A Study on the Effectiveness of Using Telepresence and Multiple Cameras in Remote Physical Therapy

Hanseul Jun*  
Stanford University

Husam Shaik†  
University of North Carolina at Chapel Hill

Michael Lewek‡  
University of North Carolina at Chapel Hill

Henry Fuchs§  
University of North Carolina at Chapel Hill

Jeremy Bailenson¶  
Stanford University

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. 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 through a user study with 11 physical therapists who had subject matter expertise and 76 participants who acted as physical therapy patients. We conducted detailed interviews of therapists at the end of their participation. Results showed that none of the main effects of using telepresence and using two cameras were supported via t-tests. However, additional analyses using linear mixed models with individual differences of participants controlled revealed a marginally significant positive effect of using two cameras on assessment scores from the therapists. The findings of this paper indicate that video fidelity of remote communication systems matters and therefore suggest telepresence systems should provide sufficient video clarity. Spatial ability of patients was found as a strong predictor of therapist assessments.

Index Terms: Human-centered computing—Interaction paradigms—Virtual reality; Human-centered computing—Human computer interaction (HCI)—Empirical studies in HCI

1 INTRODUCTION
With the advent of new media, more types of information become available for remote communication. For interpersonal communica-

Figure 1: Captures of a patient performing exercises (left and middle) and a therapist watching the patient through a VR headset (right).

tion, video conferencing (VC) followed telephony [14], which has become a widespread medium recently.

As the Media Ecology theory [20, 27] points out, that a new medium never fully replaces the existing medium, rather it is adopted in use cases where it excels over 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. For example, VC was found effective for corporate meetings with equivocality compared to emails [15] and for communication between multidisciplinary teams at hospitals when discussing issues about patients compared to telephony [4].

For remote communication, a strong candidate as the next step is the support of the 3D space. Currently, when we use VC systems, viewers see others from the perspective of the camera. Telepresence (TP) describes experiences that provide the feeling of being there [33]. TP systems provide users a 3D scene of other places that the users can freely move around.

This paper aims to examine TP as a potential successor to VC in addition to the effectiveness of using VC with two cameras to capture the space from another angle through an empirical study. Adding another camera is adopted as an easy technique to provide spatial information as adding a secondary camera shares a similar goal with TP and is not harder to adopt than using TP.

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. Previous works [22, 30] have introduced systems built 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.

Telegie 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) both of which are publicly available and easy to use.

In this paper, we describe design and implementation details of a telepresence system, report a user study examining the effects using telepresence and using two cameras, and provide implications for designing and studying telepresence systems.
2 RELATED WORK

2.1 Telepresence Systems

Among the many attempts to deliver the TP experience, we will concentrate on the ones that utilize RGBD cameras and AR/VR headsets.

In 2011, Maimone and Fuchs [18] implemented a TP system using RGBD cameras and autostereoscopic monitors. In a subsequent iteration of the system with the adoption of AR headsets [15], the autostereoscopic displays were replaced with AR headsets, using projectors to supplement the brightness of AR headsets as displays.

In 2012, Steed et al. [32] 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, to face many people on the other side.

In 2015, Robertset al. [29] introduced the withyou system. Users of their system are inside cubic immersive displays based on projectors and surrounded by multiple cameras. Also in 2015, Kowalski et al. [13] introduced LiveScan3D, a 3D data acquisition system using Kinect 2 devices. In 2017, the same scholars released an application for HoloLens devices to render point clouds from LiveScan3D.

In 2016, Room2Room [25] and Holoportation [24] 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. 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. Holoportation captures users with multiple RGBD cameras to reconstruct them into high-quality textured meshes. Through these meshes displayed on AR headsets, the system allows dyadic remote communication.

In 2018, Kolkmeier et al. [12] introduced OpenIMPRESS, a software toolkit for mixed reality remote collaboration systems. 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.

2.1.1 Applications of Virtual Reality for Physical Therapy

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 aimed to utilize VR as 3D content, but not its immersiveness.

With three adolescents with hemiplegic cerebral palsy, Golomb et al. [7] 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” [7, p. 27].

One approach for VR applications to provide physical therapy instructions is through gamification. Lange et al. [16] introduced JewelMine, a game built for rehabilitation, utilizing the Kinect to measure patient movements. Applying this VR gamification approach for rehabilitation to the older population, Rendon et al. [28] 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, 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.

VR has not only been used for providing instructions but also for building instruction programs. Camporesi, Kallmann, and Han [5] 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 further improved in subsequent research [11]. In this system, patients can see a 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. [6] compared VR-based therapy to conventional physical therapy in their 12-week study with 28 Parkinson’s Disease patients. In this study, the researchers found that the group who used VR outperformed the control group in terms of balance and gait after the 12 weeks compared to the group who received conventional physical therapy.

While the existing systems are mostly for solo use, our study will be based on a system built for remote communication between therapists and patients. Our system also leverages the immersiveness of VR headsets.

3 SYSTEM

The Telegie system consists of two applications: transmitter and viewer. The transmitter sends RGBD streams from cameras to the viewer, and the viewer renders the incoming RGBD streams to the users as a point cloud. We render each point as a quad scaled to correctly represent the field of view of the RGBD camera. The transmitter is supported for Windows 10 computers connected to Azure Kinect and iOS devices equipped with RGBD cameras. The viewer was built as a web application that can run on any device with a web browser including VR headsets.

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 second user joins the same room with their own transmitter using the obtained room ID. Finally, both users enter the room using their viewers and start communicating.

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 to exclude 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, we 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.

3.2 Distances between People

TP systems can and should set the distance between people. With recent AR studies with virtual humans finding similarity between behavior towards real people and virtual humans [17][21], personal space literature [8] sheds light on this issue finding social context as a major factor deciding which distance 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, we decided to provide a user interface to set the distances.

3.3 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 in VP8 using libvpx. For depth information, depth pixels go through optional background removal, mapping to the color camera’s coordinate system, and Temporal RVL compression \[6\]. Floor information gets extracted from depth pixels and audio information is encoded in the Opus codec \[7\] and gets sent separately from video messages.

3.4 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 \[8\], an implementation of WebRTC data channels. Unreliable data channels were used for real-time communication with lower latency. To handle packet loss, our system adopted a fountain code—Wirehair \[9\]. After splitting video messages into packets and encoding them in Wirehair, Telegie transmitters send packets with 50% of redundancy.

4 Study

The study examined remote communication systems in the context of remote physical therapy. 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.

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 a US East Coast University. All therapist participants were enrolled in a clinical Doctor of Physical Therapy curriculum. All completed coursework relevant to their role, including therapeutic exercise. The average age of the therapists was 24.455 years old (SD = 1.508). Eight of them were female and three of them were male. Every therapist was scheduled to meet eight patients.

Seventy-six participants were recruited from a US West Coast University as patients. The participants were not actual patients but received instructions from the physical therapists as they were patients. The average age was 25.711 years old (SD = 6.685) across 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 IRBs.

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 saw patients on a TV screen, and in TP conditions, therapists saw patients through a VR headset. Patients interacted with 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) but only the RGBD streams were shown to the therapist. 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. 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. Participants 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 on a tablet screen.

During the experimental sessions, the lab for the therapists had only one webcam in front of the therapists. 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 at a 90 degree angle. During the sessions, all webcams and microphones were recorded.

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 both type of 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 through the front RGBD camera and the side RGBD camera through their VR headset. Figure[1] demonstrates how patients looked like to the therapists wearing VR headsets in the TP1 condition.

During each session, the therapist gave instructions on six exercises for fifteen minutes. Figure[2] 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 exercises were chosen as they require the therapists to evaluate them from different planes and angles simultaneously and would likely benefit from having 3D spatial information. The therapists instructed two sets of each exercise for the patients. After each session, both the therapists and patients answered a post-questionnaire.

When therapists were done with all eight sessions, they were
We only asked whether they have prior experience, not the amount of prior VR experience. 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** Participants 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 [26,31]. 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 as they have to observe and evaluate body movements from multiple viewpoints simultaneously. The average score of therapists was 4.545 out of 5 (SD = 0.688), and the average score of patients was 4.500 (SD = 0.825).

**Interpersonal Communication** After each experimental session, participants were asked the level of social presence [9], communication satisfaction [23], interpersonal liking [23], and Inclusion of Other in the Self (IOS) [1] they have experienced during the session. Due to being highly correlated with each other, these four measures were examined together with their average as the interpersonal communication response (Cronbach’s $\alpha = 0.788$ for therapists; Cronbach’s $\alpha = 0.788$ for patients). When taking the average, IOS was rescaled to 1-5 from its original scale of 1-7 as other three measures were asked by a 5-point Likert scale.

**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. The perceived video clarity 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 3.829 (SD = 0.839).

**Therapist Assessments** 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 $\alpha = 0.917$), they have been merged as the therapist assessment. The mean value of therapist assessments was 4.427 (SD = 0.543).

**Physical Therapy Questionnaire for Patients** Patients were asked about their physical therapy experience after each experimental session. The 16 questions are originally from Bailenson et al. [2]. 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 $\alpha = 0.895$) and mean value was 3.623 (SD = 0.545).

**Interviews** Interviews of therapists were conducted and audio recorded. The interview questions asked for comparison between the experimental conditions.

### 4.2 Hypotheses and Research Question

Based on previous literature, it has been predicted TP to outperform VC and that using two cameras to outperform using one camera. Based on this, we predicted eight hypotheses expecting positive outcomes from TP and using two cameras to four dependent variables of interpersonal communication responses from therapists, interpersonal communication responses from patients, therapist assessments, and physical therapy evaluations from patients. For example, between TP and interpersonal communication responses from therapists, we hypothesized the level of interpersonal communication responses from therapists will be higher in TP than in VC.

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 the eight hypotheses 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?

### 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.

Figure 3 provides the distributions of all four dependent variables across experimental conditions. For the mean values and standard deviations of the accuracy and quickness of exercises from therapists, see Table 1. The accuracy and quickness of the patients were highly correlated in the positive direction ($\rho = 0.788$) based on the assessments from the therapists.

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 2 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.

| Exercises | Accuracy | Quickness |
|-----------|----------|-----------|
| Lunge     | 4.408 (0.751) | 4.368 (0.763) |
| Band      | 4.500 (0.663) | 4.408 (0.715) |
| Plank     | 4.566 (0.660) | 4.553 (0.737) |
| Ball      | 4.197 (0.966) | 4.171 (0.929) |
| Rotation  | 4.487 (0.622) | 4.526 (0.599) |
| Squat     | 4.434 (0.736) | 4.500 (0.792) |
| Average   | 4.432 (0.551) | 4.421 (0.572) |

Table 1: The mean values (and standard deviations) of therapist assessments.

In the following, 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 [3]. 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 therapist assessments are presented, followed by analyses that expand on the other three dependent variables.

5.1 Therapist Assessments

Individual differences of both therapists and patients likely influenced therapist assessments. Figure 3 visualizes the cross-therapist variation of the assessment scores. Given the cross-therapist variation was larger than the cross-condition variation (see Figure 5), 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, therapist identity added as a random effect was examined. In this model, using two cameras was found to have a marginally significant positive effect (b = 0.246, p = 0.093) on the assessment scores. The interaction between TP and using two cameras was found to have a marginally significant negative effect (b = -0.358, p = 0.085).

5.1.1 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. 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 for the distribution of video clarity per conditions. Figure 6 demonstrates the positive slopes on assessment scores from both video clarity levels.

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 (b = 0.384, p = 0.048), and the interaction between TP and using two cameras had a marginally significant negative effect (b = -0.340, p = 0.097). Using two cameras did not have a significant effect (b = 0.177, p = 0.226). The video clarity level reported by the therapists had a significant positive effect (b = 0.169, p = 0.026) on the assessment scores. Moreover, the video clarity level reported by the patients had a marginally significant positive effect (b = 0.119, p = 0.078) on the assessment scores.
5.1.2 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 therapist assessments. To examine this, the spatial ability of patients was measured during the pre-questionnaire of the study with five spatial ability questions. Figure 7 shows the positive slopes from the measured spatial ability to the assessment 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.

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 patients added as a fixed effect was tested. Noticeably, only the spatial ability was found as a significant positive effect (b = 0.155, p = 0.021). The effects from experimental conditions were no longer statistically significant with the spatial ability added to the model.

5.1.3 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 ($\rho = -0.24$) and positively correlated with therapist assessments ($\rho = 0.41$). See Figure 8 for the distributions of reported motivations levels per experimental conditions and the relationship between motivation levels and assessment scores.

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 (b = 0.275, p = 0.062) and the interaction between TP and using two cameras (b = -0.391, p = 0.060) showed the effects in the same directions they showed in the model without the motivation level controlled. 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 therapist assessments.

The effects of the genders of participants, prior VR experience, and prior PT experience were also examined with adding them as additional fixed effects. However, no statistically significant effect was found from them.

5.2 Expansion to Other Dependent Variables

In the above Section 5.1 only therapist assessments 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 therapist assessments will be expanded to the three other dependent variables.

Across Section 5.1 the influence of variables that have the potential to affect the assessments 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
Table 3: Estimated slopes and their p-values of the fixed effects from the linear mixed models as the expansion of Section 5.1 to all four dependent variables. Each column represents a dependent variable. (*: p < 0.050, †: p < 0.100)

| Additional Fixed Effect (Corresponding Section) | Fixed Effect | Therapist Assessment | Patient Physical Therapy | Therapist Interpersonal Communication | Patient Interpersonal Communication |
|-----------------------------------------------|-------------|----------------------|--------------------------|---------------------------------------|-----------------------------------|
| None                                          | TP          | 0.101                | 0.162                    | -0.130                                | -0.034                            |
|                                               | Cameras     | 0.240†               | 0.002                    | 0.088                                 | -0.163                            |
|                                               | TP × Cameras| -0.357†              | -0.093                   | -0.006                                | 0.181                             |
| Therapist Video Clarity                       | TP          | 0.403*               | 0.329                    | 0.384*                                | 0.043                             |
|                                               | Cameras     | 0.198                | -0.026                   | 0.008                                 | -0.175                            |
|                                               | Therapist Video Clarity| 0.173*          | 0.098                    | 0.297*                                | 0.045                             |
|                                               | TP × Cameras| -0.341               | -0.080                   | 0.024                                 | 0.186                             |
| Patient Spatial Ability                       | TP          | 0.039                | 0.174                    | -0.157                                | 0.004                             |
|                                               | Cameras     | 0.191                | 0.012                    | 0.064                                 | -0.137                            |
|                                               | Patient Spatial Ability| 0.155*          | -0.030                   | 0.069                                 | -0.074                            |
|                                               | TP × Cameras| -0.266               | -0.110                   | 0.035                                 | 0.138                             |
| Perceived Patient Motivation                  | TP          | 0.202                | 0.198                    | 0.088                                 | -0.056                            |
|                                               | Cameras     | 0.275*               | 0.012                    | 0.154                                 | -0.172                            |
|                                               | Perceived Patient Motivation| 0.255*         | 0.089                    | 0.436*                                | -0.054                            |
|                                               | TP × Cameras| -0.391†              | -0.103                   | -0.080                                | 0.189                             |

5.3 Open-ended Responses

5.3.1 Therapist Interviews

During their participation, all therapists were able to accumulate experience across all four experimental conditions. And before the interviews, they were immersed in our study for three hours. Leveraging this, we asked the therapists about their experience including a comparison between the experimental conditions.

The most common issue of TP the therapists mentioned is that they could not see themselves. This has been mentioned by all therapists. Therapist #10 said, “If 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 issue mentioned by 10 therapists on TP was pixelation. These comments were toward the insufficient resolution of the depth pixels, which was an issue 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 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 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 and three mentioned they are similar. 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.”

Seven therapists mentioned that they preferred VC2 over VC1, and two therapists said they prefer VC2 the most out of all conditions. 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 side, 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. 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 the participant to go on a couple different positions. But I don’t feel like there was that it took like, a couple seconds. And it wasn’t didn’t seem like too much of a hassle.”

Six therapists said they felt higher social presence in TP. One therapist 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 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 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 said the assessment questions did not work well 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.”

5.3.2 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.

6 Conclusion

In this paper, we 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 were recruited 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.

Based on previous literature, positive effects were expected for using the TP and two cameras for remote physical therapy. However, none of the eight hypotheses were statistically significant. Consequently, we further explored regarding which additional variables would affect the statistical tests conducted for the hypotheses. In Section 5.1.1, with individual differences from therapists controlled, we found a positive marginally significant effect of using two cameras and a negative marginally significant effect from the interaction between TP and using two cameras on therapist assessments.

In Section 5.1.2, we found 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. In Section 5.1.3, spatial ability of patients was found as a strong predictor therapist assessments. In Section 5.1.4, it has been found that perceived patients’ motivation level reported by therapists predicts the assessment scores.

With these analyses expanded to other dependent variables in Section 5.2, the influences on therapist assessments were partially replicated on interpersonal communication responses from therapists. When video clarity was added as an additional fixed effect to the model on interpersonal communication responses from therapists, again, TP showed a significant positive effect.

With spatial ability controlled for the model on interpersonal communication responses from therapists, unlike the model on therapist assessments, there was no significant effect. There are two possible explanations of spatial ability improving the assessment scores. One is that effects on both outcomes were coming from the general ability of individuals and the other is that spatial ability and task performance were 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 therapist assessments are actually relevant to each other.

None of the additional variables had significant influence to responses from patients on physical therapy or interpersonal communication.

After the sessions, therapists were interviewed. 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 interviews, therapists preferred to use a VC system with two cameras than one camera. This matches the results from the linear mixed model in Section 5.1.2 regarding the experimental conditions on therapist assessments. However, this preference was shown clearer in the interviews than found in the quantitative analyses. Also in the interviews, many therapists preferred VC over TP.

Using two cameras with the TP system did not result in positive outcomes compared to using one camera for TP, unlike in VC. To better understand the reason, an independent rater was asked to watch video recordings of the TP sessions with two cameras and report the ratio the therapists were watching the patients from the side. In other words, the rater checked whether the therapists actually took the advantage of having the side camera. The independent rater reported that the therapist was watching from the side 13.54% 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 the right panel of Figure 1 to see a which regions of the video recordings were counted by the raters as the therapists seeing the patients from the side. When the therapists stayed in the regions outside of the yellow markers, therapists were counted as utilizing the side camera.

6.1 Limitations

The video resolution of Telegie was lower than it could be. To provide stable streaming, we downsampled both the color and depth pixels captured by the Azure Kinect device into half per dimension.

The assessment scores from the therapists likely suffered from a ceiling effect, having their average scores above 4 out of 5. A more difficult set of exercises should have been chosen.

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 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 can be simplified. For example, calibration of the system when working with multiple cameras should be 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. With experience, for example, therapists may have better utilized the information from the side cameras in TP.

6.3 Implications

From this study, the most obvious influence was found from 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.

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.

REFERENCES

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

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

[3] D. Bates, M. Mächler, B. Bolker, and S. Walker. Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*, 2014.

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

[5] C. Camporesi, M. Kallmann, and J. J. Han. Vr solutions for improving physical therapy. In *2013 IEEE Virtual Reality (VR)*, pp. 77–78. IEEE, 2013.

[6] H. Feng, C. Li, J. Liu, L. Wang, J. Ma, G. Li, L. Gan, X. Shang, and Z. Wu. 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, 2019.

[7] M. R. Golomb, M. Barkat-Mash, B. Rabin, M. Abdelbaky, M. Huber, and G. Burdea. Eleven months of home virtual reality telerehabilitation-lessons learned. In *2009 Virtual Rehabilitation International Conference*, pp. 23–28. IEEE, 2009.

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

[9] F. Herrera, S. Y. Oh, and J. N. Bailenson. Effect of behavioral realism on social interactions inside collaborative virtual environments. *Presence*, 27(2):163–182, 2020.

[10] H. Jun and J. Bailenson. 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. IEEE, 2020.

[11] M. Kallmann, C. Camporesi, and J. Han. Vr-assissted physical rehabilitation: Adapting to the needs of therapists and patients. In *Virtual realities*, pp. 147–168. Springer, 2015.

[12] J. Kolkmeier, E. Harmsen, S. Giesselink, D. Reidsma, M. Theune, and D. Heylen. 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, 2018.

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

[14] R. E. Kraut and R. S. Fish. Prospects for video telephony. *Telecommunications Policy*, 19(9):699–719, 1995.

[15] C. T. Kydd and D. L. Ferry. Managerial use of video conferencing. *Information & Management*, 7(6):369–375, 1994.

[16] B. Lange, S. Koenig, E. McConnell, C.-Y. Chang, R. Jiang, E. Suma, M. Bolas, and A. Rizzo. Interactive game-based rehabilitation using the microsoft kinect. In *Virtual Reality Conference, IEEE*, pp. 171–172. IEEE Computer Society, 2012.

[17] M. Lee, G. Bruder, T. Höllerer, and G. Welch. Effects of unaugmented augmented reality technology for improvement of TP systems. In *Proceedings of the 29th annual symposium on user interface software and technology*, pp. 741–754, 2016.

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

[19] N. Postman. Media ecology: Communication as context. 1974.

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

[21] D. J. Roberts, A. J. Fairchild, S. P. Campion, J. O’Hare, C. M. Moore, R. Aspin, T. Duckworth, P. Gasparello, and F. Tocchia. Withyou—an experimental end-to-end telepresence system using video-based re-construction. *IEEE Journal of Selected Topics in Signal Processing*, 9(3):562–574, 2015.

[22] E. Saracee, S. Singh, K. Hendron, M. Zheng, A. Joshi, T. Ellis, and M. Betke. 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, 2017.

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

[24] A. Steed, W. Steptoe, W. Oyekoya, F. Pece, T. Weyrich, J. Kautz, D. Friedman, A. Peer, M. Solazzi, F. Tecchia, et al. Beamng: an asymmetric telepresence system. *IEEE computer graphics and applications*, 32(6):10–17, 2012.

[25] J. Steuer. Defining virtual reality: Dimensions determining telepresence. In *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, 2016.

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

[27] A. K. Mishra, M. Skubic, and C. Abbott. 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. IEEE, 2015.

[28] C. Oh, F. Herrera, and J. Bailenson. The effects of immersion and real-world distractions on virtual social interactions. *Cyberpsychology, Behavior, and Social Networking*, 22(6):365–372, 2019.

[29] S. Orts-Escolano, C. Rhemann, S. Fanello, W. Chang, A. Kowdle, Y. Deptyarve, D. Kim, P. L. Davidson, S. Khamis, M. Dou, et al. Holoporation: Virtual 3d teleportation in real-time. In *Proceedings of the 29th annual symposium on user interface software and technology*, pp. 741–754, 2016.

[30] T. Pejza, J. Kantor, H. Benko, E. Ofek, and A. Wilson. 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, 2016.