbach's $\alpha = 0.90$) and mean value was 3.62 (SD = 0.54).

Interviews Interviews of therapists were conducted and audio recorded. The interview questions asked for comparison between the experimental conditions. See Appendix D for the full interview questions.

Due to being highly correlated with each other, social presence, communication satisfaction, interpersonal liking responses, and IOS were examined together with their average as the interpersonal communication response (Cronbach's $\alpha = 0.76$ for therapists; Cronbach's $\alpha = 0.79$ for patients). When taking the average, IOS was

rescaled to 1-5 from its original scale of 1-7.

4.2 Hypotheses and Reserach Question

Based on previous literature, it has been predicted TP to outperform VC and that using two cameras to outperform using one camera. Given that, the following hypotheses were examined using t-tests:

- H1: The level of interpersonal communication responses from therapists will be higher in TP than in VC.
- H2: The level of interpersonal communication responses from patients will be higher in TP than in VC.
- H3: The level of physical therapy evaluations from therapists will be higher in TP than in VC.
- H4: The level of physical therapy evaluations from patients will be higher in TP than in VC.
- H5: The level of interpersonal communication responses from therapists will be higher with an additional camera.
- H6: The level of interpersonal communication responses from patients will be higher with an additional camera.
- H7: The level of physical therapy evaluations from therapists will be higher with an additional camera.
- H8: The level of physical therapy evaluations from patients will be higher with an additional camera.

Additionally, whether the inclusion of extra variables—prior VR experience, prior physical therapy experience, spatial ability of the participants, video clarity of the media, or level of motivation of the patients perceived by the therapists—to the previous hypotheses can change the answers to the hypotheses was tested. This forms the following research question:

- RQ1: Does including the following variables in the analyses corresponding to H1-8 affect their results: prior VR experience, prior physical therapy experience, spatial ability of the participants, video clarity of the media, or level of motivation of the patients perceived by the therapists?

Chapter 5

Results

At the beginning of this chapter, there will be an overview of the measures, followed by the tested hypotheses. In Section 5.1, the research question will be explored. In Section 5.2, interviews with therapists and open-ended responses from patients will be summarized.

For the overview of the individual measures with the mean values and standard deviations, see Table 5.1, 5.2, E.1. While in some contexts, accuracy can trade-off with speed (Duarte & Freitas, 2005), it was not the case in this study. The accuracy and quickness of the patients were highly correlated in the positive direction ($\rho = 0.79$) based on the evaluations from the therapists. See Appendix E for more descriptive statistics and box plots of the individual measures.

All eight hypotheses proposed in Section 4.2 were tested, but none of them were supported. In brief, no positive effects of TP or using two cameras were found. See Table 5.3 for the results of the statistical tests for the hypotheses. Each row of the table corresponds to one of the four dependent variables we are examining. The columns TP and Cameras contain cells with t and p-values from t-tests corresponding to hypotheses.

	Therapist	Patient
Social Presence	3.63 (0.62)	3.51 (0.70)
Communication Satisfaction	3.36 (0.97)	3.37 (0.87)
Interpersonal Liking	3.73 (0.78)	3.61 (0.84)
IOS	3.08 (1.30)	2.46 (1.10)

Table 5.1: The mean values (and standard deviations) of social presence, communication satisfaction, interpersonal liking, and IOS.

Exercises	Accuracy	Quickness
Lunge	4.41 (0.75)	4.37 (0.76)
Band	4.50 (0.66)	4.41 (0.72)
Plank	4.57 (0.66)	4.55 (0.74)
Ball	4.20 (0.97)	4.17 (0.93)
Rotation	4.49 (0.62)	4.53 (0.60)
Squat	4.43 (0.74)	4.50 (0.79)
Average	4.43 (0.55)	4.42 (0.57)

Table 5.2: The mean values (and standard deviations) of physical therapy evaluations scores from therapists.

	TP	Cameras
Therapist Interpersonal Communication	t(73.45)=-1.02 p=0.31	t(72.95)=0.41 p=0.68
Patient Interpersonal Communication	t(73.06)=0.25 p=0.81	t(72.24)=-0.45 p=0.65
Therapist Physical Therapy Evaluations	t(72.06)=-0.58 p=0.56	t(73.98)=0.41 p=0.69
Patient Physical Therapy Evaluations	t(73.83)=0.85 p=0.40	t(73.87)=-0.43 p=0.67

Table 5.3: The summary of statistical tests corresponding to hypotheses H1-H8.

5.1 Research Questions

By testing the hypotheses, we found that the initial expectations were not supported by the statistical tests conducted on the collected data. In the t-tests for the hypotheses, neither TP nor using two cameras showed a positive effect on interpersonal communication or physical therapy.

In this section, to find the reason the initial expectations were not supported, linear mixed models are used to explore the data. For the computation of linear mixed models, we used lme4 1.1-27.1 (Bates, Mächler, Bolker, & Walker, 2014). Given all hypotheses were not supported, additional variables were added to these models to understand why the hypotheses were not supported.

First, the analyses of the evaluations of physical therapy sessions from the therapists are presented, followed by analyses that expand on the other three dependent variables. This is done to present the more impactful results first and avoid presenting redundant details of the analyses.

5.1.1 Therapist Physical Therapy Evaluations

Individual differences of both therapists and patients likely influenced therapists' evaluation of the remote physical therapy sessions. Figure 5.1 visualizes the cross-therapist variation of the evaluation scores. See Figure 5.2, which contains the distributions of evaluation scores without any control variables.

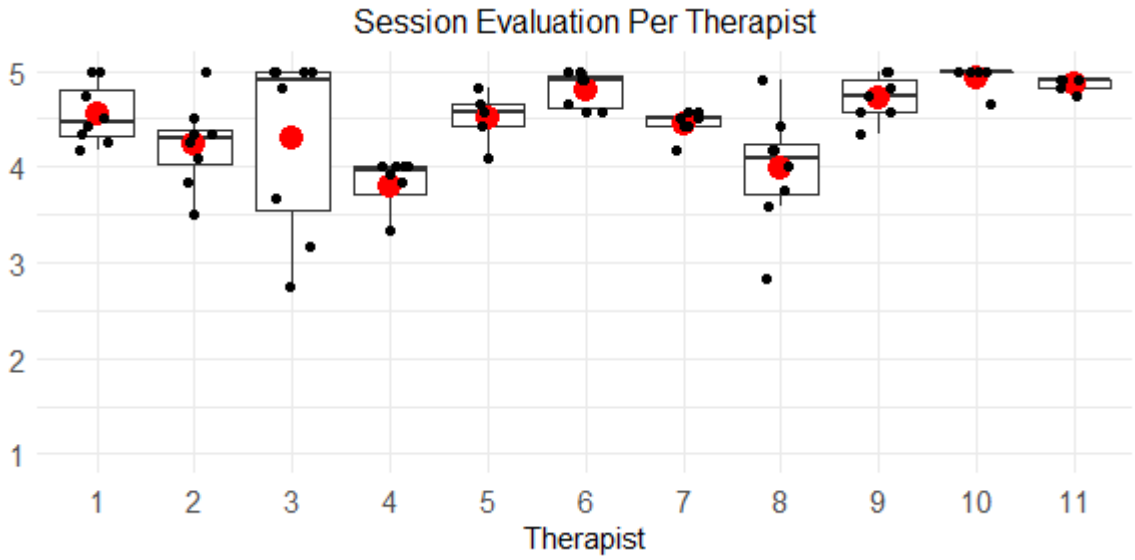


Figure 5.1: The distributions of evaluation scores from each of the therapists.

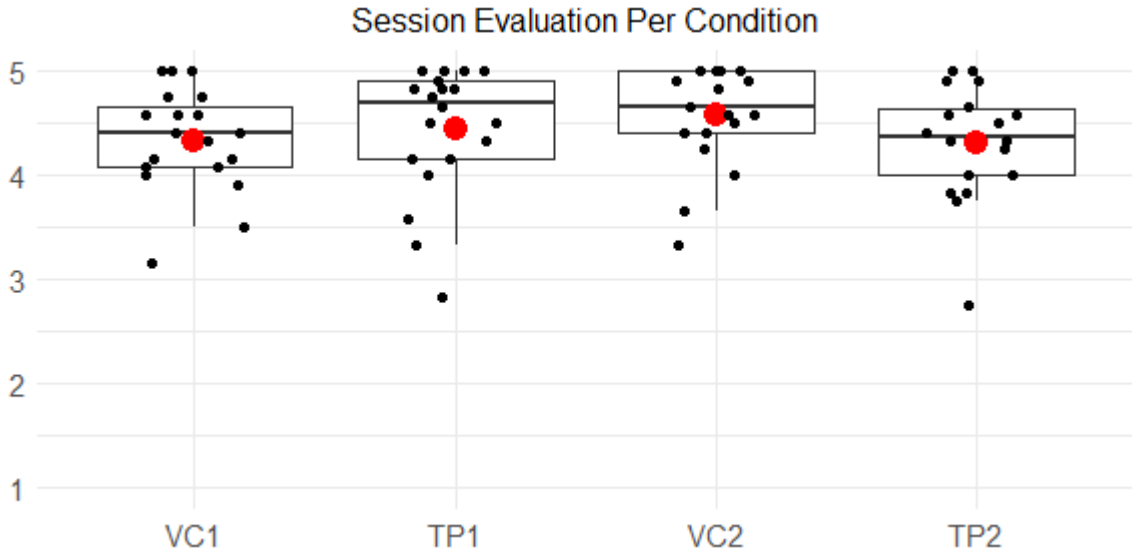


Figure 5.2: The distributions of evaluation scores per experimental condition.

Given the cross-therapist variation was larger than the cross-condition variation, to understand the effects of experimental conditions, there is a need to control the cross-therapist variation. Therefore, a linear mixed model with the experimental conditions as the fixed effects and therapist identity added as a random effect was examined. See Table 5.4 for the estimated slopes of this linear mixed model. In this

model, using two cameras was found to have a marginally significant positive effect on the evaluation scores. The interaction between TP and using two cameras was found to have a marginally significant negative effect. In other words, participants in TP2 performed worse than the prediction made based on how TP1 improved VC1 and how VC2 improved VC1. The linear model predicts TP2 to perform better than VC1 matching the summation of both improvements, but TP2 did not live up to that.

	Estimated Slope (b)	p-value
TP	0.10	0.48
Cameras	0.25	$0.09^{\dagger}$
TP × Cameras	-0.36	$0.08^{\dagger}$

Table 5.4: Estimated slopes and their p-values of the linear mixed model with conditions as the fixed effects and therapist identity as the random effect. (*: $p < 0.05$, †: $p < 0.10$)

5.1.1.1 Video Clarity

In the comparison between the TP system and the VC system that were used for this study, the gap between video resolutions was noticeable. The TP system (i.e., Telegie) had a lower resolution than the VC system (i.e., Google Meet). As this difference was not an inherent limitation of TP systems, but due to the lack of technical maturity of the system, there is value in controlling the video clarity levels of the two media.

As this video resolution difference was apparent prior to conducting the study, a question about the video clarity level was included in the post-questionnaire. Theoretically, this was based on photographic realism of the Social Influence model and fidelity as a psychological mechanism behind this study. As expected, the video clarity level reported by the therapists was highly correlated in the negative direction (ρ = -0.75) with TP. The level reported from patients did not show any correlation (ρ = 0.05), as patients were always watching the tablet regardless of the experimental condition. See Figure 5.3 for the distribution of video clarity per conditions. Figure 5.4

demonstrates the positive slopes on evaluation scores from both video clarity levels.

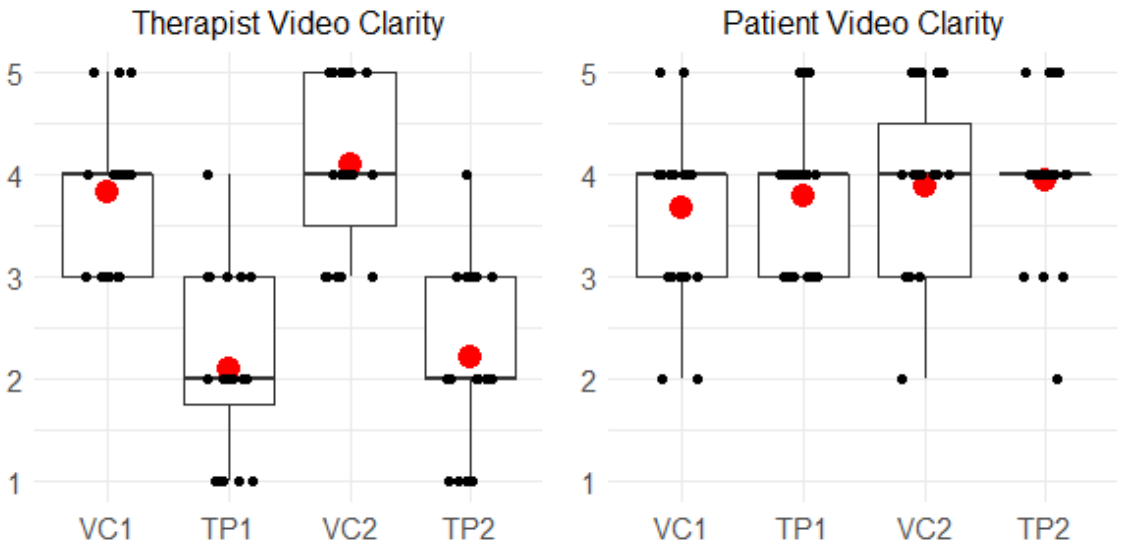


Figure 5.3: The distributions of video clarity levels reported by therapists and patients per experimental condition.

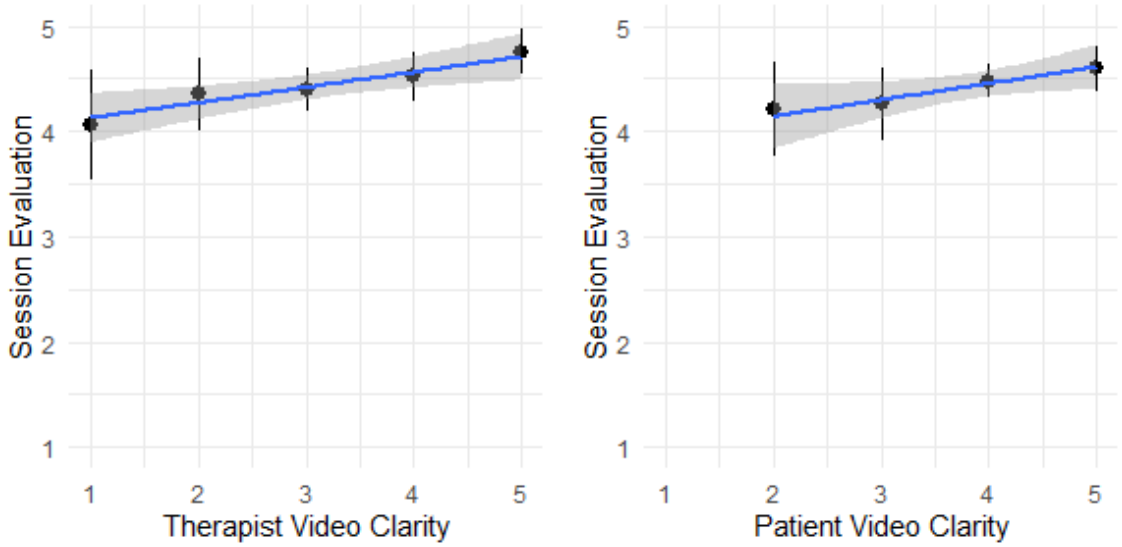


Figure 5.4: Visualization of the linear models from video clarity levels reported by therapists and patients on the evaluation scores from therapists.

The effects of experimental conditions with the influence of video clarity controlled were examined with a linear mixed model with video clarity levels as additional fixed effects. In this model, TP had a significant positive effect, and the interaction between

TP and using two cameras had a marginally significant negative effect. Using two cameras did not have a significant effect. The video clarity level reported by the therapists had a significant positive effect on the session evaluation scores. Moreover, the video clarity level reported by the patients had a marginally significant positive effect on the evaluation scores. See Table 5.5 for the estimates of the slopes of the model.

	Estimated Slope (b)	p-value
TP	0.38	0.05*
Cameras	0.18	0.23
Therapist Video Clarity	0.17	0.03*
Patient Video Clarity	0.12	$0.08^{\dagger}$
TP × Cameras	-0.34	$0.10^{\dagger}$

Table 5.5: Estimated slopes and their p-values of the linear mixed model with video clarity level from both the therapists and patients as additional fixed effects. (*: $p < 0.05$, †: $p < 0.10$)

5.1.1.2 Spatial Ability

Individual abilities matter to task performance. Since the patient was different for every physical therapy session of this study, the individual ability of the patients should have influenced the evaluation of the sessions from the therapists. To examine this, the spatial ability of patients was measured during the pre-questionnaire of the study with five spatial ability questions. Figure 5.5 shows the positive slopes from the measured spatial ability to the evaluation scores. While the spatial ability of therapists could have also influenced results, the measurement of the spatial ability of the therapists lacked variation. There was only one therapist who scored 3 out of the 11 therapists. Given this, we concentrate on the spatial ability of the patients in the below analysis.

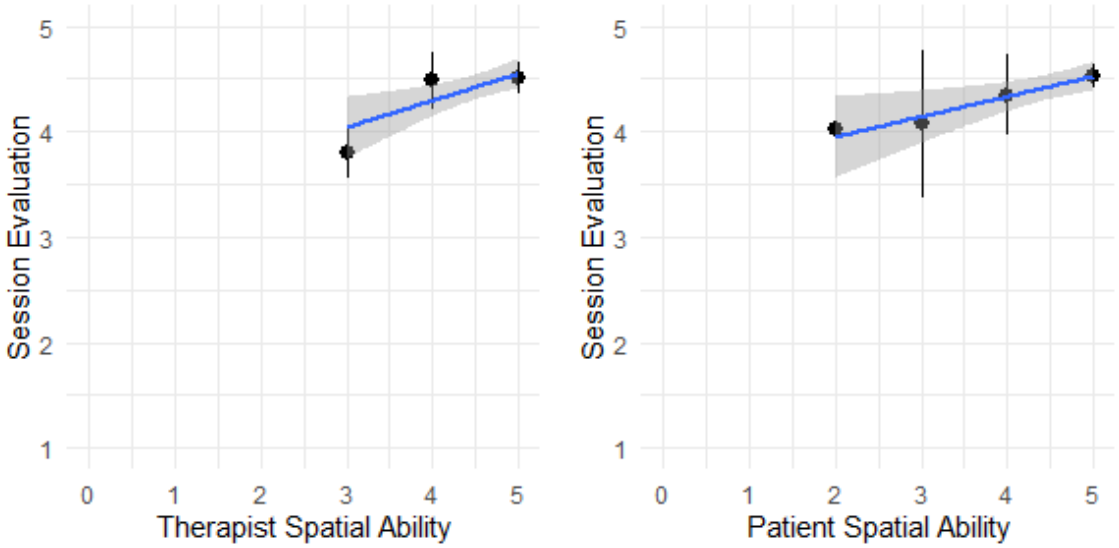


Figure 5.5: Visualization of the linear models from spatial ability levels of therapists and patients to evaluation of physical therapy sessions from therapists.

To examine the effects of experimental conditions with the level of spatial ability of the patients controlled, a linear mixed model with the spatial ability of the patients added as a fixed effect was tested. Noticeably, only the spatial ability was found as a significant positive effect. The effects from experimental conditions were no longer statistically significant with the spatial ability added to the model. See Table 5.6 for the estimated slopes of the fixed effects.

	Estimated Slope (b)	p-value
TP	0.04	0.78
Cameras	0.19	0.18
Patient Spatial Ability	0.16	0.02*
TP × Cameras	-0.27	0.19

Table 5.6: Estimated slopes and their p-values of the linear mixed model with spatial ability level of the patients as the additional fixed effect. (*: $p < 0.05$, †: $p < 0.10$)

Since spatial ability would matter more in a more spatially sophisticated task, the interaction between the spatial ability level and using two cameras for the physical therapy session was also tested. With this interaction also added as a fixed effect, only the interaction between the spatial ability level and using two cameras was a

marginally significant positive effect. See Table 5.7 for the estimated slopes.

	Estimated Slope (b)	p-value
TP	0.08	0.59
Cameras	-0.77	0.19
Patient Spatial Ability	0.17	0.42
TP × Cameras	-0.28	0.16
Cameras × Patient Spatial Ability	0.22	$0.09^{\dagger}$

Table 5.7: Estimated slopes and their p-values of the linear mixed model with the interaction between the spatial ability level of the patients and using two cameras added as another fixed effect. (*: $p < 0.05$, †: $p < 0.10$)

5.1.1.3 Perceived Patient Motivation

When the individual ability of a patient influences a physical therapy session, not only their objective capability (e.g., spatial ability) but also their motivation level during the session would matter. For example, a patient with sufficient capability to learn an exercise may not show motivation, and due to this, have difficulty learning the exercise. In their post-questionnaire, the therapists were asked to report how motivated the patients seemed. The perceived patient motivation level reported from therapists negatively correlated with TP (ρ = -0.24) and positively correlated with session evaluations ($\rho = 0.41$). See Figure 5.6 for the distributions of reported motivations levels per experimental conditions. See Figure 5.7 for the relationship between motivation levels and evaluation scores.

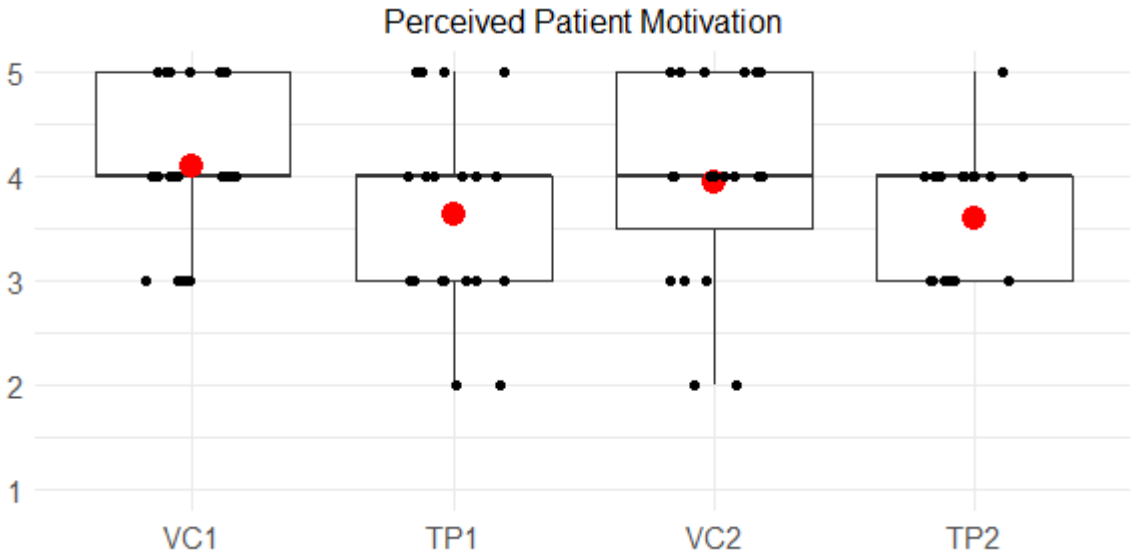


Figure 5.6: The distributions of perceived patient motivation levels reported by therapists per condition.

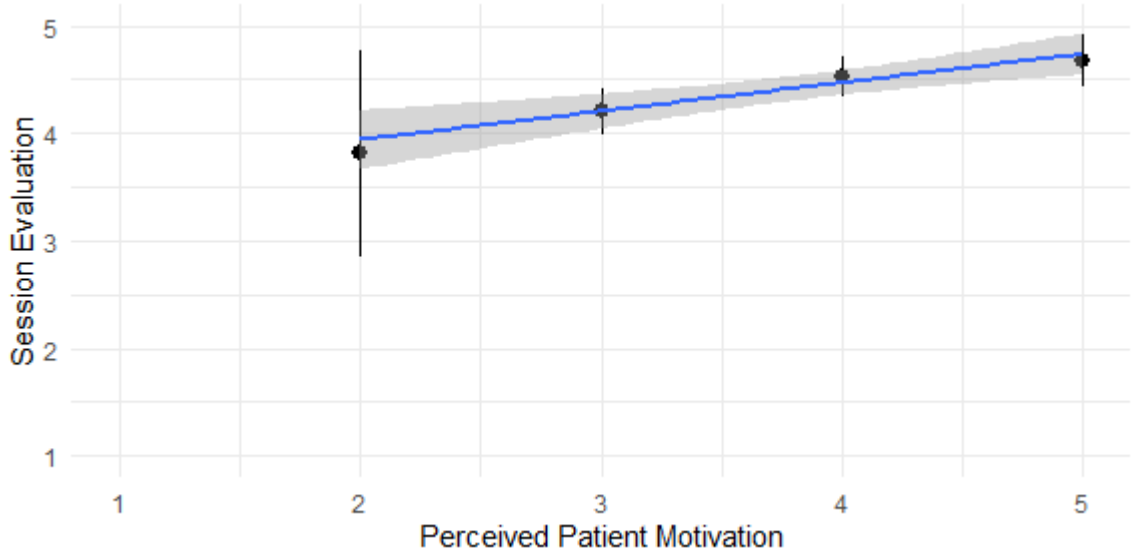


Figure 5.7: The linear model from perceived patient motivation levels reported by therapists to evaluation on physical therapy sessions from therapists.

To examine the effects of the experimental conditions with this perceived motivation level controlled, a linear mixed model with therapists' perceived motivation level as an additional fixed effect was examined. In this model, using two cameras and the interaction between TP and using two cameras showed the same effects they showed

in the model without the motivation level controlled—a marginally significant positive effect and a marginally significant negative effect. The perceived motivation level had a significant positive effect. In other words, adding perceived patient motivation to the model did not make a difference for the experimental conditions as effects, while itself was found to have a significant positive effect on the evaluations scores from therapists. See Table 5.8 for the estimated slopes of the model.

	Estimated Slope (b)	p-value
TP	0.20	0.17
Cameras	0.28	$0.06^{\dagger}$
Perceived Patient Motivation	0.23	0.01*
TP × Cameras	-0.39	$0.06^{\dagger}$

Table 5.8: Estimated slopes and their p values of the linear mixed model with perceived motivation level as the additional fixed effect. (*: $p < 0.05$, †: $p < 0.10$)

5.1.1.4 Gender Effect

In this study, the gender of the participants was asked in the pre-questionnaire as a demographic variable. As the gender of the therapists and patients are individual traits that have the potential to influence the evaluation scores, their influence was examined in the section. Figure 5.8 shows there was not much variation from this dyadic trait when the therapists were evaluating the sessions.

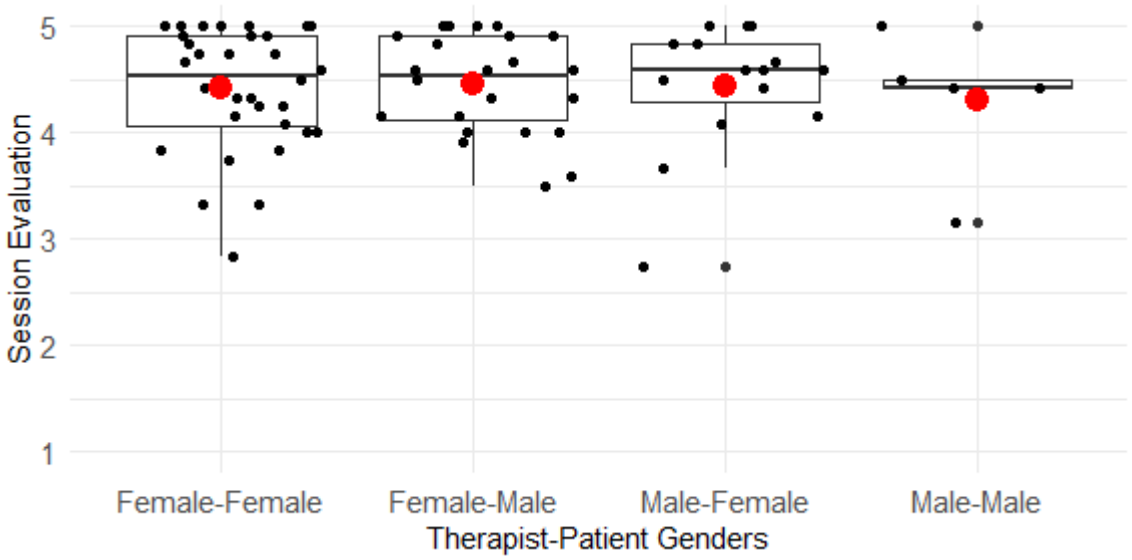


Figure 5.8: The distributions of therapists' evaluation on physical therapy sessions per genders of the therapists and patients.

To understand how the genders of the therapists and patients influenced the session evaluations from the therapists, a linear model with the gender as fixed effects were tested. In this model, the addition of the genders as fixed effects did not affect the evaluation scores. See Table 5.9 for the estimated slopes of this model.

	Estimated Slope (b)	p-value
TP	0.11	0.45
Cameras	0.25	$0.09^{\dagger}$
Female Therapist	0.03	0.90
Female Patient	0.09	0.43
TP $\times$ Cameras	-0.35	$0.09^{\dagger}$

Table 5.9: Estimated slopes and their p-values of the linear mixed model with genders of the therapists and patients as additional fixed effects. (*: $p < 0.05$, †: $p < 0.10$)

Since this model cannot capture the effect of having the same gender dyads, another linear mixed model with whether the genders of the therapists and patients were the same was tested. Again, whether the genders were the same did not have a significant effect. Table 5.10 shows the estimated slopes of this model.

	Estimated Slope (b)	p-value
TP	0.08	0.58
Cameras	0.25	$0.08^{\dagger}$
Same Gender	-0.15	0.18
TP × Cameras	-0.32	0.12

Table 5.10: Estimated slopes and their p-values of the linear mixed model with whether the genders of the therapists and patients were the same as an additional fixed effect. (*: $p < 0.05$, †: $p < 0.10$)

5.1.1.5 Order Effect

One potential source of influence to the session evaluation scores is the order of the sessions from the therapist's perspective. In this study, the therapists were asked to participate in eight sessions with a different patient per each session. During these consecutive sessions, the therapists could have experienced fatigue over the sessions or become better at conducting the sessions. Figure 5.9 shows the distribution of evaluation scores from therapists along the order of sessions that probes this order effect.

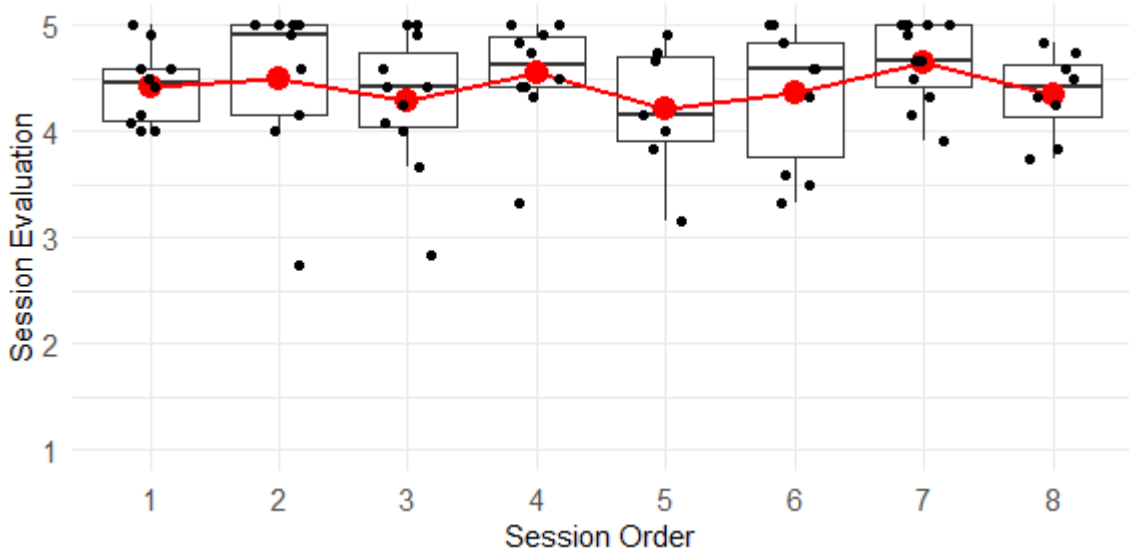


Figure 5.9: The distribution of evaluation scores per the order of sessions from the perspective of each therapist.

A linear mixed model with this order of sessions as an additional fixed effect was

tested to examine this potential effect of the session order on the evaluation scores. Resonating with the lack of variation seen in Figure 5.9, the linear mixed model did not find the order of sessions to have a significant effect. See Table 5.11 for the estimated slopes of this model.

	Estimated Slope (b)	p-value
TP	0.10	0.49
Cameras	0.24	$0.10^{\dagger}$
Order	0.01	0.63
TP $\times$ Cameras	-0.36	$0.09^{\dagger}$

Table 5.11: Estimated slopes and their p-values of the linear mixed model with the order of sessions from the perspective of each therapist as an additional fixed effect. (*: $p < 0.05$, †: $p < 0.10$)

Since the order effect of session order could have differed between the media types, especially since TP was a novel medium to the therapists, the order effect was examined again per the media types. Figure 5.10 shows the session evaluations across the session order counted separately for TP and VC. To examine the existence of the novelty effect of using new media, t-tests between the first and other three session orders have been conducted. For both TP ($t(28.33) = 1.42$, $p = 0.17$) and VC ($t(25.38) = 0.61$, $p = 0.55$), no significant novelty effect was found.

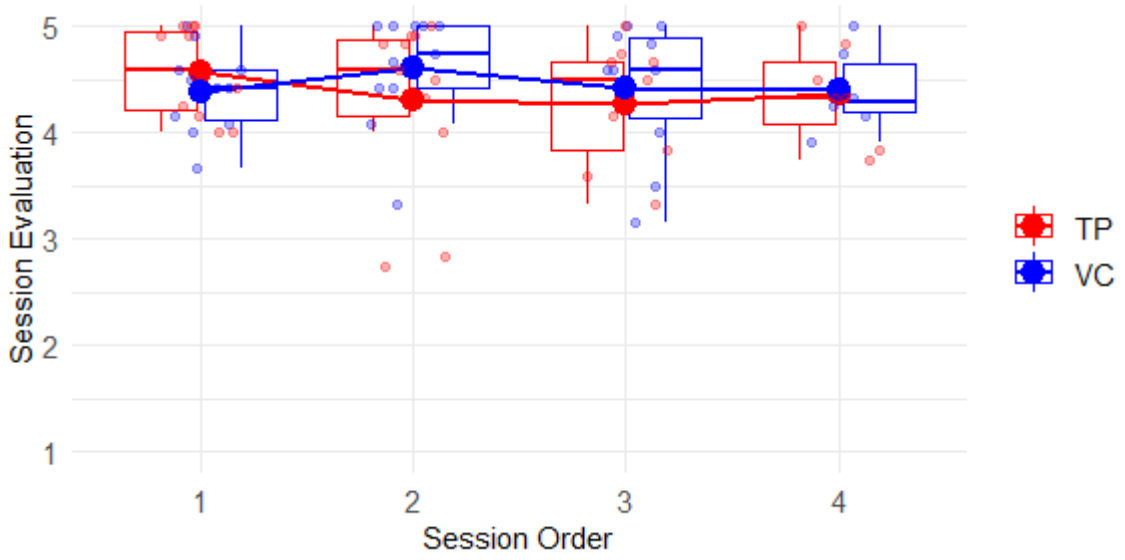


Figure 5.10: The distributions of therapists' evaluation on physical therapy exercises per order of patients from the therapists' perspective within TP and VC conditions.

Another potential source of influence on the evaluation scores is the type of exercises the therapists taught to the patients. In this study, therapists were asked to evaluate each exercise, and due to this, there may have been an influence from the types of exercise. Given the exercises were taught in a fixed order, the test on the effect from the exercise types belongs to a test of an order effect. Figure 5.11 shows the evaluation scores from therapists per each exercise they have taught to the patients.

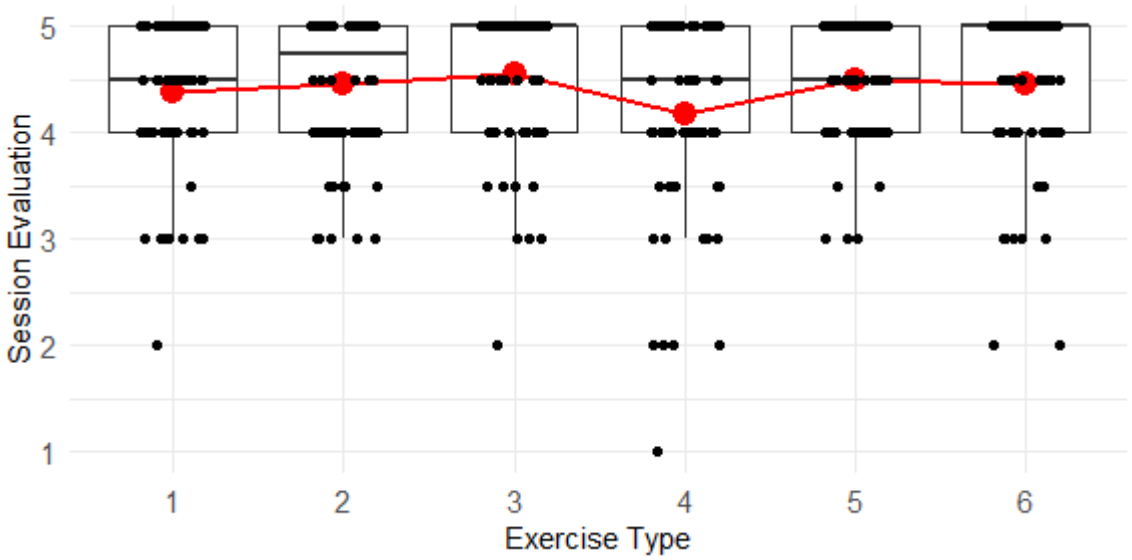


Figure 5.11: The distributions of evaluation of physical therapy exercises from therapists.

For further analysis of the influence of the exercise types, the linear mixed model with exercise types as an additional fixed effect was tested. Since exercise type is a per-exercise construct, not a per session construct, instead of the session evaluation scores from the therapists, the per exercise version of it—an average of accuracy and quickness for each exercise—was used. To control the individual differences between patients, patient identity was added as a random effect. Similarly, based on this model, no order effect is found. See Table 5.12 for the estimated slopes.

	Estimated Slope (b)	p-value
TP	0.10	0.48
Cameras	0.25	$0.09^{\dagger}$
Exercise Type	0.01	0.72
TP $\times$ Cameras	-0.36	$0.08^{\dagger}$

Table 5.12: Estimated slopes and their p values of the linear mixed model with the type of exercise as an additional fixed effect. (*: $p < 0.05$, †: $p < 0.10$)

5.1.1.6 VR Experience

Since the TP system utilizes VR, a medium that not everyone has prior experience using, different therapists could have had different levels of experience using VR. Thus, whether prior VR experience influenced the evaluation scores are examined in this section. Figure 5.12 shows the session evaluation scores divided into groups with or without prior VR experiences for both therapists and patients. As expected, therapists with prior VR experience showed a higher mean value than the therapists without prior VR experience. Patient groups did not show such differences as they did not directly use VR in this study.

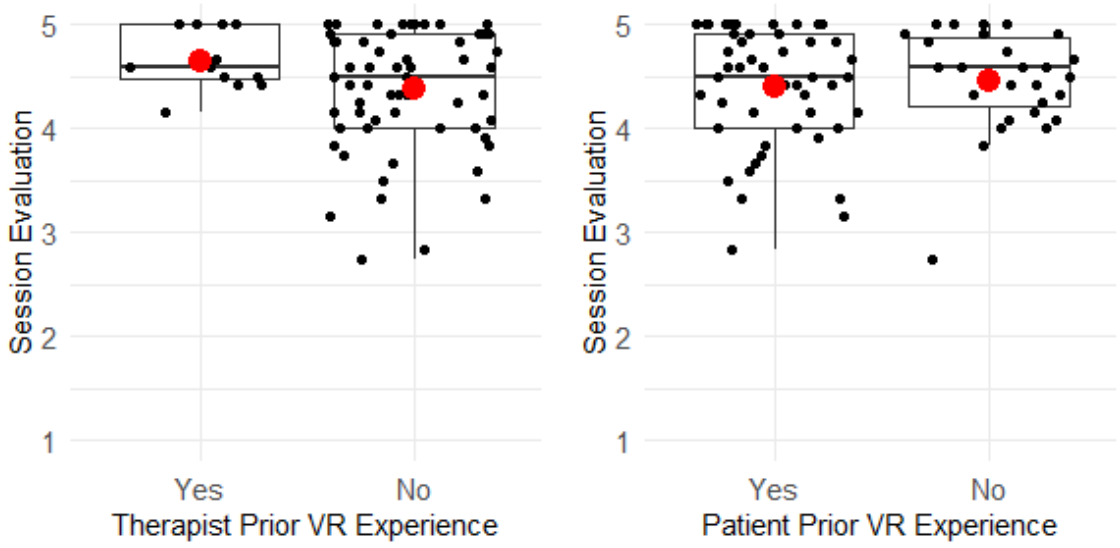


Figure 5.12: The distributions of evaluation scores divided into groups with or without prior VR experiences for both therapists and patients.

To further examine the effect of prior VR experience, given the prior VR experience of therapists matter much more than the prior VR experience of the patients, a linear mixed model only with the prior VR experience of the therapists as the additional fixed effect was tested. In the model, no significant effect from the prior VR experience of the therapists was found. See Table 5.13 for the estimated slopes of the model.

	Estimated Slope (b)	p-value
TP	0.09	0.51
Cameras	0.24	$0.10^{\dagger}$
Therapist VR Experience	0.26	0.39
TP × Cameras	-0.35	$0.09^{\dagger}$

Table 5.13: Estimated slopes and their p-values of the linear mixed model with the prior VR experience of the therapists as the additional fixed effect. (*: $p < 0.05$, †: $p < 0.10$)

Because it is possible for prior VR experience to have an interaction effect with TP, as TP conditions were when prior VR experience mattered, another linear mixed model with this interaction added as a fixed effect was tested. Again, there was no significant effect found. See Table 5.14 for the estimated slopes of this model.

	Estimated Slope (b)	p-value
TP	0.09	0.56
Cameras	0.24	$0.10^{\dagger}$
Therapist VR Experience	0.24	0.48
TP × Cameras	-0.35	$0.09^{\dagger}$
TP × Therapist VR Experience	0.04	0.88

Table 5.14: Estimated slopes and their p-values of the linear mixed model with the interaction between the prior VR experience of the therapists and TP as another fixed effect. (*: $p < 0.05$, †: $p < 0.10$)

5.1.1.7 Physical Therapy Experience

In the same way prior VR experience of the therapists can influence evaluation scores due to relative familiarity with the TP system, prior physical therapy experience of the patient can influence the evaluation scores since patients with prior physical therapy experience have the potential to outperform the patients without prior physical therapy experience. To examine this, in Figure 5.13, the distributions of session evaluation scores are compared between the groups of patients with and without prior physical therapy experience.

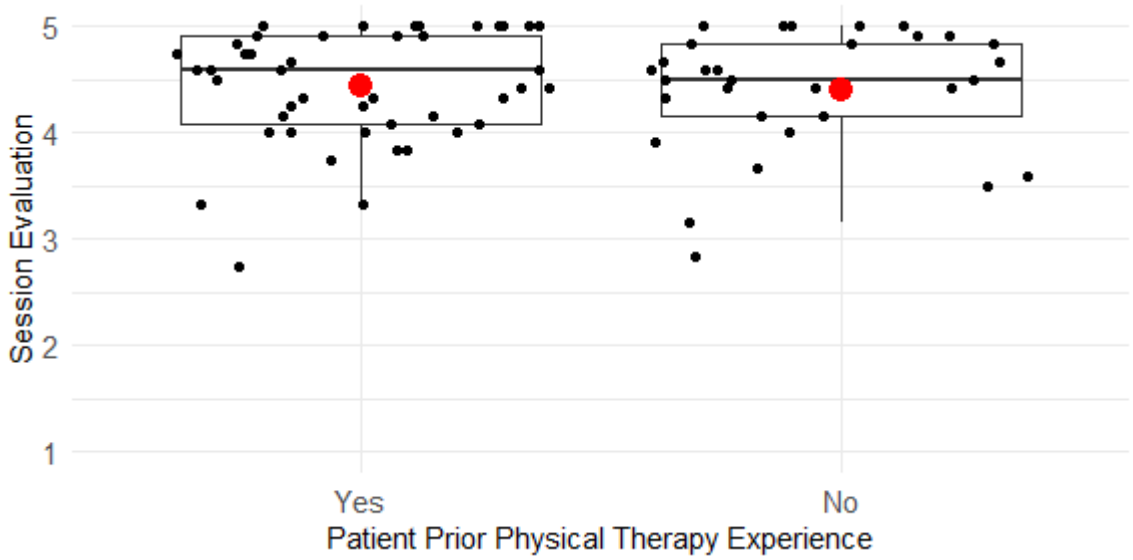


Figure 5.13: The distributions of evaluation scores on physical therapy sessions of patients with and without prior physical therapy experience.

Further examination was done using a linear mixed model with prior physical therapy experience of the patients added as the additional fixed effect. In this model, the prior physical therapy experience did not show a significant effect on the evaluation scores. See Table 5.15 for the estimated slopes.

	Estimated Slope (b)	p-value
TP	0.13	0.38
Cameras	0.23	0.12
Patient PT Experience	0.10	0.38
TP × Cameras	-0.37	$0.08^{\dagger}$

Table 5.15: Estimated slopes and their p-values of the linear mixed model with the prior physical therapy experience of the patients as the additional fixed effect. (*: $p < 0.05$, †: $p < 0.10$)

5.1.2 Expansion to Other Dependent Variables

In the above Section 5.1.1, only the physical therapy evaluation scores from the therapists were analyzed leaving three other dependent variables to be examined: physical therapy evaluations from the patients, interpersonal communication responses from

the therapists, and interpersonal communication responses from the patients. In this section, the analyses on the physical therapy evaluations will be expanded to the three other dependent variables. Beginning this expansion, Figure 5.14 provides the distributions of all four dependent variables across experimental conditions, expanding Figure 5.2.

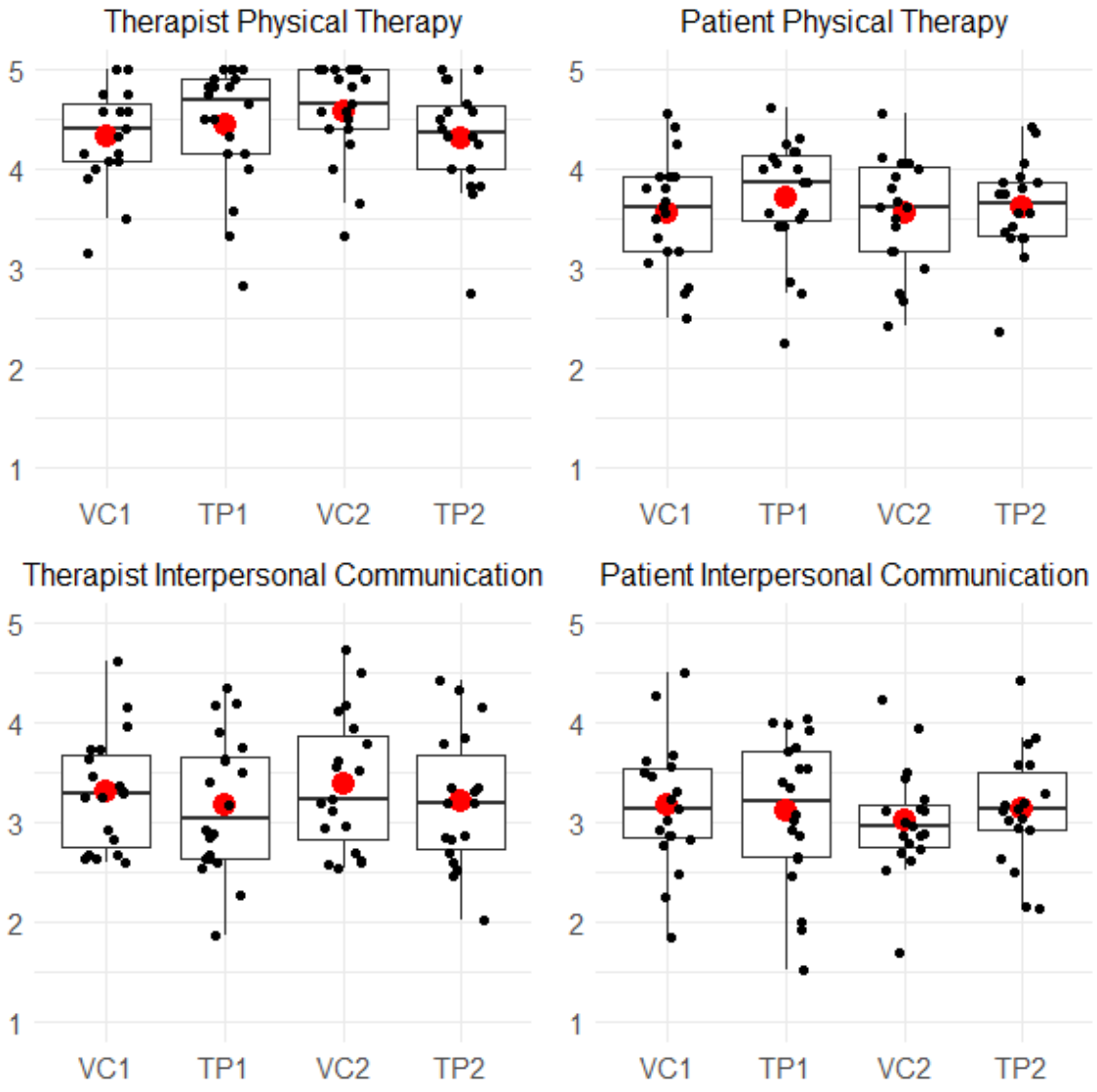


Figure 5.14: The distributions of the four dependent variables across the experimental conditions.

Across Section 5.1.1, the influence of variables that have the potential to affect the evaluation scores were examined using linear mixed models with the variables

added as additional fixed effects. This approach to examining influences on the original model that only contains the effects of experimental conditions on the evaluation scores was expanded to the three other dependent variables. See Table 5.16 for the estimated slopes of all linear mixed models that resulted from this expansion. Video clarity level from therapists was found to have the same influence on the interpersonal communication responses from therapists as it did on the physical therapy evaluation scores. Interestingly, the spatial ability level of patients did not influence the interpersonal communication responses from therapists whereas their perceived motivation level did. No significant effect was found for both patient responses to physical therapy and interpersonal communication as dependent variables.

Additional Fixed Effect (Corresponding Section)	Fixed Effect	Therapist Physical Therapy	Patient Physical Therapy	Therapist Interpersonal Communication	Patient Interpersonal Communication
None (5.1.1.)	TP	0.10	0.16	-0.13	-0.03
	Cameras	0.25†	0.00	0.09	-0.16
	TP × Cameras	-0.36†	-0.09	-0.01	0.18
Therapist Video Clarity (5.1.1.1.)	TP	0.40*	0.33	0.38*	0.04
	Cameras	0.20	-0.03	0.01	-0.18
	Therapist Video Clarity	0.17*	0.10	0.30*	0.05
	TP × Cameras	-0.34	-0.08	0.02	0.19
Patient Spatial Ability (5.1.1.2.)	TP	0.04	0.17	-0.16	0.00
	Cameras	0.19	0.01	0.06	-0.14
	Patient Spatial Ability	0.16*	-0.03	0.07	-0.07
	TP × Cameras	-0.27	-0.11	0.04	0.14
Perceived Patient Motivation (5.1.1.3.)	TP	0.20	0.20	0.09	-0.06
	Cameras	0.28†	0.01	0.15	-0.17
	Perceived Patient Motivation	0.23*	0.09	0.50*	-0.05
	TP × Cameras	-0.39†	-0.10	-0.08	0.19
Therapist and Patient Genders (5.1.1.4.)	TP	0.11	0.17	-0.12	-0.02
	Cameras	0.25†	0.00	0.10	-0.15
	Female Therapist	0.03	0.07	0.55	0.03
	Female Patient	-0.09	-0.02	-0.12	-0.17
	TP × Cameras	-0.35†	-0.09	0.01	0.20
Same Gender (5.1.1.4.)	TP	0.08	0.15	-0.15	-0.06
	Cameras	0.25†	0.00	0.09	-0.16
	Same Gender	-0.15	-0.06	-0.11	-0.15
	TP × Cameras	-0.32	-0.08	0.02	0.22
Session Order (5.1.1.5.)	TP	0.10	0.16	-0.13	-0.04
	Cameras	0.24†	0.00	0.09	-0.17
	Session Order	0.01	0.00	0.02	0.01
	TP × Cameras	-0.36†	-0.09	0.00	0.18
Therapist VR Experience (5.1.1.6.)	TP	0.10	0.18	-0.13	-0.02
	Cameras	0.24†	0.01	0.09	-0.16
	Therapist VR Experience	0.26	-0.28	-0.16	-0.35
	TP × Cameras	-0.35†	-0.11	-0.01	0.17
Patient PT Experience (5.1.1.7.)	TP	0.13	0.17	-0.13	-0.02
	Cameras	0.23	0.00	0.09	-0.17
	Patient PT Experience	0.10	0.01	-0.01	0.06
	TP × Cameras	-0.37†	-0.09	-0.01	0.17

Table 5.16: Estimated slopes and their p-values of the fixed effects from the linear mixed models as the expansion of Section 5.1.1 to all four dependent variables. Each column represents a dependent variable. (*: $p < 0.05$, †: $p < 0.10$)

5.2 Open-ended Responses

5.2.1 Therapist Interviews

After participating in up to eight sessions in a row with a different patient for each session, therapists were interviewed by the experimenter. The interview questions focused on the comparison between the conditions and were audio-recorded. Each therapist experienced all four experimental conditions before their interviews.

The most common issue of TP the therapists mentioned is that they could not see themselves. This has been mentioned by all therapists (#1, #2, #3, #4, #5, #6, #7, #8, #9, #10, #11). Therapist #10 said, "[I]f I was asked to do like a completely new exercise, it would be very hard." Therapist #11 said, "I will say that, surprisingly, easier to like, show the exercises without the VR because I could also see myself in the zoom."

Another very popular issue mentioned by 10 therapists on TP was pixelation (#1, #2, #3, #4, #6, #7, #8, #9, #10, #11). These comments were toward the insufficient resolution of the depth pixels, which was an issue they have more severely experienced than usual users of the TP system would have since the resolution was downsampled into half for both width and height to get low network latency guaranteed across the whole study. Therapist #3 said, "around the knees, like I can really only see like a knee cap." Therapist #6 said, "one thing that I noticed when we are in the pixelation was you like can't that felt weird to me when she can't see their face facial expression or anything."

Two therapists (#7, #8) compared the issues of not being able to see themselves and pixelation. Both therapists found pixelation a larger issue than not being able to see themselves. Therapist #8 said, "I think the pixelation was more frustrating. Because I just couldn't really fully see them. And like, what all their different

positions."

Four therapists (#4, #7, #8, #10) said VC was better than TP. For example, Therapist #4 said " "And I will say that with a VR, it definitely was harder. Like with it being pixelated, it was harder from like, my mindset of like a therapist to pick out the things that I want to fix. Because it just wasn't quite as clear to be like, oh, like, I can't tell if I need to have them move their feet a little bit further apart, or if that like positioning wise, or if that's what I'm seeing or what they're actually doing."

Five therapists said TP2 is better than TP1 (#2, #4, #6, #8, #9) and three mentioned they are similar (#1, #10, #11). Describing TP2 as more preferred, therapist #6 said, "I did notice the two cameras system. One, even though we're still having them turn to their side it was it was a clear image." Finding little difference between TP2 and TP1, Therapist #1 said, "To be honest, I didn't really notice that much of a difference."

Seven therapists (#1, #3, #4, #5, #6, #9, #11) mentioned that they preferred VC2 over VC1, and two therapists (#2, #11) said they prefer VC2 the most out of all conditions. One of the therapists who showed preference to VC2 (#11) mentioned VC2 felt more impersonal than VC1. Therapist #3, mentioning their preference of VC2 over VC1, said, "Like when I was first doing the two cameras, I was almost exclusively looking at head on camera. So I feel like I was under utilizing the second camera. Especially with the static movements. Once we were getting into like the squat and lunge and stuff like that. It was kind of nice to have that second camera to see the signing, which can be adjusted. You can just make the patient rotate. Yeah, exactly. But you can't see real time. So I think it was helpful to have the second camera but maybe not like essential." Describing VC2 as impersonal, Therapist #11 said, "I also think the one felt more impersonal, surprisingly. And I don't have a specific reason for that other than the fact that's heating them on to just felt a little

bit more like I was there than seeing them just on like one screen. Yeah,"

Two therapists said they did not find VC2 much better than VC1 (#7, #10). Therapist #7 said, "Um, honestly, I was fine with the one video like the one camera. I don't think the second camera added all that much. Like there was a couple times where, you know, it was helpful to have like ask the participant to go on a couple different positions. But I don't feel like there was that it only took like a couple seconds. And it wasn't didn't seem like too much of a hassle."

Six therapists (#1, #3, #4, #5, #6, #11) said they felt higher social presence in TP. One therapist (#9) said they felt lower social presence in TP due to the headset covering their face. Therapist #5 said, "[Y]ou almost want to like reach towards the patient." Therapist #9, mentioning lower social presence, said, "because of just of the pixels and the fact that I didn't like that half of my face was covered and goggles when I was trying to talk to the patient."

Five therapists (#1, #4, #8, #9, #11) mentioned having higher resolution would improve TP. Therapist #9 said, "I don't know if that's like, like a cost and benefit type thing. But I did think about that. If it could be more high resolution that would be that would be even better."

Five therapists (#1, #3, #5, #6, #7) said VR experience may help using TP. Therapist #3 said, "I think it was a lot more effective towards the end from having practice. But also, yeah, I think I just felt more comfortable with the VR in general."

One therapist (#7) said the physical therapy questions were not ideal since many patients already knew the exercises. The therapist said, "I think just a, maybe just a broad comment about the survey. It did seem like a lot of the participants came in with a pretty good understanding of mostly exercises. And so the question asking how, how, like, well, they, they learned the exercise didn't seem to be very representative of the situation because I like if they already came in knowing the exercise, there wasn't a great way to answer like, how, how, like, well, or quickly, they learned it."

5.2.2 Patient Open-ended Responses

Patients were asked to leave additional comments, if they have any, at the end of their post-questionnaire. As each patient only experienced one condition, their comments did not include the comparison between conditions.

Nineteen out of the 76 patients pointed out the tablet they used for watching the therapist was too small or was placed too low. Two out of the 76 patients mentioned the audio quality between them and their therapists was poor.

Chapter 6

Discussion

None of the eight hypotheses were statistically significant. Consequently, in this dissertation, RQ1 has been deeply explored regarding which additional variables would affect the statistical tests conducted for the hypotheses. The main dependent variable for this analysis was the evaluations of exercise sessions by the therapists and Section 5.1.1 examines the effect of individual differences from therapists on evaluation scores. In this linear mixed model that has been used as the starting point for other analyses, a positive marginally significant effect of using two cameras and a negative marginally significant effect from the interaction between TP and using two cameras were found. In other words, the condition using two cameras for a VC system produced better patient outcomes compared to the other three conditions. Consequently, in all following linear mixed models used for analyses, individual differences from therapists were controlled by setting therapist identity as a random effect. Individual differences from patients were controlled by conducting the analyses on the session-level, by taking the average score across the six exercises of a session, not on the exercise-level.

In Section 5.1.1.1, with a linear mixed model with video clarity as an additional fixed effect, it was found out that using TP had a significant positive effect when the level of video clarity was controlled. Also, a positive significant effect from video clarity itself was found. This finding resonates with the prediction made from Section 2.1.

In Section 5.1.1.2, spatial ability of patients was a strong predictor of how ther-

apists evaluated the sessions with the patients. Higher spatial ability measured by the questions predicted higher evaluation scores in a statistically significant manner. Surprisingly, being a strong predictor, adding spatial ability of patients turned the marginally significant effects from using two cameras and the interaction between TP and using two cameras into statistically insignificant.

In Section 5.1.1.3, it has been found that perceived patients' motivation level reported by therapists predicts the evaluation scores from the therapists. Given both variables were measured by the therapists at the same time during post-questionnaires ($\rho = 0.41$ between perceived motivation level and session evaluation), this finding can be considered less of a contribution than patients' spatial ability, which is clearly a separate construct from performance.

In the later parts of Section 5.1.1, the gender of participants, the order of patients and exercises, prior VR experience of therapists, and prior physical therapy experience of patients were examined via adding them as additional fixed effects to linear mixed models. None of them showed a statistically significant effect to the evaluation scores on exercises from the therapists.

These analyses using linear mixed models with additional fixed effects were expanded to other dependent variables in Section 5.1.2. In this section, with interpersonal communication responses from therapists, the influences on physical therapy evaluations were partially replicated. With no additional fixed effects, neither of the experimental conditions had a significant effect on the interpersonal communication responses.

When video clarity was added as an additional fixed effect to the model on interpersonal communication responses from therapists, not only video clarity itself but also TP showed a significant positive effect. This finding, again, resonates with the prediction made from Section 2.1.

With spatial ability as an additional fixed effect to the model on interpersonal

communication responses from therapists, unlike the model on physical therapy session evaluations, there was no significant effect from spatial ability of patients to interpersonal communication responses from therapists. There are two possible explanations of spatial ability improving the session evaluations. One is that both are coming from the general ability of individuals and the other is that both are actually relevant to each other. This lack of an effect from spatial ability to interpersonal communication responses supports the latter explanation that spatial ability and physical therapy session evaluations are actually relevant to each other.

In the model with perceived patient motivation level added to the model as a fixed model, the perceived motivation level showed a significant effect on interpersonal communication response from the therapists. The effect size from perceived motivation level was larger to the interpersonal communication responses than to physical therapy evaluations.

Linear mixed models on interpersonal communication responses from therapists with gender, order of sessions, prior VR experience of therapists, or prior physical therapy experience as additional fixed effects were also tested. There was no statistically significant effect found.

None of the fixed effects across the linear mixed models had a significant effect to responses from patients on their physical therapy or interpersonal communication responses. See Table 6.1 to see a summary of the expanded analysis that took place in Section 5.1.2.

After the sessions, both the therapists and patients were asked their opinions but in different formats. Therapists were interviewed while the audio was being recorded. Patients were asked to leave open-ended comments at the end of their post-questionnaire. As a result of having different approaches in asking their opinions and only the therapists having the cross-condition experience, the opinions coming from the therapists were much more informative than the opinions of the patients.

Additional Fixed Effect	Therapist Physical Therapy	Patient Physical Therapy	Therapist Interpersonal Communication	Patient Interpersonal Communication
Only Conditions	∘			
Video Clarity	∘		∘	
Spatial Ability	∘			
Patient Motivation	∘		∘	

Table 6.1: Summary of the extended analyses with linear mixed models having additional fixed effects. The linear mixed models are marked with ∘ if the additional fixed effects were found statistically significant. For the only conditions row, the cell is marked ∘ if the experimental conditions have resulted in a marginally significant effect.

To summarize the therapist interviews, most of them found the low resolution (10 out of 11 therapists) and the inability to see themselves (all 11 therapists) in TP systems an issue. Four therapists stated they prefer TP over VC. Five therapists said they prefer using two cameras for TP having one camera and seven therapists said they prefer having the additional camera for VC. In their open-ended responses, 19 out of the 76 patients left found the size or position of the tablet they used for seeing the therapist an issue.

The preference to use a VC system with two cameras that the therapists mentioned during their interviews matches the direction of the results from the linear mixed model of Section 5.1.1 from the experimental conditions on session evaluations from the therapists. However, the preference was shown stronger in the interviews than found in the quantitative analyses. Also, while VC did not outperform TP in our quantitative analyses, in the interviews, many therapists preferred VC over TP. The interview responses pointing out low resolution of TP as an issue agrees with the influence of video clarity. As adding video clarity level to linear mixed models made TP to have positive effects, in the interviews, therapists mentioned that increasing resolution of TP will improve user experience.

Using two cameras with the TP system did not result in positive outcomes com-

Figure 6.1: Visualization of regions the independent rater counted the therapists as watching the patients from the side camera.

pared to using one camera for TP, unlike the prior expectations. To better understand the reason, based on observations from experimenters, an independent rater was asked to watch video recordings of the TP sessions with two cameras and report the ratio the therapists using TP with two cameras were watching the patients using the side camera. In other words, the rater checked whether the therapists using two cameras for TP actually took the advantage of having the additional camera. The independent rater reported that the therapist was watching from the side 13.55% of the time during the sessions. During seven sessions out of the eighteen TP sessions with two cameras, therapists utilized the side camera less than 10% of the time during the sessions. See Figure 6.1 to see a visualization of which regions the rater counted as where the therapists were able to see the patients using the side camera. If the therapists stayed in the colored regions in video recordings, the rater counted the therapists as relying on the side camera.

6.1 Limitations

The video resolution of Telegie was low. Due to the networking performance of Telegie, which was less than ideal, I had to downsample both the color and depth pixels captured by the Azure Kinect device. This was mainly due to the usage of WebRTC, and to be more specific the data channels of WebRTC for sending packets from cameras to headsets. Data channels are not built for real-time video communication, but were still adopted for supporting Telegie through web technology. This limitation can be addressed by building Telegie as a native application for the headsets and writing the networking module of Telegie to directly use UDP. Data channels are implemented on top of SCTP, which are usually implemented on top of UDP.

Therapists were not able to see themselves in the TP condition. This has been the case since setting up an RGBD camera also in front of therapists for the study was found unrealistic due to the complexity of the setup process. In the future, by making the Telegie system easier to use, especially for setting up, letting therapists see themselves should be realistic. This would let patients also see the therapists and themselves by wearing VR headsets.

The sample size of the study could have been larger. Due to the difficulty of setting up a remote dyadic study across different time zones with one in the dyad needing to be a physical therapist, the sample size of the study is lower than ideal. With a larger sample size, for example, the marginally significant effects in Section 5.1.1 could have been analyzed in a clearer way.

6.2 Future Directions

In this study, due to the difficulty of setting up the TP system, we did not find a bidirectional TP condition realistic. As a result, only the therapists wore VR headsets.

The setup of the TP system has space to be made easier. For example, calibration of the system when working with multiple cameras should be more automatic.

While we found patients' spatial ability to predict their task performance evaluated by therapists, we did not measure other abilities of the patient that are not spatial ability. Other types of abilities should be measured to tell whether it was the spatial ability in particular that predicted the evaluation scores from therapists.

Measuring the impact of providing sufficient TP experience to the therapists can also be seen as a direction to explore. In our study, most of the therapists were not only new to the TP system, many of them did not have prior experience using VR. As this might be a potential reason why the expected hypotheses were not supported, as the expectations were made by researchers with ample VR experience, the effect of having VR and TP experience can be seen as an intriguing direction for future research.

The evaluation scores from the therapists likely suffered from a ceiling effect, having their average scores above 4 out of 5. In retrospect, the therapists should have been asked to be more critical when evaluating the patients or on a larger scale range that would allow for the therapists to further differentiate the sessions. Alternatively, a more difficult set of exercises should have been chosen.

6.3 Implications for Theories and Practices

In Section 2.1, we have pointed out five psychological mechanisms: agency, stereoscopy, fidelity, comfort, cognitive load. From this study, the most obvious influence was found from fidelity, which corresponds to video clarity levels reported from the participants. With video clarity level controlled across the experimental variables, TP had a positive effect, while there was no positive effect without such control. Also, 10 out of the 11 therapists mentioned low resolution as an issue of TP in their interviews.

From these observations, it is clear that fidelity is a factor that TP systems should not take lightly.

This issue that the TP system faced with fidelity is dismissed in the Media Richness theory (Daft & Lengel, 1986) as the theory only looks for which channels are available without looking at the quality of the channels. While this is understandable as a theory, whose goal is to abstract the real world and extract patterns out of it, for developers of TP systems, this would be a cautionary tale for applying abstract media theories. While having additional spatial information from TP was considered valuable by therapists in their interviews and was found as valuable in linear mixed models with video clarity level added as a fixed effect, due to its lack of video clarity compared to VC, therapists did not find the experience of using TP better than VC. This can be compared to what would happen if a VC system has worse audio quality than a telephone system. When aiming for improvement by providing a richer medium, the richer medium should maintain the qualities of the existing communication channels while adding a new one.

Based on the Social Influence model (Blascovich, 2002), the lower interpersonal and physical therapy outcomes from lower fidelity can be seen as results of low photographic realism leading to the virtual humans of patients not providing enough social influence. Virtual humans rendered by systems with lower video fidelity provide lower photographic realism. With this lower photographic realism, according to the Social Influence model, virtual humans are less likely to cause social influence to people facing the virtual humans. Given interpersonal communication in general or especially physical therapy instruction are forms of interaction that require social influence, lower video clarity leads to lower interpersonal or physical therapy outcomes.

Notice that Blascovich (2002) found behavioral realism more important than photographic realism saying, "photographic realism does not equate with behavioral realism and is, in fact, less important" (p. 131). Given this, one may say fidelity should

not have mattered in our study according to the Social Influence theory. However, in this specific case, the fidelity was low enough that it did affect behavioral realism of the patients from the therapists side. This is mentioned in the interviews from therapists who said the pixelation of TP did not allow them to see the facial expressions of the patients or could only see the kneecap of the whole leg. Perhaps the goal in terms of visual fidelity for TP systems should be having enough fidelity to not harm behavioral realism.

Agency was likely another mechanism that has influenced the participants given the results from the linear mixed models with the video clarity levels controlled. In this case, using TP was reported as having positive effects for both physical therapy and interpersonal communication responses from the therapists, and this is likely due to enhanced levels of agency coming from the TP system as the system allowed the therapists to move around inside the virtual environment with the patients inside. From the perspective of the Social Influence model, by having agency in terms of choosing which perspective to see virtual humans, the viewer may become more certain about the behavioral and anthropomorphic realisms of the virtual humans as the viewer can verify, for example, the behavior of the virtual human from more angles. By further verification, it is possible that this viewer perceives higher behavioral realism and this leads to higher social influence from the virtual human. And this social influence can lead to, for example, more positive responses to interpersonal communication responses.

Comfort and cognitive load did not play a large role in this study as no therapists mentioned them. None of the therapists reported discomfort of using the VR headsets or the cognitive load required for interpreting information coming from two cameras.

As a system for remote physical therapy, the goal should be having enough resolution to capture facial expressions and exact positions of the limbs. Another lesson from the interviews is the importance of letting the TP users see themselves. This

suggests usage of augmented reality technology for improvement of TP systems.

Chapter 7

Conclusion

In this dissertation, I have examined the effect of using a TP system compared to using a VC system and using two cameras to better capture the patient for remote physical therapy. A large group of participants participated given the study required the presence of a physical therapist and two sites synchronously prepared to run the study. There were 76 patients and 11 therapists.

Telegie, the TP system used for the study, has been introduced in this dissertation. The system was designed based on a set of design criteria proposed based on previously introduced TP systems. The system operates with commodity hardware and does not require experts for its operation.

Based on previous literature, positive effects were expected for using the TP and two cameras for remote physical therapy. Between the two interventions, using TP was expected to have a larger positive effect than using two cameras. However, the hypotheses based on such expectations were not supported in the context of both physical therapy and interpersonal communication.

In a deeper look using linear mixed models to control variables besides the experimental conditions, many effects between the quantitative reports were discovered. With the individual differences between the therapists and patients controlled, the condition using the VC system with two cameras was found to be better than other experimental conditions. With reported levels of video clarity controlled, using the TP system was found as better than using the VC system. Surprisingly, patients' spatial ability level performed very well as a predictor of physical therapy session

evaluations from the therapists. Perceived motivation levels of the patients reported by therapists significantly explained physical therapy evaluations from therapists. Gender of participants, order of patients and exercises, and prior VR and physical therapy experience of participants did not explain physical therapy evaluations from therapists.

In the extension of analyses on therapists' physical therapy evaluations to other dependent variables, with video clarity levels controlled, using TP had a positive effect also on patients' responses on interpersonal communication quality. Patients' spatial ability did not improve interpersonal communication, which can be interpreted as in the right direction as spatial ability must be more relevant to performance as a physical therapy session than as a communicating individual. Perceived patient motivation level reported by therapists also had a positive effect on interpersonal communication level reported by therapists as it did to physical therapy evaluations.

In their interviews, therapists have pointed out technical challenges for TP systems. They suggested having better resolutions that would allow users to see facial expressions and other body movements more clearly. They have also suggested having a mechanism that allows the users to see themselves while they are wearing VR headsets.

References

Aron, A., Aron, E. N., & Smollan, D. (1992). Inclusion of other in the self scale and the structure of interpersonal closeness. *Journal of personality and social psychology*, *63*(4), 596.

Ayres, R. U. (2021). Electronic broadcast media: Radio and tv. In *The history and future of technology* (pp. 367–396). Springer.

Bailenson, J., Patel, K., Nielsen, A., Bajscy, R., Jung, S.-H., & Kurillo, G. (2008). The effect of interactivity on learning physical actions in virtual reality. *Media Psychology*, *11*(3), 354–376.

Bailenson, J. N., Swinth, K., Hoyt, C., Persky, S., Dimov, A., & Blascovich, J. (2005). The independent and interactive effects of embodied-agent appearance and behavior on self-report, cognitive, and behavioral markers of copresence in immersive virtual environments. *Presence*, *14*(4), 379–393.

Baños, R. M., Botella, C., Rubió, I., Quero, S., García-Palacios, A., & Alcañiz, M. (2008). Presence and emotions in virtual environments: The influence of stereoscopy. *CyberPsychology & Behavior*, *11*(1), 1–8.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.

Beck, S., Kunert, A., Kulik, A., & Froehlich, B. (2013). Immersive group-to-group telepresence. *IEEE transactions on visualization and computer graphics*, *19*(4), 616–625.

Bertrand, J., Dukes, L. C., Dukes, P., Ebrahimi, E., Hayes, A., Mack, N., ... others (2013). Serious games for training, rehabilitation and workforce development. In *2013 ieee virtual reality (vr)* (pp. 195–196).

Billinghurst, M., Poupyrev, I., Kato, H., & May, R. (2000). Mixing realities in shared space: An augmented reality interface for collaborative computing. In *2000 ieee*

international conference on multimedia and expo. icme2000. proceedings. latest advances in the fast changing world of multimedia (cat. no. 00th8532) (Vol. 3, pp. 1641–1644).

Billinghurst, M., Weghorst, S., & Furness, T. (1998). Shared space: An augmented reality approach for computer supported collaborative work. *Virtual Reality*, *3*(1), 25–36.

Blascovich, J. (2002). Social influence within immersive virtual environments. In *The social life of avatars* (pp. 127–145). Springer.

Bolle, S. R., Larsen, F., Hagen, O., & Gilbert, M. (2009). Video conferencing versus telephone calls for team work across hospitals: a qualitative study on simulated emergencies. *BMC Emergency Medicine*, *9*(1), 1–8.

Bracken, C. C., & Skalski, P. (2009). Telepresence and video games: The impact of image quality. *PsychNology Journal*, *7*(1).

Bystrom, K.-E., & Barfield, W. (1999). Collaborative task performance for learning using a virtual environment. *Presence*, *8*(4), 435–448.

Camporesi, C., Kallmann, M., & Han, J. J. (2013). Vr solutions for improving physical therapy. In *2013 ieee virtual reality (vr)* (pp. 77–78).

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and instruction*, *8*(4), 293–332.

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: the perception–behavior link and social interaction. *Journal of personality and social psychology*, *76*(6), 893.

Cummings, J. J., & Bailenson, J. N. (2016). How immersive is enough? a meta-analysis of the effect of immersive technology on user presence. *Media psychology*, *19*(2), 272–309.

Daft, R. L., & Lengel, R. H. (1986). Organizational information requirements, media richness and structural design. *Management science*, *32*(5), 554–571.

Dennis, A. R., & Kinney, S. T. (1998). Testing media richness theory in the new media: The effects of cues, feedback, and task equivocality. *Information systems research*, *9*(3), 256–274.

Duarte, M., & Freitas, S. M. (2005). Speed–accuracy trade-off in voluntary postural movements. *Motor control*, *9*(2), 180–196.

Eaves, D. L., Breslin, G., Van Schaik, P., Robinson, E., & Spears, I. R. (2011). The short-term effects of real-time virtual reality feedback on motor learning in dance. *Presence: Teleoperators and Virtual Environments*, *20*(1), 62–77.

Fender, & Holz, C. (2022). Causality-preserving asynchronous reality. In *Chi conference on human factors in computing systems* (pp. 1–15).

Fender, & Müller, J. (2018). Velt: A framework for multi rgb-d camera systems. In *Proceedings of the 2018 acm international conference on interactive surfaces and spaces* (pp. 73–83).

Feng, H., Li, C., Liu, J., Wang, L., Ma, J., Li, G., ... Wu, Z. (2019). Virtual reality rehabilitation versus conventional physical therapy for improving balance and gait in parkinson's disease patients: a randomized controlled trial. *Medical science monitor: international medical journal of experimental and clinical research*, *25*, 4186.

Gamelin, G., Chellali, A., Cheikh, S., Ricca, A., Dumas, C., & Otmane, S. (2021). Point-cloud avatars to improve spatial communication in immersive collaborative virtual environments. *Personal and Ubiquitous Computing*, *25*(3), 467–484.

Golomb, M. R., Barkat-Masih, M., Rabin, B., Abdelbaky, M., Huber, M., & Burdea, G. (2009). Eleven months of home virtual reality telerehabilitation-lessons learned. In *2009 virtual rehabilitation international conference* (pp. 23–28).

Hayduk, L. A. (1983). Personal space: where we now stand. *Psychological bulletin*, *94*(2), 293.

Herrera, F., Oh, S. Y., & Bailenson, J. N. (2020). Effect of behavioral realism on

social interactions inside collaborative virtual environments. *Presence*, *27*(2), 163–182.

Jones, B., Sodhi, R., Murdock, M., Mehra, R., Benko, H., Wilson, A., ... Shapira, L. (2014). Roomalive: Magical experiences enabled by scalable, adaptive projector-camera units. In *Proceedings of the 27th annual acm symposium on user interface software and technology* (pp. 637–644).

Jones, B., Zhang, Y., Wong, P. N., & Rintel, S. (2020). Vroom: virtual robot overlay for online meetings. In *Extended abstracts of the 2020 chi conference on human factors in computing systems* (pp. 1–10).

Jun, H., & Bailenson, J. (2020). Temporal rvl: a depth stream compression method. In *2020 ieee conference on virtual reality and 3d user interfaces abstracts and workshops (vrw)* (pp. 664–665).

Jun, H., Bailenson, J. N., Fuchs, H., & Wetzstein, G. (2018). An easy-to-use pipeline for an rgbd camera and an ar headset. *PRESENCE: Virtual and Augmented Reality*, *27*(2), 202–205.

Kahai, S. S., & Cooper, R. B. (2003). 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.

Kallmann, M., Camporesi, C., & Han, J. (2015). Vr-assisted physical rehabilitation: Adapting to the needs of therapists and patients. In *Virtual realities* (pp. 147–168). Springer.

Kim, K. J., & Sundar, S. S. (2013). Can interface features affect aggression resulting from violent video game play? an examination of realistic controller and large screen size. *Cyberpsychology, Behavior, and Social Networking*, *16*(5), 329–334.

Kolkmeier, J., Harmsen, E., Giesselink, S., Reidsma, D., Theune, M., & Heylen, D. (2018). With a little help from a holographic friend: The openimpress mixed reality telepresence toolkit for remote collaboration systems. In *Proceedings of*

the 24th acm symposium on virtual reality software and technology (pp. 1–11).

Kowalski, M., Naruniec, J., & Daniluk, M. (2015). Livescan3d: A fast and inexpensive 3d data acquisition system for multiple kinect v2 sensors. In *2015 international conference on 3d vision* (pp. 318–325).

Kraut, R. E., & Fish, R. S. (1995). Prospects for video telephony. *Telecommunications Policy*, *19*(9), 699–719.

Kydd, C. T., & Ferry, D. L. (1994). Managerial use of video conferencing. *Information & Management*, *27*(6), 369–375.

Lange, B., Koenig, S., McConnell, E., Chang, C.-Y., Juang, R., Suma, E., ... Rizzo, A. (2012). Interactive game-based rehabilitation using the microsoft kinect. In *Virtual reality conference, ieee* (pp. 171–172).

Lasswell, H. D. (1948). The structure and function of communication in society. *The communication of ideas*, *37*(1), 136–139.

Lee, M., Bruder, G., Höllerer, T., & Welch, G. (2018). Effects of unaugmented periphery and vibrotactile feedback on proxemics with virtual humans in ar. *IEEE transactions on visualization and computer graphics*, *24*(4), 1525–1534.

Ling, Y., Brinkman, W.-P., Nefs, H. T., Qu, C., & Heynderickx, I. (2012). Effects of stereoscopic viewing on presence, anxiety, and cybersickness in a virtual reality environment for public speaking. *Presence: Teleoperators and Virtual Environments*, *21*(3), 254–267.

Lok, B., Naik, S., Whitton, M., & Brooks, F. P. (2003). Effects of handling real objects and self-avatar fidelity on cognitive task performance and sense of presence in virtual environments. *Presence*, *12*(6), 615–628.

MacLean, A., Young, R. M., Bellotti, V. M., & Moran, T. P. (1991). Questions, options, and criteria: elements of design space analysis. *Human-Computer Interaction*, *6*(3), 201–250.

Maimone, A., & Fuchs, H. (2011). Encumbrance-free telepresence system with real-

time 3d capture and display using commodity depth cameras. In *2011 10th ieee international symposium on mixed and augmented reality* (pp. 137–146).

Maimone, A., Yang, X., Dierk, N., State, A., Dou, M., & Fuchs, H. (2013). General-purpose telepresence with head-worn optical see-through displays and projector-based lighting. In *2013 ieee virtual reality (vr)* (pp. 23–26).

Markowitz, D. M., Laha, R., Perone, B. P., Pea, R. D., & Bailenson, J. N. (2018). Immersive virtual reality field trips facilitate learning about climate change. *Frontiers in psychology*, *9*, 2364.

McLuhan, M. (1964). *Understanding media: The extensions of man*.

Miller, M. R., Jun, H., Herrera, F., Yu Villa, J., Welch, G., & Bailenson, J. N. (2019). Social interaction in augmented reality. *PloS one*, *14*(5), e0216290.

Minsky, M. (1980). Telepresence.

Mishra, A. K., Skubic, M., & Abbott, C. (2015). Development and preliminary validation of an interactive remote physical therapy system. In *2015 37th annual international conference of the ieee engineering in medicine and biology society (embc)* (pp. 190–193).

Mogan, R., Fischer, R., & Bulbulia, J. A. (2017). To be in synchrony or not? a meta-analysis of synchrony's effects on behavior, perception, cognition and affect. *Journal of Experimental Social Psychology*, *72*, 13–20.

Murtza, R., Monroe, S., & Youmans, R. J. (2017). Heuristic evaluation for virtual reality systems. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 61, pp. 2067–2071).

Nguyen, D., & Canny, J. (2005). Multiview: spatially faithful group video conferencing. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 799–808).

Oh, C., Herrera, F., & Bailenson, J. (2019). The effects of immersion and real-world distractions on virtual social interactions. *Cyberpsychology, Behavior,*

and Social Networking, *22*(6), 365–372.

Ophir, E., Nass, C., & Wagner, A. D. (2009). Cognitive control in media multitaskers. *Proceedings of the National Academy of Sciences*, *106*(37), 15583–15587.

Orts-Escolano, S., Rhemann, C., Fanello, S., Chang, W., Kowdle, A., Degtyarev, Y., ... others (2016). Holoportation: Virtual 3d teleportation in real-time. In *Proceedings of the 29th annual symposium on user interface software and technology* (pp. 741–754).

Pejsa, T., Kantor, J., Benko, H., Ofek, E., & Wilson, A. (2016). Room2room: Enabling life-size telepresence in a projected augmented reality environment. In *Proceedings of the 19th acm conference on computer-supported cooperative work & social computing* (pp. 1716–1725).

Peters, M., Laeng, B., Latham, K., Jackson, M., Zaiyouna, R., & Richardson, C. (1995). A redrawn vandenberg and kuse mental rotations test-different versions and factors that affect performance. *Brain and cognition*, *28*(1), 39–58.

Popescu, V. G., Burdea, G. C., Bouzit, M., & Hentz, V. R. (2000). A virtual-reality-based telerehabilitation system with force feedback. *IEEE transactions on Information Technology in Biomedicine*, *4*(1), 45–51.

Postman, N. (1974). Media ecology: Communication as context.

Postolache, O., Hemanth, D. J., Alexandre, R., Gupta, D., Geman, O., & Khanna, A. (2020). Remote monitoring of physical rehabilitation of stroke patients using iot and virtual reality. *IEEE Journal on Selected Areas in Communications*, *39*(2), 562–573.

Rendon, A. A., Lohman, E. B., Thorpe, D., Johnson, E. G., Medina, E., & Bradley, B. (2012). The effect of virtual reality gaming on dynamic balance in older adults. *Age and ageing*, *41*(4), 549–552.

Rhee, T., Thompson, S., Medeiros, D., Dos Anjos, R., & Chalmers, A. (2020). Augmented virtual teleportation for high-fidelity telecollaboration. *IEEE Transac-*

tions on Visualization and Computer Graphics, *26*(5), 1923–1933.

Roberts, D. J., Fairchild, A. J., Campion, S. P., O'Hare, J., Moore, C. M., Aspin, R., ... Tecchia, F. (2015). withyou—an experimental end-to-end telepresence system using video-based reconstruction. *IEEE Journal of Selected Topics in Signal Processing*, *9*(3), 562–574.

Saraee, E., Singh, S., Hendron, K., Zheng, M., Joshi, A., Ellis, T., & Betke, M. (2017). Exercisecheck: remote monitoring and evaluation platform for home based physical therapy. In *Proceedings of the 10th international conference on pervasive technologies related to assistive environments* (pp. 87–90).

Sellen, A. J. (1995). Remote conversations: The effects of mediating talk with technology. *Human-computer interaction*, *10*(4), 401–444.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, *171*(3972), 701–703.

Short, J., Williams, E., & Christie, B. (1976). *The social psychology of telecommunications*. Toronto; London; New York: Wiley.

Slater, M., & Wilbur, S. (1997). A framework for immersive virtual environments (five): Speculations on the role of presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, *6*(6), 603–616.

Smets, G. J., & Overbeeke, K. J. (1995). Trade-off between resolution and interactivity in spatial task performance. *IEEE Computer Graphics and Applications*, *15*(5), 46–51.

Souchet, A. D., Philippe, S., Lévêque, A., Ober, F., & Leroy, L. (2021). Short- and long-term learning of job interview with a serious game in virtual reality: Influence of eyestrain, stereoscopy, and apparatus. *Virtual Reality*, 1–18.

Steed, A., Steptoe, W., Oyekoya, W., Pece, F., Weyrich, T., Kautz, J., ... others (2012). Beaming: an asymmetric telepresence system. *IEEE computer graphics and applications*, *32*(6), 10–17.

Steuer, J. (1992). Defining virtual reality: Dimensions determining telepresence. *Journal of communication*, *42*(4), 73–93.

Suh, K. S. (1999). Impact of communication medium on task performance and satisfaction: an examination of media-richness theory. *Information & Management*, *35*(5), 295–312.

Szalavári, Z., Schmalstieg, D., Fuhrmann, A., & Gervautz, M. (1998). "studierstube": An environment for collaboration in augmented reality. *Virtual Reality*, *3*(1), 37–48.

Trajkova, M., & Ferati, M. (2015). Usability evaluation of kinect-based system for ballet movements. In *International conference of design, user experience, and usability* (pp. 464–472).

Valli, S., Hakkarainen, M., & Siltanen, P. (2021). Advances in spatially faithful (3d) telepresence. In *Augmented reality.* IntechOpen.

Van Cauwenberge, A., Schaap, G., & Van Roy, R. (2014). "tv no longer commands our full attention": Effects of second-screen viewing and task relevance on cognitive load and learning from news. *Computers in Human Behavior*, *38*, 100–109.

Van Schooten, B. W., Van Dijk, E. M., Zudilova-Seinstra, E., Suinesiaputra, A., & Reiber, J. H. (2010). The effect of stereoscopy and motion cues on 3d interpretation task performance. In *Proceedings of the international conference on advanced visual interfaces* (pp. 167–170).

Wang, Z., He, R., & Chen, K. (2020). Thermal comfort and virtual reality headsets. *Applied Ergonomics*, *85*, 103066.

Wiltermuth, S. S., & Heath, C. (2009). Synchrony and cooperation. *Psychological science*, *20*(1), 1–5.

Yan, Y., Chen, K., Xie, Y., Song, Y., & Liu, Y. (2018). The effects of weight on comfort of virtual reality devices. In *International conference on applied human factors and ergonomics* (pp. 239–248).

Yang, U., & Kim, G. J. (2002). Implementation and evaluation of "just follow me": An immersive, vr-based, motion-training system. *Presence*, *11*(3), 304–323.

Zibrek, K., Martin, S., & McDonnell, R. (2019). Is photorealism important for perception of expressive virtual humans in virtual reality? *ACM Transactions on Applied Perception (TAP)*, *16*(3), 1–19.

Zibrek, K., & McDonnell, R. (2019). Social presence and place illusion are affected by photorealism in embodied vr. In *Motion, interaction and games* (pp. 1–7).

Appendix A

Spatial Arrangement

A.1 Introduction to Spatial Arrangement

Spatial arrangement is a missing piece in our understanding of remote communication systems. In a typical face-to-face communication, we can see our communication partners, and they can see us. We know where they are, and they know where we are. If we are in a group conversation, we know where all others are. In other words, we know how the whole group is arranged in space–which I define as spatial arrangement. From another perspective, people in face-to-face conversations can perceive the spatial arrangement in which they are included.

In face-to-face communication, there are two types of spatial arrangements to notice. First, the physical one is how people are positioned in the real world. Second, the perceptual one is the perceived version of the physical one. While there is exactly one physical spatial arrangement for each group conversation, the number of perceptual spatial arrangements matches the number of people in the group.

The relationship between physical and perceptual spatial arrangements is bidirectional. While, apparently, a physical spatial arrangement affects perceptual spatial arrangements, the opposite also happens as people possessing the perceptual spatial arrangements can move and modify the physical spatial arrangement they are inside.

A.2 Spatial Arrangement as a Message

The next step is interpreting spatial arrangements as a type of communication message. Based on Lasswell's model of communication (Lasswell, 1948), speaking and hearing are sending and receiving utterances as messages. In a similar manner, I propose interpreting moving and seeing how others move as sending and receiving spatial arrangements. As the spatial arrangements being sent as messages are coming from people, the spatial arrangements considered as messages are perceptual spatial arrangements. From this perspective, as air is the medium for utterances, the physical spatial arrangement can be seen as the medium for (perceptual) spatial arrangements. From now, when not explicitly mentioned as physical, a spatial arrangement will mean a perceptual spatial arrangement.

The main advantage of this interpretation of spatial arrangements as messages is its extensibility as a thought framework. Since now the physical spatial arrangement is a medium that can be replaced by other media, an analysis of remote communication systems without an equivalent to a physical spatial arrangement can take place. For example, telephone and VC systems both do not include a physical spatial arrangement. While they may fail to send spatial arrangements as messages, still we may analyze at least the failure. An attempt to analyze these remote communication systems in a way that requires an equivalent to a physical spatial arrangement, when there is none, is unlikely to be fruitful.

A.3 Definition of Spatial Arrangement

The current naive definition of spatial arrangement–where other people are–is likely to cause ambiguity when used with other terms, such as nonverbal behavior and gestures. To avoid this, spatial arrangement needs to be better defined.

The spatial arrangement of a group will consist of a position (i.e., x, y, z) and an orientation (i.e., yaw, pitch, roll) value for each person in the group. In terms of degrees of freedom, spatial arrangement will contain 6 degrees of freedom per person. To provide examples, a person walking or rotating their whole body will affect the spatial arrangement, but facial expressions and hand gestures will not.

Appendix B

Pre-questionnaire

Demographic

How old are you?

- Open ended (numeric)

What gender do you identify as?

- Female • Male • Other • Decline to Answer

VR Experience

Do you have any previous experience in virtual reality?

- Yes • No

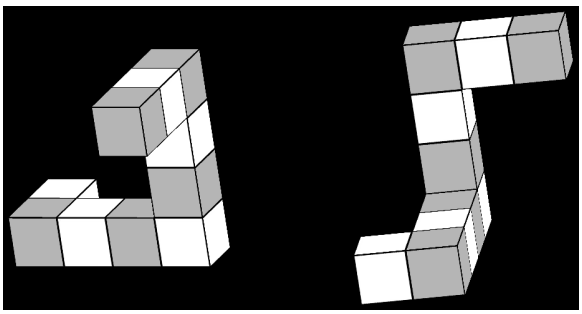
Physical Therapy Experience (Patient Only)

Do you have any previous experience in physical therapy?

- Yes • No

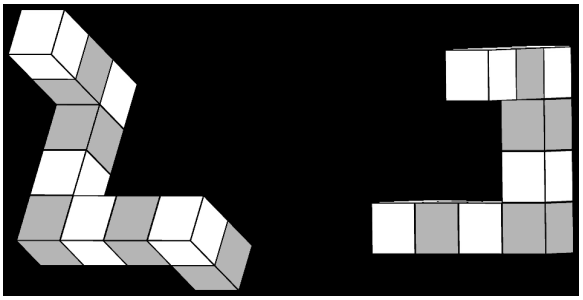
Spatial Ability

Are these two figures the same except for their orientation?



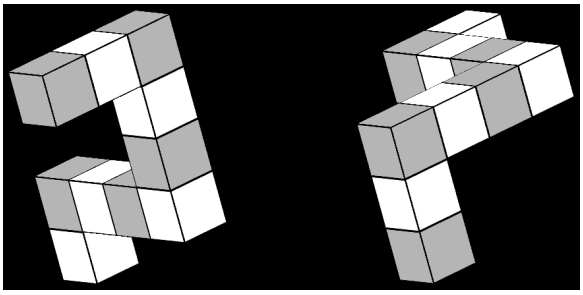
• Yes (Correct Answer) • No

Are these two figures the same except for their orientation?



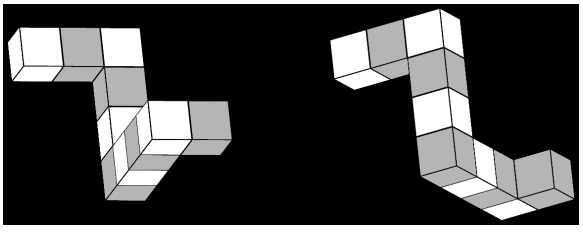
• Yes • No (Correct Answer)

Are these two figures the same except for their orientation?



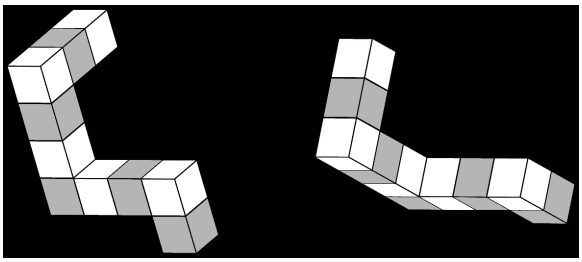
• Yes (Correct Answer) • No

Are these two figures the same except for their orientation?



• Yes • No (Correct Answer)

Are these two figures the same except for their orientation?



• Yes (Correct Answer) • No

Appendix C

Post-questionnaire

Social Presence

How much did you feel like you were face-to-face with your [partner type]?

- Extremely • Very • Moderately • Slightly • Not at All

How much did you feel like you were in the same room as your [parter type]?

- Extremely • Very • Moderately • Slightly • Not at All

How much did you feel like your [partner type] was watching you?

- Extremely • Very • Moderately • Slightly • Not at All

How much did you feel like your [parter type] was aware of your presence?

- Extremely • Very • Moderately • Slightly • Not at All

How much did you feel like your [parter type] was present?

- Extremely • Very • Moderately • Slightly • Not at All

Communication Satisfaction

How much would you like to have another conversational session like this one?

- Extremely • Very • Moderately • Slightly • Not at All

How satisfied were you with the conversation?

- Extremely • Very • Moderately • Slightly • Not at All

How much did you enjoy the conversation?

- Extremely • Very • Moderately • Slightly • Not at All

How much did the conversation flow smoothly?

- Extremely • Very • Moderately • Slightly • Not at All

Interpersonal Liking

How much do you like your [partner type]?

- Extremely • Very • Moderately • Slightly • Not at All

How much would you like to get to know your [partner type] better?

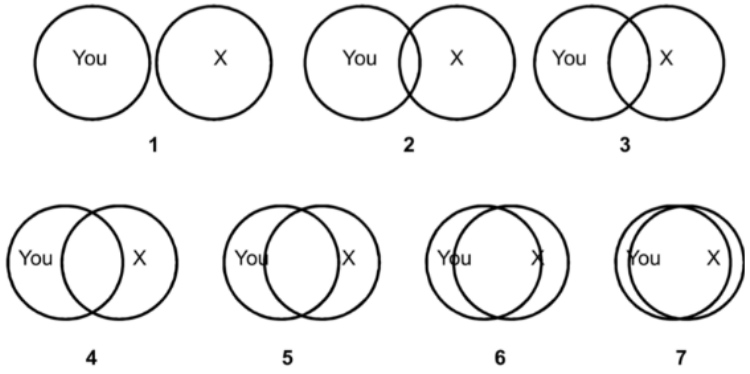
- Extremely • Very • Moderately • Slightly • Not at All

How much do you think your [partner type] would be popular with their friends?

- Extremely • Very • Moderately • Slightly • Not at All

Inclusion of Other in the Self

Which picture below best describes the relationship between you and your [partner type]?



Video Clarity

How clear was the video stream?

- Extremely • Very • Moderately • Slightly • Not at All

Perceived Patient Motivation (Therapist Only)

How motivated was the patient?

- Extremely • Very • Moderately • Slightly • Not at All

Physical Therapy Questionnaire (Therapist Only)

How accurately did the patient learn the exercises?

	Not at All	Slightly	Moderately	Very	Extermely
Lunge					
Bilateral horizontal abduction with elastic band					
Plank					
Bridge upper back with ball					
Side lying external rotation					
Squat					

How quickly did the patient learn the exercises?

	Not at All	Slightly	Moderately	Very	Extermely
Lunge					
Bilateral horizontal abduction with elastic band					
Plank Bridge upper back with ball					
Side lying external rotation					
Squat					

Physical Therapy Questionnaire (Patient Only)

Q1: How much were you aware of the therapist's intentions/wishes in this task?

- Extremely • Very • Moderately • Slightly • Not at All

Q2: How pleasant did you find this task?

- Extremely • Very • Moderately • Slightly • Not at All

Q3: How much did you experience this task as something that you did together/jointly with the instructor?

- Extremely • Very • Moderately • Slightly • Not at All

Q4: How difficult was this task? (Reversed the scores from 1-5 to 5-1 for analysis.)

- Extremely • Very • Moderately • Slightly • Not at All

Q5: How difficult was it to move around in the environment? (Reversed the scores from 1-5 to 5-1 for analysis.)

- Extremely • Very • Moderately • Slightly • Not at All

Q6: How personal was your experience in the learning environment?

- Extremely • Very • Moderately • Slightly • Not at All

Q7: How social was your experience in the learning environment?

- Extremely • Very • Moderately • Slightly • Not at All

Q8: How lively was your experience in the learning environment?

- Extremely • Very • Moderately • Slightly • Not at All

Q9: How pleasant was your experience in the learning environment?

- Extremely • Very • Moderately • Slightly • Not at All

Q10: How much did you find the therapist to be close, not distant?

- Extremely • Very • Moderately • Slightly • Not at All

Q11: How much did you find the therapist to be responsive?

- Extremely • Very • Moderately • Slightly • Not at All

Q12: How much did you find the therapist to be active, not passive?

- Extremely • Very • Moderately • Slightly • Not at All

Q13: How much did you find the therapist to be warm, not cold?

- Extremely • Very • Moderately • Slightly • Not at All

Q14: How much did you find the therapist to be helpful?

- Extremely • Very • Moderately • Slightly • Not at All

Q15: How much did you find the therapist to be realistic, not fake?

- Extremely • Very • Moderately • Slightly • Not at All

Q16: How much did you find the therapist to be an expert, not a novice?

- Extremely • Very • Moderately • Slightly • Not at All

Open-ended Questionnaire (Patient Only)

Is there anything else you would like to mention about this study?

- Open-ended

What do you think the purpose of this study was?

- Open-ended

Appendix D

Interview Questions

An interviewer asked the following three questions to each therapist. Additional follow-up questions were asked between the three questions.

- How do you compare VC to TP?
- How do you compare 1VC to 2VC?
- How do you compare 1TP to 2TP?

Appendix E

Descriptive Statistics Figures

E.1 Interpersonal Communication Responses

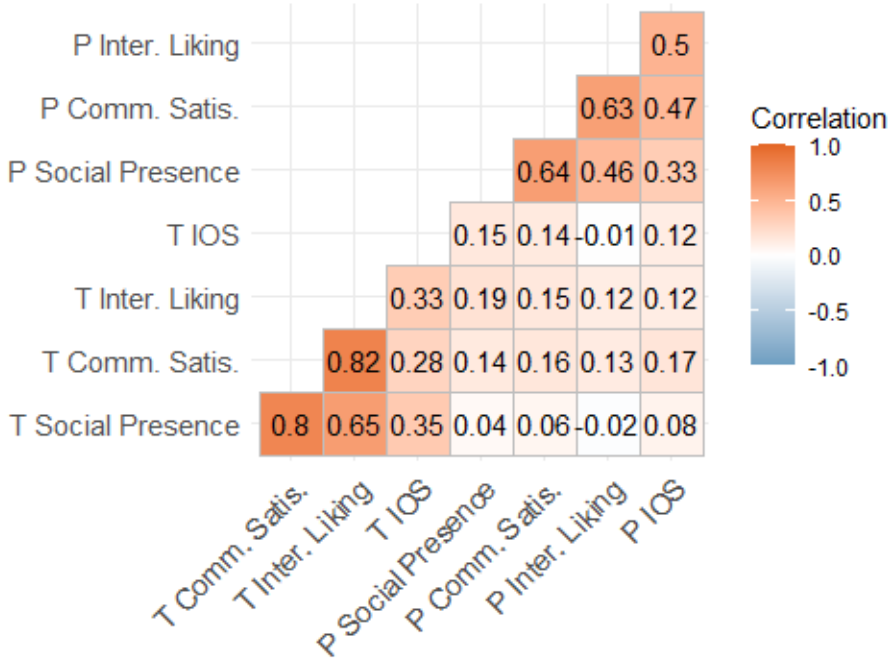
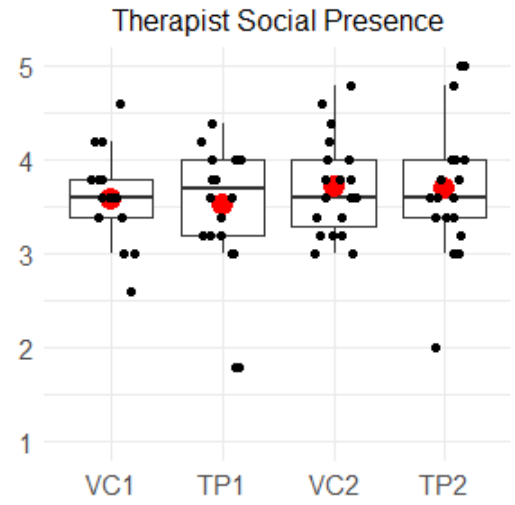
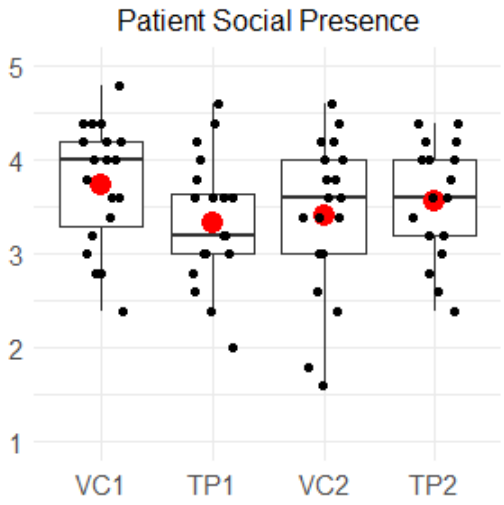
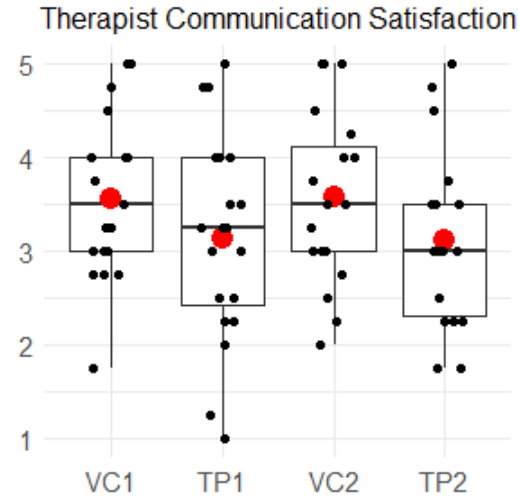
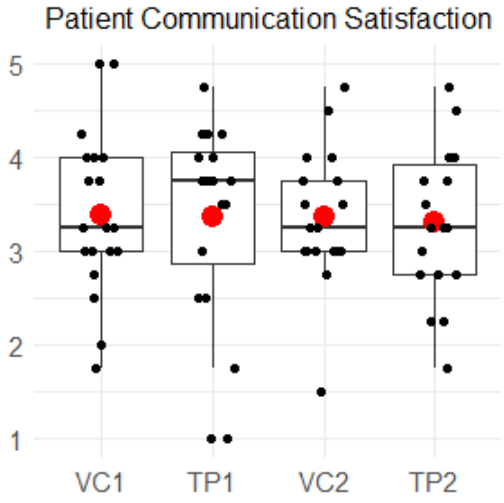


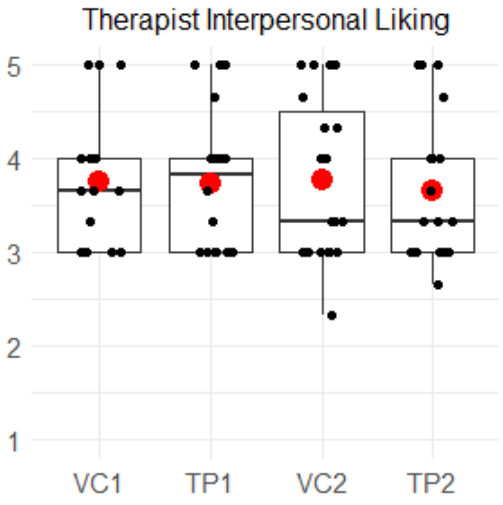
Figure E.1: Correlation coefficients between therapist and patient responses on social presence, communication satisfaction, interpersonal liking, and IOS.

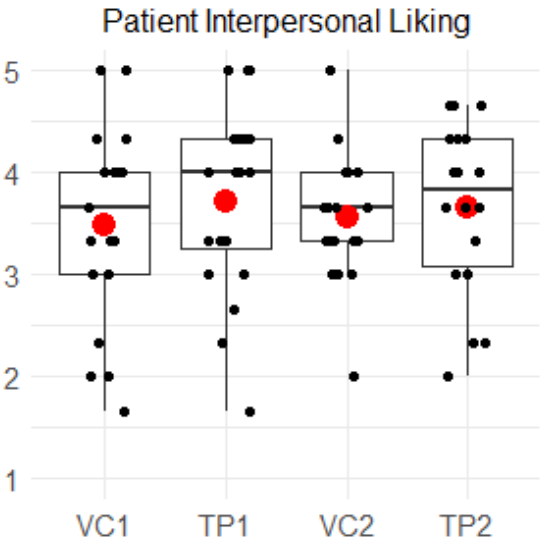

Therapist Social Presence
5
4
3
2
1
VC1
TP1
VC2
TP2


Patient Social Presence
5
4
3
2
1
VC1
TP1
VC2
TP2


Therapist Communication Satisfaction
5
4
3
2
1
VC1
TP1
VC2
TP2


Patient Communication Satisfaction
5
4
3
2
1
VC1
TP1
VC2
TP2


Therapist Interpersonal Liking
5
4
3
2
1
VC1
TP1
VC2
TP2


Patient Interpersonal Liking
5
4
3
2
1
VC1
TP1
VC2
TP2

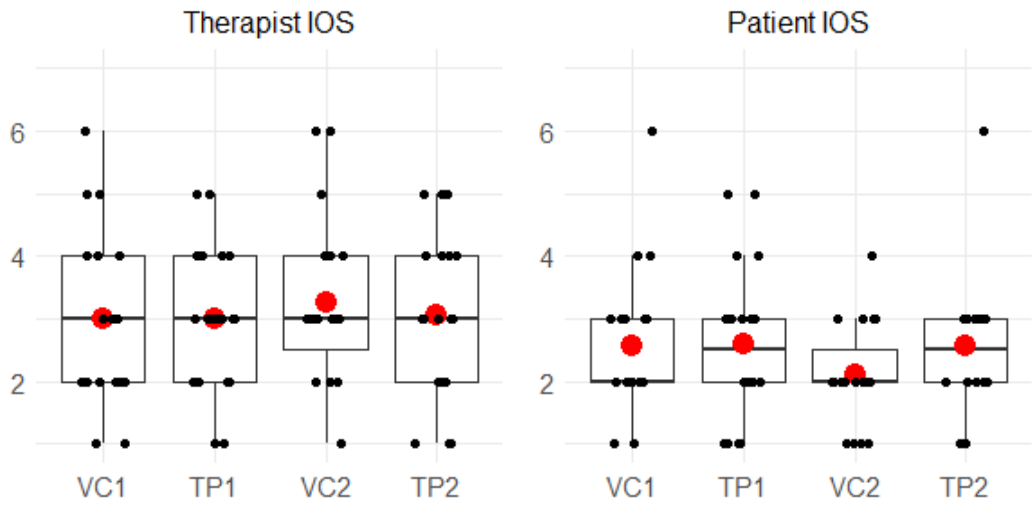


Figure E.2: Distributions of social presence, communication satisfaction, interpersonal liking, and IOS levels of therapists and patients. Each column matches an experimental condition. Each column contains a box plot that visualizes the quartiles, a point for each session, and a larger red dot indicating the mean value.

E.2 Therapist Physical Therapy Evaluations

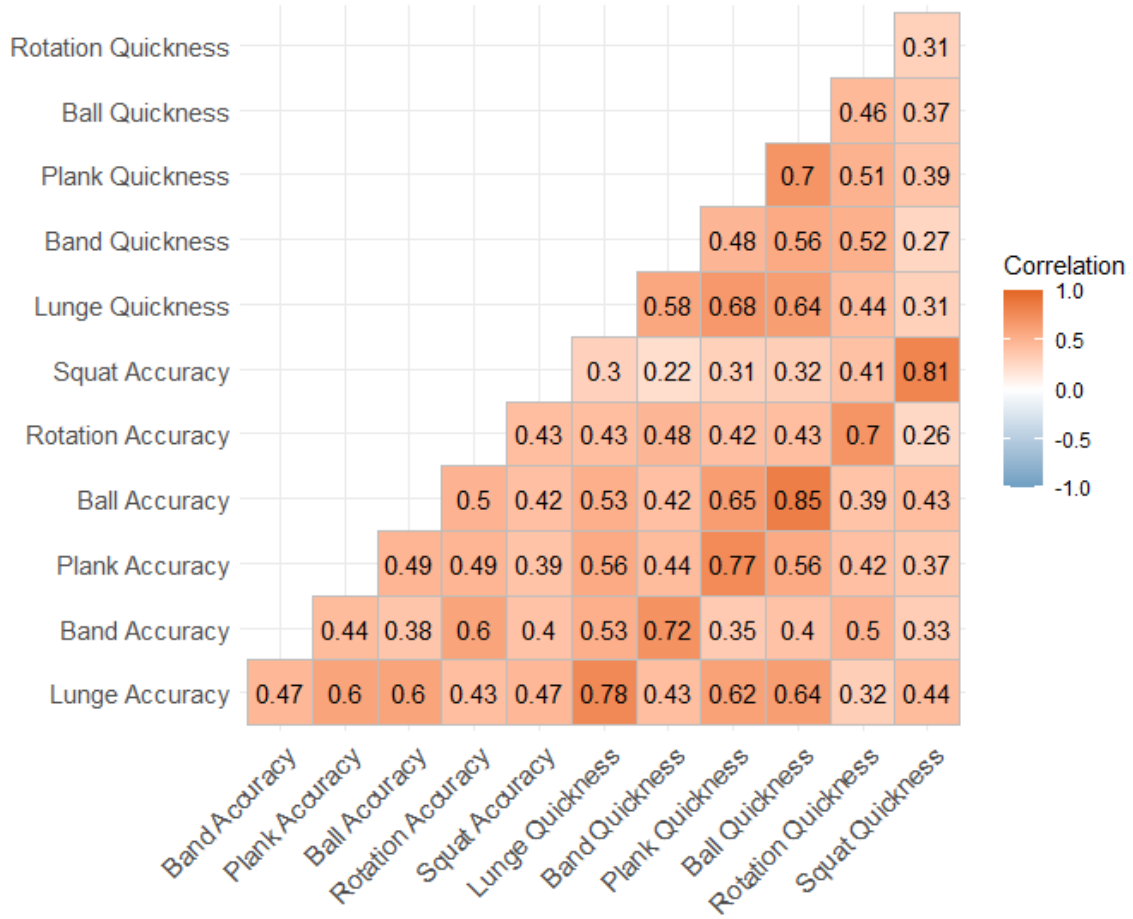
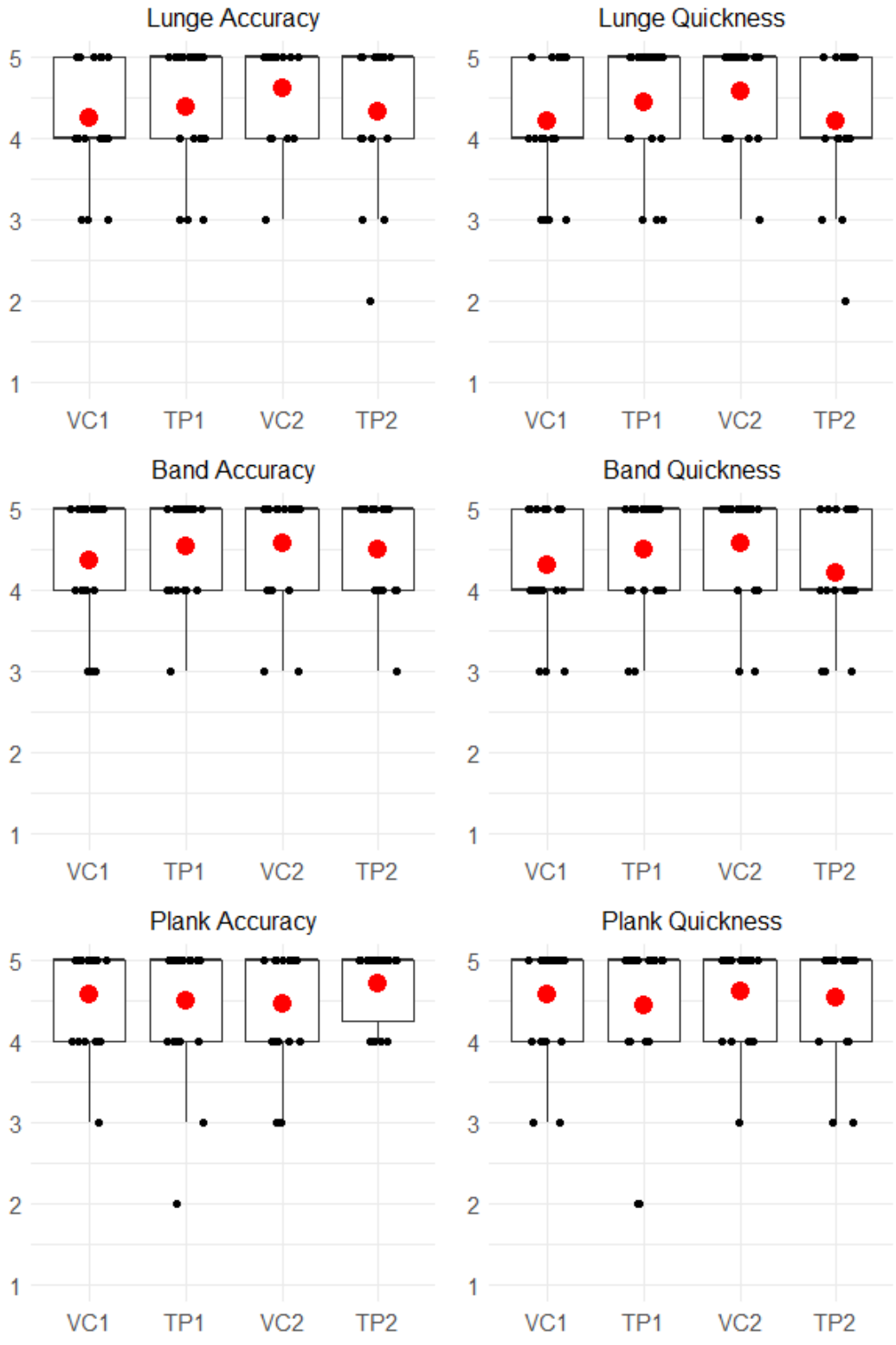


Figure E.3: Correlation coefficients between therapists' evaluations on physical therapy sessions.


Lunge Accuracy
Lunge Quickness
Band Accuracy
Band Quickness
Plank Accuracy
Plank Quickness
5
4
3
2
1
VC1
TP1
VC2
TP2

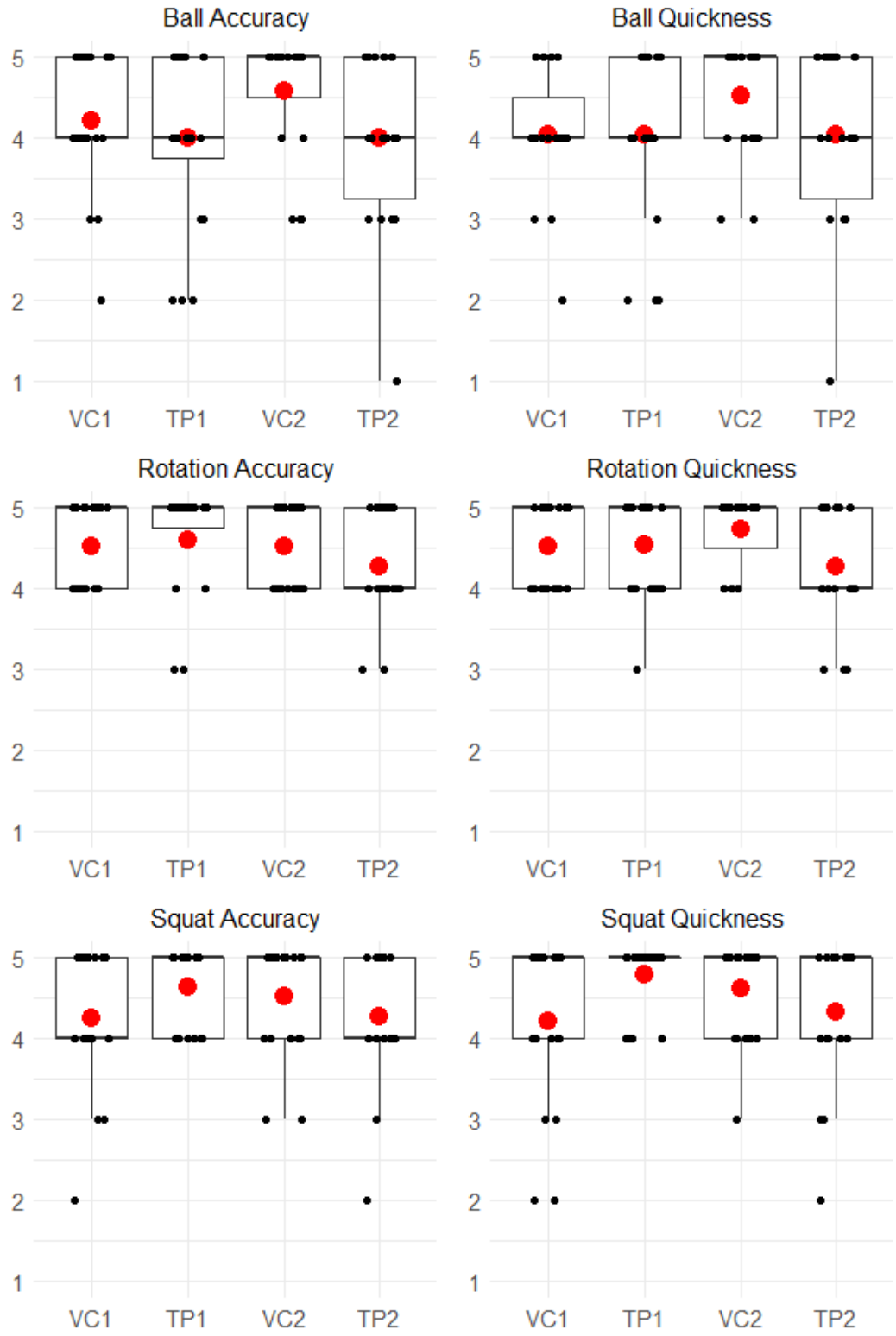


Figure E.4: Distributions of therapists' evaluations on physical therapy sessions. Each column matches an experimental condition. Each column contains a box plot that visualizes the quartiles, a point for each session, and a larger red dot indicating the mean value.

E.3 Patient Physical Therapy Evaluations

Patient Physical Therapy Questions (Short Description)	Mean Value (Standard Deviation)
Q1 (Therapist's Intentions)	4.05 (0.88)
Q2 (Pleasant Task)	3.65 (0.74)
Q3 (Together with Instructor)	3.74 (0.94)
Q4 (Difficult Task; Reverse Coded)	3.93 (0.77)
Q5 (Difficult to Move; Reverse Coded)	4.62 (0.75)
Q6 (Personal Learning Environment)	2.91 (1.00)
Q7 (Social Learning Environment)	2.58 (0.97)
Q8 (Lively Learning Environment)	2.86 (0.96)
Q9 (Pleasant Learning Environment)	3.36 (0.90)
Q10 (Close Therapist)	2.80 (0.94)
Q11 (Responsive Therapist)	3.92 (0.83)
Q12 (Active Therapist)	4.00 (0.83)
Q13 (Warm Therapist)	3.84 (0.86)
Q14 (Helpful Therapist)	3.86 (0.87)
Q15 (Realistic Therapist)	4.20 (0.77)
Q16 (Expert Therapist)	3.66 (0.92)
Average	3.62 (0.54)

Table E.1: The mean values (and standard deviations) of physical therapy evaluations from patients. Questions 4 and 5 were reverse coded with "Not at All" responses for the questions as 5 and "Extremely" as 1.

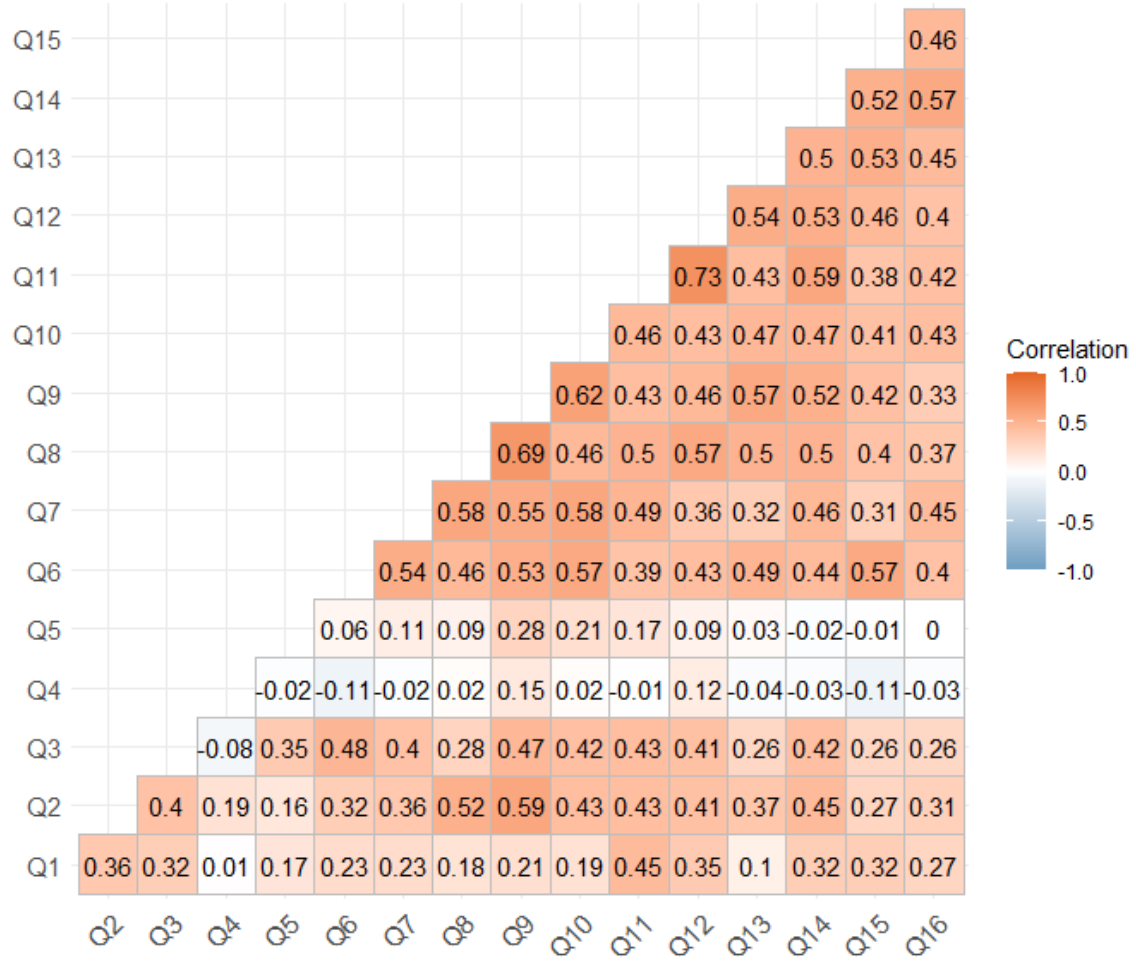
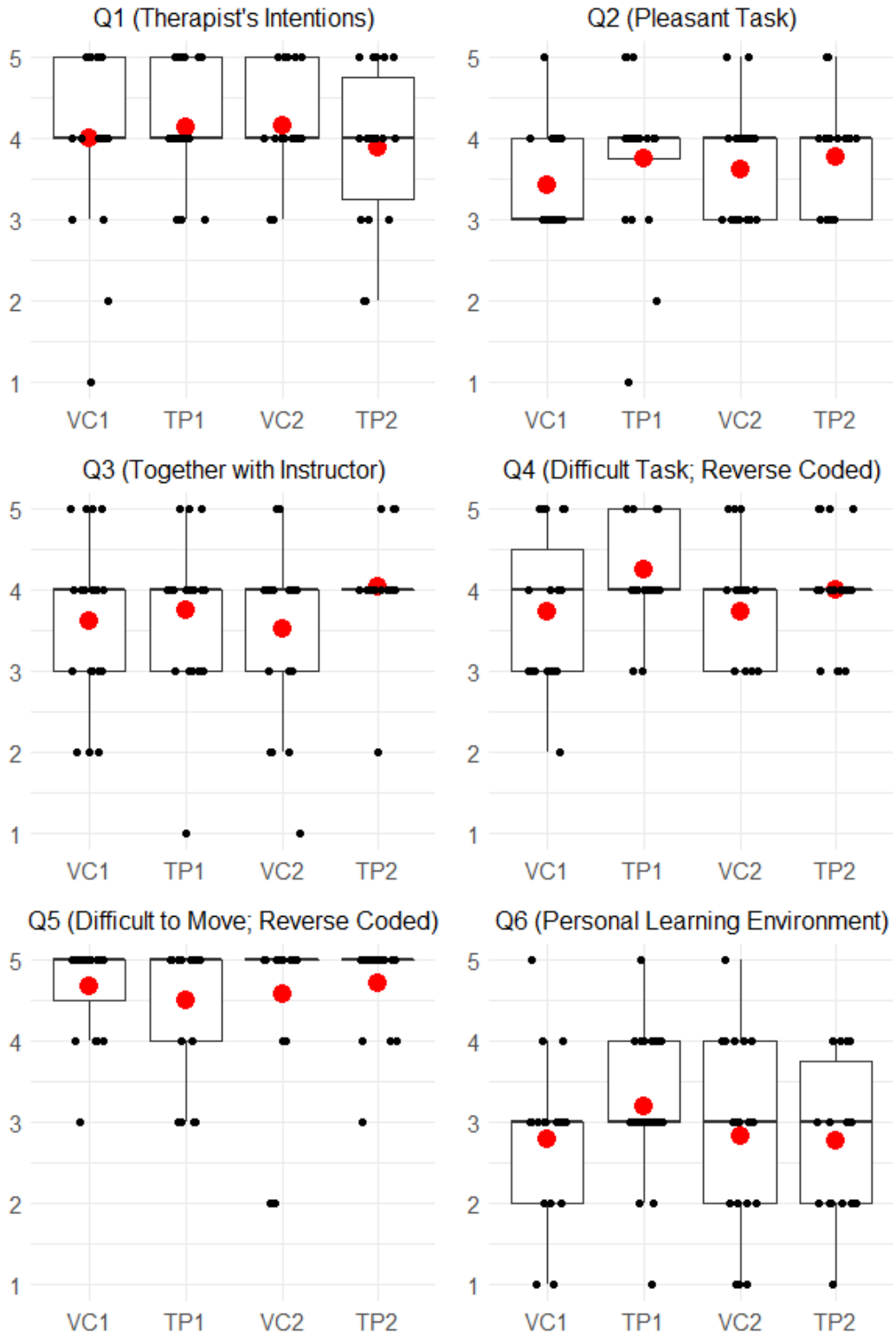
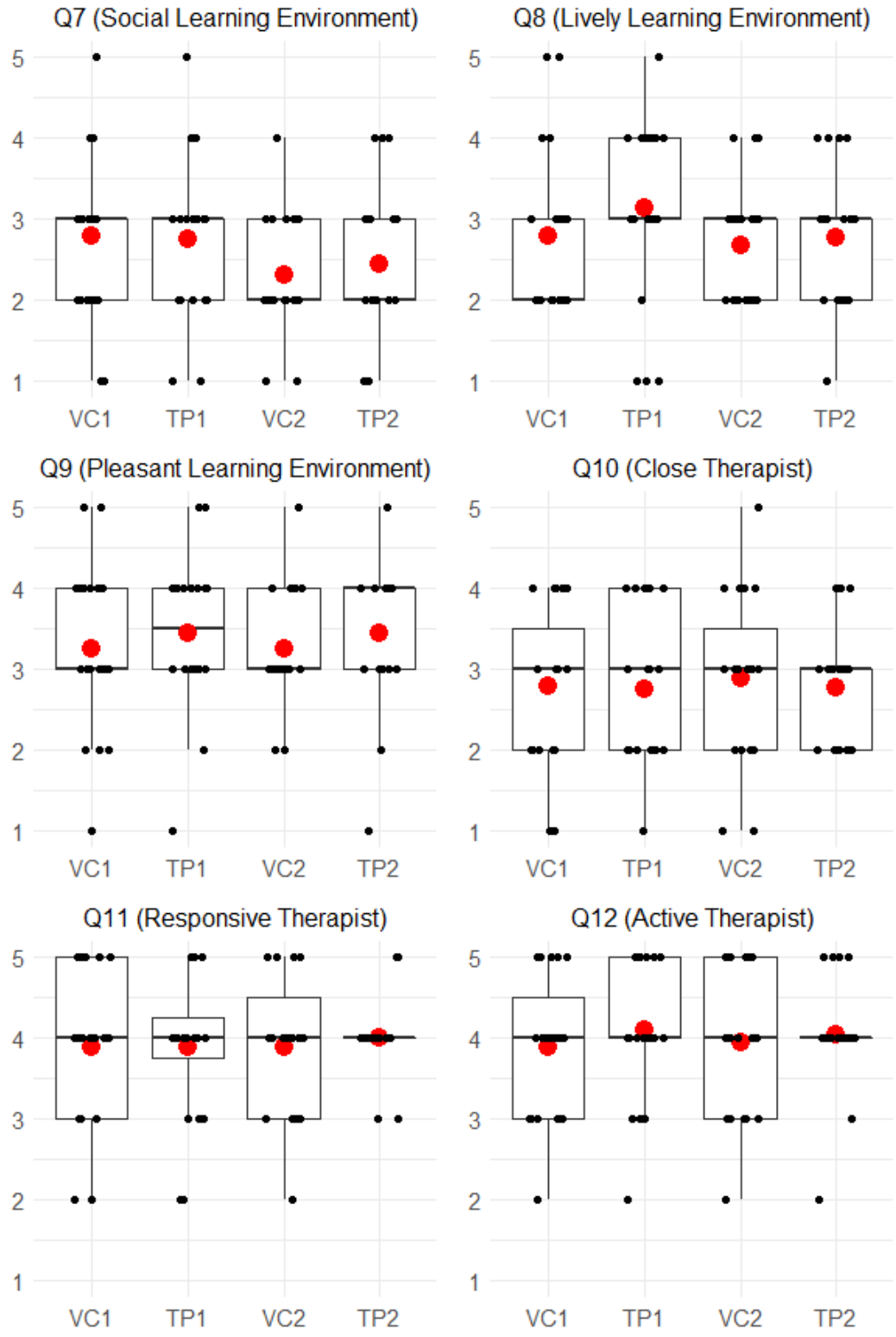


Figure E.5: Correlation coefficients between patients' evaluations on physical therapy sessions.


Q1 (Therapist's Intentions)
Q2 (Pleasant Task)
Q3 (Together with Instructor)
Q4 (Difficult Task; Reverse Coded)
Q5 (Difficult to Move; Reverse Coded)
Q6 (Personal Learning Environment)
5
4
3
2
1
VC1
TP1
VC2
TP2


Q7 (Social Learning Environment)
Q8 (Lively Learning Environment)
Q9 (Pleasant Learning Environment)
Q10 (Close Therapist)
Q11 (Responsive Therapist)
Q12 (Active Therapist)
5
4
3
2
1
VC1
TP1
VC2
TP2

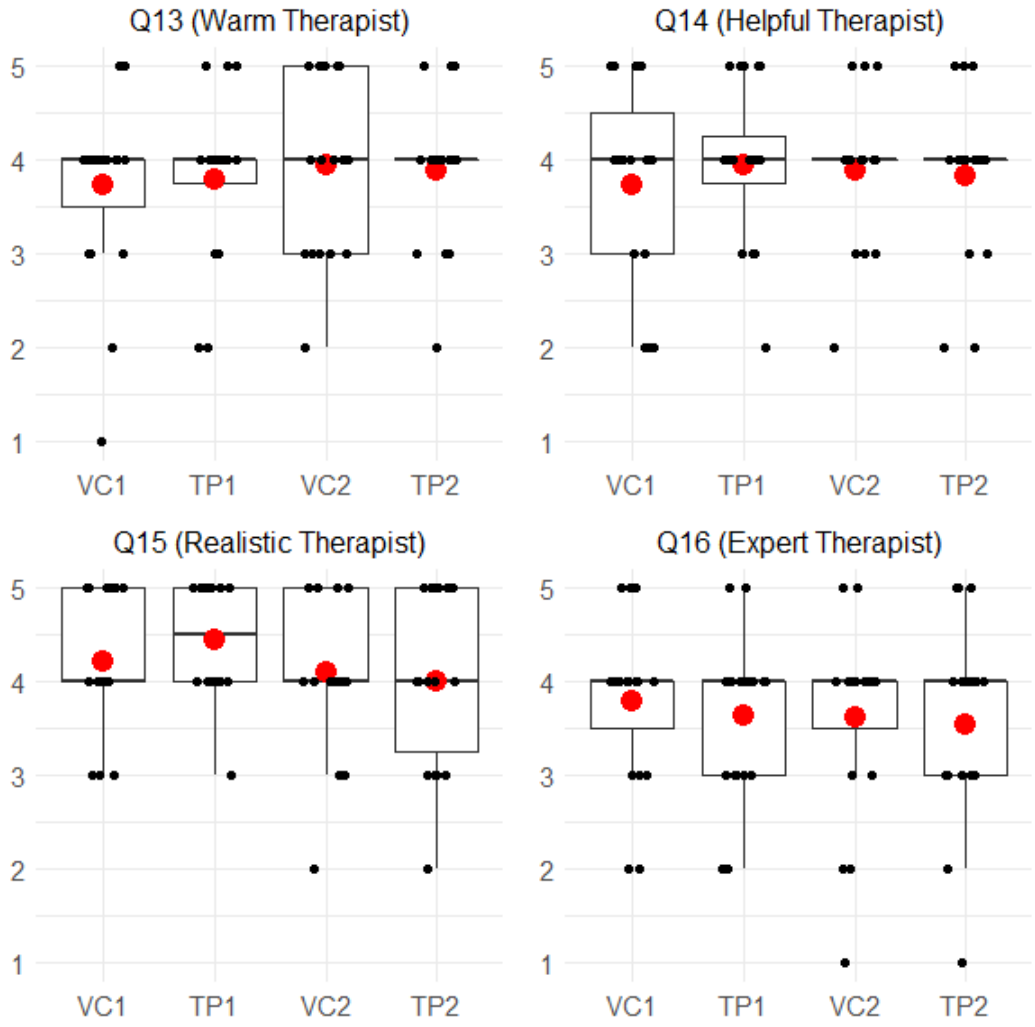


Figure E.6: Distributions of patients' evaluations on physical therapy sessions. Each column matches an experimental condition. Each column contains a box plot that visualizes the quartiles, a point for each session, and a larger red dot indicating the mean value.