BALANCING USER EXPERIENCE FOR MOBILE ONE-TO-ONE INTERPERSONAL TELEPRESENCE

by

KEVIN PATRICK COLEMAN PFEIL
M.S. University of Central Florida, 2013
B.S. University of Central Florida, 2010

A dissertation submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy
in the Department of Computer Science
in the College of Engineering and Computer Science
at the University of Central Florida
Orlando, Florida

Spring Term
2022

Major Professors: Joseph J. LaViola and Pamela Wisniewski

ABSTRACT

The COVID-19 virus disrupted all aspects of our daily lives, and though the world is finally returning to normalcy, the pandemic has shown us how ill-prepared we are to support social interactions when expected to remain socially distant. Family members missed major life events of their loved ones; face-to-face interactions were replaced with video chat; and the technologies used to facilitate interim social interactions caused an increase in depression, stress, and burn-out. It is clear that we need better solutions to address these issues, and one avenue showing promise is that of Interpersonal Telepresence.

Interpersonal Telepresence is an interaction paradigm in which two people can share mobile experiences and feel as if they are together, even though geographically distributed. In this dissertation, we posit that this paradigm has significant value in one-to-one, asymmetrical contexts, where one user can live-stream their experiences to another who remains at home. We discuss a review of the recent Interpersonal Telepresence literature, highlighting research trends and opportunities that require further examination. Specifically, we show how current telepresence prototypes do not meet the social needs of the streamer, who often feels socially awkward when using obtrusive devices.

To combat this negative finding, we present a qualitative co-design study in which end users worked together to design their ideal telepresence systems, overcoming value tensions that naturally arise between Viewer and Streamer. Expectedly, virtual reality techniques are desired to provide immersive views of the remote location; however, our participants noted that the devices to facilitate this interaction need to be hidden from the public eye. This suggests that 360° cameras should be used, but the lenses need to be embedded in wearable systems, which might affect the viewing experience. We thus present two quantitative studies in which we examine the effects of camera placement and height on the viewing experience, in an effort to understand how we can better design telepresence systems. We found that camera height is not a significant factor, meaning wearable cameras do not need to be positioned at the natural eye-level of the viewer; the streamer is able to place them according to their own needs.

Lastly, we present a qualitative study in which we deploy a custom interpersonal telepresence prototype on the co-design findings. Our participants preferred our prototype instead of simple video chat, even though it caused a somewhat increased sense of self-consciousness. Our participants indicated that they have their own preferences, even with simple design decisions such as style of hat, and we as a community need to consider ways to allow customization within our devices. Overall, our work contributes new knowledge to the telepresence field and helps system designers focus on the features that truly matter to users, in an effort to let people have richer experiences and virtually bridge the distance to their loved ones.

Ad Majorem Dei Gloriam

ACKNOWLEDGMENTS

Thanks to my dissertation committee for their support, help, and guidance to make this work the best it could be: Dr. Joseph LaViola (co-chair); Dr. Pamela Wisniewski (co-chair); Dr. Gregory Welch; and Dr. Carman Neustaedter. I am particularly grateful to Dr. Wisniewski for expanding my research horizons, pushing me to think more deeply, and not giving up on me.

I'd like to thank all members of the Interactive Computing Experiences (ICE) Research Cluster, past and present, for their support and camaraderie, especially Dr. Kate Kapalo, Dr. Eugene Taranta, Dr. Corey Pittman, Dr. Seng Lee Koh, Andres Vargas, Brian Williamson, and Sina Masnadi. A significant portion of my life was spent within the round walls of the lab, and I will *never* forget our time together. I'd like to especially thank Dr. Kapalo for being one of my closest friends. I've learned so much from you — you are my academic hero. Sincerely, thank you for counseling me, for teaching me, for listening to me, for just being a friend I could depend on. I couldn't have gotten here without you!

I'd also like to thank all members of the Socio-Technical Interaction Research (STIR) Lab, past and present, for their support and guidance as I ventured into qualitative research. I want to especially thank Dr. Karla Badillo-Urquiola and Dr. Afsaneh Razi for helping me stay focused in the final semester as we worked on our slides and writing — we did it!

Thanks to Christy LaPerriere for her graphical design of Figure 4.2, and thanks to Dr. Ravi Kattoju, Neeraj Chatlani, Jacob Belga, and Jose-Valentin Sera-Josef for their help with writing and running user study sessions.

I want to thank my family for keeping me sane — my parents, George and Paddy, my sister Kerry, my brother Sean, and my adorable nieces Aine and Hayden. You light up my life and are constant

sources of inspiration! Lastly, I want to thank "Uncle" Bill and "Aunt" Joanne Brown for being true guardian angels. Thank you for sending me to Catholic school, deepening my faith, and helping me to value education. I hope to someday make an impact on others' lives as you have made on mine. Now, my real work begins!

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

This chapter is based on part of a work previously published:

Kevin P. Pfeil, Neeraj Chatlani, Joseph J. LaViola, and Pamela Wisniewski. 2021. Bridging the Socio-Technical Gaps in Body-worn Interpersonal Live-Streaming Telepresence through a Critical Review of the Literature. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 120 (April 2021), 39 pages. DOI:https://doi.org/10.1145/3449194

In the past few years, we have experienced a phenomenon like none other in recent memory — the Novel Coronavirus has created the need for "social distancing," where people must physically distance themselves and limit the size of social gatherings [5], while still trying to function as a society and remain connected to one another. Both non-essential work and school environments have moved online as people struggled to keep a safe distance, while avoiding complete social isolation. Some countries did shutdown completely, disapproving of larger social gatherings which could contribute to spreading the virus, fining individuals who broke shutdown laws [6, 214]. Meanwhile, we pondered how this situation affected interpersonal relationships, as remote technologies tend to emphasize work and school as the primary use cases for their target audiences; for instance, depression and anxiety levels have reportedly increased on a global scale [33, 47, 85], due to a perceived sense of loneliness [20], as people are desperate to go back into the world, yet fear contracting the virus. Additionally, users report how repeatedly interacting with video chat applications can cause "Zoom Fatigue" [158]. It is in times like these that Human-Computer Interaction (HCI) researchers can make a difference by designing novel technologies that can address some of the new challenges facing the world. We might re-imagine telepresence, or the ability to perceive and/or interact with a remote environment as if actually there [159], for more immersive and intimate uses, such as orchestrating a playdate or being able to share invaluable time with an ill family member who is physically bound to their home or bed. By transforming existing technologies or imagining new ones, we might be able to bridge the socio-technical gap in a way that facilitates one-to-one telepresence experiences which maintain and strengthen interpersonal relationships that are increasingly mediated through technology.

In 1980, Minsky hypothesized that we could implement telecommunications in order to let workers remotely manipulate a robotic platform to perform a task over great distance [159]. We are seeing this idea begin to blossom; for example, stationary robots are being used by doctors for remote surgery [121, 193], and research efforts are planning robotic uses for space expeditions [83, 136, 137]. For everyday use cases, simple, wheeled robots have been developed to let a user explore and interact with a remote environment via video chat (e.g. [98, 182]). More recently, telepresence has expanded to include interpersonal experiences through telecommunications platforms (e.g., Skype, Zoom, etc.) that connect two or more parties, typically within home environments (e.g., [96, 97, 178, 179]). However, with the advent of affordable cameras with live-streaming capabilities and sufficient networking technologies, there is a new trend in research which people wear a camera on their body to share their mobile experiences with another person. This form of telepresence, which we refer to as interpersonal telepresence, is starting to gain traction, in multiple areas.

For instance, streamers are broadcasting to vast audiences comprised of users who each have a viewing device [72, 170, 244]. Twitch, Periscope, and TikTok, for example, are popular live-streaming platforms, where content creators have started to share their live experiences with audiences as they visit different places [50], perform tasks, or engage in other activities [71, 144, 145]. Meanwhile, doctors are live-streaming first-person views of surgery to enhance medical students' training [62]. In these cases, a single user broadcasts to many recipients to share the experience, rather than to build or maintain existing relationships with them.

Yet, there is an opportunity for immersive telepresence to be leveraged in a one-to-one paradigm, where there is only one streamer and one viewer, to not only share one's experience but to socially connect to another human being. This more intimate use of telepresence technology creates unique opportunities for interpersonal interactions that build relationships and trust through shared goals, such as a remote user directing an on-site user to perform specific tasks or interact with their environment [82, 225, 249], or to simply allow for a socially-distant meet-up where it would be otherwise impossible, or dangerous to do so.

Telepresence has been used in work-related scenarios and in the context of Human-Robot Interaction (HRI) [125], but with consumers being given increased access to telepresence-capable technology, such as 5G streaming and mobile HD cameras, one-to-one telepresence has also found several recreational uses. At the beginning of the 2010's, for instance, smartphones and mobile video chat applications (e.g. Skype and FaceTime) emerged, setting up the infrastructure for mobile, one-to-one, interpersonal telepresence [255]. However, the implementations of mobile telepresence are still in their infancy, as researchers are working to identify the best practices which offer optimal user experience to both sides of the interaction. Recent implementations of mobile interpersonal telepresence in the academic community have been met with apprehensions and scrutiny, and leading up to the ACM CHI 2019 conference, there was significant discourse surrounding the body-worn paradigm as a way to provide remote attendance to conferences [54, 180]; for instance, some felt that the nature of the interaction invited social harm to the person wearing a device on their body, such as leading to uncomfortable situations or exploitation, and women have expressed that wearing a camera or screen on their body would invite people to stare inappropriately.

These concerns have caused telepresence researchers to reconsider the social implications of this new interaction paradigm. In essence, these concerns stem from what Ackerman calls a "socio-technical gap" [7], in which there is a disconnect between what we should socially support and the technical capabilities for doing so. In a mobile telepresence experience, the local user, remote user, and third parties in the immediate vicinity each have social needs. Addressing these needs give rise to the technical implications of future body-worn telepresence designs. These concerns inspire us to work towards creating a better telepresence interaction for all its potential users. Therefore, we apply this socio-technical lens to a field that tends to be more technical in nature, as to bring about a better understanding of social considerations for this interaction.

Thus, the purpose of this dissertation is to address these socio-technical needs through a series of studies that give more power to the user. We describe research efforts aimed at understanding what users desire from an interpersonal telepresence interaction, and focus on realizing those needs. To do so, we utilize the Value Sensitive Design (VSD) framework.

Value Sensitive Design (VSD) Framework

In this dissertation, we apply the VSD framework, which is an iterative approach to identifying stakeholders, the benefits and harms that those stakeholders encounter, and the human values that conceptually envelop those benefits and harms [56]; see Figure 1.1 for illustration. Three types of efforts are used within this framework: Conceptual Investigations, with which researchers can leverage existing knowledge to identify the human problem space and ways to tackle it; Empirical Investigations, with which researchers can identify human values through in situ observations; and Technological Investigations, with which researchers can learn how technology can foster these human values [56]. In this dissertation, we leverage each type of investigation to isolate and foster human values for future interpersonal telepresence designs. We triangulate how individual differences *must* be considered when designing new prototypes in this problem space; whereas past research does not engage with individual differences and attempts to build devices to fit the "average" user, we demonstrate that this strategy is detrimental. We provide a set of considerations to help guide interpersonal telepresence design decisions, concluding that to truly give power to the user, we must provide them with the tools to realize their own telepresence prototypes.

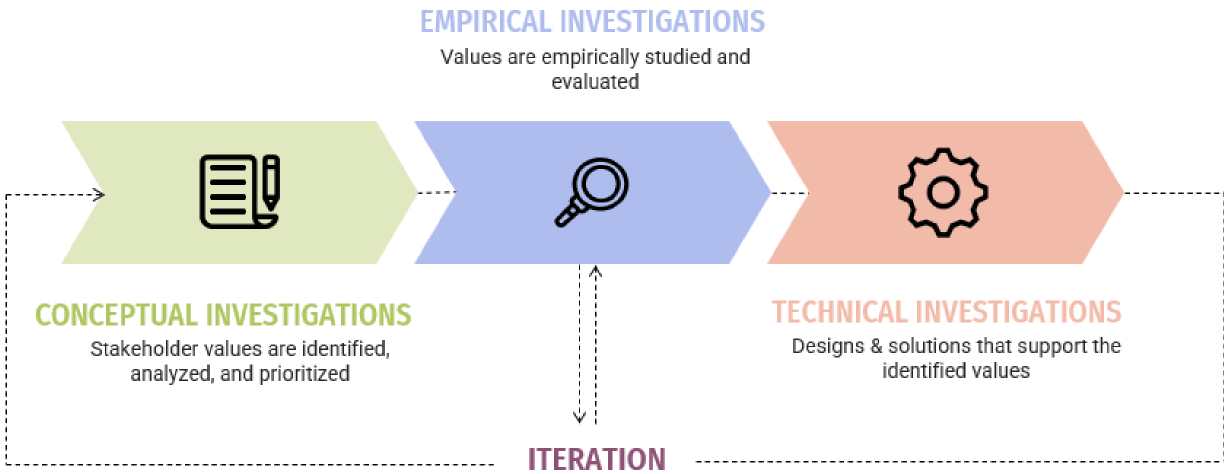


Figure 1.1: Illustration of the Value Sensitive Design framework. Three types of investigations comprise the framework: Conceptual, Empirical, and Technological.

Thesis Statement

Mobile interpersonal telepresence devices can be designed in a fashion that increases the viewer's sense of presence while reducing the streamer's social cost.

Contributions

This dissertation makes the following contributions:

- Application of the VSD framework in a field that is traditionally more technical in nature, to identify human values that should be embedded in future interpersonal telepresence design
- A conceptual investigation into stakeholders and human values through a systematic review of the mobile, one-to-one, interpersonal telepresence literature, including gaps and opportunities in the current research (Chapter 3)

- A technological investigation into the value conflicts that arise from simple design decisions such as bodily camera placement and height, demonstrating how camera placements affect the quality of the viewing experience, and how men and women prefer different egocentric camera heights (Chapters 4 and 5)
- An empirical investigation of human values that should be embedded into prospective interpersonal telepresence prototypes, by giving a voice to users in the design process through co-design (Chapter 6)
- A technological investigation of a replicable, affordable, and unobtrusive interpersonal telepresence prototype based on user input, with a study that compares this prototype to the norm of video chat (Chapter 7)

Organization

The remainder of this dissertation is organized into major chapters that follow the VSD framework; see Figure 1.2 for illustration. Chapter 2 discusses the relevant literature in this space, focusing on general telepresence usage, user perception of virtual environments, and emerging recommendations / best practices for live-streaming. In Chapter 3, we present a systematic review of the mobile, one-to-one, interpersonal telepresence literature, demonstrating how prior work mainly focuses on technological contributions for this paradigm, and how there is a lack of literature to address the socio-technical considerations. Chapter 4 details a user-centered experiment in which we find how camera placement significantly impacts a viewer's experience, and Chapter 5 shows that egocentric camera height is not particularly a major factor to consider as long as it falls within three standard deviations of natural human height, though preferred height does differ between men and women. Chapter 6 illustrates a study in which we identify human values that should be considered when designing interpersonal telepresence prototypes, and Chapter 7 describes an experiment in which we deploy a prototype that fostered these values, where participants compared this prototype to standard voice chat applications. Finally, Chapter 8 synthesizes design considerations for interpersonal telepresence, and provides a roadmap for future work to tackle in this research space.

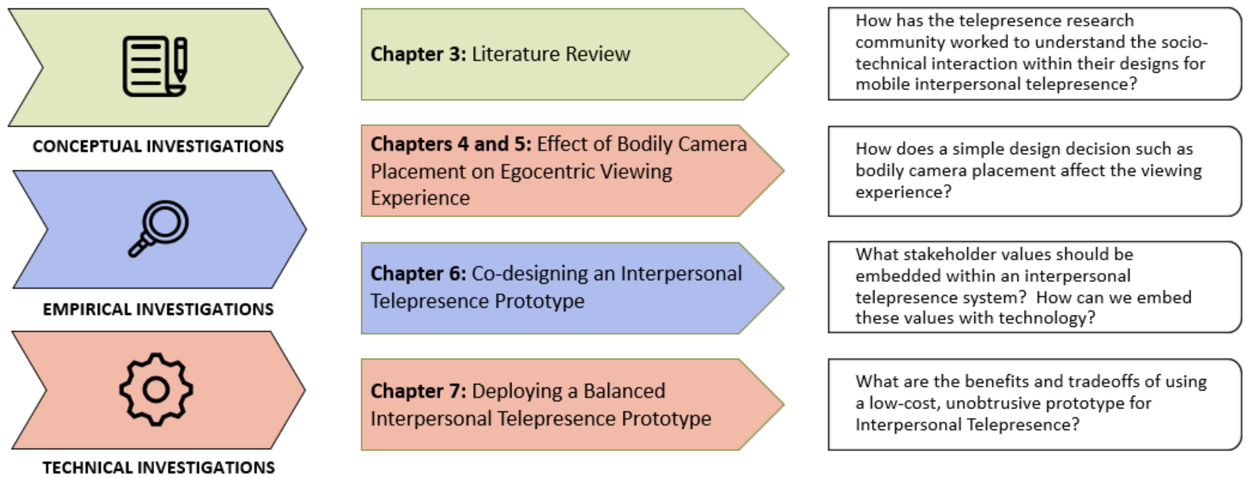


Figure 1.2: Overview of the dissertation. Using the VSD framework, we conduct conceptual, empirical, and technological investigations.

CHAPTER 2: RELATED WORK

There have been numerous research efforts in the field of telepresence, in which the goal is to provide one user with a way to feel as if they were in a remote environment (this is known as *spatial presence* [271]). With interpersonal telepresence, the goal has evolved to let two users feel as if they are with each other (this is known as *social presence* [187]). In this section, we briefly discuss how telepresence design originated in and evolved with human-robot interaction (HRI), and how the field has branched out to this mobile one-to-one paradigm. As researchers have progressed this style with wearable devices that give a first-person view, we find that bodily camera placement impacts the quality of view; and with the inclusion of a 360° camera which grants immersive, egocentric viewpoints, virtual reality (VR) head-mounted displays (HMD) allow the viewer to freely explore an environment at will. However, there has been significant research that shows how human perception of virtual environments — and even real environments when using an Augmented Reality (AR) device — differs from that of the real world, pointing to a potential negative user experience for telepresence cases. We review the extant literature in this space and discuss the implications for proper telepresence design. Lastly, in light of these considerations, we discuss design recommendations that have emerged from the Cinematic VR research, noting how the literature in this space does not approach the problem from a human-centered point of view, instead recommending "one-size-fits-all" design considerations.

The Progression of Telepresence

In 1980, Marvin Minsky conceived the idea of telepresence as a means of remotely controlling robotic platforms over long distances, through the use of immersive sensory feedback, which would allow the controlling user to feel as if they were actually in that environment as they accomplished specific tasks [159]. At a remote workstation, the worker would receive a multi-sensory, egocentric point of view from their robot, and be able to react as if they were actually there; for instance, they could use natural arm movements to remotely grip an object and move it somewhere else [131]. Since this conception, we have seen numerous research efforts emerging from the field of HRI under the moniker of mobile robotic telepresence (MRP) [125]. 3D user interfaces (3DUI) have enabled the direct manipulation of small-scale robots via the use of hand gestures to give commands to these robots [186, 194], and in the field of healthcare, a larger robot — the Da Vinci Surgical System — is used to facilitate minimally invasive surgeries while being controlled by a surgeon via a remote console [247]. Other robots allow the elderly and disabled to interact with family members, healthcare attendants, and doctors, through the use of mobile platforms which provide a natural sense of presence [118, 251]. Although mobile robotic telepresence has been shown to enhance interaction between two friends or family members, end users have also commented that these robotic platforms have major shortcomings; for instance, wheeled robots have difficulty navigating rough terrain [78], and small aerial vehicles (or drones) are not permissible in many public areas [4]. Further, some people have expressed anxiety when collocated with robots [43], and others have commented that these platforms are impersonal, especially in sensitive contexts [34].

Telepresence has begun to evolve past HRI in social contexts, as a means of enhancing common telecommunication by providing face-to-face meetings between two or more remote parties. Typical teleconferencing systems aim to provide a static "window" to a remote location via a direct video feed [96, 181, 275], and researchers have endeavored to create systems which incorporate greater feelings of spatial presence. Microsoft's Room2Room technology affords this by video-capturing a person's body and projecting it into another person's remote location [192]. An evaluation revealed that this method provides a more tangible representation of each party member, enabling more natural communication (i.e. body language), which can be lost in window-type systems. Telepresence has also provided means for more intimate methods of teleconferencing, as a way for families to keep in touch over long distances, more closely emulating a form of constant presence through "always-on" video teleconferencing [96, 97, 179]. Researchers have deployed technology probes in the home, using stationary installations to facilitate a "window" into another household [96, 97]. However, since mobile technologies have advanced rapidly in the previous decade, we have begun to see examples of non-stationary telepresence outside of the home, which we highlight in the next section.

Mobile, One-to-One, Interpersonal Telepresence

As mobile devices such as smartphones and tablet computers have become commonplace, so too has the ability for telepresence. FaceTime and Skype released video chat applications for these mobile devices in 2010. Around the same time, consumer-grade telepresence robots using these devices became available (c.f., [78, 125, 182, 240, 274]), mostly taking the form of "tablets on wheels," although other platforms such as quadcopter drones have been leveraged to provide an unrestricted viewpoint to a remote user [229]. These afford a user with a physical embodiment that they can use to interact in social settings with their remote environment. However, they have been criticized for being impersonal both in professional and social contexts [34], and people have been observed keeping their distance from various robot platforms [43, 258].

At the same time, live-streaming via mobile devices has gained popularity as a way for people to share their immediate experiences with another, who would see the shared environment from the perspective of the device [217]. In lieu of using robotic telepresence platforms, researchers have begun adopting a relatively new approach that leveraged live-streaming mobile devices, where the sharing user wears the device and situates it in such a way that the recipient can experience the remote environment from their perspective. This provides an interesting sense of embodiment and avoids the most common complaints associated with MRP, namely a lack of freedom of movement and a sense of impersonality experienced by those interacting with the robot [50, 134]; even small drones, which are seemingly unrestricted in movement, are not optimal for indoor use, but more importantly, they are prohibited in many places — for instance, drones are not allowed to fly over sports arenas during an event in the United States, where there are regulations set by the federal government regarding legal drone use [4]. New uses of telepresence have emerged that uniquely utilize this one-to-one interpersonal setup. One use scenario is the sharing of a novice worker's perspective to a remote expert, who has the ability to provide guidance through Augmented Reality (AR) cues, overlaying the novice's physical perception with the expert's context-sensitive virtual information [8, 267, 280]. This paradigm has appeared in multiple fields, including medicine [268] and manufacturing/maintenance [28, 168, 249].

Live-streamed video from a worker's first-person perspective also gave rise to new opportunities in manufacturing and maintenance. For instance, expert workers could train, instruct, or provide guidance to novice workers without the need to travel [99]. Using AR cues, trainers can even annotate a worker's view, in order to depict actions that need to be taken in an assembly task [8, 114, 246]. This form of remote guidance leverages telepresence for production-based environments, and this could, in the future, become a norm in the industry. However, this interaction will occur behind closed doors, and there is not a significant social element to it. This is unlike mobile telepresence experiences, in which a user could navigate around a large environment. For instance, the Faroe Islands provided live, remote tours to anyone interested in exploring the islands during COVID-19, using telepresence technology[1]; in a more populous destination, the streamer might encounter third party members who may or may not know that this interaction is occurring.

As such, there are different considerations that need to be made for stationary versus non-stationary

[1]https://visitfaroeislands.com/remote-tourism/

(i.e., mobile) contexts. Progression in mobile, interpersonal telepresence has given researchers pause as they begin to consider the needs of their users, the different social contexts in which they seek to use telepresence, and the different ways that telepresence is used for their interaction [75]. With this change in telepresence paradigm comes a new set of considerations that must be made, especially given the addition of a new human in the telepresence loop (i.e. the on-site user sharing their environment), who has their own socio-technical needs that must be met while functioning as an intermediary for another person's interaction with the remote environment. For instance, placement of the live-streaming camera directly affects the viewing experience, but it also plays a direct role in the streaming user's image and self-image. In the next section we briefly discuss relevant literature that applied cameras in different bodily placements.

Wearable Camera Placements

There has been much work in the HCI community to test different bodily camera placements for telepresence scenarios. Companies like GoPro sell a variety of mounts and straps to support multiple placements, including backpack-based poles, head and helmet mounts, chest and shoulder straps, hand and wrist straps, and more [1, 2]. Many recent academic telepresence projects used placements such as on the shoulder [106, 116, 221, 253], the chest [14, 87], the abdomen [250], the crown of the head [100, 101, 105, 133, 134], over the head (via a pole) [243], in a backpack behind the Streamer [9], and even on the Streamer's face [161, 162].

Every bodily camera placement has benefits and drawbacks. Cameras worn on the chest, abdomen, or shoulder are not affected by the streamer's head turns, but the field of view is potentially reduced. Cameras worn on the head allow for a less occluded view, but as the camera and head are coupled, the viewer can be affected by head movements. To combat this, it is possible to implement image stabilization, [105, 120], but this can introduce latency or reduction of image size. An overhead vantage point via backpack, mounting pole, and gimbal stabilizer could provide stabilization without latency work, but that means the streamer would need to wear additional gear. For a given streamer, though, placing the camera in these locations affects the view height, and it also changes the quality of view. If the camera is a 360° and the viewer watches with VR equipment, then the height of the camera potentially impacts the perception of the environment. Leyrer et al. [138] found that varying viewpoint heights for a given scene can affect distance judgment as well, as is also implied by Kuhl et al. [126]. Banakou et al. [16] studied this effect when participants were given a child avatar body. A recent study suggested putting the camera at a constant height of 4'11" (150cm), but was not in the context of body-worn telepresence [113]. In a robotic telepresence setting, it is possible to move the camera up and down [152], but this is not as feasible when the camera is worn on a person's body, without changing other variables. As such, we expect that camera height and bodily camera placement must be balanced to provide optimal user experience. However, we understand that these factors may affect the viewer's perception of the remote environment. For instance, there is a significant body of literature that documents how humans tend to underestimate distances (and therefore heights) in VR. In the next section we discuss the relevant literature in this space as it relates to a viewer's perception of a remote environment.

Distance Underestimation in Virtual and Augmented Reality

For over two decades, researchers have studied how users perceive distances in VR, and a common result is that users tend to significantly underestimate distances in VR compared to the real world [211]. There are a variety of factors that have been identified as contributors to a user's perception of their virtual environment. For instance, Mohler et al. investigated the inclusion of self-avatars (a virtual representation of the user), and found that these provided a significant frame of reference, resulting in more accurate distance judgements [167]. Likewise, Leyrer et al. show that camera height manipulations in a virtual environment also affects a user's perception; for instance, by increasing the camera height in relation to the user's actual height, the authors found that users tend to underestimate distances [138]. Additionally, Kunz et al. show that the quality of computer graphics inside a VE affects judgements, such that lower-quality textures were more conducive to poorer distance estimations [127]. While these articles are pivotal to the VR distance underestimation literature, we note that these factors (presence of self-avatars, eye height, quality of graphics) can be constant in the real world, and thus constant for immersive telepresence scenarios.

In the extant literature, researchers have utilized multiple procedures in order to measure distance estimations, including verbal judgments, blind walking, timed imagined walking, and blind throwing [211]. Verbal judgments have been used to draw out a user's estimation of depth with a simple procedure, but accuracy tends to decline as targets are located further away [127, 143, 211]. Perhaps the most popular method, blind walking consists of a participant viewing a target, becoming blindfolded, and then walking until they believe they reached the target. This procedure has been used in both real and virtual environments, and historically, participants viewing a VE have walked significantly shorter distances than those viewing the real world [60, 166, 248, 270]. One of the weaknesses of this technique, as reported by Jones et al., is that participants might be able to peek at the ground through the gap between the face and the HMD to use optical flow as feedback [94]. However, a more recent work by Jones et al. suggests that modern VR HMDs alleviate this problem, as wider FOVs fill the periphery of the viewer [95]. Similar to blind walking, timed imagined walking consists of a user viewing a target and becoming blindfolded, but instead of walking to the target, they imagine walking; then they tell the researcher when they believe they would have arrived, if they walked normally. Previous results have shown that this method elicits responses comparable to blind walking [67, 199]. Lastly, blind throwing measures a user's depth estimation while allowing them to remain stationary [207, 218]. Sahm et al. used blind throwing and found that responses were comparable to blind walking; they conclude that this procedure is a suitable measurement for when blind walking is not usable [218].

As noted in a recent literature survey by El Jamiy and Marsh, there is not as much work regarding Augmented Reality (AR) distance estimation compared to that in VR [44], but there are some findings that have emerged in this area. Regarding device type, some researchers have studied the effects of wearing an Optical See-Through (OST) AR display. For instance, Grechkin et al. [67], Jones et al. [93, 94], and Livingston et al. [142] used the NVIS nVisor ST device, which could be used as a VR or AR display. As the participants could see the real world normally through the lenses, the authors were able to directly study if the limited FOV or additional weight of a device caused a difference in depth estimation. Grechkin et al. found no statistical difference between wearing the HMD and normal viewing [67]; however, Jones did find a significant difference in depth estimation, such that users wearing the HMD underestimated distances [93]. This same result was confirmed in another experiment by Jones et al.; the authors found similar findings when using a within-subjects and a between-subjects study design [94]. Livingston et al. compared OST AR distance estimation between indoor and outdoor environments, in the presence of AR cues, and found that participants underestimated distance in an indoor hallway environment, but overestimated distances in an outdoor parking lot environment [142]. Swan et al. used a different OST AR device — the Sony Glasstron — and found that users underestimated distance with this device as well [242].

Fewer researchers still have studied how Video See-Through (VST) displays affect distance judgments of a physical environment. Vaziri et al. used a custom VST device and implemented software techniques to transform the view into a live, non-photo-realistic representation of the environment [259]. Here, the authors were able to study the effects of varying graphical representations of the real world, and found that the non-photo-realistic conditions were conducive to significantly different responses compared to viewing the real world through a VST; however, they did not make a direct comparison to real-world viewing without a VST device. In a separate study, Kyto et al.

studied depth estimation while the user saw AR cues [128]. Instead of measuring how well a user perceived the physical environment through a VST HMD, they used an action-based measurement to understand perception of AR cue depth, with monoscopic and stereoscopic devices. Naturally, the authors found that stereoscopic rendering of AR cues was more conducive to understanding distance. In a recent study, Gagnon et al. used the HTC Vive with ZED Mini attachment in order to study user estimations of lengthier distances in a virtual environment [58]. Using a verbal reporting procedure, they found a trend where users moderately overestimated shorter distances (25m–200m), but then significantly underestimated larger distances (300m–500m). Lastly, we have also shown how field of view directly impacts a viewer's perception of distance in a given environment [151, 197]. When a viewer wears an HMD — or watches a live-stream on their computer — they are not able to perceive the remote environment in the same manner as the streamer. However, there have been some design recommendations to help provide the best possible view. In the next section, we review some of these recommendations that have emerged from Cinematic VR and 360° Videography.

Cinematic VR and Panoramic Camera Videography Design Recommendations

A common feature of virtual reality films generated with 360° cameras is the inclusion of a static, unmoving tripod to capture surroundings. This allows a user to explore a particular view without any occlusion, and prevents simulator sickness caused by a disagreement between the visual and vestibular systems [112]. Though effective, the video is typically constrained to previously-recorded footage. For a more personal experience, people can live-stream a video feed to another while moving about an environment.

360° cameras are still in their infancy and are on the way to becoming more affordable. As such, there is a lack of definitive guidelines to help panoramic content creators optimize their shots.

That said, the current trend is the recommendation of a "one-size-fits-all" value, not accounting for individual differences of users. Many divergent recommendations have been issued for static shots or pictures, such as keeping the camera "eye-level with your subject" [237], "at chest-level" [130], or at "person-height" [200]. As far as distance from the subject is concerned, various resources suggest keeping the camera at a balanced distance [46, 49, 201, 220, 222]. Keskinen et al. found that optimal camera height is around 150cm [113], and Rothe et al. suggested a height of 156cm if the user is standing [216], but these were for static, unmoving cameras.

Telepresence with body-worn cameras cannot completely subscribe to these tips, because there will often be more than one subject in the scene, the streamer will often walk around, and the camera viewpoint is often coupled with the height of the streamer. For instance, a 6'0" user placing the camera over their head would introduce an even taller view. To accommodate a more natural height for a shorter viewer, that streamer could place the camera on the shoulder. This is a seemingly simple design considerations to make, yet — as we will see in this dissertation — it has somewhat drastic consequences. In the next chapter, we will demonstrate how the noticeability of a telepresence interaction causes the streamer to sometimes feel self-conscious, among other social considerations that must be considered when designing a telepresence prototype.

Summary

In this chapter, we discussed a brief history of telepresence and how the mobile, one-to-one paradigm has emerged from previous work. One of the goals of any interaction paradigm is to achieve simplicity of use, and wearable cameras help to achieve this goal; but, a simple design consideration such as this does indeed have more impact than we would expect. From a technical perspective, bodily camera placement and camera height has a direct impact on the viewer's perception of the remote environment, a phenomenon rooted in a significant body of literature that documents how humans tend to underestimate distances in virtual and augmented reality. However, as we will discuss in the next chapter, technical considerations only cover one part of the socio-technical framework; we must also consider the social implications of our design decisions.

CHAPTER 3: IDENTIFYING THE GAPS AND OPPORTUNITIES FOR MOBILE INTERPERSONAL TELEPRESENCE

This chapter is based on work previously published: Kevin P. Pfeil, Neeraj Chatlani, Joseph J. LaViola, and Pamela Wisniewski. 2021. Bridging the Socio-Technical Gaps in Body-worn Interpersonal Live-Streaming Telepresence through a Critical Review of the Literature. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 120 (April 2021), 39 pages. DOI:https://doi.org/10.1145/3449194

Introduction

In order to understand the future of interpersonal telepresence, we first look to the past, with a goal of identifying trends and gaps in the literature. In this chapter, we describe a systematic literature review of interpersonal telepresence works in the span of 2010–2019. In particular, we look for emerging themes that help us identify the human values that are embedded in the current state-of-the-art telepresence devices which can be addressed with future work, in order to enhance user experience for the stakeholders of interpersonal telepresence. Thus, in this chapter, and in accordance with the VSD framework, we present a conceptual investigation into the human values that emerge in the interpersonal telepresence space, through a systematic literature review of 52 articles describing mobile interpersonal telepresence (see Figure 3.1). In this chapter, we ask the following research questions:

- **RQ1:** How has the telepresence research community worked to understand the socio-technical interaction within their designs for mobile interpersonal telepresence?
- **RQ2:** What considerations have been made to accommodate the social needs of all parties

involved in mobile interpersonal telepresence scenarios?

- **RQ3:** What are the technical design implications that arise from changes in the design of mobile interpersonal telepresence, in order to accommodate these social needs?

To answer these questions, we performed a systematic review of the mobile interpersonal telepresence literature emerging from the research community from 2010–2019. We performed an exhaustive search for articles that focused on some form of real-world interpersonal telepresence, where the streamer had the ability to change locations, thus creating a mobile telepresence experience. Additionally, we searched for interpersonal scenarios and not work-related ones, as our focus is on the socio-technical implications of current telepresence design. From there, we conducted a content analysis with open and emergent coding, categorizing our findings under technical or social factors. Overall, we found that the telepresence research community has made an attempt to design for and measure a sense of both spatial presence and social presence (RQ1). However, this was almost entirely done through the use of custom measurements, highlighting an urgent need

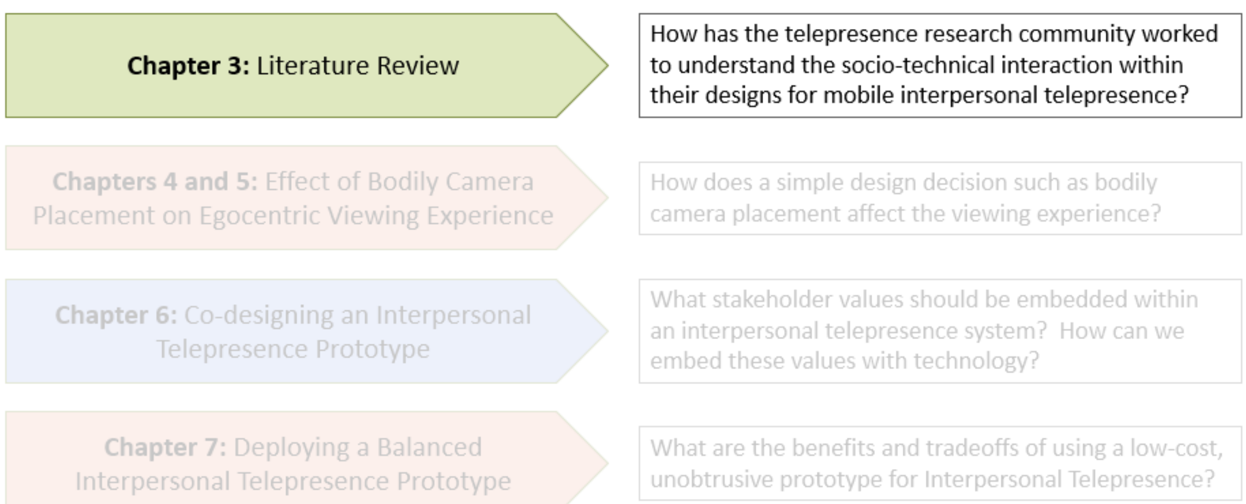


Figure 3.1: In this chapter, we perform a conceptual investigation into human values through a review of the literature from 2010–2019.

for telepresence researchers to create and adopt standardized, validated instruments for measuring remote presence.

While progress is being made to improve the viewer's sense of social presence, which enhances their overall experience, we found that the streamer has been generally disregarded as an important party in the paradigm, with their own specific set of social needs (RQ2). This is especially relevant given that in most telepresence scenarios, the streamer is the one primarily interacting with the remote environment, either on the viewer's behalf or as themselves within the telepresence scenario. Many streamers also complained that the body-worn technology led to socially awkward situations due to how noticeable and conspicuous the technology was. In cases where efforts were taken to hide the body-worn technology, third parties in the streamers' environments became confused. As our understanding of the social needs of all human parties involved in telepresence scenarios grows, we note that researchers have worked to create systems which allow for more natural communication between the viewer and streamer, including the use of 360-degree cameras with head tracking, or varied methods of camera placement depending on the prior relationship of the streamer and viewer (RQ3).

This chapter contributes an in-depth review of the current state of the interpersonal telepresence research community's socio-technical considerations. We accomplish this with a focus on understanding how researchers work to address the social needs of the viewer, streamer, and those third parties in the remote environment, while attempting to preserve the viewer's immersion and the streamer's sense of agency and safety. This chapter also provides recommendations and an agenda of future work for the telepresence community, in order to advance this novel paradigm while ensuring that all participants, whether viewing, streaming, or observing, feel a sense of social equity and enjoyment.

Methods

To answer our research questions, we conducted a systematic literature review of 52 articles which described mobile, one-to-one, interpersonal telepresence. In this section we describe our methodology, which is comprised of multiple phases including a systematic literature search, scoping criteria, and data analysis approach.

Systematic Literature Search

As we are focused on an emerging topic within the computer science discipline, we searched the ACM and IEEE Digital Libraries for relevant articles. We used a combination of search terms that intersected at the type of human interaction (i.e., "human-to-human," "personal," or "interpersonal") and the type of technological interaction ("telepresence," "streaming" "live streaming," "video," or "virtual reality") that together form a telepresence experience. Our search was limited to peer-reviewed articles published in English and between the years of 2010 to 2019. Because Skype introduced video chat to mobile devices in 2010, FaceTime was released in 2010, and consumer-grade omnidirectional cameras supporting live streaming became regular around 2013, we felt that this date range would capture the genesis of mobile, one-to-one, interpersonal telepresence. Our initial data collection was conducted during the time span of December 2019 through January 2020. In September 2020, we updated our search to include articles published in 2020. In total, our search yielded 1,087 articles, which we scoped for relevancy. In the next section, we describe our inclusion criteria.

Scoping Criteria: Interpersonal Telepresence

Though telepresence takes many forms, our research interests are focused on a specific kind mobile (non-stationary and body-worn), one-to-one, interpersonal telepresence. Therefore, we included articles based on the following inclusion criteria:

1. **Article must describe a real-world telepresence experience with, at minimum, video live-streaming (i.e. not a virtual environment).** While it is possible to foster a sense of presence through virtual environments, our research interests are specific to real-world interactions.

2. **Article must describe a human-to-human interaction (i.e., not human-to-robot or robot-to-robot).** This was included as we are only interested in human-to-human interaction which carries an entirely unique set of socio-technical implications than telepresence found within HRI; in addition, although robotics researchers are constantly improving the quality of MRP platforms, they are currently unusable in many scenarios [125, 141]. Thus, we include this criterion to allow for the broadest possible capture of telepresence scenarios to mediate interpersonal relationships where researchers forgo the use of a robot.

3. **Article must describe an experience where the sharing user is non-stationary (i.e., the physical ability to change locations)**. Interpersonal telepresence is noted for its ability to overcome mobility issues commonly associated with MRP, and so we aimed to focus on those scenarios which primarily benefit from a human's natural ability to move. Further, we expect this criterion to filter teleconferencing scenarios, which fall out of scope of our research interests.

4. **Article must describe a recreational context, and not a work-related scenario.** Our research interests align with the study of the social relationships between participating parties

in the everyday use of telepresence, and not those in isolated, task-based scenarios. A wealth of work has already been done to synthesize the research on work-related telepresence (c.f., [59, 135, 246, 249, 267]). Therefore, these contexts were held out-of-scope for the current review.

5. **Article must describe a body-worn device to facilitate the telepresence experience.** This criterion was included because we are interested in telepresence situations in which the viewing users can see the shared environment from the sharer's point of view, opposed to views which are captured from an imaging accessory such as a tripod or mount, e.g. teleconferencing. "Body-worn" devices include a wide range of devices, from cameras mounted in harnesses, to a smartphone stored in a shirt pocket, to a camera which is handheld; we include an article if the sharing user has the ability to constantly manipulate the camera, thus excluding articles if that user cannot control the camera. Since there is an ongoing debate regarding appropriate and ethical use of body-worn devices [54, 180, 279], we anticipated this criterion to also help us identify how prior work tackled this problem.

6. **Article must describe a one-to-one interaction.** This criterion was included because we wanted to focus on this unique, more interpersonal interaction style, where two people have a more intimate means of communication, which could facilitate back-and-forth coordination to accomplish some mutually beneficial goal. We exclude one-to-many and many-to-many configurations because they lack this interpersonal quality of interaction and thus fall out of the scope of our research interests.

For every article returned from the searches, we reviewed the title, abstract, and if necessary, the main body, to ensure it met all criterions listed above. If so, we added it to our dataset. Then, for every article in our dataset, we cross-referenced citations and references, adding any new articles to our final dataset if they also met our inclusion criteria. This was done for every new article in our dataset until no new articles were found. We coded 1,087 articles for relevancy. In total, we identified 52 relevant articles, which we analyzed in our review. In the next section, we describe our data analysis approach and present our final codebook based on our qualitative analysis.

Data Analysis Approach

We performed the qualitative analysis on the 52 articles, identifying emergent codes, using an iterative and grounded approach [132]. We began with open coding and noted the emergent themes found within the dataset; these typically included factors that we explicitly noted, but also include underlying concepts which we found across multiple articles. Our coding was also updated as new codes emerged. A second researcher was consulted during this process to resolve, separate, or merge ambiguous codes. After open coding, a round of axial coding was conducted where we grouped our codes into more cohesive dimensions and themes (as shown in Table 3.1). These described considerations for the interactants, as well as for the interaction as a whole. We then mapped our coded dimensions into two categories — Technical and Social — which respectively relate to our RQ3 and RQ2. The combination of these dimensions then forms the concepts required to answer our RQ1. The Technical dimensions describe the nature and setup of the technology used by the interactants, along with the design considerations made to benefit the sharing individual, the viewing individual, or the ability of the two parties to communicate. The Social dimensions describe the nature of the relationship between the two primary parties and the streamed environment. This includes the roles that the users played in relation to the activity or task being performed, the sharing party's experience in their environment while engaged in telepresence, and how the receiving user interacted with both their partner and the shared environment. In the next section, we present our results.

Table 3.1: Final Codebook Used to Guide our Qualitative Analysis

Themes	Dimensions	Description (*Codes*)
Study Meta-Data	**Primary Methodology**	The type of study conducted in terms of methodology (*Quantitative, Qualitative, N/A*)
	Study Location	Was the study conducted in the lab or the field? (*Lab, Field, N/A*)
	Study Duration	How long the study lasted (*Cross-sectional, Longitudinal, Repeated Measures*)
	Participant Quantity	Total number of participants in the study
	Study Dyads	Were participants paired in the study? (*Yes, No*)
	Country	The country where the study took place
Technical Factors	**Independent Variables**	The technical independent variables within the study
	Dependent Variables	The technical dependent variables within the study
	Symmetry	Equipment setup compared between Streamer and Viewer (*Symmetrical, Asymmetrical*)
Streamer Considerations	**Camera Type**	The video capture device used by the Streamer (*360 Camera, AR Glasses, Camcorder, Fisheye Camera, Smartphone, Tablet, Web Cam*)
	Camera Placement	The place where the Streamer's camera was situated (*Handheld, Head, Shoulder, Chest, Behind the Back*)
	View Stabilization	The method used to reduce jitter in the captured video stream (*Software, Hardware, None*)
Viewer Considerations	**Viewing Apparatus**	Type of video display used by the Viewer (*Computer Monitor, Smartphone, Tablet / Laptop, TV Display, VR HMD*)
	Communication Mode	The modalities afforded to the Viewer to communicate with the Streamer (*Audio, Live Video, Text Messages, Gestures, Gaze*)
	Gestural Communication	Technique used to let the Viewer provide hand gestures to the Streamer (*Graphic Overlay, Puppet / Avatar, Video Stream*)
	Gaze Conveyance	Technique used to let the Viewer provide their gaze heading to the Streamer (*Graphic Overlay, Physical Attribute, Spatial Sound, Vibrations, Video Stream*)
Social Factors	**Independent Variables**	The social independent variables within the study
	Dependent Variables	The social dependent variables within the study
	Proposed Relationship	The type of interpersonal relationship between Streamer and Viewer specified in the article (*Crowdsourced/Strangers, Friends/Family, Long-Distance Relationships, Generic*)
	Proposed Activity	The type of activity specified in the article which the participants were engaged in (*Classroom Attendance, Games, Shopping, Sports, Generic*)
	Agency of Streamer	The level of control negotiated by the Streamer (*Completely Cyranic, Collaborative Control, Completely Free*)
	Embodiment	Type of representation afforded to the Viewer (*None, Streamer is Embodiment, Video Embodiment, Robotic Embodiment*)
	Noticeability	The reason that brought attention to the Streamer (*Not Noticeable, Interaction Style, Wearable Device*)

Results

Our qualitative analysis synthesizes the state-of-the-art of mobile, one-to-one, interpersonal telepresence. We first report descriptive statistics of the dataset in accordance with our final codebook, followed by a presentation of the Technical and Social factors found within the articles. Finally, we discuss the interplay between our technical and social factors, describing how social considerations are integrated into technical designs, and vice versa.

Descriptive Statistics of the Articles

Our dataset (N = 52) was comprised of articles published between 2011 and 2020 (inclusive), but the majority of articles (85%) were published in 2014 or later. Fifty described artifacts, empirical studies, or a mixture of both, but we found 1 theory paper and 1 methods paper. As such, the majority of articles featured technology probes and evaluations. Thirty-five (67%) described a study as part of their contribution. Table 3.2 provides an index of codes related to these studies and the associated articles coded. All studies were cross-sectional, with only 1 exception, which was longitudinal. We did not find any studies with a repeated measures duration. The majority of these (20) were Lab-based, but 15 were Field-based studies.

In the lab-based studies, the primary methodologies were quantitative (17/20), and the remainder used qualitative methods. In the field-based studies, the authors primarily used qualitative methods (13/15), and the remainder used quantitative methods. All 19 of the quantitative studies described an artifact which was then evaluated in some way; 6 of these were evaluations of a new artifact without comparing against a control. The primary data collection approach in the qualitative articles were field notes (7) and interviews (12). Only 1 study used diary methods.

As interpersonal telepresence is contingent on an interaction between two parties, we would expect

Table 3.2: Results of study meta-data from our codebook

Dimension	Code	#	%	Reference(s)
Primary Methodology	**Quantitative**	19	37%	[13, 25, 26, 31, 69, 100, 102, 104, 105, 116, 123, 154, 162–164, 171, 190, 277, 278]
	Qualitative	16	31 %	[14, 35, 40, 86, 87, 90, 115, 160, 165, 175, 183, 202, 205, 231, 243, 262]
	N/A	17	33%	[21, 26, 101, 103, 106, 122, 134, 147, 153, 160, 176, 189, 232, 250, 253, 263, 266]
Study Location	**Lab**	20	38%	[13, 25, 27, 31, 40, 69, 90, 100, 102, 104, 105, 123, 154, 162, 164, 171, 175, 190, 277, 278]
	Field	15	29%	[14, 35, 86, 87, 115, 116, 161, 163, 165, 183, 202, 205, 231, 243, 262]
	N/A	17	33%	[21, 26, 101, 103, 106, 122, 134, 147, 153, 160, 176, 189, 232, 250, 253, 263, 266]
Study Duration	**Cross-sectional**	34	65%	[13, 25, 27, 31, 35, 40, 69, 86, 87, 90, 100, 102, 104, 105, 115, 116, 123, 154, 160, 162–165, 171, 175, 183, 190, 202, 205, 231, 243, 262, 277, 278]
	Longitudinal	1	2%	[14]
	N/A	17	33%	[21, 26, 101, 103, 106, 122, 134, 147, 153, 160, 176, 189, 232, 250, 253, 263, 266]
Study Dyads?	**Yes**	22	42%	[13, 14, 25, 27, 31, 35, 69, 86, 87, 90, 115, 116, 154, 161, 162, 183, 190, 202, 205, 231, 243, 262]
	No	13	25%	[40, 100, 102, 104, 105, 123, 163–165, 171, 175, 277, 278]
	N/A	17	33%	[21, 26, 101, 103, 106, 122, 134, 147, 153, 160, 176, 189, 232, 250, 253, 263, 266]
Article Country	**Australia**	1	2%	[134]
	Canada	10	19%	[14, 35, 87, 90, 183, 189, 190, 202, 232, 243]
	India	1	2%	[231]
	Japan	23	44%	[25–27, 31, 100–106, 147, 153, 154, 160–164, 175, 176, 250, 253]
	New Zealand	5	10%	[13, 21, 171, 277, 278]
	Singapore	1	2%	[263]
	UK	2	4%	[165, 266]
	USA	9	17%	[40, 69, 86, 115, 116, 122, 123, 205, 262]

studies to involve the use of participant dyads. We found 22 studies to do so. In these cases, the quantity of participants ranged from 4 to 68 (M = 21.5, SD = 15.8). The authors typically recruited pairs of participants who knew each other, to have a more ecologically valid study. For the 13 articles only using 1 participant at a time, the number of users ranged from 5 to 26 (M = 12.9, SD = 6.59). These were geared towards understanding a single telepresence role, or evaluating an artifact. Convenience sampling was the primary method to recruit participants.

Overall, our dataset is comprised of 31 unique research projects across all research groups, with a heavy emphasis on the technological capabilities, and a minor emphasis on how people interact with others using these technologies. 29% of the articles were authored by 11 unique research groups, but 37 articles (71%) came from 5 research groups. The University of Tokyo contributed 15 articles, all of which were under advisement of Dr. Jun Rekimoto. Two major projects emerged from this subset "JackIn Head" (9 articles [100–105, 153, 175, 176]) and "ChameleonMask" (4 articles [160–163]). Simon Frasier University contributed 10 articles, all of which were under advisement of Dr. Carman Neustaedter. Each of these articles were unique contributions, and 6 described field studies that garnered in situ feedback. Microsoft Research contributed 5 articles, authored by Dr. Kori Inkpen, Dr. John Tang, or both. Waseda University contributed 4 articles, all of which were under the advisement of Dr. Jiro Tanaka. They were all related to one major project "GoTogether" [25–27, 31]. Fx Palo Alto contributed 3 articles, all of which described the "Polly" project [116, 122, 123].

Scanning the articles for terminology, we found a wide variety of names for the user roles. The most common terms for the remote person sharing their environment, when looking at unique projects instead of articles, was "Local User"; others include Body, Surrogate, Guide, Host, Proxy, and Sharer. By the same fashion, the most common name for the person receiving the shared view was "Remote User"; other names included Ghost, Director, Operator, Consumer, Guest, and Spectator. We also found the term "Local" used to describe this party. This type of terminology depends on the point of view — after all, both users are technically remote in relation to the other. Other metaphors found across our dataset are vivid and help to convey the telepresence concept, but the various terminology might carry negative connotations. To circumvent this confusion, and to keep consistency, for the remainder of this dissertation, we use the terms Streamer and Viewer, as they are respectful to the human. Regardless of the reason or purpose, in mobile, one-to-one, interpersonal telepresence there will almost always be an individual who is sending the live-stream video (a Streamer), and an individual who is watching that video feed (a Viewer).

Technological Factors

In this section, we detail the results of the technological factors as found in our codebook. See Table 3.3 for an index of codes related to these factors, with the articles described by these codes. An emergent theme surrounding interpersonal telepresence work is that researchers tend to frame their technology probes around one of the interactants. They commonly look for new solutions to enhance the overall experience for the Viewer, or they consider new affordances that the Streamer could use to improve the social interaction. The result, however, is an increase in technological requirements to get this experience to work. In this section, we review these technological considerations which have emerged in the dataset to help understand the benefit each factor brings to the experience.

Technology-Centered Independent Variables

As previously noted, 35 articles (67%) in our dataset conducted a study, and of these, only 13 manipulated technical independent variables; we note that only 3 of these were field studies. As such, the more technical articles in our dataset generally evaluated the capabilities of their artifacts instead of deploying them to real world scenarios. One exception was in the work by Rae et al. who, as part of a general telepresence study, deployed two technology probes — one tablet propped up on a tripod, and one body-worn device. Their findings indicate that the body-worn device was logistically less demanding, and more conducive to viewing an environment (due to the

Table 3.3: Descriptive results of technical features. Per dimension, articles which did not describe their code, or which were not applicable, are excluded from this table; percentages may not sum to 100%.

Dimension	Code	#	%	Reference(s)
Symmetry	**Asymmetrical**	41	79%	[13, 21, 25–27, 31, 40, 86, 87, 90, 100, 100, 102–106, 115, 116, 122, 123, 134, 153, 154, 160–165, 171, 175, 176, 189, 205, 231, 243, 250, 253, 262, 277]
	Symmetrical	9	17%	[14, 35, 69, 147, 183, 190, 202, 232, 278]
Camera Type	**360° Camera**	16	31%	[25–27, 100–105, 134, 154, 175, 176, 232, 243, 278]
	AR Glasses	3	6%	[13, 21, 202]
	Camcorder	1	2%	[86]
	Fisheye Camera	1	2%	[153]
	Smartphone	15	33%	[14, 35, 40, 69, 87, 90, 115, 116, 122, 123, 171, 183, 190, 205, 231, 262, 277]
	Tablet	5	10%	[160–163, 250]
	Web Cam	6	12%	[31, 106, 147, 164, 165, 253]
Camera Placement	**Affixed to Prop**	1	2%	[35]
	Behind the Back	2	4%	[154, 243]
	Chest	6	13%	[14, 31, 87, 153, 205, 250]
	Handheld	12	23%	[40, 69, 86, 90, 115, 171, 183, 231, 232, 262, 277, 278]
	Head	19	37%	[13, 21, 100–105, 134, 147, 160–163, 165, 175, 176, 190, 202]
	Shoulder	9	17%	[25–27, 106, 116, 122, 123, 164, 253]
View Stabilization	**Hardware**	3	6%	[116, 122, 123]
	Software	7	13%	[101–105, 175, 176]
	None	39	75%	[13, 14, 21, 25–27, 31, 35, 40, 69, 86, 87, 90, 100, 106, 115, 134, 147, 153, 154, 160–165, 171, 183, 190, 202, 205, 231, 232, 243, 250, 253, 262, 277, 278]
Viewing Apparatus	**Computer Monitor**	14	27%	[13, 40, 87, 90, 106, 116, 122, 123, 160–163, 165, 231]
	Smartphone	8	15%	[14, 35, 69, 171, 183, 202, 205, 278]
	Tablet / Laptop	5	10%	[21, 86, 154, 243, 250]
	Television	3	6%	[100, 115, 262]
	VR HMD	16	31%	[25–27, 31, 101–105, 134, 147, 175, 176, 190, 232, 277]

Dimension	Code	#	%	Reference(s)
Viewer Communication Mode	**Audio Only**	14	27%	[14, 31, 69, 102, 103, 105, 106, 115, 153, 165, 175, 176, 232, 243]
	Audio + Gaze	4	8%	[122, 123, 123, 154, 171]
	Audio + Gestures	1	2%	[253]
	Audio + Gaze + Gestures	4	8%	[25–27, 277]
	Audio + Live Video	15	29%	[35, 40, 86, 87, 90, 147, 162, 163, 183, 190, 202, 205, 250, 262, 278]
	Audio + Text Messages	3	6%	[160, 161, 231]
	Gaze	3	6%	[21, 100, 164]
	Gaze + Gestures	1	2%	[134]
Gesture Communication	**Graphic Overlay**	4	8%	[25–27, 134]
	Puppet / Avatar	1	2%	[253]
	Video Stream	9	17%	[147, 160–163, 190, 262, 277, 278]
	None	35	67%	[13, 14, 21, 31, 35, 40, 69, 86, 87, 90, 100–106, 115, 116, 122, 123, 153, 154, 164, 165, 171, 175, 176, 183, 202, 205, 231, 232, 243, 250]
Gaze Conveyance	**Graphic Overlay**	8	15%	[21, 25–27, 100, 134, 171, 277]
	Physical Attribute	5	10%	[31, 116, 122, 123, 164]
	Spatial Sound	1	2%	[175]
	Vibrations	1	2%	[154]
	Video Stream	4	8%	[147, 190, 262, 278]
	None	30	58%	[13, 14, 35, 40, 69, 86, 87, 90, 101–106, 115, 153, 160–163, 165, 176, 183, 202, 205, 231, 232, 243, 250, 253]

mobile nature of the setup); however, both Streamer and Viewer sometimes felt awkward due to bystanders looking at them [205].

Concerning lab studies, in two articles by Kasahara et al., the authors manipulated the presence of video stabilization, specifically when the Viewer wears a Virtual Reality Head-Mounted Display (VR HMD). Naturally, Viewers scored the Simulator Sickness Questionnaire in favor of the stabilization [102, 104]. In three articles, various researchers studied user preference between regular video chat and panoramic video chat [69, 171, 277]. Each used the Networked Minds Measure of Social Presence questionnaire [22], and collectively found that the use of panoramic video streaming allows for more personable experiences with an increase of social presence. One step further, in an article by Young et al., the researchers found that viewing a 3D representation of an environment was superior to viewing normal panoramic video [278]. Matsuda et al. developed a wearable vibrotactile device that communicated to the Streamer where the Viewer was looking, and compared the use of this device against a similar setup that did not implement it; they found that their device was a convenient addition that allowed directional communication without speaking [154]. In the remainder of the studies, the authors developed multiple iterations of a technique or technology to tackle the same concept and compared them to each other. These included emotional state cues embedded in the telepresence experience [13], gaze-sharing visualization techniques [123], dual-video display type [190].

In summary, the technological IVs found within the dataset were sparse. Studies (particularly in the lab) were more evaluative of a specific piece of technology or compared the researchers' device to a control. In this way, the researchers did not manipulate an IV in their study but were still able to measure the effectiveness of their technology.

Technology-Centered Dependent Variables

Nineteen articles (37%) performed quantitative measurements to evaluate their technologies. Some of these DVs were direct measurements, such as error rate, task completion time, etc.; but others were questions in the form of psychometric scales. None of these technology-centered questions were derived from pre-validated surveys, and were instead custom items. We performed a short qualitative analysis to match these custom items to pre-validated measures based on content of the questions; we used a report by van Baren to help identify a list of related instruments [256]. We found most of the questions to map into the concepts of Usefulness, Ease of Use, and Ease of Learning; as such, the USE questionnaire would be a suitable measure [146]. The remainder mapped into the Sensory Factor concept of the Presence Questionnaire [271], by gauging audio and video clarity.

Capture and Display: Video Stream Considerations

The majority of our dataset (79%) described an interaction in which the technological setup was asymmetrical between the interacting parties. Our dataset supports the paradigm of one Streamer and one Viewer, echoing the notion that the interaction benefits the user who cannot be there in person. As such, the Streamer's apparatus mainly focuses on capturing the remote environment, and the Viewer's equipment is used to display it. A Streamer's camera setup typically falls into two categories: single lens devices such as those embedded in smartphones and tablets, and multi-lens devices such as 360-degree cameras. Because smartphones are an ubiquitous commodity and nearly every model has one or more cameras built-in, they are attractive solutions to generating video stream. By using a video chat application such as Skype and by pointing the camera outwards, any user can capture their surroundings. The Viewer, then, can view the feed on their own smartphone, or on a larger screen such as a desktop monitor. This paradigm was found in 63% of our dataset (e.g. [14, 40, 171, 250]). For a more immersive viewing experience, the Streamer can capture their environment with panoramic cameras, and the Viewer can use a VR HMD. With a 360-degree camera, a Streamer does not need to worry about camera rotation, as HMDs equipped with head-tracking capabilities grant a Viewer the ability to control the angle of their view. This paradigm was found in 31% of our dataset (e.g. [27, 105, 243]).

A small portion of our dataset (17%) described symmetrical technological setups between the interacting parties. In these cases, the common theme is that the two users are friends, families, or intimates who use telepresence to bridge the distance between them in their interpersonal relationships. Instead of only one person being unable to "be there" with the other, both users fit the bill. As such, they either take turns sharing their respective environments [14, 232], or they simultaneously share their surroundings while looking at those of their partner [35, 190, 202]. In this manner, both users are Streamers, and both users are Viewers.

Viewpoint Optimization: Stabilization

Regardless of camera type, in order to ensure success of the telepresence interaction, the video stream must be high quality in terms of viewability. There are multiple considerations that influence this viewability, including camera stabilization. As the camera is mobile and susceptible to jitter caused by the Streamer's walking motion, some authors note that camera stabilization is a necessity; however, only 19% of our dataset implemented a stabilizer, and these 10 articles only represent 2 unique projects. These projects show that there are two ways to stabilize a stream: with software and with hardware.

Kasahara et al. describe their JackIn Head project, in which a 360-degree camera is stabilized using software [104, 105]. The authors intend for their Viewers to use a VR HMD, and note that there is a high chance that they will experience discomfort referred to as simulator sickness [112]. Their stabilization technique calculates optical flow, which is then used across a series of video frames to determine how the camera shifts and rotates. By predicting these movements, a compensating transformation can be applied to a video frame before being displayed to the Viewer. The result is a smoothed and enjoyable view, but as the authors note, their technique essentially decouples the forward heading between the Streamer and Viewer; in other words, it is possible for the Viewer to lose sight of what the Streamer is looking at. However, they were able to show through a user study that their stabilization technique reduced simulator sickness [105].

While simulator sickness is typically associated with VR HMDs, Kimber et al. note that typical mobile video chat applications, e.g. on smartphones, can still produce shaky visuals that are too "unpleasant to watch" [116]. As part of a greater effort, the authors implemented stabilization in the form of a motorized gimbal. The electronic sensors detect subtle shifts in rotation, and it compensates to keep the camera upright and at a constant angle. Unlike software stabilization, this maintains coupling between the Streamer's and Viewer's forward headings. In their Polly project, the authors apply this hardware to a Streamer's setup as a wearable device that sits on the shoulder, allowing their smartphone cameras to produce a steady video feed [116, 122, 123].

Viewpoint Optimization: Streamer's Camera Placement

Depending on where the camera is situated, the contents of the stream can change. For instance, if the camera is worn on the head, the stream could be considered a first-person view of the environment. If it is worn on the shoulder, then the experience could help to facilitate a more social interaction by letting the Viewer see the Streamer's face. In our dataset, we found a variety of camera placements tried by researchers, though no article studied the effects of camera placement directly.

Handheld: A significant portion of our dataset (23%) described scenarios where the Streamer held the camera in their hands, and in most of these cases, the camera was embedded in a smartphone. In only one of these, a 360-degree lens attachment was plugged into the phone, such that the Viewer could watch using a VR HMD [232]. In one other, researchers paired a smartphone with a 360-degree camera, to provide a hybrid view of the environment [278]. In the rest, the commonality is that all of the Streamers held the camera in front of them, so to aim it towards what they wanted the Viewer to see. Jones et al. studied how Streamers manipulate the camera when held in the hand, and reported that many users simply don't operate it in an ideal manner; e.g. they hold the camera too close to their face, or fail to provide an optimal view when trying to target a particular object [90]. They did describe that one of the more positive behaviors was Streamers panning the camera in a complete circle, allowing the Viewer to see the entire surroundings. This particular behavior was encouraged by Young et al., who developed a system that turns a smartphone panorama into a spherical image that the Viewer can navigate on their own [277]. The backbone of this technique was the inclusion of a Simultaneous Localization and Mapping (SLAM) algorithm to generate this panoramic image, which could then be viewed using any kind of device [171, 277].

Head-mounted: Multiple articles in our dataset noted that while handheld cameras are effective, they limit the activities that a Streamer can perform since their hands are busy holding the device [90, 116]. To overcome this limitation, other researchers have developed wearables, which strap, mount, or otherwise couple the camera to the body. The most common wearable camera placement in our dataset is on the head (37%), exhibited in 9 unique projects. Kasahara et al. developed a custom multi-lens 360-degree camera which was worn like a crown, and they noted that this position was chosen because it provides a first-person view [105]. Similarly, Lee et al. developed a helmet which had both a 360-degree camera and a webcam attached to the top. Here, the combination of cameras let the Viewer explore the entire surroundings and simultaneously know where the Streamer was looking [134]. Procyk et al. coupled a miniature camera to a pair of sunglasses, which essentially mimicked the Google Glass device. The camera therefore provided a first-person view, though the Viewer could not explore the surroundings on their own [202]. Although head-mounted cameras provide a first-person view, we note this also causes the Streamer to be effectively invisible to the Viewer most (if not all) of the time. To accommodate a more social interaction, where the Streamer's face might be seen by the Viewer, a different placement could be utilized, such as on the shoulder.

Shoulder-Mounted: The next most common wearable placement was on or over the shoulder (17%), accounting for 5 unique projects. First, we found two unique efforts that described a miniature robotic avatar that sat on the Streamer's shoulder; the avatars were each fitted with actuated features to convey more natural communication. For example, TEROOS by Kashiwabara et al. had eyes and robotic eyebrows to convey emotion [106], and the 20DOF Humanoid described by Tsumaki et al. had little arms to convey gestures [253]. These avatars are somewhat useful in conveying the Viewer's natural communication, but most of the other works with shoulder-mounted cameras focused on conveying the Streamer's gestures and gaze. As previously mentioned, Kimber et al. developed a shoulder-worn device which stabilized a smartphone with a gimbal. The gimbal also served another purpose, which was to grant the Viewer with a way to control the rotation of the camera through a GUI [116, 123]. The unique benefit of this camera placement is that the Viewer can rotate the device towards the Streamer's face, so to have a more social interaction. This is reflected in various works by Cai et al., who used a 360-degree camera mounted over the Streamer's shoulder [25–27]. In their scenarios focusing on collaboration, this camera placement afforded the Viewer with a way to see where the Streamer was looking, as well as a way to see what their hands were pointing at, so to provide the Viewer with a better understanding of what their telepresence partner was focused on.

Chest: We found six other articles (12%) which described placing the camera somewhere on the Streamer's chest. None of these used 360-degree cameras, as the back portion of the viewpoint would be blocked by the Streamer's body; but, this position is highly conducive to understanding the Streamer's immediate environment. In our dataset, this position was chosen for a variety of reasons. Ishak et al. described how chest-mounted cameras were "not overly obtrusive" compared to head-mounted cameras [87], and further, Baishya and Neustaedter noted that by placing a camera in the Streamer's front shirt pocket, the device can be embedded in clothing while not bringing attention to the user [14]. Unique works by Chang et al. and Matsuda et al. describe how, by hanging a rig around the neck of the Streamer, the camera can rest around the chest while providing maximal comfort compared to other placements [31, 153].

Behind the Back: The final bodily camera placement found in our dataset was behind the Streamer's back, accounting for two unique articles (4%). To accommodate this position, the Streamer typically wears a backpack which is outfitted with a pole, hoisting up the camera. The resulting perspective is akin to a third-person view of the Streamer. In one article by Tang et al., the camera was a 360-degree device, allowing the Viewers to explore the entire surroundings. The authors employed a behind-the-back camera for a remote collaboration scenario, and they chose this position because it provides a Viewer with a way to understand where the Streamer is looking (by seeing their head), in addition to freeing up their hands for gesturing and manipulation [243]. In the other article by Matsuda et al., the authors developed a custom pipe rig to hold up the camera. In their work, they also developed a vibro-tactile device that conveyed 360-degree directions using vibration motors; this allowed the Viewer to see the complete surroundings and "point" to where they wanted the Streamer to look [154].

The various camera placements found in our dataset have been chosen for different reasons, though we note that none of the articles studied different camera placements directly. The Viewer's opportunity to explore the environment is enhanced with a head-worn or chest-worn camera, but their social experience may be greater when a handheld or shoulder-worn camera is used, because they can see the Streamer.

Communication Modalities from Viewer to Streamer

We found a variety of modalities used to support the Viewer's communication to the Streamer, with some articles providing a mixture of multiple channels. Unsurprisingly, verbal communication was supported in 81% of the dataset. Text messages were rarely used (6% of the dataset), but when included, they were auxiliary modes of communication [160, 161, 231]. In certain setups where the Streamer used a smartphone, a live video feed of the Viewer was displayed in the remote environment; this was found in 29% of the dataset. Interestingly, we did not find any articles that used emojis or emotes, most likely due to our scoping criteria; we focus on one-to-one interactions, where paralinguistic digital affordances are more appropriate for broad-audience interactions [76].

Besides verbal communication, humans often perform gestures such as pointing and head nodding in order to provide context to their spoken communication. As reported in a variety of articles, these cues are typically lost in telepresence experiences because the Streamers and Viewers often cannot see each other [90, 243]; as described above, some of the researchers noted that they selected a specific camera placement because it allowed the Viewer to see the Streamer's head and hands, helping to overcome one piece of the problem. The use of a video conferencing application on a smartphone also lends itself to supporting natural communication between the Viewer and any third-party members, but because the camera (and screen) are often pointed outwards, the Streamer typically cannot see the Viewer. Thus, conveying natural communication from the Viewer to the Streamer is a more difficult challenge. Many articles attempted to overcome this limitation with new technology probes to support hand gestures, convey gaze, or both.

Hand Gestures: We found fourteen articles (27%) to support the Viewer's hand gestures during the telepresence experience. One of these articles by Tsumaki et al. leveraged a miniature humanoid avatar which sat on the Streamer's shoulders, equipped with robot arms to convey gestures [253]; though not specified by the authors, it is possible to harness three-dimensional user interfaces (3DUI), e.g. the Microsoft Kinect, to capture real arm movements and map them to the robot [131]. Four unique projects across 7 articles (13%) described integrating a video stream of the Viewer's hands. In the ChameleonMask project, Misawa and Rekimoto used computer vision to detect the hand and display it to the Streamer. Here, the unique aspect of the project is that the Streamer is wearing a helmet which has both an HMD and a tablet display; within the HMD, the Streamer can see the Viewer's hand gestures. The third-party members can see the Viewer's face on the tablet, but cannot see the their hands, and are instead asked to interact with the Streamer's arms (such as shaking hands) [160–163]. Similarly, Young et al. extracted the Viewer's hands while they explore the Streamer's surroundings with an HMD (their smartphone is the display, similar to Google Cardboard); as they naturally point to an object, their phone picks up their hand and overlays it to the Streamer's feed. Though the Streamer generates the telepresence image, both users interact with it; the authors refer to this as a "shared spherical mapped environment" [277].

Four other articles (8%) used a similar technique, but instead of displaying the Viewer's actual hands, the authors instead provide a virtual representation. Lee et al. and Cai et al. discuss how they used 3DUI techniques to detect the Viewer's hand while they wore an HMD (here, both use the Leap Motion device). The hand position and rotation in the Viewer's local space is then mapped to a 3D model, which is displayed to the Streamer as a virtual overlay. In order to achieve this, the Streamer was equipped with an AR display [25–27, 134]. In these examples, the Viewer's hands must be in front of their face in order to be detected, so the hand gestures are dependent of gaze. As a natural extension, with an adequate tracking solution (e.g. HTC Vive's outside-in tracking), a Viewer's hands could be detected independent of gaze. Therefore, the virtual overlay solution might be more efficient when the Viewer wants to point at one or more objects which do not fall within their gaze heading. However, it remains to be seen which method is more efficient at communicating hand gestures, or which technique is more preferred by the Streamer.

Gaze Conveyance: Nineteen articles (37%) supported some type of gaze conveyance from the Viewer to Streamer. One strategy found in 8 articles (15%) is the inclusion of a virtual representation. Similar to cases where the authors wanted to display hand gestures using a 3D model, the Streamers wore an AR display, and certain cues were included to show where the Viewers were looking. For instance, Cai et al. displayed a virtual head which looked towards the direction of where the Viewer was gazing. Thus, depending on the rotation of this head, the Streamer could tell where their partner was looking [26, 27]. Lee et al. took a different approach, by displaying a simple rectangle that represented the region where the Viewer was looking. Additionally, if both users were looking in opposite directions, an arrow would be displayed to indicate where their partner was gazing [134]. Billinghurst et al. developed a similar approach that they call a "radar", where virtual overlays indicate the direction of the other user's gaze [21]. In these articles, the Streamer was using a 360-degree camera so that the Viewer could explore the entire surroundings; but, gaze conveyance is still beneficial to panoramic smartphone users, as shown by Muller et al. In their work, by leveraging rotational tracking of the phone, the authors were able to detect where on the panorama each user was looking [171].

Other researchers concluded that it is pertinent to be able to show the direction of the Viewer's gaze, but they did not use an AR display. Five articles (10%) spanning 3 unique projects relied on a physical attribute of the telepresence setup instead of a virtual representation. As previously described, in the Polly project, a shoulder-mounted smartphone is attached to a gimbal; using a GUI, the Viewer is able to manually control the rotation of the phone. Thus, the Streamer can look at the phone to infer where the Viewer is looking [116, 122, 123]. Likewise, Chang et al. used a webcam which was worn around the Streamer's neck; it was actuated such that the Viewer can change the direction of the camera, and the Streamer could thus determine where their partner was looking [31]. One unique project by Misawa et al. leveraged a miniature, face-shaped screen which sat on the shoulder of the Streamer. The screen was 3D and a back-projector displayed the Viewer's face. As the authors describe, this technique is superior to a flat screen since it overcomes the "Mona Lisa effect" [164].

Three articles (6%) used raw video stream in order to convey what both users were looking at. In two of these, the users acted as Streamers and Viewers simultaneously, and their viewpoints were overlapped as both users wore a VR HMD. For instance, Pan et al. used this technique in order to bridge the physical distance between partners in a long-distance relationship (LDR) [190], and Maeda et al. used this technique as part of a larger project, theorizing that the Viewer could learn motors skills by watching the Streamer's first-person view while performing some task [147]. In a different article by Young et al., a SLAM algorithm was used to reconstruct an environment in 3D which was shared between two users; then, live video streams were spatially mapped in this reconstruction, so the other user could see exactly where the other was looking.

In one article, Nagai et al. used spatial audio in order to convey where the Viewer was looking. The Streamer, equipped with a proper auditory display such as surround-sound headphones, would hear the voice of their partner coming from a specific location, which changed depending on their gaze. Using rotational tracking, it is possible to calculate which direction the audio needed to come from [176]. Finally, one article by Matsuda et al. described using an array of vibration motors resting upon the Streamer's neck in order to subtly convey where the Viewer was looking; depending on which motor was being activated, the Streamer could "feel" the Viewer's gaze direction.

Social Factors

While most of the articles in our dataset were contributions from more technical groups, we found some emergent themes within the social realm as well. The foundational goal of telepresence is to grant the feeling of "being there," but the intepersonal telepresence paradigm transcends sensory stimulation, and thus this goal cannot be achieved unless social needs of the Streamer and the Viewer are also met. In this section, we review the social factors identified from our dataset and describe how each of these have been tackled in a variety of ways. See Table 3.4 for an index of codes related to these factors, with the articles described by these codes.

Human-Centered Independent Variables

As previously noted, only 35 articles (67%) in our dataset conducted a study, and only 9 of these used social factors as independent variables. Of these nine, four described a field study, but all were cross-sectional. In other words, none studied the long-term effects of various telepresence usage. The most common social IV was Scenario, in that the researchers deployed their technology probe for different activities, but this did not seem to have an effect on the reported findings [90, 104, 105, 205]. One manipulated Acquaintanceship, where the third-party participants were asked to shake hands with either strangers or friends, in a social touching experiment [163]; here, the results indicated very little effect of this IV. Since in natural communication humans can infer

Table 3.4: Descriptive results of social aspects. Per dimension, articles which did not describe their code, or which were not applicable, are excluded from this table; percentages may not sum to 100%.

Dimension	Code	#	%	Reference(s)
Proposed Relationship	**Crowdsourced**	5	10%	[40, 87, 162, 163, 231]
	Friends/Family	14	27%	[25, 27, 35, 69, 86, 90, 115, 116, 164, 183, 202, 205, 243, 262]
	Generic Use	30	58%	[13, 21, 26, 31, 100–106, 122, 123, 134, 147, 153, 154, 160, 161, 165, 171, 175, 176, 189, 250, 253, 263, 266, 277, 278]
	Long-Distance Relationships	3	6%	[14, 190, 232]
Proposed Activity	**Classroom Attendance**	1	2%	[87]
	Games	1	2%	[231]
	Generic Use	31	60%	[13, 14, 21, 31, 69, 86, 90, 100, 102–106, 122, 123, 134, 153, 160–163, 175, 176, 189, 190, 232, 253, 263, 266, 277, 278]
	Shopping	3	6%	[26, 27, 164]
	Sightseeing / Tours	10	19%	[25, 40, 115, 116, 154, 165, 171, 205, 243, 262]
	Sports	6	12%	[35, 101, 147, 183, 202, 250]
Agency of Streamer	**Collaborative Control**	21	40%	[25–27, 31, 40, 87, 90, 106, 115, 116, 122, 123, 134, 147, 154, 171, 205, 231, 243, 250, 262]
	Completely Cyranic	5	10%	[160–163, 165]
	Completely Free	13	25%	[13, 14, 21, 35, 69, 86, 183, 189, 190, 202, 232, 277, 278]
	Indeterminate	13	25%	[100–105, 153, 164, 175, 176, 253, 263, 266]
Viewer's Embodiment	**No Embodiment**	26	50%	[13, 14, 21, 25–27, 31, 40, 100–105, 134, 153, 154, 171, 175, 176, 190, 202, 231, 232, 243, 277]
	Streamer is Embodiment	5	10%	[160–163, 165]
	Robotic Embodiment	4	8%	[106, 147, 164, 253]
	Video Embodiment	14	27%	[35, 69, 86, 87, 90, 115, 116, 122, 123, 183, 205, 250, 262, 278]

Dimension	Code	#	%	Reference(s)
Noticeability	**Not Noticeable**	4	8%	[13, 14, 100, 165]
	Interaction Style	12	23%	[40, 69, 86, 90, 171, 183, 202, 231, 232, 262, 277, 278]
	Wearable Device	33	63%	[21, 25–27, 31, 35, 87, 101–106, 115, 116, 122, 123, 134, 147, 153, 154, 160–164, 175, 176, 190, 205, 243, 250, 253]

where another is looking by observing their body language, another work attempted to see if people can still determine this by observing the physical characteristics of their custom telepresence setup Polly [123]; although the time it took to make this consideration was statistically significant in favor of a human's body language, the authors note that it was still relatively fast with their prototype. Lastly, two articles describing one study manipulated whether two users video-streamed their own surroundings simultaneously, or if one of the users was watching from home, in a bicycling scenario; the authors noted that the experience was more engaging if both users shared their own surroundings to the other opposed to one watching from home [35, 183].

Human-Centered Dependent Variables

Regarding the measurable dependent variables in the articles which report a study, 16 articles (31%) performed quantitative measurements that were human-centered in nature. Nine of these articles used pre-validated measures; these included the Simulator Sickness Questionnaire [112], Networked Minds Measure of Social Presence [22], Social Presence Questionnaire [230], Spatial Presence Questionnaire [226], and the Affective Benefits and Costs of Communication Technologies questionnaire [276]. However, we did find the strong majority of measures to be custom questions. We performed a short qualitative analysis to match these custom items to validated measures based on content of the questions, using a report by van Baren to help identify a list of related instruments [256]. We found many of these to map to the concepts provided by the

Networked Minds Measure of Social Presence [22]; but, others mapped more accurately into the Social Presence concept as measured in the GlobalED questionnaire [68]. According to Cui and as reported by Oh et al., and as manifested in our dataset, the term "social presence" seems to be an umbrella term used to describe multiple concepts [39, 187]. Other questions mapped into various concepts such as Usefulness, Satisfaction, and Ease of Use as measured in the USE questionnaire [146], the Control concept measured in Witmer's & Singer's Presence Questionnaire [271], and the Immersion concept from the Slater-Usoh-Steed questionnaire [234]. As expected, researchers in this area tend to attempt to measure social presence, as that is one of the ultimate goals of interpersonal telepresence, but the dataset indicates that there is a need for more standardized and validated questionnaires when attempting to measure these constructs.

Proposed Interpersonal Relationships

While mobile robotic telepresence affords an individual with a way to feel present in a remote environment on their own volition, interpersonal telepresence grants this presence by leveraging the Streamer. We look to our dataset in order to understand who this Streamer can be, and how they relate to the Viewer. While most of the dataset did not propose their work for a specific relationship context (58%), we did find twenty-two articles (42%) to discuss or apply their work for one, including crowdsourced/strangers, family & friends, and LDRs.

Friends and Family: The most common of these were for friends and family, represented across 14 articles (27%). Although telepresence lends itself to shortening large distances between two parties, we find it is still useful to employ this interaction for cases where one member simply cannot be with another during the day. For instance, Procyk et al. describe an activity where two users enjoy geocaching; instead of collaborating to find a single geocache, the users split up to hunt two different ones while sharing their progress to the other [202]. Chua et al. describe a similar technology probe for biking, where either two parties take separate paths and shared their surroundings with their partner, or one party would stay at home and watch the video feed of the other [35]. Kim et al. deployed a technology probe to understand how one person could shared their everyday experiences to someone who is watching from home [115]. Further, Kimber et al. discuss how telepresence can be used to let disabled users experience the world through the eyes of a Streamer [116]. Between these articles, we find that friends and family members have opportunity to enhance their relationships by performing joint activities remotely, or at least by sharing their experiences to a loved one who cannot be with them due to geographical distance or even health concerns.

Long Distance Relationships: There is a significant body of work regarding in-home computer-mediated communication for long-distance relationships, and we found 3 articles (6%) in our dataset to specify this relationship context for mobile situations. Similar to the above, these 3 articles followed 2 main applications for the technology probes. Pan et al. described a case where both users wore an HMD and could see their partner's first-person view overlapped onto their own. Though this study was performed in a lab with makeshift activities, both users reported a sense of togetherness, which is crucial in an LDR, so the concept could easily be applied to real-world activities [190]. Baishya and Neustaedter show how by leveraging "always-on video chat" two users can see what their partner is doing on a whim; so, if they cannot be together, they can at least have a sense of understanding what the other is experiencing [14].

Crowdsourced/Strangers: The final interpersonal relationship proposed by authors in the dataset was crowdsourced/strangers, or the absence of a developed interpersonal relationship. Five articles (10%) described cases where they anticipate the Viewer to have a specific goal in mind, and the Streamer helps to achieve those goals. de Greef et al. hypothesized the ability to link up a disabled or hospitalized Viewer with a random Streamer to facilitate remote sightseeing [40], and Misawa and Rekimoto discuss the possibility of a "physical body marketplace" where the Viewer can enlist the help of an unknown Streamer to perform remote tasks for them [162, 163]. In work by Singanamalla et al., Viewers tell Streamers where to go as they navigate an "Escape Room". Similar to Misawa and Rekimoto, the authors describe their vision of Streamers earning money by performing remote tasks for Viewers [231]. Ishak et al. performed a similar study, hypothesizing that in the event a student could not attend class, a Streamer could be physically present while streaming the contents to the Viewer [87]. These works demonstrate that telepresence does not necessarily hinge on both parties knowing each other prior to the interaction. Streamers could be compensated for their time, akin to ride-sharing services like Uber or Lyft. However, it remains to be seen what activities are appropriate for interactions involving strangers.

Proposed Telepresence Activities

We analyzed the articles in our dataset to understand any trends in activity that the Streamer and Viewer could be involved in as part of the telepresence interaction. Most of the dataset (61%) did not deploy their technology for a specific activity, but we did find the rest to discuss a handful of interesting applications.

Sightseeing / Remote Tourism: One of the most prominent activities in our dataset is sightseeing / remote tourism (19%). Typically geared towards friends and family, these articles helped bridge the distance between loved ones. Muller et al. described their method of sharing an interesting environment by using a smartphone to generate a panorama, which the other user could explore on their own volition [171]. Rae et al. conducted a field study in which they asked Streamers to walk in an indoor museum [205]. In a field study by Kim et al., participants used their technology probe for going to the beach, touring their city's Chinatown, and hanging out at a park, among other activities [115]. In essence, all of these articles describe a situation where the Viewer could understand the environment that their Streamer partner was in, during a slow-pace activity. However, we did find some other articles which discussed applying telepresence to more intense scenarios.

Sporting Activities: Six articles (12%) proposed the application of telepresence towards more action-based scenarios. Tobita hypothesized that the Viewer could take a passive approach by acting as a "coach," giving advice to the Streamer who is engaged in playing tennis [250], and Kasahara and Rekimoto suggested that Viewers could watch a stream of paragliding to feel like they were flying (although they did not directly try it in their article) [101]. Other researchers, using a symmetrical setup, allowed both users to act as a Streamer and Viewer simultaneously, providing a way to perform joint activities remotely. Procyk et al. used such a paradigm for joint geocaching [202], and Chua et al. used it for bicycling [35]. In a unique case, Maeda et al. hypothesized that telepresence with a first-person view could allow a Viewer to learn muscle memory by watching the Streamer. In their work, a Viewer attempted to juggle by watching how the Streamer an expert was able to perform [147].

Other Various Activities: In the remaining five articles, the authors hypothesized a variety of other activities, including classroom attendance [87], shopping [26, 27, 164], and games [231]. In the shopping scenarios, the Viewer has more input in the overall interaction; at times, the Streamer even asks for help from the Viewer to find an object on the shelf. As evident in our dataset, the current research focuses less on *what* this type of telepresence can be used for, and instead focuses more on *how* telepresence interaction can take place. Most of the articles in our dataset did not specify a particular activity, or hypothesized one or many without performing a study in that context. Although the various contexts of relationship and activity might drive telepresence design choices, we expect that any given user might have their own wants and needs. Rae et al. performed a survey to understand what end users might want general telepresence for, and they found their respondents to specify a wide variety of use cases which fall within the purview of interpersonal telepresence, including sightseeing, shopping, and sports, as shown above. Other activities not found in our dataset include conference attendance, major life events, and enjoying a meal together [205].

Agency of Streamer

An emergent theme found in our dataset is that the Streamer was tasked to behave in a variety of ways, and were sometimes asked to relinquish their free will in order to achieve the goals of the Viewer. These behaviors do not stem from a particular technological setup, but rather from social expectations regarding how the telepresence interaction should work. While we could not determine any behavior in 13 articles (25%) due to their more technology-focused nature, we found three main behaviors emerge from the rest of our dataset, following a spectrum of agency.

Completely Free: On one end of the spectrum, the Streamer has complete agency and behaves as normal; this was found in 25% of our dataset. In these cases, the user turns on their video streaming apparatus and the Viewer simply watches without giving much direction. In an intimate scenario, Baishya and Neustaedter describe the use of "always-on video chat" in order to provide members of an LDR with a first-person view of their partner's environment. Here, the users might not be aware that their partner is watching, and they go about their daily lives as normal [14]. Procyk et al. describe simultaneous video-streaming between a pair of users, so that one could see the viewpoint of the other as they both engaged in geocaching for recreation [202]. Various works discussed video-streaming various activities while a Viewer watches from home. For instance, users streamed their biking experiences in the work by Chua et al. [35, 183], and parents were able to stream their children's playtime in work by Inkpen et al. [86]. In summary, the Viewer can watch the Streamer's video feed in order to feel like they are "along for the ride;" but, for a more social experience, collaboration is often employed.

Collaborative Control: In the middle of the spectrum, twenty-one articles (40%) described letting the Streamer maintain their agency while taking Viewer feedback into consideration. In these articles, the common theme is that the Streamer and Viewer work together to complete some task. For instance, Tang et al. described a field study where participants were asked to take pictures of specific landmarks; here, the Viewer has prior knowledge of these landmarks, and provided input to the Streamer in the form of directions. The Streamer still acted normally, while taking into account the Viewer's input [243]. Likewise, Cai et al. described a scenario where a Streamer physically goes to a store, and the Viewer uses the telepresence interaction to help find specific items on the shelves. In this case, the Viewer can be providing a guidance-like service to the Streamer, but the Streamer could also be shopping for the Viewer on their behalf [26, 27]. Venolia et al. describe a field study in which one person remained home, but received a video stream as their loved ones explored a museum. Here, the Viewers were able to freely communicate with the Streamers, and the authors employed a technique which allowed both parties to take a picture together, so that it seemed like two parties were actually together. By using a camera and green screen technology, the Viewer's background was stripped, such that their image could easily be inserted into a live picture [262]. Ishak et al. describe a scenario where the Streamer attends class for the Viewer. Here, the Streamer might not even be enrolled in the class, but can be a conduit for the Viewer to virtually attend when they cannot physically be present [87]. Thus, collaboration can be used to achieve the goal of either user or both.

Completely Cyranic: On the other end of the spectrum, we found 5 articles (10%) spanning 2 unique projects that discuss how the Streamer purposefully relinquishes their agency in order to achieve the task of the Viewer (a cyranic interaction). The term "cyranic" has been used in prior work to describe a social encounter where one person acts and speaks exactly as told via discrete commands of a remote person; as described by previous research, this mirrors a concept found in the play "Cyrano de Bergerac" where the titular character secretly tells another exactly what to say (see [36]). One article in our dataset by Mitchell et al. applied a literal cyranic interaction in an effort to understand how the third-party interactants reacted to this concept [165]. A Streamer engaged in conversation with these third-party members while speaking exactly as the Viewer commanded. The authors noted that the participants started off wary of the strangely-behaving human, but eventually understood and accepted the paradigm. In one other project, Misawa and Rekimoto deployed their ChameleonMask, in which a Streamer wore a tablet device on their face which displayed the Viewer, so to provide the illusion that they were the Viewer [160, 161]. In this project, the Viewer provided discrete audio messages to the Streamer, who behaved exactly as commanded; then, the Viewer spoke using a public audio channel to the third-party interactants. The authors hypothesize that this paradigm can be used in an Uber-like service, where a Streamer is paid to let the Viewer "borrow" their body [161, 162].

Viewer Embodiment

In robotic telepresence, the Viewer is typically afforded a wheeled platform which grants them a physical representation in the remote environment; but in interpersonal telepresence, this embodiment is typically lost. We looked to our dataset to understand not only how Viewers can perceive a remote environment, but how the environment can perceive the Viewer. In our dataset, we found 26 articles (50%) to not provide any type of embodiment to the user. In these cases, the Streamer uses their camera to provide a viewpoint of their surroundings, but the Viewer's communication, if any, does not reach anyone but the Streamer. For instance, in the JackIn Head project, the Streamer wears a custom 360-degree camera on their head, and does not have a display with which to project the Viewer back into their environment [105]. This form of telepresence does not lend itself for many social interactions, but other researchers have explored a variety of ways to overcome this challenge.

Robotic Embodiment: In some of the earlier works found in our dataset, researchers tried to use robotics to provide a physical avatar which is tethered to the Streamer; this type of interface was found in 4 articles (8%). The most prominent cases were described by Kashiwabara et al., who created TEROOS a camera with eyes and actuated eyebrows [106] and Tsumaki et al., who made their humanoid robot which had actuated arms [253]. All of these articles were published prior to 2013, and the state of the art telepresence has since then shifted towards use of more commercial hardware (e.g. smartphones and video chat, or 360-degree cameras).

Video Embodiment: Fourteen articles (27%) discussed the use of a camera embedded into a smartphone or tablet to generate the live-stream. A collateral outcome of using these devices is the ability to pipe the Viewer back into the remote environment; in this way, the Viewer is embodied by the video device. In the Polly project by Kimber et al., the Viewer is not only displayed in the remote environment, but they also have the ability to control the physical device's orientation [116]. Likewise, in the work by Ishak et al., the Streamer wears a smartphone on their shoulder, allowing the Viewer to appear in a small window. Though some Viewers thought that the smartphone screen was too small (and thus did not provide them with an adequate sense of embodiment), others indicated that they wanted to keep a low profile [87]. This technique may be effective in situations where family members are geographically distributed. In a work by Inkpen et al., a remote family member could watch young children play, and the kids could see the faces of their loved ones [86].

Streamer is Embodiment: The final type of representation found in our dataset relied on the actual Streamer to serve as the Viewer's surrogate, and thus serve as their embodiment. Five articles (10%), found mainly in the ChameleonMask project, described the Streamer acting directly on behalf of the Viewer (see 4.3.5 Agency of Streamer, above). Misawa and Rekimoto discuss having the Streamer wear a tablet device on their face, and the video provides an illusion that the Viewer is the one present before a third-party interactant. Although this concept technically fits in the video embodiment category, the authors add an additional layer by having the Streamer perform physical actions as directed by the Viewer [162, 163]. Mitchell et al. describes their scenario in which the Streamer acts completely and intently on the Viewer's behalf, as a literal "avatar"[165].

Noticeability

An emergent theme in multiple field studies found in our dataset was the social awkwardness of the Streamer as they used a telepresence setup. We would anticipate this finding to manifest when body-worn devices are used, but even more modest technology probes led to this negative result. For instance, Rae et al. deployed a shoulder-worn smartphone, and some Streamers felt like they were being stared at [205]. Likewise, participants in the field study of Kim et al. noted that even simple video chat resulted in a higher level of embarrassment. The authors here note that a design implication is to make a technology probe as inconspicuous as possible [115]. We thus analyzed the proposed technologies in each article of our dataset to understand how noticeable they were. We found that 45 articles (87%) described a noticeable interaction, either due to the fact that the Streamer is wearing equipment in a conspicuous way (33 articles), or because passersby were able to see or hear the telepresence taking place (12 articles, e.g. the Viewer's audio was not localized to the Streamer). Only 4 articles had inconspicuous setups, and in each of these, the camera was embedded into the clothing of the Streamer in a way that did not draw attention to them. For instance, Baishya and Neustaedter had the users place their smartphone in their front shirt pocket [14], and Procyk et al. generated the stream by attaching a small camera to the frame of a pair of sunglasses [202]. In these examples, the clothing itself was not conspicuous, and the cameras were well-hidden. We note, however, that this does not completely solve this problem of social awkwardness; one of the bystanders in the work by Procyk et al. believed that the Streamer was talking to them (when they were actually talking to the Viewer), and a fight almost ensued [202]. We do not yet understand what the best way to convey that a telepresence scenario is occurring while balancing the needs of all parties involved.

Emergent Themes from Qualitative Findings

Although most of the articles in our dataset were contributions from the more technical side of HCI, we did find some important qualitative findings and emergent themes, which we feel warrant an echo in our work.

Viewer Dependence on the Streamer: In some articles, the Viewer participants indicated that the interaction was simply not good enough. Some users thought watching a video stream became stale after extended use [35], and while some felt that telepresence was a decent substitute, it was not as good as actually "being there" [202]. This might be due to the fact that generally, the Viewer cannot control where the Streamer goes [243]. Other participants reported that they did not enjoy being dependent on the Streamer to look around the environment, as it made them feel like they were sitting "in a wheelchair" [205]. From a Streamer's point of view, having to control the camera for their partner made users feel like a prop, resulting in them feeling "used" [205]. This theme of dependence was typical of technical setups which leveraged single-lens cameras, e.g. smartphones; though not exclusive to smartphones, we find this feeling of dependency to be reduced when using a 360-degree camera which affords the Viewer with a way to explore the environment on their own accord.

Privacy Challenges for All Users: There was a minor theme regarding privacy. Many of the Viewer participants were unable to adequately gauge the social surroundings of the Streamer, and were found to modulate the strength of their audible communication while being conscious of the words they were saying, so to not offend anyone in the remote environment [202, 205]. Further, some were not convinced that the live-stream was secured to just their use, and again were conscious with their utterances. In a longitudinal field study for members of a long-distance relationship, Baishya et al. deployed an "always-on" video chat solution, so that one person could see their partner's view on a whim; even in an intimate relationship, this on-demand interaction caused a degree of uneasiness, as the users did not have the ability to prevent their partner from seeing or hearing certain situations [14]. Regarding privacy of third-party members, we did not find much consideration, but most of the field studies were conducted in North America, where video-streaming in public seems to be more accepted than other parts of the world. However, this is certainly an issue which needs to be addressed with future work.

Concerns about Streamer Safety: We found another minor theme regarding the safety of the Streamer. As previously mentioned, a passerby felt offended by one of the participants, who was actually talking to their telepresence partner. This almost caused a fight [202]; but worse, one Streamer was too enraptured by the hand-held technology probe that they were almost hit by a car [90]. In one case, the telepresence interaction was too inconspicuous that it provoked a third-party member; in the other case, it was too involved that the Streamer nearly lost their life. As noted above, a strong portion of the articles discussed a technology probe that was noticeable to third party members. Even when engaging in simple video chat, Streamers felt a sense of awkwardness [115, 202, 205]. They were self-conscious about using the technology probes; in these cases, physical harm was not apparent, but psychological harm was. It is clear that too obvious of a technology probe may cause third-party members to take notice, but hidden devices which cause a noticeable interaction still cause confusion for these individuals. There is a clear need for balance, in order to maximize the needs and desires of all parties involved.

Preferred Streamer Characteristics: Lastly, we also found a minor theme regarding the characteristics of an "ideal" Streamer. In a survey by Ishak et al., survey respondents were asked to help identify what matters when choosing a Streamer for the hypothetical task of remote classroom attendance. They found that students did not particularly have a preference for characteristics such as gender, race, or age, and instead indicated a desire for certain performance-based traits, such as extroversion, good note-taking skills, and academic standing [87]. On the other hand, Misawa and Rekimoto, in their technology probe where a Streamer is presented like the Viewer, suggest that to enhance the illusion that the Viewer is the one interacting with third-party members, the Streamer should have identical physical traits to the Viewer [160]. For non-performance-based use cases, it seems that Streamer characteristics are not a major consideration; for instance, the purpose of remote sightseeing is to share an environment, and generally, we suspect Streamer traits will not affect user experience. However, for performance-based scenarios such as sporting or remote shopping with a stranger, certain traits such as languages spoken, physique, and knowledge of the environment might influence the success of the interaction.

Interplay between Technical and Social Factors

The previous sections describe the technical and social factors of interpersonal telepresence design separately, but in our analysis, we find that these are very intertwined. Social considerations breed technical considerations, and vice-versa, but each of these stems from an attempt to optimize a certain aspect of the overall user experience. In this section, we discuss the links between these factors and how they influence the design of telepresence experiences.

Technology Breeds Social Considerations

When designing an interpersonal telepresence experience, researchers might think that adding more physical technology can be used to improve certain aspects of the interaction. To some extent, this is certainly true. However, our dataset informs us that more body-worn technology leads to more social considerations, especially on the part of the Streamer, who must bear the additional weight of the socio-technical burden. For instance, as reported by Ishak et al., the size of a body-worn smartphone influenced how third-party members perceived the Viewer [87]. A smaller screen yields a smaller image of the Viewer, and a larger screen yields a larger image. This affects how others can communicate with them, and how they perceive themselves in this communication paradigm.

Different technical setups have been reported to affect immersion, co-presence, and social presence, and studies have shown that the Viewer's apparatus leads to differences in these psychometric scales [115, 277]. One of the main purposes of interpersonal telepresence is to let the remote user feel like they are "actually there," so it is natural that researchers have developed various artifacts in an attempt to increase these scores; but, these artifacts come with social cost. As reported in some field studies, the Streamer can feel socially awkward while using telepresence due to wearing a device on their body, or because they are engaging in a non-normal activity. In particular, researchers have posited that to free the hands of the Streamer, they can wear the camera on their body; but, this is currently not a social norm, and bystanders have been found to give "strange looks" in public [202]; in other words, the telepresence interaction is noticeable, which makes the Streamer stand out [35, 115, 202, 205]. Further, there is strong interplay between technology, social considerations, and measurable outcomes. In our dataset, we found some authors to make a call for camera stabilization. Using immersive viewing apparatus, users could succumb to simulator sickness, and software stabilization has been used to help prevent this [102, 104]. Other researchers used a hardware-based approach, integrating a wearable device into the technical setup a Streamer uses [116, 123]. This could, unfortunately, result in a greater social cost given the increased noticeability.

Social Landscape Influences Technological Considerations

In our dataset, when designing prototypes or studies, the authors often had a specific social context in mind. In the cases which proposed strangers interacting with the telepresence setup, the primary concern was to let the Viewer see the remote environment opposed to seeing the Streamer. For instance, some authors deployed a body-worn camera placement that excluded the Streamer's face [40, 87, 162, 163]. This is in stark contrast to the articles proposing friends/family and LDR interactions, where the camera was often placed on the Streamer's shoulder or was handheld, providing opportunity to capture the their face [27, 35, 69, 123]. In these cases, streaming the remote environment or activity was a co-objective to providing a more social experience. In addition to making the interaction more social, providing the Viewer with a way to see the Streamer's face unlocks natural communication. In our dataset, the purpose of many papers was to provide or demonstrate collaboration capability, and other articles noted that simple telepresence setups are conducive to the participants missing out on what the other is seeing (e.g. "did you see that?" [115]). To overcome this, the Streamer can employ a 360-degree camera, which allows the Viewer to see the remote environment and, when placed properly, the Streamer too. In various field studies such as those by Rae et al. [205], Tang et al. [243], and Jones et al. [90], the authors note that typical telepresence setups do not capture natural communication in its entirety. To provide the Viewer with a way to understand that of the Streamer, simply placing the camera in a way that allows them to see the hands and face have been shown to suffice. However, additional hardware is required for both parties in order to capture and display that of the Viewer. For example, when a 360-degree camera is used to capture the environment, to communicate hand gestures, researchers often employ specialized depth cameras such as the Leap Motion. To communicate gaze, researchers often employ head-tracking devices such as those found in the HTC Vive HMD. Then, to display these to the Streamer, researchers employ AR glasses [25–27, 134].

Discussion

Participant feedback within the field studies of our dataset reveal moderate comfort with the telepresence concept across numerous concepts, and as such there is much opportunity to improve the state of the art. In this section, we outline paths for future work in a call to action for the SIGCHI community.

The Search for Meaningful Dependent Variables. What is Success?

The interpersonal telepresence interaction style is fairly new, and we noticed that there were very few studies to investigate real-world applications. More work is needed in the form of longitudinal field studies to understand how this novel interaction is perceived by all parties involved. In particular, we need to understand how to make this acceptable over a long period of use, especially when its novelty would wear off. Additionally, as apparent by the vast quantity of custom measurements used in most of the dataset, the telepresence community needs to converge upon valid methods of measuring meaningful outcomes. We noted that most of the items found in the dataset were custom, often in the form of singletons, and these could have been replaced by pre-validated questionnaires. As illustrated in work by DeLone and McLean in the field of Information Systems [41], "success" can be measured by various dimensions and that, in order to move a field forward, researchers must converge on a subset of measures that define success. In the field of HCI, many outcomes have been measured through the use of pre-validated instruments (e.g. USE questionnaire [146], Slater-Usoh-Steed questionnaire [234]); but, we acknowledge that many of the questions being asked by the researchers did not map to these instruments, suggesting that we lack robust ways to operationalize the socio-technical outcomes we are interested in measuring in the context of interpersonal immersive telepresence. Therefore, as a research community, we need to engage with one another to reflect on the social (e.g., connectedness, support, well-being), contextual (e.g., time to complete a task, accuracy, immersion in the environment) and technical (e.g., video fidelity, ease-of-use, usefulness) outcomes that we care about when determining success in our field. Also, when future work develops new measures, researchers should make sure to psychometrically validate these instruments and share them with the community.

Regarding dependent outcomes that have been commonly studied in interpersonal telepresence, there have been two primary goals — to achieve the sense of "being there" in the remote environment, and to achieve the sense of "being there" with a remote partner — and to accurately measure how we make advances in these areas, the community should consider adopting a standard set of pre-validated instruments to measure both spatial and social presence. The Slater-Usoh-Steed questionnaire and the Presence questionnaire by Witmer and Singer have been widely adopted to measure presence in virtual environments, and they can also be used here to measure spatial presence [234, 271]. In essence, a remote environment captured by a camera for telepresence is akin to a high-fidelity virtual environment, so the questionnaires apply, although the items will need to be adapted to ensure the language makes sense for this specific context. The research community has a more difficult challenge for determining a common definition for social presence; as this interaction style has a strong social component, it is important to make this identification. Oh et al. suggested that co-presence and social presence are one in the same [187], but the most commonly used measure (the Networked Minds Measure, [22]) lists co-presence as a sub-construct of social presence. We found multiple definitions for this concept across studies of our dataset, so it is difficult to compare the results of one study to another. Cui made this same observation in 2013, noting that the literature diverges from a "commonly accepted paradigm" [39], and to our knowledge, this is still an unsolved issue.

In order to make fair comparisons (and thus be able to improve the state-of-the-art), we as a community need to work towards understanding how to best make that convergence. However, with the field of interpersonal telepresence still being in its infancy, we do acknowledge that standardized measurements may not always be able to address future research questions. It is out of the scope of this work to develop these new measures, but we wanted to bring this issue to light such that future researchers can help solve this problem. At the very least, researchers should perform an appropriate psychometric analysis on their measurement tools to ensure some level of construct validity.

This would allow future researchers to compare and contrast their technology probes against previous work in a more standardized manner, while also preserving their freedom to create custom instruments.

The Forgotten Users: Streamer Considerations are Understudied

One of the most jarring findings across our dataset is a common disregard of the Streamer. Many projects seem to posit that the Streamer is a replacement for a telepresence robot, and do not attempt to measure their experience, instead focusing just on the Viewer. The interpersonal telepresence community needs to remember that these people are stakeholders too, and we need to make explicit considerations for them. Many of the Streamer participants in the field studies found in our dataset reported a sense of social awkwardness when they performed their role, yet, the strong majority of articles described a telepresence setup which drew attention to these users, either because the interaction was not localized to them, or because they were wearing extra, conspicuous technology on their body. In order to maximize the Viewer's user experience, researchers tend to add more and more technology to the Streamer's setup; but this could be making matters worse. The existing research has already revealed that the various live-streaming techniques (e.g. smartphone and 360-degree cameras) provide the Viewer with a strong sense of presence, so it is now time to shift the community's research agenda to optimize the Streamer's experience.

Firstly, instead of just measuring the Viewer's sense of presence, we need to measure social presence for both parties. After all, a successful social interaction is contingent on both. Only one article in our dataset attempted to measure social presence of both users, and future work needs to make this a high priority. We also need to develop a way to better understand how a given telepresence setup would cause a person to feel a sense of social awkwardness, before, during, and after use. To our knowledge, there is no pre-validated measure to quantify this, but it would be valuable such that we can influence future design to help reduce it. Secondly, we need to consider that by optimizing the Viewer's experience, we might be worsening that of the Streamer. Likewise, by optimizing a Streamer's experience, we will probably reduce the Viewer's experience. In various field studies of our dataset, authors conclude that social awkwardness stems from the wearable devices [115, 205], especially as the addition of physical technology becomes more apparent to the third party. Even as wearable technology becomes more compact and less noticeable, we must remember that the technology's design and its intended use will affect a Streamer's desire to interact with it. Thus, a goal of future telepresence work should be to strive for experience optimization of both end-users, understanding that their social needs may not align in a given telepresence scenario.

Thirdly, we must consider the physical safety of the Streamer as they interact with interpersonal telepresence. In a few dataset articles, the authors noted that Streamers entered into dangerous scenarios due to the novel interaction, e.g. nearly getting into a fight with a passerby [202], or almost getting hit by a vehicle because their attention was drawn to the technology probe [90]. We also note that because they are often asked to wear cameras on their body, they stand out in public; crime reports indicate that tourists are often targeted for theft because they stand out, particularly because they have expensive camera equipment [73, 80]. Since remote sightseeing is a prominent use case found in our dataset, we acknowledge that telepresence users could be targeted too. We must remind ourselves that Streamers are not tools or robots to facilitate telepresence; they are people and we must ensure their safety as they interact with interpersonal telepresence. They take on all of the risks, and the research community must think of ways to protect them.

The telepresence community has long since focused on the needs of the Viewer, and it is now time to shift our focus towards the Streamer. Additionally, as telepresence itself becomes viable as an entertainment product, there is now a class of streaming-based workers that operate entirely via telepresence, raising questions as to the workers' rights and fair expectations of these users. We call for researchers within the telepresence field to make considerations for Streamers' well-being, especially in situations where third parties are involved in the telepresence environment. For instance, previous work regarding public live-streaming for games shows how live-streamers can become physically and mentally drained [209]. We note that these considerations change depending on the relationship between Streamers and Viewers. However, if telepresence is now itself undertaken as an occupation, serious work needs to be done to affirm the rights of these telepresence workers, exploring the various social contexts in which this work occurs, and developing a framework of social and moral norms for professional telepresence Streamers, which can also be extended as guidelines to non-professional interpersonal telepresence. We anticipate future work to emerge in the form of participatory and iterative design with real users, to brainstorm and implement technology probes which will reduce the potential harms and ethical issues Streamer users may face.

The Social Dynamics Between Streamers and Third Party

In addition to reducing the risks posed to the Streamer, we must also consider the needs of third party members. As Rae et al. note, very little of past research considers the third party as a stakeholder in a telepresence interaction [205], and we found this to be the case in our dataset as well. As a community we need to ensure that we account for this group of people as we conduct our research. In the United States and Canada, where most of the field studies in our dataset took place, video recording in public is viewed as a legal right [124, 239], so in common areas there is no reasonable expectation of privacy. In other cultures, such as Japan, it has been observed that even in public, live-streaming is frowned upon socially [70]. Researchers applying telepresence in part of a field study need to consider the culture in which they plan on conducting studies; but even if the law and culture both approve, we still need to consider the ethical implications such as socially accepted conventions of privacy.

As previously noted, an avenue for future work is to minimize the technology such that the Streamer does not stand out; this will help to reduce the social awkwardness caused by third party members staring at them. This will also reduce the third party's ability to spot that telepresence is taking place. It is questionable, then, if this is an appropriate way of conducting a telepresence experience. It is out of the scope of our work to provide ethical guidance regarding the third party members, but we pose the following questions to the research community in hopes that future work will tackle this issue: is it necessary to inform third party members that they are being captured as part of a public telepresence experience? How can we balance the rights of these people while protecting the Streamer? Answers to these questions will bring forth a more enjoyable experience for all parties involved, including those who have been thus far been an after-thought in telepresence research.

We note that even if the telepresence technology can be minimized and hidden from plain sight, some Streamers might still feel uneasy, knowing that they are potentially intruding on the privacy of other people. Likewise, we note that some Streamers will not perceive telepresence as some form of social intrusion, even if the technology probe was excessively obtrusive to wear. As noted in one field study, some participants felt socially awkward, while others "enjoyed the attention drawn" to them due to wearing a new, visible piece of technology [205]. Thus, it might be the case that individual differences cause users to have various attitudes towards acting as a Streamer. It would be beneficial to researchers and end users if we could predict their level of comfort. For instance, researchers could apply personality models such as the Myers-Briggs or Big-Five paradigms to understand the Streamer's personality as it relates to their willingness to engage in a public telepresence experience [65, 173]. We hypothesize that it is possible to make some prediction of a Streamer's comfort when using a telepresence prototype, but this will be dependent on identifying the social factors which help with this prediction. If we are able to make a distinction with high accuracy, then we could facilitate the ability to recommend a specific type of technological setup that would provide the most social comfort to a potential Streamer.

Technical Recommendations for Social Telepresence

Our dataset revealed several primary objectives of interpersonal telepresence: letting the Viewer explore the remote environment; letting the Viewer naturally communicate with the Streamer; and letting the Viewer communicate with the remote environment. The technology probes tackle one or many of these objectives, resulting in various ways they are used, and there is still opportunity to enhance the state-of-the-art by developing additional technologies.

Getting to "Actually There": Enhancements for Exploration

As is expected of a telepresence interaction, all of the technology probes discussed a way to let the Viewer explore the remote environment, to some degree. Cameras ranged from single lens devices to omnidirectional devices, and though the type of video differs, participants were still able to feel a sense of spatial presence regardless of type. This means that the more expensive equipment (360-degree cameras and VR HMDs) might not be absolutely necessary in order to fulfill this goal, though HMD use does increase immersion. None of the articles particularly discussed audio features, but microphones are commonly embedded in consumer-grade devices. Smartphones typically have a single microphone, and 360-degree cameras sometimes have directional microphones (e.g. the Ricoh Theta V[1]) which can provide a more immersive feeling when paired with a VR HMD that supports directional audio. So far, and as found in our dataset, the state-of-the-art telepresence setups facilitate the exploration of a remote environment only through visual and auditory cues, but there is opportunity to engage the other major senses to truly make the Viewer feel "actu-

[1]https://theta360.com/en/about/theta/v.html

ally there." For instance, we can leverage techniques from the VR community to simulate various types of haptic feedback. Wilberz et al. recently demonstrated an HMD-mounted device which can let a Viewer feel heat and wind of a virtual environment [269]. Brooks et al. were also able to provide an illusion of varying temperatures using the sense of smell [24]. These output devices can be paired with sensors in a Streamer's setup, to measure wind speed and temperature [64]. We suspect that these techniques will increase the Viewer's perceived sense of spatial presence, especially in outdoor environments, and future work is needed to understand the magnitude that each brings to the experience.

Integrating Natural Communication between Streamer and Viewer

Next, the majority of the articles in our dataset describe facilitating communication between the Streamer and Viewer, and we hold that this is a primary objective especially for cases where the two parties are friends, family, or intimates. In our dataset, communication between these users is typically tied to auditory cues, with a subset of technology probes using simple video chat. As noted, natural communication (e.g. hand gestures and gaze) has not been studied too much, but there are a variety of ways to convey these features. Interestingly, none of the articles in our dataset leveraged an eye-tracking device to pin-point Viewer's gaze, but that is a natural extension to the current literature. Eye-tracking hardware is available for flat screens and HMDs, and even as a mobile wearable[2], so both parties can use these to enhance their communication while exploring the same environment together.

[2] https://www.tobiipro.com/product-listing/tobii-pro-glasses-2/

Lastly, a smaller subset of our dataset discussed the ability for the Viewer to interact with the remote environment, i.e. to communicate with the third party members. In most of these articles, the Viewers were embodied via video, meaning the Streamer used a smartphone screen to show them in real time. The commonality here is that none of the Viewers used a VR HMD, because if they did, their faces would be blocked by the device. This results in a reduced sense of immersion and, potentially, social presence. To overcome this problem, we suspect that techniques such as Rekimoto's "Behind the Mask" prototype could allow the Viewer to have a video embodiment while still using a VR HMD [210]. This technique uses multiple cameras to capture an HMD wearer's face even though they have the device on their head. After stitching the different camera feeds together, the result is a reconstruction of the user's entire face, which could be sent back to the third party members in the remote environment.

There is plenty of future work that the telepresence community can perform in order to enhance the socio-technical user experience. We note, however, that it requires a better understanding of each user's physical and social needs, which may conflict at times, given the nature of the telepresence scenario and the users' relationships. It is pertinent that we realize how to provide a balance such that interpersonal telepresence will be desirable by Streamers, Viewers, and the third party, allowing for a natural, enjoyable, and equitable means of communication.

Limitations and Future Work

We conducted a comprehensive literature review of mobile, body-worn, interpersonal telepresence, used for non-work scenarios. These criteria were chosen as they describe a new form of telepresence that has only recently gained traction within the research community. However, this restrictive scoping may have limited our ability to include other interpersonal telepresence scenarios which could also hold relevance to this main topic of inquiry.

Most notably, we excluded work-related telepresence scenarios, as in most cases these are stationary, and only involve a social dynamic between the Streamer and Viewer, i.e. there is no third party. However, these dynamics are still interpersonal and vary widely, depending on the existing relationship between both parties, and can contribute to the discussion about ensuring equitable experience between parties. With large companies beginning to adopt remote technologies in the midst of the COVID-19 pandemic, including expert help via AR, more study should be conducted in order to ensure that the unique social implications of work telepresence are properly investigated. For instance, research implies that there might be a new class of employees who rely on AR for most of their day. Introducing new technology especially that which must be worn on the body raises ethical questions and human factors issues. Future work must address these considerations to ensure that all parties in the work environment are treated equitably.

Many of the articles in our review did not explicitly detail the social relationship between Streamer and Viewer, and so we were unable to properly quantify them. Future research should ensure that the nature of these relationships is more clearly articulated; but, we acknowledge that interpersonal telepresence may be used for a wide variety of use cases. In our review, we excluded telepresence scenarios that involve a one-to-many user relationship, which occur in situations such as Twitch streaming and mobile broadcasting. While these scenarios involve fewer opportunities for interpersonal communication, they nonetheless provide other forms of interaction modality that are important to consider as we explore new social dynamics within interpersonal telepresence, such as parasocial relationships. This is especially relevant given that these forms of telepresence are emerging as forms of paid entertainment, and not just for recreational purposes. In our future work, we plan on investigating proper communication modalities that can be used in this one-to-many paradigm, while providing a strong sense of social presence within these online communities.

Conclusion

Mobile interpersonal telepresence is a relatively new concept, and our review of this field has revealed many interesting projects that are pushing the technology past its current limits. We remind ourselves that we must progress not just through technical means, but by also considering the social needs of the users for whom we intend to develop this new interaction paradigm. We call upon the telepresence community to take our findings into consideration in order to enhance Viewer's experience, reduce the Streamer's social cost, and standardize the methodology we use to ensure we are making developments in the right direction. By doing so, we can achieve a more socially equitable interaction that can enhance many lives. Especially in trying times of social distancing, we can bring people together and facilitate exploration of remote environments.

This chapter revealed the current state of the literature in the interpersonal telepresence space, highlighting how researchers want to enhance the feelings of *being there*; for the Viewer stakeholders, this means helping the user have stronger spatial presence, but for both Viewers and Streamers, we must work to enhance social presence. The dataset is more technical work in nature, but field studies revealed how Streamers value more socially acceptable technologies, and that Viewers want to have the ability to explore the remote environment on their own. In essence, there is a struggle for autonomy between these two major stakeholders that is difficult to resolve. In the next chapters, we demonstrate how simple design decisions such as bodily camera placement cause value tensions; for instance, Streamers can choose a placement that follows their own preferences, but in doing so, they might limit the opportunity the Viewer has to explore.

CHAPTER 4: AN ANALYSIS OF USER PERCEPTION REGARDING BODY-WORN 360° CAMERA PLACEMENTS AND HEIGHTS FOR TELEPRESENCE

This chapter is based on work previously published: Kevin Pfeil, Pamela Wisniewski, and Joseph J. LaViola Jr. 2019. An Analysis of User Perception Regarding Body-Worn 360 Camera Placements and Heights for Telepresence. In *ACM Symposium on Applied Perception 2019* (*SAP '19*). Association for Computing Machinery, New York, NY, USA, Article 13, 110. DOI:https://doi.org/10.1145/3343036.3343120

To continue our work towards an understanding of how to balance interpersonal telepresence design using the VSD framework, we next describe a technological investigation that demonstrates how a simple design decision such as bodily camera placement can cause a value tension — in this

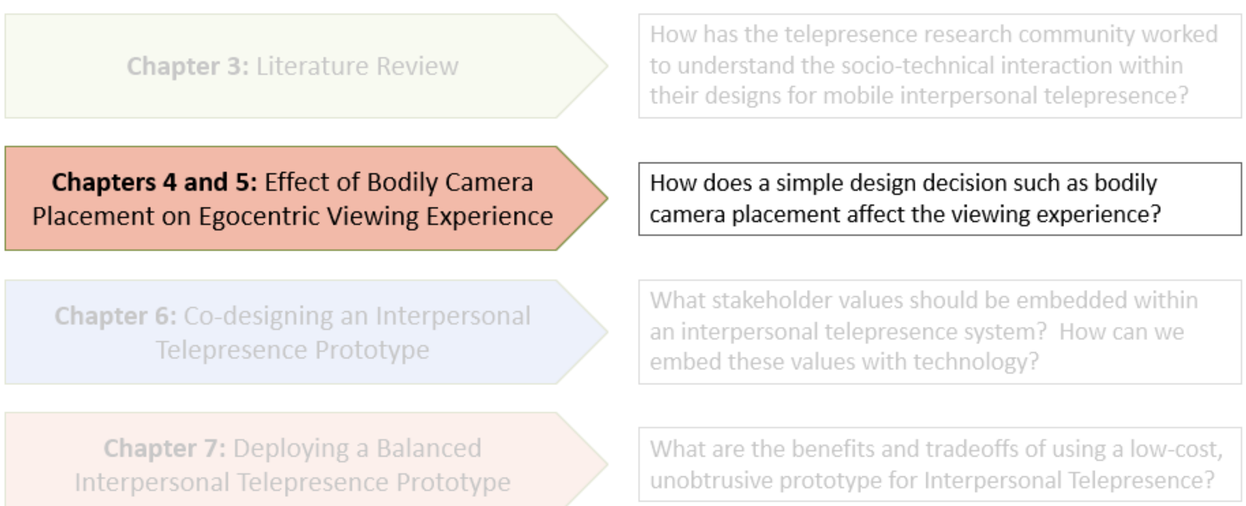


Figure 4.1: In Chapters 4 and 5, we perform a technological investigation in which we explore how a simple design decision — camera placement and height — causes a tension of values.

case, a struggle for autonomy (see Figure 4.1). In robotic telepresence, height is not arbitrary, and previous efforts have described techniques to change camera height in real-time, through mechanical actuation [152, 204]. In interpersonal telepresence, camera height is tethered to a streamer's physical height, as well as the location where they wear the camera on their body. Thus, the purpose of this study was two-fold: first, we wanted to understand how a streamer's height (and thus camera height) would influence a viewer's user experience. Secondly, as a streamer could change camera placement (but not necessarily control height), we wanted to understand how various placements influenced the viewing experience.

This chapter is divided into four major sections: Introduction, Methods, Results, and Discussion. We first describe the problem space in the Introduction. Second, we discuss our approach towards tackling the problem by designing a user study, with design decisions informed by the previous literature, in the Methods section. Third, we describe the results of our user study, including descriptive statistics and hypothesis testing, in the Results section. Lastly, we provide design recommendations and implications for future work in the Discussion section.

Introduction

Live-streamed video is an increasingly popular content medium. In 2015, Periscope hit a 10 million account milestone [245], and in 2017, Twitch.tv saw 2 million monthly broadcasters [254]. 360° video streaming is a new technology that is envisioned as the "next big thing" [15, 66]. Viewers often watch these panoramic videos via social media and streaming websites [241], but there is an opportunity to create an immersive telepresence experience by sharing live-streamed content to users of Virtual Reality (VR) head-mounted displays (HMD). There are a number of use cases where this might be useful. For instance, a geographically distributed family could reconnect with their elderly [273]. A person suffering from social anxieties could remotely explore the world with

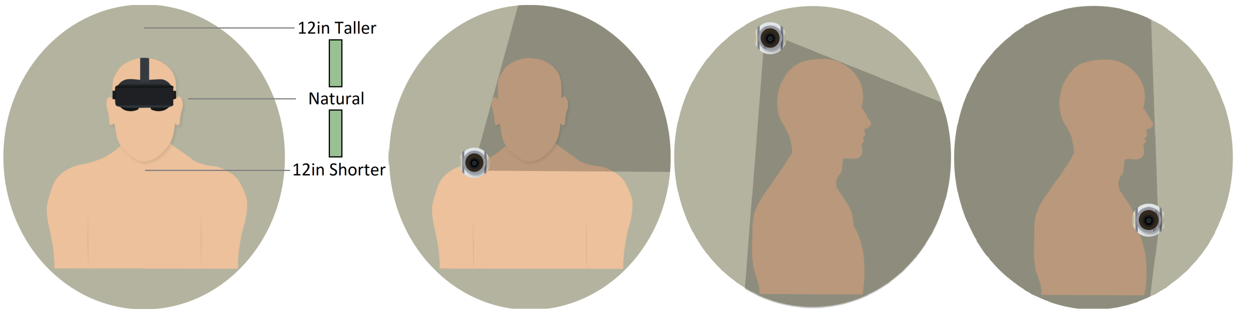


Figure 4.2: Placements for wearable 360° cameras each have pros and cons. Not only does the overall view vary, but the resulting height changes as well. In this study, we analyze user preference of three placements — shoulder-mounted, over the head, and on the chest — as well as three relative heights for each — natural, 12in shorter, and 12in taller — for a 3x3 within-subjects study.

their friends [213]. A bed-bound patient could enjoy a day sight-seeing with their loved ones [169].

We expect this kind of telepresence to become commonplace, enabling people from around the globe to connect like never before. The HCI community has begun to issue recommendations and best practices regarding panoramic video, but there is still work to be done, particularly to optimize viewpoints. In this study, we aim to identify how various body-worn camera placements and heights affect user experience. While the community has developed an array of prototypes, we are the first to specifically compare and contrast viewer experience among exemplar camera placements. Therefore, our research questions for this work include the following:

- **RQ1:** What is the optimal body-worn camera placement for reducing visual occlusions created by the video streamer?
- **RQ2:** What is the optimal body-worn camera height, in relation to the viewer's own height?
- **RQ3:** What is the best combination of body-worn camera placement and height to enhance a video viewer's overall experience?

To help answer these questions, we conducted a 3x3 within-subjects experiment with 26 participants in a VR simulation, varying the factors of *Camera Placement* and *Camera Height*. Our main dependent variable was user response regarding satisfaction with the experience. Overall, we found a significant main effect of Camera Placement, such that the video viewers disliked the 360° camera worn on the shoulder, since occlusion due to the streamer's head was prevalent. However, we also uncovered an interesting effect where camera placement affected perception of view height. When the camera was placed over the head of the streamer, even if the height was actually shorter than normal, the participants still felt like the view was tall. The work described in this chapter contributes the following:

- A novel user study examining the impact of camera Placement and Height on a user's viewing experience of body-worn 360° video using an HMD
- An understanding of how vantage points and environmental stimuli affect user perception of camera height
- Design recommendations for content creators, system designers, and camera accessory developers who will prototype future body-worn telepresence experiences

In the next sections, we detail our methods, including study design and procedure, followed by the results of our user study. We then discuss the design implications that arose from this research, and discuss a roadmap for future research efforts in this space.

Methods

In this section, we describe our user study aimed at understanding the effects of camera placement and camera height on a viewer's interpersonal telepresence experience. We first describe the motivation behind our study design decisions, followed by hypotheses based on these conditions. Next, we describe the study subjects, apparatus, and reproducible procedure.

Study Design

We designed a 3x3 within-subjects study varying *Camera Placement* and *Camera Height*. We used a simulated virtual environment (VE) for our study. VEs have been used in research for a variety of reasons, including to provide a controlled study space, reduce external and potentially confounding variables, and maintain study feasibility, with little sacrifice to realism [18, 43, 206]. The use of a VE for our study allowed us to maintain variable constancy, such as the streamer walking gait, level of action, camera steadiness, latency, networking hiccups, etc.

Camera Placement had 3 levels representative of prior literature and common action camera placements - over the head ("Overhead") [243], on the chest near the shirt pocket ("Pocket") [14, 87], and on the shoulder ("Shoulder") [106, 116, 221, 253]. For each of our conditions, there are different ways of achieving similar levels of occlusion while being able to manipulate height. For instance, in the Overhead conditions, the streamer's body occludes the bottom part of the view; similar views are commonly achieved using a hand-held selfie stick or telescopic pole. In the Pocket conditions, the streamer's body blocks the back portion of the view; similar views can be achieved by placing the camera near the abdomen [250] or by using a neck-worn camera (e.g. those used by law enforcement officers). In the Shoulder conditions, the streamer's body blocks a piece of the bottom part of the view, and the head blocks the view opposite of the mounted shoulder. The view can be manipulated by including actuators to adjust the camera with 6 degrees of freedom [116, 152].

Camera Height had 3 levels we felt would give us a good range of exploration - the participants' natural eye height ("Normal"), their eye height plus 12 inches ("Taller"), and their eye height minus 12 inches ("Shorter"). We chose 12 inches because the difference between an average male and average female in the United States is 6 inches, and two standard deviations of height is 6 inches [57]. Thus, a 12 inch step covers likely ground. We also acknowledge that live streaming is becoming increasingly popular, even at the extremities of average adult human height. For instance, consider the National Basketball Association's VR app [177]. Currently, the app enables viewers to watch live games with a VR HMD, from the view of a static, court-side camera. It seems plausible that the NBA would, in the future, live-stream feeds worn by the players, whose average height in the 2018-2019 season was 6'7" with a standard deviation of 3.3 inches [208]. Additionally, while we do not advocate for minors to wear cameras, we acknowledge that this is a plausible scenario [48, 111], especially as life-logging tools are becoming more readily available. As such, our height conditions represent a wide range of plausible use cases.

These levels totaled 9 conditions, which were randomized and counter-balanced in a Latin Square design. Our study received IRB approval, and we obtained informed consent from participants before they participated in our study.

Research Hypotheses

We expect viewers to desire a camera height that matches their own height, but in practice this is not always feasible. Average adult height varies significantly between men and women, and between people from different countries [30, 55, 57, 139, 260]. To compensate for a difference in height, the streamer could move the camera to a different part of the body; but, this may result in an unnatural or occluded view. We would expect an optimal experience to have a natural viewing height while also affording the clearest viewpoint possible. We thus conducted our study considering these factors, hypothesizing the following:

- H1: We expect a main effect of camera placement such that users will prefer an unoccluded view.
- H2: We expect a main effect of camera height such that users will prefer a viewing angle similar to their natural height.

Subjects

A priori power analysis using G*Power indicated that we needed a minimum of 22 users to detect a medium effect size [53]. We recruited 26 participants for our study from the student body of the University of Central Florida. 21 were male and 5 were female. Their age ranged from 18 to 29 (M = 20.8; SD = 2.74). We measured participant height; the range was 4’9” / 1.44m to 6’3” / 1.91m (M = 5’9” / 1.75m; SD = 4.21in / 10.7cm). All participants had normal vision, or they wore corrective lenses during the study. We asked participants how often they watch 360° videos, and the Median response was “Rarely.” Similarly, the Median response for how often they use VR was “Rarely.”

Apparatus

We created our virtual environment using the 2017.3.0f3 version of Unity3D. We ran the study on a laptop with Windows 10, Intel core i7-7700HQ at 2.8GHz, with 12GB of RAM, with an Nvidia GeForce GTX 1060. The HTC Vive HMD was used to run the scenario.

Figure 4.3: Sample snapshot of the environment used to simulate an interpersonal telepresence experience. Stimuli were distributed and organized in various placements.

Virtual Environment

We retooled a VE found on the Unity Asset Store to create a virtual art museum (see Figure 4.3, retooled from [228]). A museum tour is a plausible example use case, and we were influenced by Tang et al. [243], who performed a real-life task in which participants needed to search for art during a virtual tour. Our room was rectangular with a dividing wall through the middle, lengthwise. On the walls were famous paintings. The very center of each painting was approximately 5’10” / 1.8m off the floor. Between paintings, there were pedestals that held sculptures. The sculptures rested approximately 2’3” / .7m off the floor. On the floor, ornate rugs were laid out. Scattered in the room were digital human museum-goers who stayed in-place.

For each stimulus type, we proposed a “real” and a “counterfeit” (see Procedure). We tried to balance subtlety with objectivity for the counterfeits, such that the participants would need to inspect the objects yet be able to recognize that something has changed. For the paintings, we performed a web search with the phrase “famous paintings” and selected a subset. To find the counterfeits, we performed a web search for each one, including the word “parody.” We were able to pair every real painting with a parody. For instance, we used Starry Night, and the counterfeit featured Darth Vader [61]. For the sculptures, we first defined a list of objects that typically have one size in the real world, in hopes of giving visual cues for perceiving camera height. For the counterfeits, we used properties of association. For instance, we used a soccer ball, and the counterfeit was a basketball. For the carpets, we performed a web search for “ornate rugs,” and found various patterns. For the counterfeits, we inverted the color scheme but maintained the patterns.

The stimuli were mapped symmetrically. There was a box-shaped path that the virtual streamer walked, and both sides of the path were as equal as possible in terms of stimuli count and placement. One side of the museum had a door, and the opposite side had a large painting. All paintings and sculptures were scaled to life-size. In total, there were 15 famous paintings, 12 sculptures, 6 rugs, and 4 museum-goers. To simulate common visual artifacts of live-stream cameras, we constructed a virtual Camera Rig that consisted of six 90° cameras, each using the following Unity3D post-processing techniques and values:

- **Antialiasing** (Fast Approximate Anti-Aliasing)
- **Ambient Occlusion** (Intensity = 1; Radius = 0.3; Sample Count = Medium; Downsampling Enabled)
- **Motion Blur** (Shutter Angle = 270; Sample Count = 10; Frame Blending = 0)
- **Grain** (Intensity = 0.5; Luminance Contribution = 0.8; Size = 0.7; Colored Enabled)

We mapped the output of the camera rig to the Unity3D Skybox. This, combined with the post-processing techniques, resulted in the "stitching" artifact that can be found in multi-lens cameras. The virtual streamer was the "Ethan" model that is found in Unity3D's tutorials, modified to have a constant walking speed regardless of model size. For our study, we assume camera stabilization, i.e. the streamer's walking gait had no effect on the camera, except for when the model turned 90 degrees on the path.

Procedure

Recruited participants first reviewed an informed consent form. We did not collect signatures. Users were seated in a chair inside of the HTC Vive play area; the play area was scaled down, so participant head translations did not have an effect on the viewing experience. Next, we gave a demographics survey, noting the height of our participant, and subtracting the difference between their eyes and the top of their head (approx. 5in). Following, we described the study to our users. We explained that the objective was to explore the museum by inspecting all paintings, sculptures, and rugs. To ensure the user adequately and intently explored the museum each time, we randomly replaced one of the paintings, sculptures, and rugs with a counterfeit - the user's task, therefore, was to identify these counterfeits every condition. To prevent a learning effect, the counterfeit objects and their locations were randomized every run. Before the user entered VR, we first showed them a simple website with pictures of the *correct* stimuli, allowing as much time as needed for them to become familiar. We then ran the user through a practice trial, in which the camera floated in air (there was no streamer). The length of this practice run - and each trial - was approx. 2.5 minutes.

The user was then run in the first condition with the streamer visible. We did not tell the user which condition was being run. During the task, we hand-recorded the user's audible feedback. After the run was complete, the user filled out a questionnaire. We then told the user how tall the camera was, to untangle confusion which could affect our final survey. We then loaded the next condition and repeated these steps. After all conditions were completed, we gave the user one final questionnaire, to rank all conditions. The time to complete the study was approximately 50 minutes, and participants were given 10USD in cash.

Dependent Variables

We administered a questionnaire after each condition to measure if the user thought the view was *Free from Occlusions* and if it had a *Natural Height*. The measures consisted of the following items, on a 7-point Likert scale (1 = strongly disagree, 7 = strongly agree):

- Questions regarding **Free-from-Occlusions (*FfO*)**:
  - The camera placement allowed me to see everything I needed to see
  - Nothing blocked my view to the point where I became disoriented
  - My field of view was clear, so I could perform the task
- Questions regarding **Natural Height (*NH*)**:
  - The height of the camera felt natural to me
  - The camera height let me view the environment with ease
  - I liked the height of the camera placement
- All things considered, please give a score to that camera placement (1 = Terrible, 7 = Excellent)

After each condition, we also asked *"Was the view shorter, taller, or equal to your natural height?"* Users were also asked to explain what they liked or disliked about each condition. After all conditions were complete, the user was asked to rank all of the conditions from best to worst. They were also asked to tell us the rationale for why they selected the best and worst conditions.

Data Analysis Approach

In order to assess the construct validity of the dependent variables, we first tested for internal consistency by calculating Cronbach's alpha [38]. *FfO* and *NH* were both above the 0.7 threshold for reliability (*FfO*: $\alpha = .889$; *NH*: $\alpha = .824$). Thus, we averaged the values together to form an index per construct, per condition. As a single-item measure, we did not assess *Overall Rating* for construct validity. Since the DVs were not normally distributed, we used non-parametric tests to test our hypotheses. We anticipated a possible interaction effect between camera placement and height, so we chose to use the Aligned Rank Transform (ART) tool [272], so that we could run a repeated measures ANOVA. For our post-hoc tests, we controlled Type I errors by performing Holm's Sequential Bonferroni Adjustment [81]. For qualitative feedback, we used an open coding process to identify emerging themes. We present illustrative quotes to help unpack some of the nuance around the experimental results of our study.

Results

We first present the descriptive characteristics of our data, followed by the results of our hypothesis testing, and a summary of our findings.

Table 4.1: Descriptive Statistics of All Conditions by Dependent Variable

Placement	Height	Free-from-Occlusions	Natural Height	Overall Rating
Overhead	Taller	M = 5.923, SD = 1.356	M = 5.551, SD = 1.447	M = 5.846, SD = 1.084
	Natural	M = 6.103, SD = 1.014	M = 5.821, SD = 1.246	M = 6.000, SD = 0.849
	Shorter	M = 5.949, SD = 1.183	M = 5.154, SD = 1.620	M = 5.269, SD = 1.430
Pocket	Taller	M = 6.000, SD = 1.227	M = 5.564, SD = 1.456	M = 5.654, SD = 1.294
	Natural	M = 6.205, SD = 1.333	M = 6.308, SD = 0.958	M = 6.269, SD = 0.919
	Shorter	M = 6.013, SD = 1.222	M = 5.231, SD = 1.494	M = 5.385, SD = 1.235
Shoulder	Taller	M = 3.397, SD = 1.854	M = 4.551, SD = 1.649	M = 3.731, SD = 1.185
	Natural	M = 3.000, SD = 1.683	M = 4.538, SD = 1.633	M = 3.385, SD = 1.359
	Shorter	M = 3.244, SD = 1.692	M = 4.321, SD = 1.640	M = 3.423, SD = 1.419

Descriptive Statistics

The descriptive statistics of our DVs can be found in Table 4.1. The following sections describe the results of repeated measures ANOVAs, as shown in Table 4.2.

ANOVA Results

H1: Main Effect of Camera Placement

An ANOVA revealed a significant effect of Camera Placement on each of our dependent variables; see Table 4.2. Post-hoc t-tests revealed significant differences (see Table 4.3); for the *FfO* construct, there were differences between Overhead and Shoulder, as well as Shoulder and Pocket. Most of our participants indicated that the Shoulder placement was annoying, frustrating, or generally negative because the streamer's head blocked the right side. See Figure 4.4.

We also found a significant main effect of Camera Placement on *NH*. Post-hoc t-tests again revealed

Table 4.2: Repeated Measure ANOVA Results

Construct	ANOVA Result
Main Effect of Camera Placement	
FfO	$F(2,50) = 71.50, p < .001, \eta_p^2 = .741$
NH	$F(2,50) = 16.58, p < .001, \eta_p^2 = .399$
Overall	$F(2,50) = 58.96, p < .001, \eta_p^2 = .702$
Main Effect of Camera Height	
FfO	$F(2,50) = 0.531, p = .591, \eta_p^2 = .021$
NH	$F(2,50) = 5.946, p < .005, \eta_p^2 = .192$
Overall	$F(2,50) = 2.707, p = .077, \eta_p^2 = .098$
Interaction Effect of Camera Height * Camera Placement	
FfO	$F(4,100) = 1.808, p = .133, \eta_p^2 = .067$
NH	$F(4,100) = 1.182, p = .324, \eta_p^2 = .045$
Overall	$F(4,100) = 1.438, p = .227, \eta_p^2 = .054$

differences between Overhead and Shoulder, as well as Shoulder and Pocket. We would not expect this, as the varying levels of height were consistent between all camera placements. Due to this unanticipated finding, we later present a post hoc analysis of how participants' perception of height varied based on camera placement; see Figure 4.5.

We also found a significant main effect of Camera Placement on *Overall Rating*. Post-hoc t-tests again revealed differences between Overhead and Shoulder, as well as Shoulder and Pocket. This result compounds with the previous t-tests; occluding the entire right side (effectively 90° of the entire viewpoint) detracted from the overall experience.

H2: Main Effect of Camera Height

An ANOVA revealed a significant main effect on the *NH* construct, but not on *FfO* or on *Overall Rating*. Post-hoc t-tests reveal significant differences between Natural and Shorter heights, as well as Taller and Shorter heights; see Table 4.3. Expectedly, participants found their natural height to be favorable, but unexpectedly, they were unfazed by the taller placement. They did, however, find the shorter camera heights to be less natural. See Figure 4.5 for illustration.

*Interaction Effect of Camera Placement * Camera Height*

An ANOVA did not reveal a significant interaction effect between Camera Placement and Camera Height; see Table 4.2.

Table 4.3: Significant Post-hoc T-Test Results by Main Effect

Construct	Condition A	M	SD	Condition B	M	SD	Result
				Effect of Camera Placement			
FfO	Overhead	5.99	1.19	Shoulder	3.21	1.75	$t(25) = 11.4, p < .001$
FfO	Pocket	6.07	1.26	Shoulder	3.21	1.75	$t(25) = 10.1, p < .001$
NH	Overhead	5.51	0.94	Shoulder	4.81	0.95	$t(25) = 2.99, p < .001$
NH	Pocket	5.70	0.76	Shoulder	4.81	0.95	$t(25) = 5.13, p < .001$
Overall	Overhead	5.71	0.88	Shoulder	3.51	1.08	$t(25) = 8.83, p < .001$
Overall	Pocket	5.77	0.87	Shoulder	3.51	1.08	$t(25) = 9.73, p < .001$
				Effect of Camera Height			
NH	Natural	5.556	1.502	Shorter	4.902	1.632	$t(25) = 3.004, p < .05$
NH	Taller	5.560	1.447	Shorter	4.902	1.632	$t(25) = 2.756, p < .05$

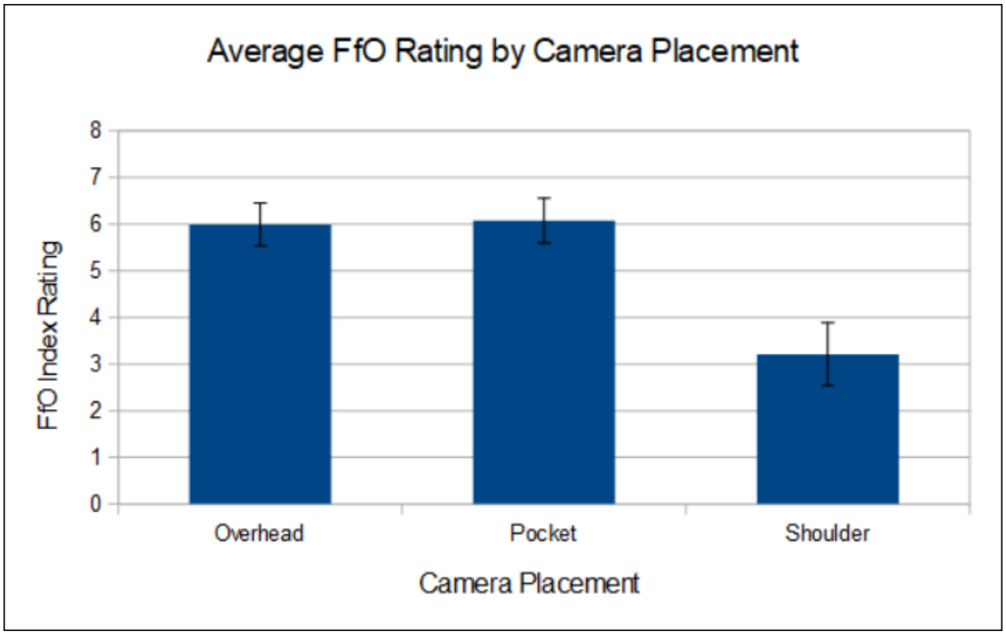


Figure 4.4: Free-from-Occlusions (FfO) Index Rating with 95% confidence, by camera placement. The camera being worn on the shoulder resulted in a significant portion of the scene being occluded, detracting from user perception.

Perception of Camera Height

Due to the significant main effect of Camera Placement on the *NH* construct we conducted a post hoc analysis to help understand why this result emerged. Using our qualitative feedback, we found an interesting result regarding how tall each condition made the participants feel; see Figure 4.8. During the Overhead conditions, participants often believed the viewpoints were taller than what they actually were, simply because the streamer's head was seen underneath them. If we treat the results of this question as a pass/fail item, then only 54% Overhead, 71% Pocket, and 51% Shoulder heights were correctly perceived. Considering our VE, there are only a handful of items with which users could really infer height - paintings, sculptures / pedestals, museum-goers, and the streamer. The Overhead and Shoulder placements naturally forced the users to see the avatar's head either below or at eye-level, respectively; but for the Pocket placement, the users had to go

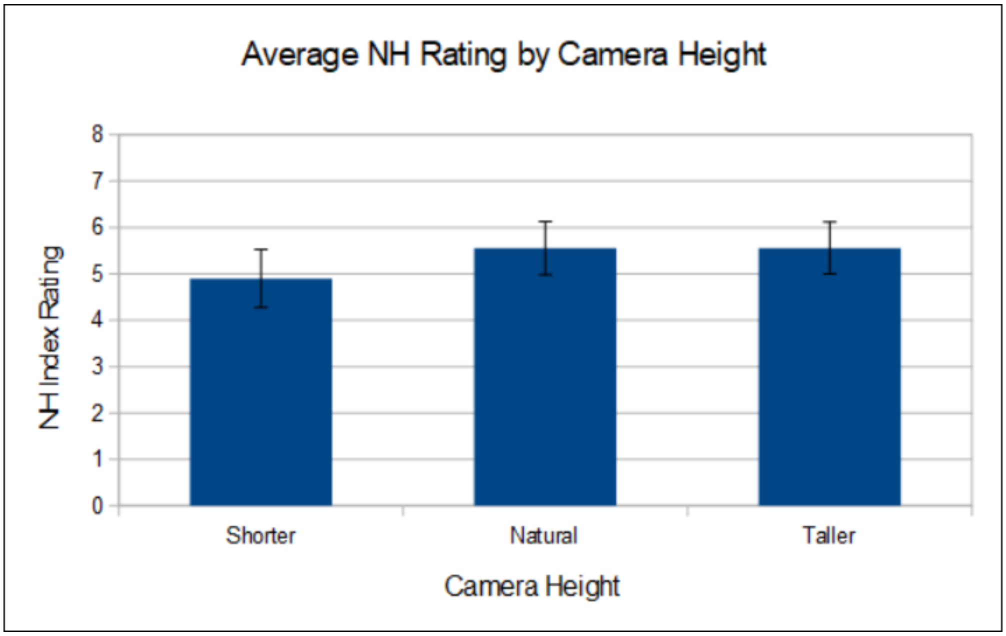


Figure 4.5: Natural Height (NH) Index Rating with 95% confidence, by camera height. Participants naturally indicated Medium and Taller heights as their favorites.

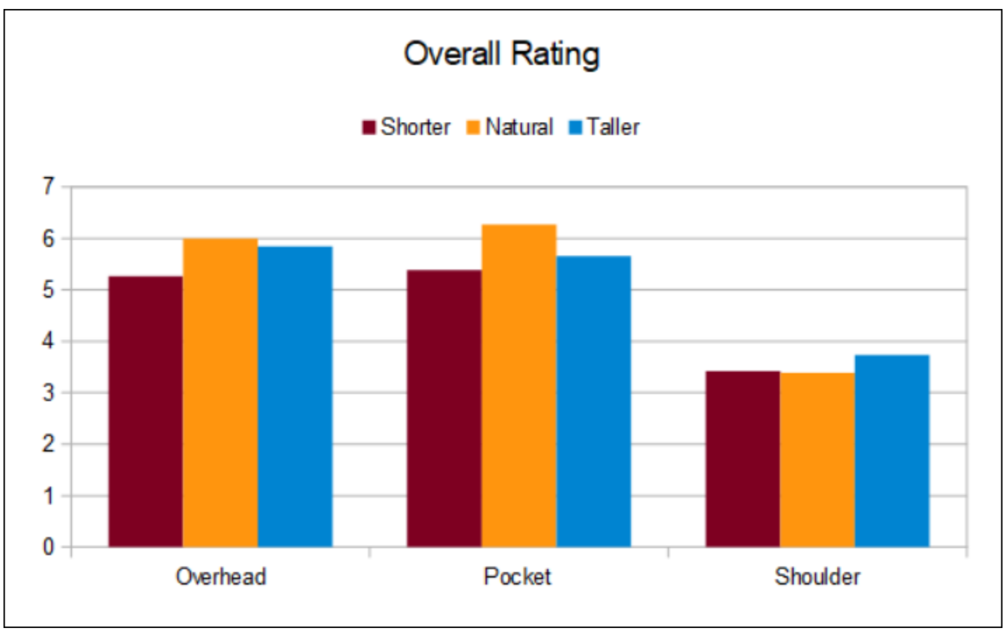


Figure 4.6: Average Overall Rating by Condition. The Shoulder placement was rated poorly, and in general, participants more preferred the Natural Height conditions.

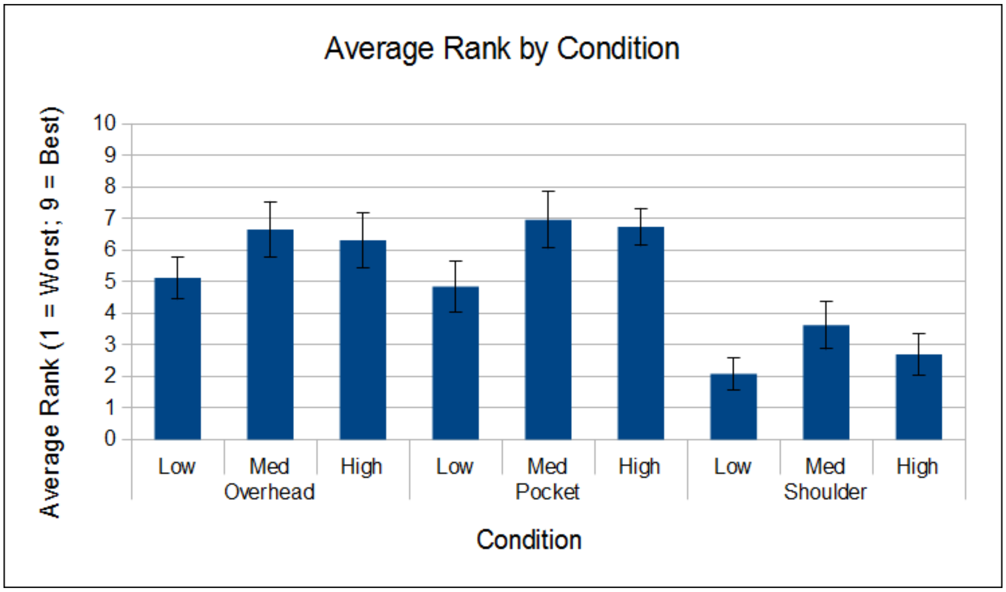


Figure 4.7: Average Rank by condition with 95% confidence. The Overhead and Pocket placements were well regarded compared to the Shoulder placement. In general, users most preferred the camera to be at their natural viewing height.

out of their way to see anything other than the avatar's legs. Supporting evidence can be found in comments. For the Overhead placement, some participants felt taller simply because there was a head beneath them:

- *"I feel taller because I see the dude's head below."*
- *"The viewing angle was a bit high, looking down on everything."*
- *"I feel like the overall view is lower, but because the avatar is under me, I still feel tall."*
- *(As part of positive feedback for the Shorter condition) "Im tall so the farther off the ground I am the more comfortable"*

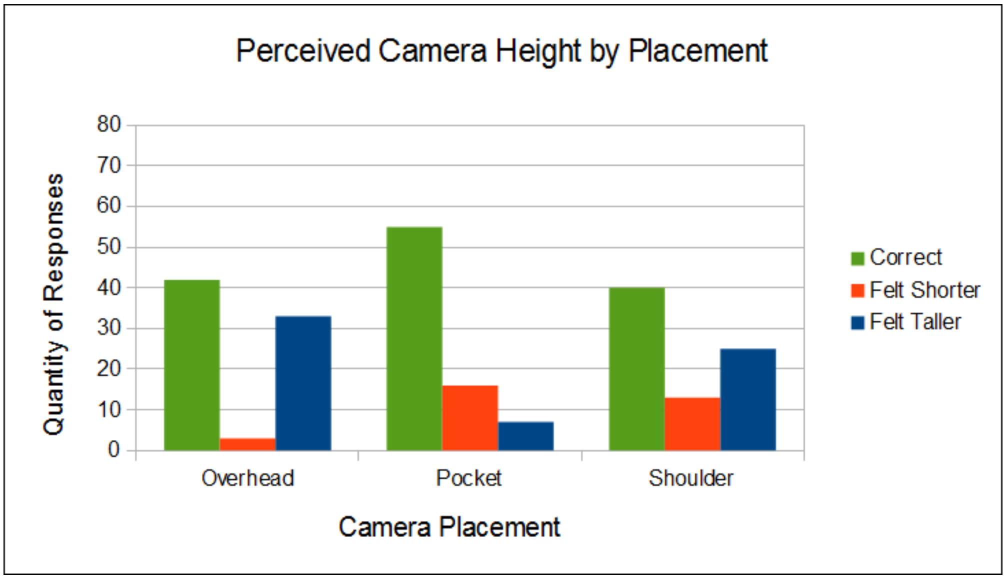


Figure 4.8: Perceived Camera Height by Placement. Participants were in-tune with height while using the Pocket placement, but for the Overhead conditions, participants often felt taller regardless of height.

For the Pocket placement, the participants indicated that the camera wearer did not affect them, and they really only needed to look forward and to their sides, rarely commenting on anything past their immediate view:

- *"I didn't do it, but if I looked back, I would've seen the avatar."*
- *"The model rarely got in the way of my view..."*
- *"This was a great, unobscured view of almost everything."*
- *"Nothing from the avatar impeded my vision..."*
- *"I barely noticed the avatar."*
- *"The front pocket gives a very clear view of everything in the front."*

Based on this user feedback, it is clear that the virtual streamer was a visual cue that participants used to infer height.

Feedback from Participants

Next, we analyzed the questionnaire item regarding subjective height perception, and we coded the open response questions in order to determine which placements were regarded positively and negatively, and why.

Ranking Data

Looking at the Ranking data (Figure 4.7), we find that Shoulder was indeed the worst position of the three; Overhead and Pocket were very positive and comparable. For all placements, we find that the Natural height was viewed as best, followed by a Taller height. This feedback compounds with the quantitative results. Out of all twenty-six (26) users, none of them ranked the Shoulder best, and twenty (20) thought it to be the absolute worst. See Figure 4.7 for illustration.

There were a number of themes that emerged regarding factors that influenced user satisfaction. A majority of the participants (65%) indicated that a field of view which was free from occlusions was the main benefit of their favorite condition. Forty-two percent (42%) indicated that their favorite placement felt natural to them. Only nine (9) participants responded that height was a major detriment in their least preferred condition. Interestingly, six (6) participants responded that visual features of the virtual avatar was a major drawback:

- *"I didn't like his head being so close to me."*

- *"...I felt uncomfortable with the head bouncing."*
- *"...I felt like the avatar was bouncing too much."*
- *"The avatar's head was extremely distracting."*
- *"It feels a little odd with someone's head right under my chin..."*

In the end, however, occlusion (or lack thereof) is the main contributor to the success of the camera placements.

Virtual Presence and Viewpoint Metaphors

While we often found that our participants simply did not want to see the avatar, some of the users did provide us with interesting feedback, revealing comments that indicate they gained some sense of immersion or presence in the virtual environment. Others provided colorful metaphors within their negative comments for the various camera placements:

- **Overhead**:
  - *"It felt like I was riding a horse."*
  - *"...it was like piggy-backing on someone's back."*
- **Pocket**:
  - *"It let me view my feet and legs easily..."*
  - *"Finally know what its like to be tall..."*
  - *"Made me feel like I was actually walking through."*
  - *"I felt like I was being held like a baby."*

- *"This is what its like being in my girlfriend's body."*

- **Shoulder**:
  - *"You can see the characters face, and have kinda like a first/third person view."*
  - *"I was able to view paintings as if I were really there..."*
  - *"...I don't like being on someones shoulder."*
  - *"I felt like a Siamese twin [sic]."*

Discussion

In this section, we discuss design implications based on our results. In particular, we describe how viewers mainly want a viewpoint that offers them maximal exploration capability, and that camera height was not a significant factor for overall user experience. Additionally, we note how user perception of camera height changes with camera placement. Subsequently, we discuss a future research agenda based on our findings.

The Clearer the View, the Better

Our results indicate a strong disdain for camera placements which block a significant portion of the front hemisphere. The Shoulder only had approximately 90° of the environment blocked, and the Pocket had 180° blocked; but the right side was more important to our users than the back. Our users were able to complete inspection before the avatar walked past stimuli, plus they were seated for the entirety of the study (so it was difficult to turn their head around). We would expect our finding to hold true in cases where important stimuli enter the view from the front. For complex environments where stimuli moves or appears from behind, we would anticipate a placement similar to Overhead being most desirable, to provide the most opportunity for exploration. For instance, the multi-lens camera setup in Kasahara et al. seems to meet user needs [100, 101, 105]. While our users did not like the Shoulder camera placement, interestingly, prior researchers did find a similar rig to be well-received [116, 122, 123]; they implemented a gimbal device which can be directly manipulated by the viewer via a GUI system, so perhaps it was this sense of control that helped users perceive it positively [116].

Users Don't Want to See the (Virtual) Streamer

Our participants often pointed out that the virtual streamer had strange, unattractive, or undesirable traits, or simply didn't want to see the model bouncing around while walking. We acknowledge the Ethan model looks somewhat strange, and this could have made the participants perceive the avatar negatively. But, user comments also indicated that camera placement affected overall perception. The Pocket offered the most unoccluded view, and users had to go out of the way to see the avatar's features; but even in these conditions we found users pointing to negative features of the avatar. This shows that our participants, for our task and in our environment, did not want the avatar to appear in their view; as such, our study cannot answer to a setup where the streamer is someone that the user knows, e.g. a friend, family member, or celebrity streamer. We would expect users to perceive camera placements more positively if a loved one was in the view [116], but users could still communicate with each other regardless [14, 87].

Placement Matters More than Height

Based on these results, it seems that camera height may not be a driving factor for success; while the Shorter height was the least natural to our users, it didn't seem to detract from the experience as much as camera placement. This is contradictory to previous research, which suggests that lower camera heights are more acceptable to higher heights [215]. We suspect that there are external variables not identified here which cause a difference in height preference; for instance, culture may be such a variable. We acknowledge that our study only analyzed 3 relative heights, but +/- 12in covers a large range. If we inspected an even wider range, we would begin including positions alongside the extremities of adult human height. Our results corroborate with previous findings: our participants had trouble identifying the "correct" camera height, which is expected - humans do not excel at judging VR heights and distances [12, 60, 138, 166]. Telepresence between taller and shorter individuals may not be harmed by the disparity in user heights. A drastic difference may result in a drop in user satisfaction, but our results imply that it would need to be severe.

Different Placements Give Different Sense of Story

User feedback indicated varying experiences through the metaphors they provided in their comments. It is important to convey a sense of presence to the user as that allows them to feel as if they were "actually there" [23, 77, 227], which is one of the goals for telepresence. Metaphors have been used to help describe telepresence setups, such as the user assuming the role of a parrot sitting on the shoulder [116], or a ghost watching the world from another person's view [100], or even borrowing another person's body [162]; but our users sometimes felt like they were taking a ride on an animal or being carried around like a child. We find that the Pocket placement helped conveyed a sense of active exploration because the users felt like *they were the avatar*, due to the character being behind the camera, whereas the Overhead and Shoulder conditions conveyed a sense of passive exploration because the users felt like *they were watching the avatar* from a third-person viewpoint.

Content streamers have opportunity to give two different types of experiences, simply by wearing the camera in different spots - one where the viewer is the "star of the show," and one where the viewer watches the events transpire. In a case where the streamer and viewer are strangers, we would recommend a placement similar to Pocket. When the streamers know each other and want to have a communicative experience, an Overhead or even a Shoulder placement could suffice, as the viewer can then clearly see non-verbal social cues such as upper body gestures or facial features.

Limitations and Future Work

While we believe our virtual environment consisted of a good number of stimuli via the paintings and sculptures, the carpets were not too interesting, and the counterfeits were easy to spot. While carpets make sense for a museum environment, more complex stimuli such as sidewalk chalk art could be better suited for an inspection task, and therefore may have revealed a difference in preference between Overhead and Pocket. We are confident, however, with our results being representative for environments with much stimuli that is "eye-level." We also acknowledge that our virtual environment may be perceived differently than a "real-life" setup. In the future, we plan on taking learned outcomes from this study and applying them to a variety of real-world scenarios. For instance, 360° videos are commonly employed for action sports such as skiing and skateboarding, but our test case was a milder indoor exploration task. While our results cannot generalize to all types of telepresence interaction, we plan on using the lessons learned here and analyzing the effects of different walking gaits, speeds, and levels of activity, as well as analyzing how these qualities are perceived by participants. Further, while our users found the Shoulder placement negative here, it may be a good placement when the streamer is a friend or family member, so the viewer could see their face. Our study did not contain an aspect of verbal communication, but we plan on taking the current findings and applying them to a real-world task, e.g. a dyadic exploration of a remote environment between friends. We also plan on studying different environments and scenarios in which streamers have begun to explore, to analyze how viewers perceive varying levels of action.

Conclusion

Telepresence is the culmination of recent technological breakthroughs that we expect to become commonplace in the near future. Our novel investigation contributed a study that revealed how users of similar systems perceive various body-worn 360° camera placements, and we found that there are both positive and negative qualities which can be adjusted for optimal usage. We recommend, if possible, that shorter streamers use an Overhead camera placement, and taller streamers use a Pocket-esque placement, to comfortably accommodate their audience. Our results indicate that this kind of interaction can be used between people of varying heights, in a variety of different cultures, communities, and environments. It is our hope that our investigation will lead streamers onto a path which will provide the best possible experience for their viewers.

This study demonstrated that Viewers do have a preference for where a camera should be situated on a Streamer's body — a place that offers as much exploration opportunity as possible. However, different placements convey different senses of story, so the context in which interpersonal telepresence is being applied needs to be considered. Lastly, we found that users had a difficult time estimating the height of the camera, and we wanted to investigate this further. As such, we continued our investigation by exploring the effects of individual differences — specifically, variance in human physiology — on the viewing experience. Therefore, in the next chapter, we discuss this follow-up study which we conducted to help understand what users prefer in terms of camera height.

CHAPTER 5: THE EFFECTS OF GENDER AND THE PRESENCE OF THIRD-PARTY HUMANS ON TELEPRESENCE CAMERA HEIGHT PREFERENCES

This chapter is based on work previously published: Kevin Pfeil, Pamela J. Wisniewski, and Joseph J. Laviola Jr. 2020. The Effects of Gender and the Presence of Third-Party Humans on Telepresence Camera Height Preferences. In *ACM Symposium on Applied Perception 2020* (*SAP '20*). Association for Computing Machinery, New York, NY, USA, Article 13, 110. DOI:https://doi.org/10.1145/3385955.3407924

In the previous chapter, we demonstrated how a viewer would want the streamer to wear a camera on their body in a way that would grant maximal exploration of the environment. The viewing experience is thus directly affected by a simple design decision such as where on the body the camera should be worn. However, what we were unable to identify is the potential limits of camera height — for instance, if the height is too tall or too short, that might result in the view feeling unnatural. Further, we found that participants were unable to fully understand the height of the camera, as has been shown with the problem of VR distance underestimation. As such, this chapter describes a user study in which we ask participants to directly set the height of a camera. Knowing how tall or short a camera could be situated helps identify, in part, how individual differences of human physiology can affect the viewing experience. This knowledge is integral to interpersonal telepresence, because it would indicate *who* could be a streamer, for any given viewer. Therefore, in this chapter, we discuss a technological investigation that lets users directly manipulate camera height — in line with the theme of autonomy — to help understand how individual differences interact with human values.

This chapter is divided into four major sections: Introduction, Methods, Results, and Discussion.

We first describe the problem space in the Introduction. Second, we discuss our approach towards tackling the problem by conducting a user study, with design decisions informed by the previous chapter and prior literature, in the Methods section. Third, we describe the results of our user study, including descriptive statistics, hypothesis testing, and an in-depth analysis of user responses, in the Results section. Lastly, we provide design recommendations and implications for future work in the Discussion section.

Introduction

Immersive Virtual Reality (VR) is making a resurgence with the release of more affordable, consumer-grade hardware such as the HTC Vive head-mounted display (HMD); the advent of 360° / panoramic cameras such as the Ricoh Theta V; and the ease of access to content hosting platforms such as YouTube. With these tools, even novice users are able to create immersive content and share these experiences with others. For instance, there is a recent trend in which people (or "streamers") wear cameras on their body, and live-stream their environment while walking around their physical surroundings. This setup allows a viewer to receive an immersive tour of any given remote location that has a streamer within it. As such, this interaction is a form of telepresence, which is the ability to perceive and/or interact with a remote environment, as if being there [159]. This kind of telepresence has been applied for activities such as shopping [27], running [9], and sightseeing [25], among others.

A primary objective of immersive VR and telepresence is to make the experience as natural as possible, as if the viewer was actually present in the remote or virtual environment. One factor that can greatly influence the users' experience is the point-of-view in which the viewer interacts; depending on how tall the streamer is, the view may or may not be optimal for their audience. It is thus increasingly important to understand what factors contribute to an enjoyable experience for all. As such, researchers have studied optimal camera placement on the body (e.g., overhead, shoulder, or chest) [195], and have found that those which have the most unoccluded view (e.g. overhead rather than on the shoulder) accommodate the most opportunity for exploration, and are thus preferred most.

Meanwhile, others have studied the effect of third-party avatars on distance perception within a given virtual environment (VE). Some have found that these avatars do not have an effect [11, 191], but Langbehn et al. revealed that they help to provide a sense of scale within a VE [129]. Given the importance of camera height, camera placement, and the presence of human avatars within a virtual environment, we aim to understand how these factors influence one another. Additionally, we are interested in how individual differences, such as a users' height and gender, play a role in their optimal camera height preference. Previous efforts suggest that camera height is relatively arbitrary, and that the optimal height for all users is approximately 150cm [113, 216]. Building upon these previous works, in this study we allow users to manually set the camera height in various conditions. We ask the following research questions:

- **RQ1:** What factors (camera placement and avatar presence) are important for determining optimal camera height based on user preference?
- **RQ2:** What individual differences (gender, user height) play a role in determining user preference of camera height?
- **RQ3:** What is the acceptable range (shortest, preferred, tallest) of user-selected camera heights based on the factors and individual differences above?

To answer these research questions, we conducted a 2x2 mixed-design experiment with 93 participants, with a within-subjects factor of Camera Placement, and a between-subjects factor of Avatar Presence. While viewing an immersive VE that simulated a body-worn 360° camera, participants were asked to manually set the height of the camera according to what they felt was most natural to them. Additionally, we asked them to identify their shortest and tallest preferred camera heights. Results show that Camera Placement was not a significant factor that influenced height preference, but human avatar presence was significant; when avatars were present, participants preferred taller camera heights. The average camera height preference for men when avatars were present was around 160cm; but, women preferred a shorter height — approximately 140cm. When avatars were not present, average camera height preference was approximately 145cm and 120cm, respectively. Actual user height was not found to play a role in height preferences. Lastly, the average thresholds for acceptable height were approximately +/- 40cm from a user's natural height selections. Our work contributes a user-centric approach to understanding camera height preferences within an immersive VE, with a call to consider individual differences instead of choosing a "one-size-fits-all" solution. Our main contributions include the following:

- Our work is the first to examine user preference of camera height relative to camera placement and the presence of avatars.
- We found that the virtual context (presence of human avatars), rather than the physical placement of the camera on the host, had a significant influence on users' preferred camera height.
- We also highlight the importance of taking into account individual differences, specifically gender, when customizing virtual telepresence user experiences to meet users' needs.

Methods

Previous research efforts have shown how various factors impact user experience, including camera placement, and the presence of human avatars [11, 129, 138]. We based our study on these factors and applied a user-centric approach by letting participants identify their own preferred camera height. On demand, our participants could increase or decrease the height of the camera in a simulated telepresence experience.

Study Design

We conducted a 2x2 mixed-design experiment with independent variables (IVs) of *Camera Placement* and *Avatar Presence*. The *Camera Placement* variable was a within-subjects factor and had 2 levels based on the previous chapter — "Overhead," where the camera is mounted above the avatar's head, and "Pocket," where the camera is mounted on the avatar's chest. Both were first-person views, i.e. the user could rotate the view by simply moving their head. The *Avatar Presence* variable had 2 levels — Present and Absent — and was between-subjects to prevent a learning effect. In the Avatars Present condition, all third-party human avatars were visible. In the Absent condition, they were invisible, with the exception of the streamer. We used this variable because we anticipated participants making height judgment calls using human avatars as a frame of reference, as per Langbehn et al. [129]. We randomly assigned participants to their corresponding group and condition order, in a counter-balanced design. Our study received IRB approval.

Using Virtual Reality to Simulate Immersive Telepresence

Our study is comparable to those of Keskinen et al. [113] and Rothe et al. [216], who used real 360° footage from different heights (approximately 10cm steps) from a tripod-mounted camera. Their research was related to static 360° video, and thus they used real, prerecorded footage. Our research interests focus on mobile telepresence, and here we simulate a live-streamed 360° camera using a VE, due to the many external and possibly confounding factors which can hinder our measurements.

First, as we are interested in telepresence, we must consider how the camera is worn by a moving person instead of static on a tripod. With a single actor, it is nigh impossible to maintain a constant walking gait for a number of video clips (i.e., people have different walking motions [219], which could confound a study). Second, we must also consider the height of the streamer. In order to preserve camera placement while manipulating height, we would need to enlist multiple actors of incremental height, which confounds visual identity of the streamer in addition to exacerbating the walking gait problem. We considered alternatives to help alleviate this issue — one idea was the development of an actuated, mobile platform on which the same person could stand (thus changing their height while preserving visual identity); another was shrouding actors with a sheet, to hide their faces and bodies. Both of these do not completely solve the above issues, and perhaps worse still, depart from any semblance of telepresence. We thus decided to simulate a 360° camera in a VR setting, allowing us to reasonably overcome these aforementioned problems.

We acknowledge that users are generally less adept at estimating distances while wearing a VR HMD, and we suspect that this phenomenon will translate to immersive telepresence experiences. We thus recall these previous findings and, as appropriate, eliminate them as variables in our study. We use a high-resolution HMD (HTC Vive) and change IPD as needed per participant; and, as a self-avatar is similar to a telepresence streamer, we keep one visible at all times. We also manipulate third-party avatar presence as an independent variable to help understand how camera height preferences shift with varying social environments. Since various contexts can have an effect on optimal camera heights for telepresence, it is pertinent to understand the influence third party members have on an experience.

Apparatus

We used a “Showroom Environment” found on the Unity Asset Store [228] as a template to build our VE. The showroom comes with hangars for paintings and low pedestals for sculptures. Sightseeing tours (real or virtual) are common VR experiences [150, 243], so we used the environment to build a fictitious one. See Figure 4.3 for an illustration of the environment.

On the hangars, we placed life-size images of famous paintings, such as Mona Lisa, The Starry Night, and The Kiss, finding a mixture of small, medium, and large paintings. The center of each painting was at a height of 177cm measured from the floor. In total we used 15 famous paintings. On the pedestals, we placed a variety of everyday objects which we believed would provide familiarity and thus a frame of reference for size. For instance, we added a can of Pepsi, a soccer ball, and a Mac computer. The pedestals were each approximately 69cm tall, measured from the floor, and the sculptures were of varying size. In total, we used 12 sculptures on the pedestals. On the floor, we laid out ornate rugs to provide more interesting stimuli below eye level, using various textures found online. In total, there were 6 ornate rugs.

In addition to these stimuli, our environment had two male and two female human avatars. These avatars did not move except for a simple “breathing” animation. All stimuli were mapped symmetrically in the environment. In our environment, a virtual human (the Ethan model from Unity3D tutorials) provided a social telepresence experience to our users. This simulated streamer walked the scene at a constant speed, in a box-shaped path around the room. In this way, we were able to keep navigation constant in our study. If the user looked around, they would be able to see the streamer’s face and body.

There are numerous 360° cameras available for professional and consumer purchase, and we created a 6-camera rig in Unity3D. This, combined with the 3D effect afforded by VR, is not unlike the

Vuze+ stereo 3D camera [1]. Further, we applied the following Unity3D post-processing techniques (to be consistent with the study by Pfeil et al. [195]), resulting in the common visual artifacts seen when live streaming multi-lens camera feeds:

- **Antialiasing** (Fast Approximate Anti-Aliasing)
- **Ambient Occlusion** (Intensity = 1; Radius = 0.3; Sample Count = Medium; Downsampling Enabled)
- **Motion Blur** (Shutter Angle = 270; Sample Count = 10; Frame Blending = 0)
- **Grain** (Intensity = 0.5; Luminance Contribution = 0.8; Size = 0.7; Colored Enabled)

We used the 2017.3.0f3 version of Unity3D to display our VE. It was displayed using a VR-ready laptop with Windows 10, Intel core i7-7700HQ at 2.8GHz, with 12GB of RAM, with an Nvidia GeForce GTX 1060. The HTC Vive was used for our study, and the users were given an HTC Vive controller to provide inputs as appropriate.

Procedure

Recruited participants were seated inside of the HTC Vive play area. We collected demographics and explained the purpose of the study. We noted the height of the participant and subtracted the difference between their eyes and the top of their head (approx. 15cm) to record their Eye Height. Following, we explained that their objective was to manipulate the height until it felt the most natural to them, and to identify how tall and how short the camera could go, before it negatively affected their experience.

[1] https://vuze.camera/camera/vuze-plus-camera

We randomized and logged the initial camera height; the HTC Vive play area was minimized such that participants' head motions (aside from rotation) would not cause a change in camera placement. Thus, user height did not cause a change in display. Using the random initial height, we correspondingly scaled the virtual streamer, such that the camera remained in the same place on its body during the study. For purposes of our study, we assumed camera stabilization; i.e., the streamer's walking gait had no effect on the camera, except for when the avatar turned a corner. Inside the VE, the streamer walked around the museum one full rotation before we asked any questions, to let users familiarize themselves with the environment. After the streamer completed the round, we gave the participant an HTC Vive controller. By clicking on the top or bottom of the thumb pad, they could increase or decrease the camera height by 2.54cm (1 inch), respectively. The virtual streamer scaled with the camera, so to be visually consistent. When the user increased the camera height, the streamer became taller. When the user decreased the height, the streamer became shorter. The streamer continued walking in the museum at a constant speed, regardless of scale.

We then asked the user to adjust the camera until it arrives at the most natural height for them; we logged this value. Next, we asked them to slowly increase the camera height until the view started to become too tall; we logged this value. We then manually reset the camera height to the value corresponding to the "most natural" height, and then asked the participant to slowly lower the view until it started to become too short; we logged this value. After collecting the data, we reset the environment and toggled the *Camera Placement* to the next condition, repeating these steps. The time to complete the study was approximately 15 minutes, and participants were given 5 USD in cash.

Research Hypotheses

There are a number of factors which can play a role in determining a user's preferred camera height. In the previous chapter, we discussed that different camera placements on a streamer's body affect user perception of height; multiple researchers found that third-party avatars help to provide a sense of scale [129, 207] in a VE; and, expectedly, camera height preferences should scale with a user's actual height. As such, we hypothesized the following:

- **H1**: Participants will prefer significantly lower camera heights when the camera is placed overhead.
- **H2**: Participants will prefer significantly lower camera heights when avatars are not present.
- **H3**: A participants height will be significantly associated with their preferred camera height.
- **H4**: Females will prefer a significantly lower camera height than males.

Dependent Variables and Data Analysis Approach

As noted in our procedure, we collected data points from each user per condition. These data points were how tall the camera was in the environment, in centimeters. Per condition, we measured the user's height selections for *Natural*, *Tallest*, and *Shortest*. We used a Shapiro-Wilks test of normality to determine if our dependent variables were normally distributed, and they were (Natural: $p = .099$; Tallest: $p = .186$; Shortest: $p = .139$). To address our research hypotheses, we conducted an analysis of covariance (ANCOVA), with *Camera Placement* and *Avatar Presence* as fixed factors, and participant height and gender as covariates. Additionally, we use descriptive statistics in order to detail the acceptable range of camera heights.

Subjects

We recruited 97 college students from the University of Central Florida using an email blast, but we identified 4 who gave us outlier data points (values outside of three standard deviations); we removed them. Our resulting participant pool consisted of 63 males and 30 females. Their age range was from 18 to 44 (M = 22.2; SD = 4.17). We measured participant height; the range was 135cm to 180cm (M = 159cm; SD = 10.7cm), following a normal distribution; see Figure 5.1 for an illustration of participant height. All participants had normal vision, or they wore corrective lenses during the study. On a 5pt scale (1 = Never, 5 = Very Frequent), the mean VR experience among our users was 2.1 (SD = 1.3).

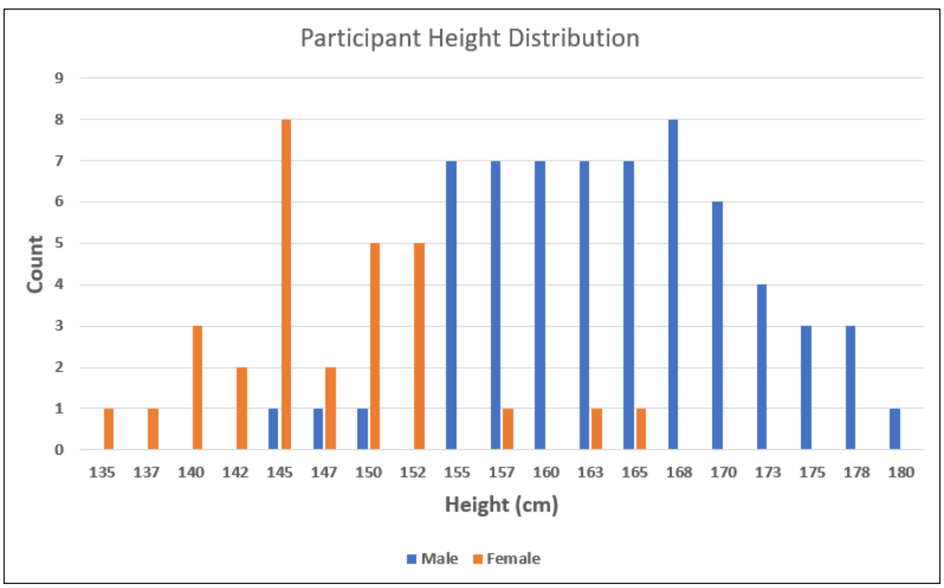


Figure 5.1: Histogram of participant height, by Gender.

Table 5.1: Descriptive Statistics of Preferred Camera Height (cm) by Camera Placement, Avatar Presence, and Gender

Camera Placement	Avatar Presence	Gender	Preferred Camera Height
Pocket	Absent	M	M = 142.7, SD = 30.23
		F	M = 118.2, SD = 32.26
	Present	M	M = 161.0, SD = 17.91
		F	M = 140.7, SD = 24.41
Overhead	Absent	M	M = 147.9, SD = 29.82
		F	M = 117.5, SD = 28.72
	Present	M	M = 166.1, SD = 20.49
		F	M = 147.0, SD = 20.02

Results

In this section, we first present the descriptive characteristics of our data, followed by hypothesis testing using a repeated-measures ANCOVA.

Descriptive Statistics

Tables 5.1 and 5.2 describe the mean heights and standard deviations for our DVs. On average, when human avatars were absent, participants set the height of the camera closer to the floor. Further, the participant responses for preferred camera height has far less variance when avatars were present in the scene; see Figure 5.2. Additionally and expectedly, men set the camera heights taller than women. Further, our participants mainly set the camera only slightly higher in the Overhead conditions.

Table 5.2: Descriptive Statistics of Preferred Camera Height (cm) by Camera Placement and Gender, and Avatar Presence and Gender

Camera Placement	**Gender**	**Preferred Camera Height**
Pocket	M	M = 145.1, SD = 29.30
	F	M = 129.5, SD = 30.35
Overhead	M	M = 157.8, SD = 26.63
	F	M = 132.2, SD = 28.58
Avatar Presence	**Gender**	**Preferred Camera Height**
Absent	M	M = 145.3, SD = 29.88
	F	M = 117.9, SD = 30.02
Present	M	M = 163.6, SD = 19.27
	F	M = 143.8, SD = 22.17

Hypothesis Testing Results (RQ1 & RQ2)

We performed hypothesis testing using a repeated measures ANCOVA; Table 5.3 summarizes our statistical findings. *Camera Placement* was a within-subjects variable, and *Avatar Presence* was a between-subjects variable. Participants' gender and eye height were treated as covariates.

To test H1, which was that participants would compensate for seeing a streamer under them in the Overhead condition, we tested the main effect of *Camera Placement*. We did not find a significant main effect of *Camera Placement* ($F(1, 89) = 0.541, p = .464, \eta_p^2 = .006$), which means that users' preference of camera height did not vary based on the camera being placed overhead or on the streamer's chest.

To test H2, which was that third-party avatars would provide a sense of scale and thus lead to users selecting a camera height closer to their own height, we tested the main effect of *Avatar Presence*. We found a significant main effect ($F(1, 89) = 18.12, p < .001, \eta_p^2 = .169$), such that the absence of avatars lead to users setting the camera height lower to the ground. This also indicates that the

Table 5.3: Results from a Repeated Measures ANCOVA regarding effects of Camera Placement (CP), Avatar Presence (AP), User Gender (UG), and User Height (UH) on Natural Height Preference

Effect on Natural Height	ANCOVA Result
Main Effects	
CP	$F(1, 89) = 0.541, p = .464, \eta_p^2 = .006$
AP	$F(1, 89) = 18.12, p < .001, \eta_p^2 = .169$
UG	$F(1, 89) = 12.48, p < .001, \eta_p^2 = .123$
UH	$F(1, 89) = 0.331, p = .567, \eta_p^2 = .004$
Interaction Effects	
CP * AP	$F(1, 89) = 0.269, p = .606, \eta_p^2 = .003$
CP * UH	$F(1, 89) = 0.391, p = .533, \eta_p^2 = .004$
CP * UG	$F(1, 89) = 0.664, p = .417, \eta_p^2 = .007$

presence of avatars helps users to select heights closer to their own.

To test H3, which was that participants would prefer a camera height that is closer to their own, we tested the main effect of User Height. We did not find a significant main effect ($F(1, 89) = 0.331, p = .567, \eta_p^2 = .004$), which indicates that a user's height did not directly influence their camera height selections.

To test H4, which was that females would prefer a lower camera height than males, we tested the main effect of Gender. We found a significant main effect ($F(1, 89) = 0.331, p = .567, \eta_p^2 = .004$), such that women did select lower camera heights than men.

We did not find any significant interactions among our variables.

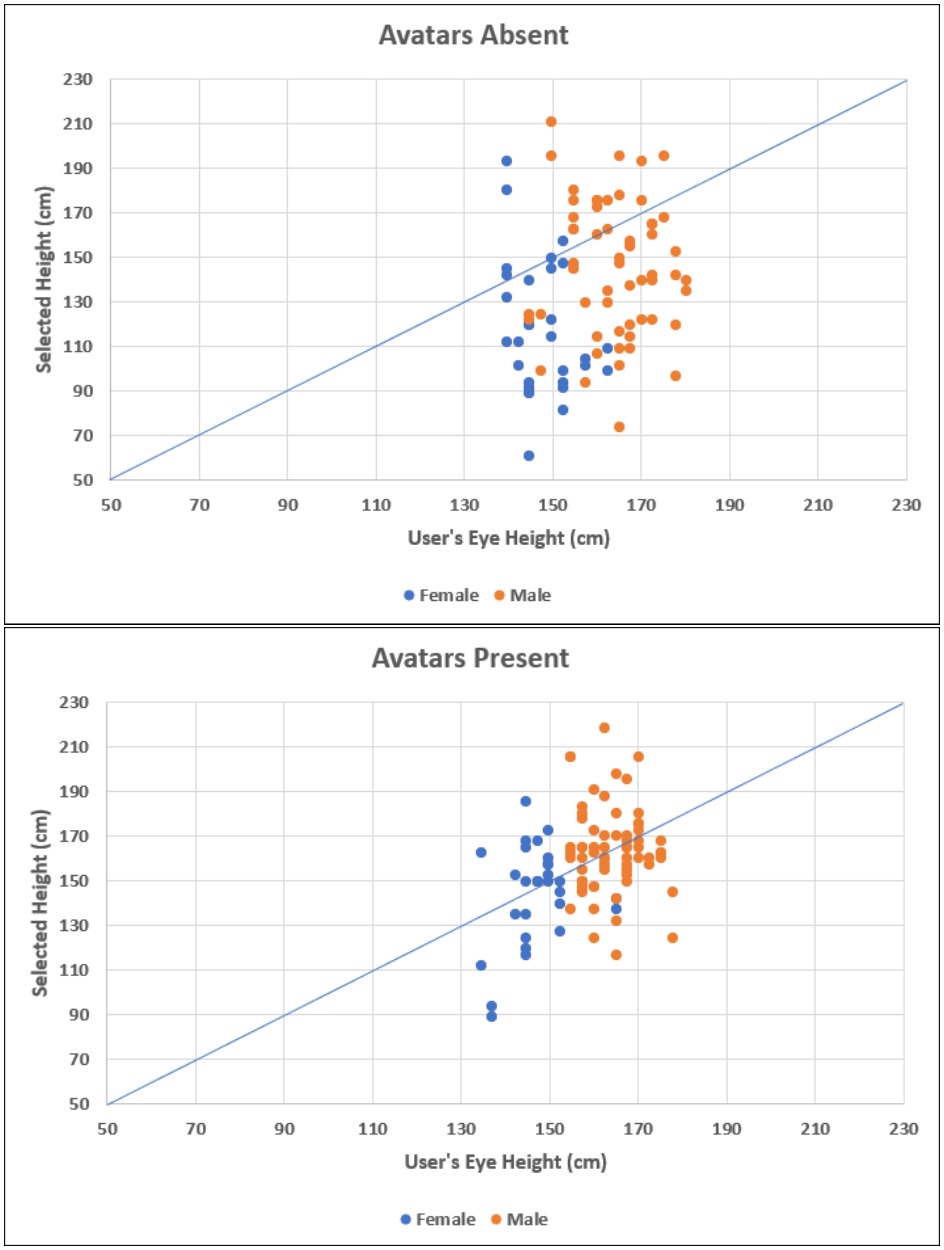


Figure 5.2: Scatterplots of user preferred camera height to actual eye height, by Gender. Top: Avatars Absent conditions. Bottom: Avatars Present Condition. Blue lines illustrates 1:1 ratio. In the presence of avatars, our participants clustered their responses closer towards their actual height than when avatars were absent.

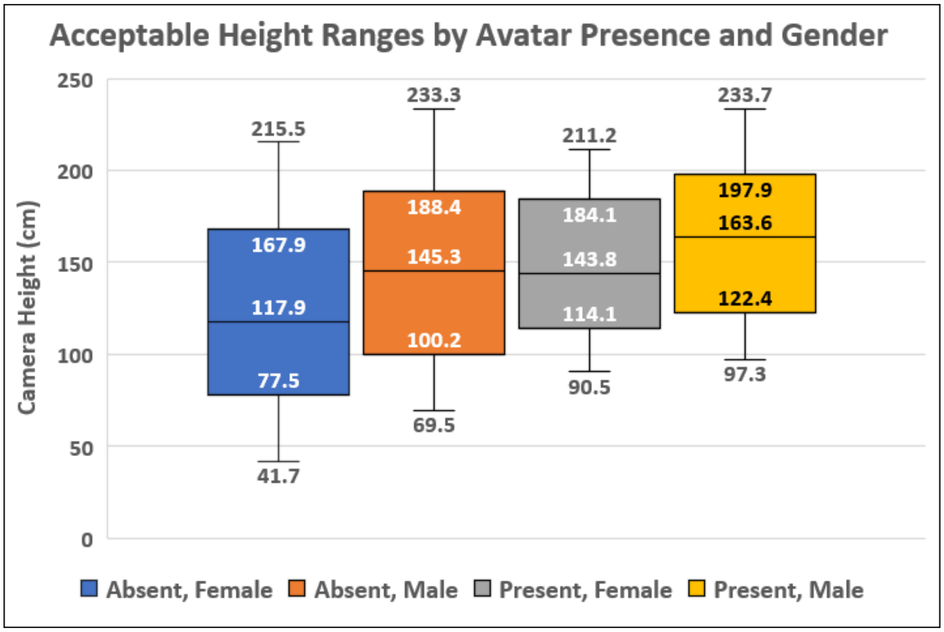


Figure 5.3: Average range of acceptable camera heights, by avatar presence and user gender. There is a very wide buffer surrounding the average camera height preferences which still affords an acceptable view. Black bars represent 1 σ.

Analysis of Height Preference Thresholds (RQ3)

To address our RQ3, we analyzed the responses given by our participants for their Tallest and Shortest heights. This assists us in determining the average range of acceptable camera heights; see Figure 5.3 for illustration. When avatars were absent, the average Tallest height thresholds were 167.9cm for females (SD = 47.6cm), and 188.4cm for males (SD = 44.9cm), and the average Shortest height thresholds were 77.5cm (SD = 35.8cm) and 100.2cm (SD = 30.7cm), respectively. This results in an average range of acceptable heights that is approximately 90cm wide for both men and women.

When avatars were present in the scene, the average Tallest height thresholds were 184.1cm for females (SD = 27.1cm), and 197.9cm for males (SD = 35.8cm), and the average Shortest height thresholds were 114.1cm (SD = 23.7cm) and 122.4cm (SD = 25.1cm), respectively. This results in an average range of acceptable heights that is approximately 72cm wide for both men and women.

Both of these results indicate that the acceptable range of camera heights is *very wide*. These ranges completely encompass average human height in the United States minus 3 standard deviations [57]; but, only in the Avatars Present conditions do we find the range to encompass average human height plus 2 standard deviations. This means that any given viewer should have an agreeable user experience, if the streamer wears the camera at their eye level or below.

Discussion

Our results offer various insights into optimizing camera height for user experiences. In this section, we describe our recommendations for telepresence designers, which include considerations for individual differences and various social environments.

Contextual Cues within the Environment Matter More Than Camera Placement

Our results indicate that avatars in view of the camera provide a significant frame of reference, strongly affecting a user's preferred camera height selection (H2). It is important to consider the social environment when conducting an immersive experience. If it is known to have people walking around (e.g. a popular art museum, concert, or sports venue), it becomes more appropriate to set the camera height closer to the viewer's natural height. In a socially empty scene (e.g. nature trail), camera height becomes more arbitrary. Additionally, our study uses a novel approach in the light of VR distance underestimation, and though our task was unique compared to previous efforts, our results corroborate with prior findings that avatars indeed provide a strong frame of reference [129, 207].

We also found that bodily camera placement on a streamer is not a significant factor that influences user height selections. This contradicts prior work in which an Overhead camera placement influenced participants to believe the camera was taller than it actually was [195]. Our intuition tells us that this finding stems from a difference of study procedure. In the previous study by Pfeil et al., participants were shown three unique camera heights (users' natural height and +/- 30cm) from the perspective of three bodily placements [195]; in our study, participants had the ability to change height at will, at 2.54cm increments. By allowing participants to explore heights in-depth, we expect that our study elicited more accurate personal preferences.

Gender Differences Matter

Although we anticipated our user-centered procedure to draw out user height as a significant factor (H3), our results corroborate with previous literature in that it is not [113]. However, we did find that user gender is a significant factor to consider when developing an immersive experience. Expectedly, men chose heights approximately 20cm taller than women prefer (H4), and while this is a seemingly obvious finding (as height is correlated with gender), to our knowledge, no recommendations in prior literature for optimizing camera height have considered this gender difference. Our results show that this it is an important consideration. In both academic literature and online articles pertaining to 360° film, it is currently suggested to set the camera at a static height [113, 130, 200, 216], but our findings indicate females would want a different view compared to males, to provide a more natural feeling. As such, we recommend immersive experience designers to consider this difference, to provide optimal user experiences for all, instead of a one-size-fits-all approach.

In our study and in previous studies ([113, 216]), the majority of participants were men. As we have found the statistical importance gender has on camera height preference, an implication for future work is to design for both sub-populations and thus sample as equally as possible. In addition, we need to consider other individual differences not identified by our work. As immersive telepresence is projected to be used by a wide variety of people, it is important to understand how different populations will have different camera height preferences.

Identifying Appropriate Telepresence Streamers

Our study points to a large buffer of acceptable camera heights for telepresence. This means that for any given user viewing a camera feed in VR, there is a wide range of acceptable streamers who can provide a positive viewing experience; and, since camera placement does not seem to influence preferred view height, a given streamer can adjust a body-worn camera as needed, without worrying about negatively affecting the view quality. The main consideration, then, is the ability to freely explore the environment unoccluded. In general telepresence use, we are confident that average human height will not detract from the experience. However, we acknowledge that there are prominent use cases in which the streamers will push past these thresholds. For instance, the National Basketball Association (NBA) has been ramping up court-side live-streaming for VR [177], and we expect that in the future, we will see players wear cameras to provide first-person viewpoints of a game. Since the average player height in the 2018-2019 season was over 200cm [208], which is outside the "acceptable" range of camera heights, we suspect a drop-off in terms of user experience. More work is needed to help understand how people react to large gaps in height between viewer and streamer.

Limitations and Future Work

We would like to acknowledge the limitations of our work that can inform future research. First, our study merged related work in immersive body-worn telepresence within a virtual reality context. Therefore, our findings inform researchers and designers on understanding users' camera height preference in the context of a simulated virtual environment. However, while this work might inform future work on interpersonal immersive telepresence, we cannot verify the generalization of our findings to the real world. As such, an area of future inquiry would be to replicate this study in a real context, instead of a virtual environment, to see if our results withstand.

Additionally, within the VR perception community, distance estimation tasks are usually constructed through blind-walking, bean-bag tossing, or verbal judgment calls [110, 166, 191, 207]. We are more interested in general scenarios moving towards immersive telepresence, and here offer a more user-centric measurement. However, our study did not include a measure for user satisfaction. Rather, we used self-selected camera height as a proxy for their personal preference. In our future studies, we would like to incorporate both user-selected choice and measures of user satisfaction to better compare our results to the previous findings of Keskinen et al. [113].

Conclusion

Previous research on setting camera height in 360° film and immersive VR tried to find "one-size-fits-all" values, but we have shown that different heights are warranted for the different genders, particularly because of the natural height difference between men and women. Additionally, these target heights change depending on the presence or absence of third-party humans in the environment, from both a VR and telepresence point of view. To accommodate these differences, a streamer can naturally adjust the camera by wearing it on different parts of the body, without affecting quality of view in terms of height. Overall, in a social telepresence setup, there seems to be a wide buffer of acceptable camera heights when worn on a streamer's body. Therefore, we do not anticipate a loss of quality should the streamer and viewer be of differing heights.

Our study contributes knowledge to various communities. First, we find that the presence of human avatars in a virtual environment does influence the selection of preferred camera height. In the VR community, this knowledge can help designers overcome the distance underestimation problem, when their goal is to create environments that allow users to feel like they are "actually there." For the telepresence community, this indicates that varying social landscapes influence users to desire a camera height closer to their own eye-height. However, our analysis *also* indicates that, while preferences do exist, the acceptable range of heights is vast. Therefore, telepresence designers should strive to keep their camera height in the range of average human height, but should not encounter issue if the camera deviates from a user's eye height.

In the previous two chapters, we demonstrate how proper body-worn camera placement affects the quality of view, and that streamer height is not expected to play a significant role in viewing experience; however, we do find that preferred camera height differs between men and women. As such, individual differences must be accounted for in the design of interpersonal telepresence experiences. We are confident that future designs can meet the needs of the viewer, but the question of self-consciousness still remains. We thus now shift our attention to the Streamer's needs, a necessity per our literature review. Towards this end, we describe in the next chapter work towards identifying human values that should be embedded in future prototypes to support this stakeholder. Whereas much of the literature in this space presents new devices and provides an evaluation, we give users a voice early in the design process by conducting a co-design study. In this way, we can better identify capabilities to support what is actually desired.

CHAPTER 6: CO-DESIGNING FOR INTERPERSONAL TELEPRESENCE: NEGOTIATING THE VALUE-TENSIONS BETWEEN VIEWERS AND STREAMERS

To begin our work towards better telepresence experiences for Streamers, we conducted a study that used a series of co-design sessions with diverse groups of individuals. The purpose of this chapter is to describe this study, in which we identify human values that should be embedded in a future telepresence setup; and, whereas much of the work in this space describes technology probes, we instead conduct an empirical investigation per the VSD framework, where we give more power to end-users by garnering their feedback early in the design process (see Figure 6.1).

This chapter describes the methods we used to conduct these co-design sessions, which were held during COVID-19 lockdown. At the time of work, it was not feasible to conduct these sessions in-

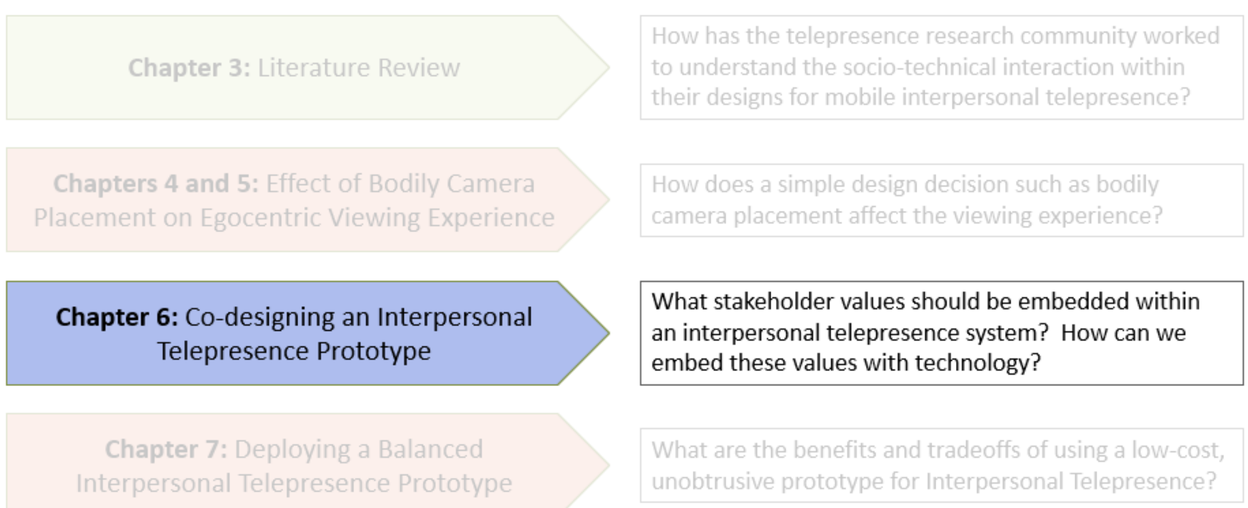


Figure 6.1: In this chapter, we perform an empirical investigation in which we identify human values that should be embedded into future interpersonal telepresence designs.

person, so we utilized a virtual "paper dolls" technique on an online collaboration platform. The items found in the study's materials represent technologies that were described in our literature review (Chapter 3). This chapter is divided into major sections — first, we ground the work with a brief introduction; then, we discuss the methods used to perform our co-design sessions; following, we describe our results; and lastly, we provide a discussion and implications for future work.

Introduction

As shown in Chapter 3, many of the prototypes used to facilitate interpersonal telepresence are typically designed without end-user input. The resulting devices tend to be created without specific consideration for the Streamer, and instead only showcase advanced capabilities that benefit the Viewer. A disconcerting finding among relevant studies is that the Streamers sometimes feel socially awkward using the developed devices [202, 205], which indicates a specific need to focus efforts on designing a prototype that would be socially acceptable. Another potential pitfall of creating new prototypes without incorporating user input in the design process is that many capabilities may not even be desirable; and as we identified, many studies in this area only describe laboratory experiments, and do not deploy their solutions for field studies using real stakeholders.

As such, there is a specific need to capture and incorporate stakeholder feedback, to direct future development of interpersonal telepresence prototypes. In this chapter, we therefore ask the following research questions:

1. **RQ1:** What stakeholder (Streamers and Viewers) values should be embedded within an interpersonal telepresence system?
2. **RQ2:** What features should be included in an interpersonal telepresence system to embed these values?

3. **RQ3:** How can we resolve the value tensions between Streamers and Viewers to translate these values into a real-world system that optimizes both users' experiences?

To answer these research questions, we conducted a participatory design (PD) study in which we ask a diverse group of individuals to work together and identify the features that should be incorporated into a desirable prototype, using a "Bag of Stuff" technique [51]. Our participants were split into two groups — one was asked to design a system from a Streamer's point of view, and the other was the same, but from a Viewer's point of view. After completing a design activity, our participants then merged to create a design that incorporated feedback from both groups — a balanced prototype. Overall, we found that our stakeholders, though prompted to take a greedy approach during the design sessions, made considerable effort to ensure that the other party would still be satisfied. The Streamer groups wanted to ensure a positive viewing experience, and the Viewer groups generally wanted to account for the streamer's social well-being. Though the selected devices were fairly different between the groups, we did note that the reasoning behind their choices had significant overlap. Through this research, we make the following unique contributions:

1. A user-centered co-design study that allowed stakeholders to make direct design decisions for interpersonal telepresence
2. Identification of the values that drive expectations for new prototypes in the interpersonal telepresence space
3. Interpersonal telepresence design recommendations that embed user-centered stakeholder values which future prototypes can reference.

In the following sections, we describe our study procedure and data analysis approach. We follow with the results of our study, including codebooks that captured emergent themes from participant feedback, followed by a discussion that includes design considerations for interpersonal telepresence.

Methods

Previous work on interpersonal telepresence systems shows that the design of these systems prioritize the Viewer's needs over the Streamer's user experience. Therefore, we wanted to help identify how to achieve a balance of power that favored both of these primary stakeholders. Prior work in this space mainly queried participants to evaluate specific prototypes, and as such, there is a lack of work that applied a user-centric lens to develop these prototypes. We thus devised a co-design study that would allow potential stakeholders to have a voice in the system design process, by utilizing the VSD framework to guide our work. In the following sections, we discuss the co-design activities that were aimed at identifying features that should be included in a telepresence prototype, and the values that influenced the inclusion of these features.

Co-Design Session Overview

We conducted 5 co-design sessions with 26 participants to design interpersonal telepresence system prototypes. Due to the ongoing COVID-19 pandemic, we chose to hold our sessions online, through a private Zoom[1] web-conferencing call. Before recruiting participants, we received Institutional Review Board (IRB) approval to conduct our study. Since the study was conducted online, the IRB required consent to be obtained electronically prior to the study sessions. Each of session followed the same agenda — first, researchers and participants introduced themselves and verified that their microphones worked properly. Sharing video was not required. Then, four

[1] https://zoom.us

design activities were conducted in the following order: 1.) Brainstorming Design Scenarios, 2.) Designing an Initial Prototype, 3.) Presenting Big Ideas and Design Critiques, and 4.) Designing a Balanced Prototype. After the final activity was completed, participants were instructed to fill out a final online survey. The survey allowed participants to provide further insight into their design decisions, identify strengths and weaknesses of the final prototype, add any design changes they would make on a personal level, and state anything they were not comfortable sharing with the group. Participants were issued a $20 Amazon gift card as compensation for participating in the co-design session. All activities were audio and video recorded for analysis. In the following sections, we describe each of the activities in detail.

Activity 1: Brainstorming Scenarios

First, the lead researcher described the purpose of the study and introduced the concept of interpersonal telepresence. Then each participant was asked to share a recent live-streaming experience in an ice-breaker activity. They were given two prompts:

- Think about the last time you watched an "in-real-life" mobile live-stream. What was the scenario?
- Think about an activity you would like to watch in an "in-real life" mobile live-stream. What is it?

Participants were encouraged to verbally share as much information as they wanted, and the lead researcher followed up with various questions as necessary, to dig deeper into the participants' disposition to their experiences. This activity was performed to help participants feel more comfortable sharing their thoughts in the sessions, and to set the stage for the remainder of the session.

Table 6.1: Design prompts issued to our participants. The scenario was constant, but each group was asked to greedily design from either the Streamer or Viewer perspective.

Viewers	Streamers
You and a friend wanted to go to Disney Springs together. Disney Springs is a Disney-themed entertainment, dining, and shopping area in Orlando, Florida. However, due to a health concern, you need to stay at home. So, your friend is putting together a new live-streaming prototype that will let you feel like you are actually there. **They want to design it in a way that will let you have the best possible viewing experience. Our job is to design this prototype.**	You and a friend wanted to go to Disney Springs together. Disney Springs is a Disney-themed entertainment, dining, and shopping area in Orlando, Florida. However, due to a health concern, you need to stay at home. So, your friend is putting together a new live-streaming prototype that will let you feel like you are actually there. **However, they tend to feel self-conscious in public. They want to design the prototype in a way that will let them feel most comfortable when using it. Our job is to design this prototype.**

Activity 2: Initial Prototype Design

Participants were split into two pseudo-random groups; these groups were balanced by gender and experience level with live streaming technology. Each group was given their own private breakout room and asked to design a prototype for a given scenario. One group was tasked with greedily designing a Streamer's live-streaming prototype, but from the Viewer's perspective; the other group was tasked with greedily designing the same prototype, but from the Streamer's perspective. See Table 6.1 for exact design prompt verbiage.

We used these prompts because Disney Springs is a popular tourist destination that our participants would know, and we wanted to choose an experience that would warrant virtually bringing along a loved one who was unable to physically attend. We specifically chose the design prompt language because we wanted to explicitly understand how participants would tackle the problem of user self-consciousness, which has been documented in many related field studies [115, 198, 202, 205].

To design their prototypes, participants were given access to a slide package hosted on Google Drive[2], so that they could work on the same design synchronously. Each group had access to a unique file. The slide package contained 4 slides. The first slide had a blank mannequin in the middle, and the 3 subsequent slides depicted various live-streaming technologies, based on a collection of prior literature (See Appendix C for full list). Each technology had a picture representing it, a caption that described it, and a link to an example picture. We selected these items because they have been represented by academic works in this area, or because they are commonly used in live-streaming setups — our goal was to provide the participants with a comprehensive list of technologies representative of both academic and consumer-grade devices. This collection of technologies is akin to a "Bag-of-Stuff", which is a commonly-used co-design technique that allows users to visually design a prototype using arts and crafts [51]. Participants were able to copy/paste items from this list of technology onto the blank mannequin, but they were also informed that they could find their own images from the internet, draw with primitive shapes, or simply describe an item with text. Each group had 25 minutes to work together to build their prototype, discuss their ideas, and describe why they selected or rejected items.

Activity 3: Big Ideas and Design Critiques

Once time was up, both groups came back together to present their designs. The first group to present was randomized. Each group presented their ideas along with their justification of why items were chosen. The lead researcher also asked clarifying questions related to the planned capabilities exhibited in the design. After a group finished presenting, the other group was asked to identify strengths and weaknesses for the given design; after, the two groups swapped roles. This activity lasted approximately 25 minutes.

[2]https://www.google.com/intl/en/drive/

Activity 4: Designing a Balanced Prototype

Finally, participants were asked to consider the strengths and weaknesses that were identified in Activity 3 to design a final, balanced prototype. This activity resembled the design technique "Mixing Ideas" in which participants share in large groups their respective inputs [264]. The lead researcher assembled the design in a new template. Participants were encouraged to discuss rationale for selected items, to reach a consensus in the event of conflict, and to work together to find appropriate compromises. This activity lasted for approximately 25 minutes.

Participant Demographics

We distributed information about our study through university channels as well as online message boards in Orlando, Florida. Interested individuals were asked to fill out an online form to ensure eligibility and indicate which sessions they would be available to join. Participants were required to be 18 years old or older, speak English, have a computer with Zoom capabilities, and a stable internet connection. We asked participants to provide their age, gender, and to self-report, using a 5-point Likert scale (1 = No Experience and 5 = Very Experienced), their level of experience with the following: Videography, Filmmaking, Live Vlogging, Virtual Reality, Augmented Reality, 360° Cameras, and DSLR Cameras.

We had a total of 26 individuals participate in our study (P5, P20, and P27 left the session early, but are still included in the analysis). We had 13 females and 13 males, and participants were between 18 to 32 years old (M = 22.4, SD = 4.11). Table 6.2 provides a summary of our participants.

Table 6.2: Participant Demographics. We had five unique sessions, and participants were pseudo-randomly assigned sub-groups for the break-out activities.

Session ID	Participant ID	Group	Age	Gender	Expertise (of 5)
S1	P1	Streamer	28	F	3.6
S1	P2	Streamer	21	M	1.1
S1	P3	Viewer	20	F	3.6
S1	P4	Viewer	19	M	2.1
S1	P5	Streamer	20	M	3.6
S1	P6	Viewer	22	F	2.1
S2	P7	Viewer	18	M	2.3
S2	P8	Viewer	24	M	3.6
S2	P9	Streamer	26	F	1.1
S2	P10	Streamer	20	F	2.6
S3	P11	Viewer	20	M	2.0
S3	P12	Viewer	20	F	1.1
S3	P13	Streamer	23	F	2.9
S3	P14	Streamer	27	F	2.9
S3	P15	Viewer	31	M	3.3
S4	P16	Streamer	30	M	3.0
S4	P17	Streamer	29	F	1.9
S4	P18	Viewer	19	F	1.4
S4	P19	Viewer	28	M	4.3
S4	P20	Viewer	24	F	2.4
S4	P21	ST	19	M	1.0
S5	P23	Streamer	19	F	2.9
S5	P24	Streamer	20	M	2.3
S5	P25	Viewer	18	M	2.7
S5	P26	Viewer	20	M	3.1
S5	P27	Viewer	18	F	1.7

Data Analysis Approach

We recorded all sessions, which were manually transcribed. The session artifacts (prototypes from Activities 2 and 4) were saved as digital images. Using participant utterances and digital images, we conducted three different qualitative analyses. First, we conducted a thematic analysis, using the transcriptions from Activity 2, to understand what *values* were important to our users when designing technologies suited for telepresence interaction [56]. We used open coding to log participants' explicit and implicit values from their utterances. Then we used axial coding to form major themes across the participants. The final codebook summarizing our themes and codes is shown in Table 6.3.

Next, we conducted a feature analysis to understand what technologies should be included in a mobile telepresence prototype that would support the values identified by our participants. We used both the digital artifacts from Activity 2 and the transcripts from Activities 2 and 3 for our analysis. The final codebook summarizing the selected features is shown in Table 6.4.

Lastly, given the nature of our study, we expected to find value tensions and feature conflicts between the two groups; it is important to understand how to reconcile these conflicts. We thus conducted a final thematic analysis to identify the negotiated features that were agreed upon by both groups. Using the dialogue and prototype images from Activity 4, we performed apriori coding, with the codebook found in Table 6.4 as a guide.

Findings: Identifying Human Values

We begin presenting our findings by reporting the emergent themes and codes from our value-sensitive thematic analysis. In this section, we present the human values that guided our participants' interpersonal telepresence prototype designs. In this section, the unit of analysis is participant; N = 26.

Table 6.3: Final Codebook used for our stakeholder Values analysis.

Theme	Codes	Exemplar Quote
Streamer Experience		
Reduce Burden on Streamer	Physical Comfort	"*This is comfortable for the person wearing it too, in a way, 'cause there's no harnesses on the body, you know, it's just one camera on the ear.*" - P9, Streamer
	Minimalist Design	"*I also think that in terms of how many devices the person has, like, maybe minimize that so the person can enjoy some kind of experience as well.*" - P20, Viewer
	Personalization	"*If the viewer wants, it's just a preference... Should have both options. Like you have that option, if you are using the VR, or you can just control like hand, and move it any other place.*" - P16, Streamer
	Easy Set-up/Removal	"*It would be nice if at like lunch break or something, then when you take everything off and just like rest, they can put everything in the backpack... or like if they're using the restroom, they can just put everything away quickly, and then put it back on, once they come out*" - P13, Streamer
	Cost	"*You also have to consider price, and that most people, when they do a live-streaming thing, they would want to use it with a device they already have...*" - P21, Streamer
Support Social Needs of Streamer	Privacy	"*I don't know if the wearable speakers are going to be too much, but if they work, he could listen to you pretty clearly, but people around you will also listen to you clearly, so I don't know if you want to keep that privacy aspect just for your friend, or you want to share that to basically everyone.*" - P8, Viewer
	Safety	"*If you're walking at [Disney], someone walks in front of you, there's a little kid that runs in front of you, like just for safety reasons, for the streamer and their environment, they should still be mostly tapped in with where they are physically.*" - P26, Viewer
	Social Acceptance	*"I think it's just one of those things where like, if you're in [Disney], you're gonna be out in public, right? So you don't want to look ridiculous with all the stuff..."* - P25, Viewer
Viewer Experience		
Allow Viewer Freedom to Explore	Natural Sight	*"We don't want it too low too cuz then you know the person seeing it - we need a good placement for the camera for them to see."* - P9, Streamer
	Autonomy	"*I don't have to rely on my friend to move...if I wanted to see which is on the left or on the right or on my backside, I don't have to ask my friend to make a turn or look in that direction.*" - P19, Viewer

Theme	Codes	Exemplar Quote
	Physical Representation	*"I was maybe thinking, if there's like any way to merge maybe the backpack with some sort of system in which you could have the robot next to you... cuz I was thinking of doing like an extension to the person who was not physically there."* - P8, Viewer
Viewer & Streamer Interaction		
Support Together-ness	Being Side-by-Side	*"You would have the same view as them, but I think it might be better if you were like, there next to them, as a camera they're holding, or something, rather than, uh, the camera being mounted on their head..."* - P12, Viewer
	Shared Experience	*"I think we should add that kinda of microphone to speak to the other people as well, and then the 360 camera, and that kind of thing, so we have the option. Its kinda we are together there, and then he has some freedom, but then kinda we cooperate to have experience together."* - P16, Streamer

Identifying Values for the Streamer's Experience

We first describe the values participants expressed related to the Streamer's experience (N = 26), highlighting areas where concurrences and contentions occurred.

Reducing the Burden on the Streamer

Most of the participants (N = 23; 11 Streamers, 12 Viewers) shared values that would help ensure no affliction would be caused towards the person physically located in the target environment (i.e., the Streamer). These participants focused on making sure the Streamer's needs were met.

More than half of these participants (N = 18; 10 Streamers, 8 Viewers) were concerned for the ***physical comfort*** of the Streamer. They felt that the Streamer would be doing a lot of walking, therefore, the equipment would start to get heavy and uncomfortable.

> *"Having that on you all day could feel like kind of clunky. Always walking around with it on you, having to take it off when you're getting on rides or something... but I think*

that [gimbals are] probably too much for just a normal person just walking around.”
P4, Viewer

They wanted the wearable devices to be lightweight so that the Streamer could have an easier time handling the equipment. They also suggested distributing the placement of the devices so weight did not pull on one side of the body. Some even suggested including a harness or some sort of back support to help the Streamer with the weight of the equipment.

“One thing I’d like to see here is a harness that’s attached to the backpack, and maybe it would give some support to the items that will be attaching to it, because with time it’s going to get heavy and probably uncomfortable for the person who’s wearing it.”
P8, Viewer

There were even some participants who considered the heat emitted from the electronics. They felt the devices would be used for long periods of time or would have to be recharged, so that could make them overheat. The participants thus made sure the Streamer would not place the devices directly on their body.

“I think another thing to consider, with the battery, is it would get hot, especially as you’re using the phone while it was being charged... so for that, it’d probably be better to put it inside the [backpack] instead of on your shirt.” P2, Streamer

Seventeen participants (9 Streamers, 8 Viewers) noted that the ***easy set-up/removal*** of the prototype was of utmost importance. Participants realized that the interaction paradigm is bound to have both planned and unplanned incidents that would require the attention of the Streamer. These include situations where a pause in the interaction is necessary (e.g., a lunch or bathroom break), or where a disruption of service will be unavoidable (e.g., a device’s battery power will deplete).

Half of the participants (n = 13; 9 Streamers, 4 Viewers) discussed including devices users already have. They valued affordability. Many of these participants specifically mentioned that if they could use their own personal smartphone for video capture and earbuds to facilitate a conversation, then that would save them from having to spend large amounts of money on new equipment.

> *"Would we even bother with headphones? I feel like everybody has earbuds. It's more accessible."* P6, Viewer

Yet, this sentiment came from mostly the Streamers. Viewers tended to choose the new and exciting devices, thinking these would give them the best viewing experience.

Almost half of the participants (N = 12; 5 Streamers, 7 Viewers) valued a ***minimalist design***, in this case, reducing the quantity of devices used in the prototype. Participants felt that if the streamer had to juggle too much equipment, it would interfere with their experience and make them stand out from the crowd. They felt that managing too many devices would prevent the Streamer from participating in activities.

> *"I don't want my friend to carry too many devices, I want his hands to be free, because if you go to Disney Springs, you will have a lot of shops there, and you would also experience the kind of stuff they're selling, and you'd like your friend to hold those things in his hand, or try those things, or try some food there."* - P19, Viewer

Participants also emphasized that only the devices that were absolutely necessary (must-haves) should be included. This means, that any tech not required (nice-to-haves) should be kept out of the design. Hence, participants preferred wearable devices. Many of these participants were also those who expressed that these wearables should be lightweight or comfortable to use.

There was also a number of participants that felt that the prototype should allow for ***personalization*** (N = 7; 4 Streamers, 3 Viewers). These participants were against a "one size fits all" approach. For the Streamer participants, this meant that the on-site individual should consider taking an array of cameras (e.g., smartphone, AR glasses with embedded camera, action camera) and swap between them as needed. For instance, a head-mounted camera could be used when the hands need to be free, and a handheld camera could be used when more precision is needed.

Support the Social Needs of the Streamer

Another major theme was the desire to maintain the Streamer's self-image, in an effort to ensure they would not experience a feeling of self-consciousness. A total of 21 participants (11 Streamers, 10 Viewers) expressed some consideration of this value.

The majority (N = 16; 9 Streamers, 7 Viewers) expressed that the ideal telepresence prototype would be ***socially accepted*** by others in proximity with the Streamer. Overall, participants felt that the Streamer would not want others to know a live-streaming interaction was taking place. They also felt that some devices were more socially accepted than others.

> *"The multi-lens 360-degree camera... that one might be the best, since you'd be able to look around wherever; but then it might look a little ridiculous when the person is walking around."* - P11, Viewer

Therefore, they felt that the ideal prototype would be one that is hidden from the view of the public. Yet, participants recognized that this is not always feasible; so, they suggested that any visible technology should be ones that are socially acceptable and used by the majority of the public (e.g., a smartphone). Alternatively, they suggested designing these devices in a way that is visually attractive (i.e., fashionable).

There were some participants (N = 4; 3 Streamers, 1 Viewer) that were interested in preserving the ***privacy*** of the Streamer. In these cases, participants deliberated over choosing between speakers, headphones, and ear buds. All of these participants wanted to ensure the conversations between the Streamer and Viewer could not be overheard by others.

> *"Just have some headphones for talking like these airpods, because it's private as well, right?"* - P16, Streamer

There were 3 participants (1 Streamer, 2 Viewers) who valued the ***safety*** of the Streamer. The Viewers were most concerned with the types and quantity of devices that the Streamer would be using during their experience. They felt that some technologies may require too much mental demand, which can pose a physical safety concern on the user. For example, if the user is distracted by the need to operate the devices, they could enter into a dangerous situation (e.g., trip and fall). Likewise, the type of audio device the Streamer was using could pose a risk. Therefore, they wanted to make sure to include devices that *"don't block external sounds."*

> *"If I'm that person walking around Disney, I want to make sure, obviously, I'm being safe and not running into people and things, so I don't want a full headset, but, the person at home can obviously have that full headset experience if they'd like.* - P12, Viewer

In contrast to the Viewers, P17 (Streamer) also took into consideration the theme park's security. They realized there may be some regulations or additional considerations that may need to be taken into account when using certain devices, specifically drones.

Identifying Values for the Viewer's Experience

Next, we share the values participants (N = 26) expressed related to the Viewer's experience.

Allow the Viewer the Freedom to Explore

A total of 19 participants (9 Streamers, 10 Viewers) directly expressed a desire to ensure that the person receiving the remote experience could explore the environment in a natural way.

The most prominent consideration (N = 13; 8 Streamers, 5 Viewers) was ensuring that the camera angle captured a viewpoint that could give ***natural sight***. While designing the prototype, some of the Streamer participants thought that a 360° view would actually be a detriment to the user. Here, the participants felt that since the Streamer would be unable to control where their partner looked, they might miss out on what they are trying to show them. However, the Viewer participants believed that a 360° view is necessary, in order to maximize the ability to explore the environment, i.e. to look at what *they* would want to look at. Other participants recognized that depending on where the camera is positioned, the experience changes. For instance, most participants wanted to use a wearable device — but if the camera is on the shoulder, it might be too low when the streamer is sitting. If it is on top of the head, then it might be too high up — but it also depends on the environment itself. Thus, being able to react to a changing landscape to provide a positive viewing angle is a priority.

Ten participants (6 Viewer, 4 Streamer) opined that the viewer should have some sort of ***autonomy*** during the interaction. Similar to the above, these participants expressed that the Viewer should be able to choose their own viewing angle, but for various reasons. Streamer participants noted that this would reduce the number of times the Viewer would need to ask their partner to make corrections to the camera's angle. In essence, giving control to the viewer offloads work that the streamer needs to perform, while simultaneously making the viewing experience more robust.

> *"I don't have to rely on my friend to move...if I wanted to see which is on the left or on the right or on my backside, I don't have to ask my friend to make a turn or look in that direction"* - P19, Viewer

Nine participants (7 Viewer, 2 Streamer) indicated that they would want a ***physical representation*** of the Viewer, so that they felt like they were *actually there* in the remote environment during their interaction. Here is where we find one of the major discrepancies between groups. Whereas our Streamer participants succeeded at designing prototypes that would let a user *see* the remote environment, the Viewer participants were the ones who more often designed prototypes to let them *experience* it.

Identifying Values for Viewer & Streamer Interaction

Lastly, we found differing opinions between stakeholder roles for what *type* of experiences should be conveyed using interpersonal telepresence devices.

Support Togetherness

A total of 11 participants (6 Streamer, 5 Viewer) thought about what the actual interaction should feel like between the two partners using a potential setup.

Four Viewers (No Streamers) expressed the idea that a telepresence prototype should allow the two partners to feel as if they were ***side-by-side*** during the interaction. In general, this feeling could be facilitated by allowing both users to see each other, opposed to just letting the viewer see the streamer as captured by the camera. However, this also means that the streamer needs a way to see their partner as well. Use of existing applications such as video chat can help bridge the gap.

Seven participants (6 Streamer, 1 Viewer) expressed that it is more preferable for the viewer to see exactly what the streamer sees to provide a ***shared experience***. They wanted to see *only* what the streamer sees. Instead of affording the viewer with a way to see whatever they want in the remote environment, these participants wanted to restrict the view, but for selfless reasons. Some indicated that a 360° camera offers *too much* of a view, and that would be a major detriment; others felt that if the view was not first person, then it invites view misalignment that prevents the two users from understanding where the other is looking — after all, typical streaming platforms do not allow users with a way to incorporate natural body language, such as pointing and head nods. Therefore, restricting the view such that it always matches the streamer's heading ensures that both users will see the same stimuli.

Findings: Identifying Features that Embed Human Values

We next describe the results from our Feature Analysis. We wanted to understand what features Streamers and Viewers considered important and facilitated their values. Therefore, in this next section, we focus on the actual hardware and software each group of stakeholders chose for their ultimate prototypes. In this section, the unit of analysis is group; N = 10.

Overall, all groups indicated a need for live video streaming and features that support verbal communication between the streamer and viewer, with some additional features to better support the stakeholders. We describe each of their chosen features in greater detail. Illustrations of the final design artifacts are presented by group; see Figures 6.2, 6.3, 6.4, 6.5, and 6.6.

Table 6.4: Final Codebook used for our Feature Analysis.

Categories (N = 10)	Capabilities (N = 10)	Exemplar Quote
Live Streaming Video (100%)	Panoramic Video (60%)	*"I feel like a 360 camera would be better, just because that way you aren't forced to look in one direction"* - P26, Viewer
	Normal Video (50%)	*"We have a couple options, like the ear-mounted camera is pretty low-key, a smartphone is pretty typical, everyone's there taking photos so that would fit in"* - P13, Streamer
	Multiple Cameras (40%)	*"If he wants like a better experience, he could switch out for one of the other cameras. Or maybe like the AR glasses you could keep in the shirt pocket.*" - P24, Streamer
	Face-to-Face Video (30%)	*"I was kinda looking at more of the gimbal stabilizer with a smartphone, so we could use that for you to like see each other.*" - P26, Viewer
	Stabilization (30%)	*"Definitely the gimbal stabilizer would be pretty awesome, cuz you know how shaky it can get right?*" - P8, Viewer
Audio Communication (100%)	Synchronous Voice (100%)	*"We definitely need a way to communicate my voice definitely, so definitely need some headphones...*" - P8, Viewer
	Ambient Noises (40%)	*"I know [Disney] has like music playing when you go there, so I think having like something that can [relay] audio to you during the experience would make it better*" -P24, Streamer
Viewer Feedback (80%)	Tangible Device (40%)	*"Can we put like the strip - maybe just like something just like super tiny on the camera? And then... it would turn green when they're facing the right way."* - P23, Streamer
	Visual Augmentation (30%)	*"The AR glasses would kinda be your FaceTime, and there would be like low overhead, like you wouldn't have to look down at your phone or anything.* - P11, Viewer"
	Text Messaging (20%)	*"The phone would be the next accessible thing for a chat to be on the phone while they're walking around.*" - P4, Viewer

Categories (N = 10)	Capabilities (N = 10)	Exemplar Quote
Streamer Considerations (70%)	Physical Storage (60%)	*"It would be nice if at like lunch break or something, then when you take everything off and just like rest, they can put everything in the backpack."* - P13, Streamer
	Hands-Free Mount (30%)	*"In terms of having like a pole or something like that to hold the camera, if its simple... that would help a lot the person that is in the place as well".* - P20, Viewer
	Ambient Noise Passthrough (20%)	*"It might be better to have like for the person at Disney... to have earbuds that don't block out external sounds"* -P12, Viewer
	Shrouding (20%)	*"[Discussing wearable 360 cameras] I would say probably have the streamer wear a hat as well, just for their sake.*" - P26, Viewer
	Weight Alleviation (10%)	*"One thing I'd like to see here is maybe like a harness that's attached to the backpack, and maybe it would give some support to the items that will be attaching to it."* - P8, Viewer
Viewer Considerations (70%)	Virtual Reality Display (60%)	*"Know what would be cool? Is if you got that going on, but then your buddy is sitting at home with VR goggles on.*" - P4, Viewer
	2D Display (20%)	*"[Discussing displays] I would say a computer, or a phone, like I don't think they would have any other technology.*" - P13, Streamer
	Remote Control (20%)	*"We can give the person that is at home control to the drone, so that person can remote control the drone.*" - P16, Streamer
	Physical Embodiment (20%)	*"[Discussing a Robotic Avatar]I was thinking of doing like an extension to the person who was not physically there.*" - P8, Viewer
Additional Considerations (50%)	Battery Charging (40%)	"*There should be the power banks probably to charge all the items.*" - P8, Viewer
	Network Connection (30%)	*"The smartphone has built in streaming services as well, so you don't have to have some like encoder, and then a broadcaster, and all those things.*" - P5, Streamer
	Drone Tethering (10%)	*"The streamer should be able to set parameters, through an app on their phone.*" - P21, Streamer

Live Streaming Video

All groups specified the need for live video capture within their designs (N = 10, 100%), but the types of cameras — and the resulting features — were different between the two stakeholder groups. We go into further detail below.

Panoramic Video

One of the clear divergences between groups was the type of camera. As 360° cameras empower the Viewer to explore the remote environment on their own accord, panoramic video was thus a popular choice among the Viewer groups; it was chosen in all of these groups (N = 5). Four of these selected a multi-lens 360° solution that sits on the head of the streamer (akin to the JackIn Head project [105]), while 1 chose a dual-lens 360° camera that could be mounted upon a pole and moved with their hands. No groups chose to include a wide-angle lens camera.

Only one Streamer group specified use of a panoramic camera, but this group also brainstormed the idea of a quadcopter drone that would follow the streamer around, and capture the environment using a mounted camera. This not only lets the Viewer look around on their own accord, but it also reduces the amount of effort the streamer needs to exert. Overall, the Streamer-focused groups generally felt that panoramic cameras were not conducive to a positive streaming experience, whereas a regular video feed would be more preferable. Next, we dig deeper into the reasons for including normal video capture devices instead of panoramic devices.

Normal Video

Instead of selecting a wide angle camera or a multi-lens device, the Streamer groups tended to choose normal, single-lens device that could facilitate live-streaming (N = 4). Three of these groups chose a smartphone; two chose an ear-mounted camera; and one chose a head-worn camera incorporated in an Augmented Reality Glasses device. Note that some groups decided to include multiple cameras in the same setup; for instance, one Viewer group chose to include a smartphone device to supplement their selected panoramic camera. We dive deeper into rationale for multiple cameras in the following section.

The Streamer groups did not unanimously select a single-lens video capture device, but a smartphone with built-in camera seemed to be a popular choice. Participants identified a variety of benefits to using such a device; for instance, modern smartphones are essentially miniaturized computers, so they can serve multiple purposes (e.g. network tethering, video calls, texting). Further, participants suggested that since smartphones are common in today's society, they are an obvious inclusion, as it would reduce the overall cost of the system. The groups that selected an ear-mounted camera found such a device to be a positive inclusion due to its minimalist form and

Figure 6.2: Design artifacts from Session 1. Left: Streamer. Middle: Viewer. Right: Balanced.

Figure 6.3: Design artifacts from Session 2. Left: Streamer. Middle: Viewer. Right: Balanced.

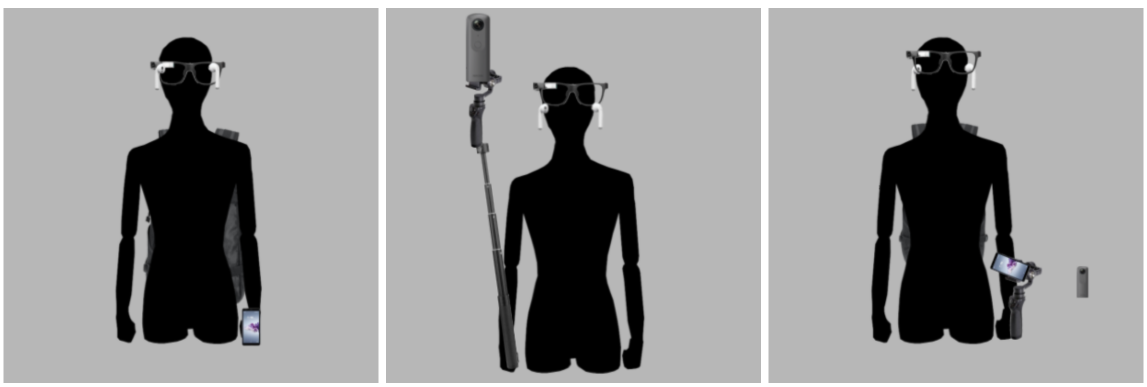


Figure 6.4: Design artifacts from Session 3. Left: Streamer. Middle: Viewer. Right: Balanced.

Figure 6.5: Design artifacts from Session 4. Left: Streamer. Middle: Viewer. Right: Balanced.

Figure 6.6: Design artifacts from Session 5. Left: Streamer. Middle: Viewer. Right: Balanced.

inconspicuous nature. Further, since it is tied to the streamer's head rotation, the two users easily have an aligned view.

One group mentioned that this device would satisfy an "eye view" criteria, which was that the camera allowed for both users to have the same exact view. This was echoed by another group who offered insight that panoramic cameras would give *too much* of a view.

For these reasons, the inclusion of normal video capture devices was prominent in Streamer groups. However, a subset of groups also chose to include multiple video capture devices. In the next section, we dig into the reasoning behind this choice.

Multiple Cameras

Three groups in total (2 Streamer, 1 Viewer) indicated that the ideal prototype would include multiple cameras instead of just one. The common reason for including multiple cameras was so the users could adapt to different environments or switch the focus of the streamed experience; for instance, one group noted that sometimes the desired interaction is face-to-face video chat, and other times the Viewer might want to see more of the remote environment. Akin to work by

Kim et al., using multiple cameras simultaneously also allows a Streamer to showcase a remote environment while maintaining a face-to-face video chat [115].

However, the availability of multiple cameras also affords the Streamer with ways to adapt to a changing environment. For instance, one group suggested that a Streamer might want to switch between devices, for reasons such as arm fatigue, a change in scenario that requires the Streamer's hands to be free, or because the visual landscape shifts. Thus, incorporation of multiple cameras allows the Streamer to maintain a positive experience for both parties.

Face-to-Face Communication

Although both groups received the same basic goal, which was to let the streamer and viewer share an experience and feel like they were together, only three groups felt the need to include face-to-face communication; all were Viewer groups. To facilitate this style of interaction, the groups chose to include either smartphones for video chat, or AR Glasses that would display a video graphic in the eyes of the Streamer. AR Glasses were preferable because they allow the Streamer to keep their hands free, and the graphic overlay could always be seen regardless of head orientation. However, the use of a smartphone was also a preferable choice because video chat applications are ubiquitous, being supported by all major smartphone platforms. As such, a Streamer would not look out of place while holding their phone in their hand.

Video Stabilization

The final Live Video Streaming feature included by our participants was video stabilization. Multiple groups identified the immediate benefit of stabilization, but only three groups (2 Viewer, 1 Streamer) decided to include this feature. The Viewer groups elected to use a physical gimbal stabilizer. These groups thought the Streamer's walking gait would introduce too much jitter, and thus the physical stabilizer could offset any motions that would otherwise cause a detriment to the viewing experience. Other groups discussed the physical stabilizer's benefit, but ultimately decided that the scenario we posited was not intense (i.e. it was not a sports scenario that required running or jumping), and therefore the stabilization was not necessary.

The lone Streamer-focused group who elected to include stabilization took a software-based approach, as gimbal hardware conflicted with their body-worn design. According to these participants, a physical stabilizer requires the users' hands to hold it upright, but it would also be too bulky and stand out too much. For these reasons, a physical device is not preferable. Our "bag of stuff" did not include software solutions, and therefore the previous reasons may be the cause for the other 4 Streamer groups not including any type of stabilization in their designs.

Audio Communication

All 10 participant groups identified the need to capture live audio within their telepresence prototypes. There were two main considerations — synchronous voice, where the two users could speak to each other in real time, and inclusion of ambient noise, where the viewer is presented with high quality environmental sounds in addition to their partner's voice.

Synchronous Voice

All 10 participant groups indicated that the viewer and streamer should be able to talk to each other during the telepresence experience, and the specific technologies to facilitate this interaction only slightly varied. All 5 Streamer-focused groups, as well as 3 Viewer-focused groups, selected earbuds that have a built-in microphone; these would be either wired or connected to a user's personal mobile device via Bluetooth. The earbuds were generally selected because they are lightweight, and they are extremely common — thus the Streamer does not need to purchase any additional equipment, and would not look out of place using them.

Although headphones might be a heavier accessory, one Viewer group chose them because they offer higher quality audio; this particular group also indicated that audio quality is extremely important to them, so a pair of noise-canceling headphones would allow the two partners to communicate more clearly with each other and thus have a more positive experience. Lastly, one Viewer group thought that the head-worn AR Glasses, such as Google Glass, would provide an adequate audio solution; this is because the ear would not be covered up by headphones or earbuds, and thus the user could also hear the surrounding noise.

Although our participants clearly identified the need for verbal communication between partners, there was less explicit desire to let the Viewers hear the environmental noises. In the next section we describe how some groups specifically wanted the users to hear the remote environment.

Ambient Noise

Four total groups (2 Streamer, 2 Viewer) specified a desire to let the Viewer hear sounds that occur from the remote environment. We saw three unique approaches to facilitate this feature; two groups — one of each type — elected to have the streamer use wearable microphones in the form of a shoulder-worn device in the same fashion as Wearable Speakers (see our "Bag of Stuff"). Such a microphone array could be worn under the Streamer's clothing, therefore improving the Viewer's sense of spatial presence while not socially burdening the Streamer.

One Viewer group elected to include a high-quality microphone in their design. Whereas most of the groups chose to rely on earbuds with built-in microphones, this group identified that the quality would be too grainy trying to record the environment. Their solution is a boom microphone similar to those found in film production, so that the Streamer could control which sounds are being recorded in the live-streaming experience. Lastly, one Streamer group designed a drone that would have a built-in microphone and speaker; through this design, a Viewer could have complete control over the experience and be able to feel as if they were "actually there".

Streamer Considerations

In addition to the basic video and audio streaming capabilities, seven groups (3 Streamer, 4 Viewer) expressed interest in at least one feature that should be included to assist the Streamer during their experience. These include storage space, mounting equipment, ambient noise presentation, and others that generally enhance the streaming experience.

Physical Storage

Four Viewer groups and two Streamer groups identified the need for storage space. Since the scenario targeted a location that includes a shopping complex, a place to put purchases and a way to keep the hands free was desirable. Additionally, some participants noted that the streamer might take breaks between live-streaming sessions, and would need a place to store all of the items. Lastly, participants were generally interested in ensuring the Streamer was prepared and had all the necessary equipment, so a bag to store spares or battery packs was an obvious choice. Although our "Bag of Stuff" solely included a backpack, participants expressed that the storage could be in the fashion of sling bags, waist bags, or purses — any fashionable accessory that would serve the function.

Hands-Free Mount

Two Viewer groups and one Streamer group designed their prototype using a device on which a user could mount another piece of equipment, in an effort to free up the Streamer's hands. One Viewer group chose to use a pole on which a high quality microphone could be mounted, because the user's hands would already be busy controlling the camera. Another Viewer group elected to use a small humanoid robotic avatar on which the camera would be mounted, thus providing a view that allows the two partners to have a side-by-side experience. Lastly, the lone Streamer group decided to include a shoulder strap with a camera clip, akin to a design by Rae et al. [205]. Here, the camera is mounted across the Streamer's chest, but they could remove it and manually operate it to give their partner a better view.

The other groups either did not want to include a mount, or the devices were innately body-worn (such as a multi-lens $360°$ camera on the head). The main reason for *not* including a mount device is the fact that they are not particularly prevalent in society, and thus the Streamer would look too conspicuous wearing one; this notion was found among Streamer and Viewer groups alike.

Ambient Noise Pass-through

One unanticipated feature specified by two Viewer groups (but no Streamer groups) was the ability to let the Streamer hear their surroundings with ease. As most of the groups chose to include ear-buds or headphones in their designs to facilitate voice communication with the Viewer, the ability to hear ambient noises is reduced, and thus safety becomes a concern. These two groups both specified the AirPods Pro brand, which sport a "transparency" feature. This allows the Streamer to have a high-quality audio channel with their partner, but it also affords them with a way to hear important noises — such as announcements or approaching vehicles — that could otherwise be lost. Although the majority of groups did not express a desire for this feature, we believe it is in part because they simply did not think of it. No groups expressed negative sentiment for ambient noise pass-through, so we expect that inclusion is implied to add a layer of safety to a telepresence prototype.

Shroud

Two groups (1 Streamer, 1 Viewer) identified the need for an item or technique to hide another piece of technology. In both of these cases, the participants wanted to ensure that the Streamer did not experience a feeling of self-consciousness. In the lone Streamer group, the prototype called for the use of a microphone array, to let the Viewer receive high quality audio of the environment; but, since the device is large, that group wanted to hide it underneath the Streamer's shirt. The lone Viewer group brainstormed the use of a head-worn multi-lens 360° camera, but since that sort of device is not socially acceptable, it would be hidden underneath a hat.

Though we expected the other Streamer groups to incorporate shrouds into their designs, they ultimately elected to take a minimalist approach to their prototypes, leaving off the more obtrusive devices in favor of smaller and more socially acceptable ones. We are pleased to see, however, that some Viewer groups explicitly took the Streamer's well-being into account with their designs.

Weight Alleviation

Only one Viewer group, and no Streamer groups, explicitly called for a way to reduce the physical load on the streamer. As above, the Streamer groups took a minimalistic approach to their designs, and thus their prototypes were not so heavy. However, the Viewer groups typically added more devices to their designs. This one group noted that it is unfair to load up their partner with devices without finding a way to enhance their comfort. This group was unable to reach a unanimous decision on the device they wanted to include, but they brainstormed an electronic back brace (similar to an exoskeleton) as well as a robotic arm that could hold devices for the user.

Viewer Feedback

Although every group added synchronous audio communication between the two users, participants sometimes identified that it is difficult to communicate spatial requests via voice. To help combat this problem, 8 participant groups (5 Streamer, 3 Viewer) thought of a few ways which would allow the viewer to give better context when talking to their remote partner. These included use of a physical device for pointing, visual augmentation through AR lenses, or simple text messaging.

Tangible Device

Four groups (3 Streamer, 1 Viewer) devised the use of a physical object to help the viewer communicate with their partner. Because of the difficulty expressing directions — especially when both users are looking different ways — these groups thought it would be best to provide a way to help the users reach an aligned heading. Two Streamer groups incorporated LED lights into their solution; in one design, a light positioned next to the ear-worn camera would turn green when the streamer is looking in the direction that the viewer wanted to see. In another, the lights would be fashioned into a circular bracelet, and they would light up to express a direction.

The remaining two groups each included a robotic avatar that would accompany the streamer. The single Viewer group designed a wearable humanoid torso that would embody the remote user during the interaction; by using a VR 3D spatial interface, the user would point using their hands, and the robot arms would accordingly point in the intended direction. This gives the Streamer a physical indicator of where their partner wants to go or look. In one Streamer group, a quadcopter drone would fly next to the Streamer and serve as the Viewer's avatar. Here, the Streamer could understand where their partner was looking by simply observing the direction it was facing.

Visual Augmentations

Instead of using a physical device that might draw attention to the Streamer, three groups (2 Streamer, 1 Viewer) decided to use AR Glasses to provide enhanced communication. These participants indicated that AR glasses are preferred because they are fashionable, so the Streamer would not look out of place while using them. The interesting feature of visual augmentation is that it is nearly boundless, in that almost any graphical element can be displayed. One group suggested that a small window of live video should appear within the lens, so that the Streamer can either see the face of their partner, or see a video thumbnail of where their partner looks within the 360° video. Another suggested that the AR glasses display the face of their partner. The final group thought that displaying additional information about the surrounding environment would benefit both users; for instance, a popup graphic could explicitly tell the user what a certain building was, and it would also relay this information to their partner, so that both users could have the same understanding of the environment.

Text Messaging

Lastly, two groups (1 Streamer, 1 Viewer) chose to include text messaging between partners. The lone Viewer-focused group chose to facilitate this feature through a smartphone, as that is already a well-established mode of communication, but the Streamer-focused group chose to display text on an AR Glasses device. Here, the text messages are more complimentary to the existing modes of communication, and displaying the text directly in front of the streamer removes the need to hold a smartphone in their hands. In this way, the information is received more quickly.

Viewer Considerations

Although our activities asked our participants to only design a prototype that the streamer would use, seven groups (5 Viewer, 2 Streamer) specifically designated some features that should be afforded to the Viewer. These included the types of video display they would use, as well as remote control over some device in the streamer's setup.

Virtual Reality Platform

Six groups (5 Viewer, 1 Streamer) were specifically interested in providing the Viewer with a VR platform. Naturally, these groups were also the ones that chose to include a 360° camera in their setup. By using a VR headset with head tracking, the viewer could look around the environment with natural camera control. One group also included use of VR controllers that would map to a robot avatar's arm motion. Although it is possible to display normal video within a VR display, the benefits of such a device are lost, so it does not make sense to include it outside of 360° cameras. Thus, some groups also expressed direct interest in a 2D display.

2D Display

Two Viewer groups, but no Streamer groups, explained that having a non-VR display would be easier to use than a VR display; this is because it takes a physical toll to actually control the camera feed with head rotations. These groups concluded that the viewer should have access to both VR and non-VR displays, and therefore have the option to choose the viewing device as they saw fit. As VR displays might also induce simulator sickness, including a 2D device in the loop can help alleviate this issue. A 360° image can be transformed into an equirectangular, pannable image, so it will still look natural on a flat screen. While none of our streamer-focused groups explicitly designated a viewing device that would be used by the viewer, researchers asked them what they would anticipate being used. Responses included personal devices such as smartphones, tablets, or laptop screens — whatever device the viewer wanted to use.

Remote Control

Three groups (2 Streamer, 1 Viewer) specified remote control capability, and two of these, as above, were for controlling robotic avatars. These two groups incorporated VR controllers which would either map to humanoid robot arms, or move a quadcopter drone within a bounded area. One other Streamer group brainstormed the idea of allowing the viewer to remotely manipulate the camera orientation. Here, the camera was an ear-mounted device that would be tied to the streamer's head movements, but it would incorporate a mechanical hinge. Thus, the viewer could angle the camera's pitch and yaw.

Physical Embodiment

Two groups — one of each type — chose to incorporate a physical device that could be controlled by the viewer at home. The lone Streamer group decided to include a quadcopter drone into the prototype setup, so that the viewer had complete control over the camera, as well as an entity that declared to other people that there was a remote person engaging in a telepresence interaction. Since the viewer would be in complete control of the drone, the physical toll on the streamer is significantly reduced; however, our participants also thought it would be prudent to logically restrict the drone's movements through a smartphone application, such that it must always be within range of the streamer, and follow rules set before it.

The single Viewer group interested in a physical embodiment chose to incorporate a small humanoid robot that had functional arms; this would be an avatar of sorts, but would sit on the shoulder of the streamer. Here, the robot would wear a 360° camera instead of the actual streamer, so that the viewpoint would be coupled with the avatar's head; then, the viewer at home would be able to control the robotic arms through their own natural arm movements, via VR controllers and inverse kinematics. In this way, the remote user would receive a first-person view and would be able to physically point out objects of interest as if they were actually there.

Other Considerations

Lastly, there were a few other features identified by 5 participant groups (3 Viewer, 2 Streamer) that need to be considered. These include battery charging, stable network connections, and a single case of drone tethering. Thematically, these features either highlight the importance of preparedness, or are included to ensure the experience runs smoothly.

Battery Charging

Four groups (3 Viewer, 1 Streamer) expressed that the streaming setup should include a way to charge or replace batteries for the various electronic devices within the prototype. Our participants tended to include backpacks within their designs, and specified that replacement batteries would be stored in it. The single Streamer group, however, thought it would be more appropriate to bring a portable power bank. By running extension cables through the shoulder strap of a bag, and by leveraging device mounts, the streamer could easily plug a device into the power bank without interrupting the live-streaming experience.

Network Charging

Three groups (2 Viewer, 1 Streamer) identified the need for stable network connection, so that the two users could enjoy a positive experience together without latency or hiccups in communication. Although this feature is implied for live-streaming experiences, these participants explicitly wanted to ensure that network communication was sufficient. To accommodate this feature, users noted that smartphones could act as a tethering device, or a high-quality air card could be used.

Drone Tethering

The final feature, identified by the lone Streamer group which elected to use a quadcopter drone, is a smartphone application that the streamer would use to set parameters that the drone would follow. These include maximum height that the drone could fly, and an automated feature that would ensure the drone could not fly too far from the streamer; as the participants noted, the streamer is the one on-site, so they would still be responsible for their remote friend's behavior.

Findings: Negotiation of Features

Having investigated our individual participants' values and the sub-groups' desired features within our telepresence context, we were able to identify where the two major stakeholders' opinions diverge. It is important to find compromise between these groups to ensure that both are satisfied with a given telepresence experience. Thus, using the same codebook as the initial Feature Analysis (Table 6.4), we conducted a similar analysis of the final group designs from Activity 4. In this section, we discuss how our participant sub-groups found common ground in order to construct a telepresence prototype that all members would be happy to use. As such, the unit of analysis for this section is Group (N = 5).

Live Streaming Video: Maximizing the View, Minimizing the Awkwardness

In Activity 2, the Viewer groups primarily wanted to use a 360° camera that could be worn on the head of the user, whereas the Streamer groups wanted to use a normal camera, to reduce a feeling of social awkwardness. Naturally, then, all 5 of our groups included live video streaming as a feature in Activity 4; but since one of the major divergences was the **type** of camera to be used, our groups had to find common ground to balance viewing experience and streamer's self-image. Our groups devised a few plans to reach this agreement.

One group decided to forego the panoramic camera and simply use a smartphone device, but include a couple of clip locations to support the camera, such that the streamer could adjust the view as the environment or scenarios shift. This group suggested using a shoulder strap with a clip, a hat with an action camera mount on the brim, and simply maneuvering the camera with their hands as their desired camera locations. Their main concern was the streamer's self-image, and they felt that this solution was socially acceptable.

Another singular group ultimately decided to implement a 360° camera worn on the head. This group included the streamer sub-group that designed the drone from Activity 2, so these participants were generally more interested in the capabilities than they were with the streamers self-image. However, this group did decide to take a minimalist approach, only incorporating the technology that was absolutely necessary to a telepresence experience; in their eyes, the 360° camera was the bare necessity, so they did not include much after that.

The remaining 3 groups also included a panoramic camera, but the participants were also very concerned with the streamer's self-image. Thus, they devised solutions to keep this value a priority while ensuring that the viewer also received a panoramic view. Two of these groups chose a head-worn 360° camera, but thought of ways to hide it from the public eye — to do so, they thought it would be prudent to shroud it with a different accessory. For instance, one group thought the camera lenses could be embedded inside a Mickey Mouse Ears hat, which would be socially acceptable (as our design prompt was for a Disney experience). The other group brainstormed the possibility of embedding the lenses inside a necklace, but ultimately decided against that idea in favor of a fashionable hat or sweatband, which, in addition to being socially acceptable, would also help the user if they would be outside in the heat. The final group that chose a panoramic camera was also concerned that the streamer's setup should be socially acceptable, but it should also allow the streamer to choose the device that they would be comfortable using. As such, this group wanted to accommodate individual preferences. Here, the participants described that an ideal system would make use of a smartphone camera in instances where the two users wanted to have a face-to-face interaction, akin to video chat; meanwhile, the smartphone's camera would also stream the remote environment to the viewer. However, when the environment becomes full of visual stimuli, the streamer should be able to swap the smartphone for a 360° camera (or, perhaps, a 360° attachment could be plugged into the phone).

Even though the streamer and viewer groups originally differed regarding the types of video capture device, they were able to settle the differences and negotiate a balance between their values and desired live-streaming video capture features. Next, we analyze the selected features for transmitting and receiving audio communication.

Audio Communication: Maintaining a Direct Line

Individually, all sub-groups noted that an ideal telepresence setup would include a synchronous voice feature, and this did not change for the final activity. All groups indicated that earbuds would be sufficient to support a conversation, discrete enough to prevent the user from standing out, and common enough such that potential users should already have a pair of them. One group did suggest that headphones might offer a higher quality sound, but this group did say that the audio playback device should be chosen based on individual preferences, as headphones are also very bulky and would make a user stand out.

Two groups suggested that the telepresence setup should also be able to capture ambient noises, such that the viewer can have a more immersive experience of the remote environment. Both of these groups indicated that the streamer's setup should include a high-quality microphone that could be held in their hand. One of these groups wanted to use a hand-held boom microphone, and the other wanted to include a directional microphone that could plugin to a handheld tablet (such as an iPad).

Naturally, all of our groups wanted to include a feature to let the two users communicate audibly with each other, but our participants tended to identify that verbal communication is not enough to convey certain messages. Since non-verbal communication is lost over voice chat, alternative modes of communication are valuable. In the next section, we discuss how our groups designed ways for the viewer to provide more feedback to the streamer.

Viewer Feedback: Augmenting Interpersonal Communication

Based on prior experiences, some of our participants were able to immediately identify that voice chat is simply insufficient to convey important messages related to the environment; a live-streaming setup does not innately support a method for the viewer to use non-verbal communication that they could use if they were physically collocated with their remote partner. In Activity 2, sub-groups were able to brainstorm a variety of techniques that could allow the viewer to convey non-verbal cues, including tangible devices, visual augmentation, and text chat. In the final activity, though, the majority of our groups converged on one main method — visual augmentation.

A single group decided that the streamer should have a mobile device (smartphone or tablet) to hold in their hands. The mobile device would have special software that pairs with the viewer's display, such that when they want to direct their partner to an object of interest, a still-frame would appear on the mobile device. This would give the streamer a visual cue of where to go or what to inspect. This group also suggested that if the environment became too loud, then the mobile device could also be used for text chat, so no information would be lost.

The remaining 4 groups decided to include a pair of AR glasses (e.g. Google Glass) in their final prototypes. In keeping with the value of maintaining a positive streamer self-image, no group chose a larger display such as the Hololens or Magic Leap; they were comfortable sacrificing field of view in order to provide the streamer with a comfortable experience. Our participants also suggested that AR glasses could look exactly like a pair of normal eyeglasses, so that it would be impossible for a bystander to tell that the streamer is using an atypical piece of equipment.

Although these groups selected the AR glasses for their prototype, they brainstormed different applications for the device. All 4 groups identified that this device could be used to show a live video feed of their partner's face, granting the user with a way to see their partner and identify facial expressions. Further, the groups that selected a 360° camera thought that AR glasses could also be used to indicate where the viewer was looking, using directional arrows. From Activity 2, the group that brainstormed a map translated that feature to the final prototype, so that the viewer could relay navigational information to the streamer; and, one other group thought it would be critical to display battery information of the various devices in a heads-up display for the streamer to identify when devices needed attention.

Our participants, having originally brainstormed multiple devices to convey non-verbal communication, ultimately converged on a low-profile, high-utility solution in the form of AR glasses. We note that none of the groups desired bulkier devices such as the Hololens or Magic Leap, instead selecting the natural eyeglass form factor. In the future, we might see even more alternative devices to support social acceptance, such as AR contact lenses. In the next section, we discuss the selected features to support the Streamer's experience.

Streamer Considerations: Preparing for Every Scenario

In Activity 2, our Streamer participants generally wanted to include only the devices and materials that were necessary to facilitate the telepresence interaction; though our Viewer groups felt the same, they also recognized that our design scenario involves fluid situations, and that it is necessary for the on-site streamer to have the ability to react accordingly. In part, this was a reason why our Viewer groups selected more affordances for the Streamer in their designs.

Through Activity 4, we find that our groups ultimately wanted to provide additional affordances, but in a way that would not draw attention to the user. Four of our final groups selected physical storage for the streamer to wear, and although our only "Bag of Stuff" storage item was a backpack, *none* of our groups wanted to use a full backpack — they indicated that it is more appropriate to use waist bags, purses, or even a cargo vest with pockets — whatever is large enough to hold the items the streamer needs to store without being overly obtrusive.

We also note during this final activity that only 1 group wanted to use a mounting device to free up the streamer's hands. This particular group was the only one that did not want to use a panoramic camera, instead opting to leverage a smartphone as a video capture device, and thus decided that including mounts in various places (on a shoulder strap, on the brim of the hat) is necessary. The remaining groups chose a setup that did not warrant a mounting device, mainly because their devices were worn on the body in a natural way (e.g. on the head like a hat).

All groups designed a constant audio communication channel for the Streamer and their partner, and all groups decided that earbuds would be sufficient to provide this interaction; but, one group also thought the streamer should have the option to use noise-cancelling headphones. This was the only group, then, that agreed to include an audio pass-through feature in their audio playback device, so that the streamer could still hear their surroundings. Streamer safety was the main concern, as they need to be aware of announcements and potential hazards.

The final streamer consideration that appeared in the designs of 2 groups was a shroud for the camera. In these groups, a head-worn multi-lens 360° camera was selected, and since this style of wearable computing is not socially acceptable in its own right, the participants decided to embed the device within normal clothing; one group chose Mickey Mouse ears (which is normal for the design scenario), and the other thought the camera lenses could be embedded into a sweatband. In this way, the functionality of the panoramic cameras persists, but the social acceptability increases dramatically. Having identified features to generally support the Streamer, we next turn to those that can support the Viewer.

Viewer Considerations: Relinquished Control, Maximized View

In Activity 2, some of our participants designed features that would provide the viewer with a high sense of spatial presence (i.e. the feeling of "being there" [212]). These features manifested as physical devices that they would have complete control over, such as robotic platforms, or in one case, an actuated camera hinge that the viewer could maneuver.

In Activity 4, we note that none of these features made it to the final designs, partly due to the negative impact they would have on a streamer's feeling of social acceptability. Thus, none of the final designs include a physical embodiment, and none of them include a remote control option. The streamer, therefore, would have complete control over the interaction scenario, and as they would be the only individual of the two who is at social risk, this is understandable.

However, this does not mean the viewer is unable to interact with their partner; as noted above, all 5 groups included a means for the viewer to provide non-verbal feedback to the streamer, and most of the groups chose visual augmentation. Therefore, although our groups wanted to ensure that the streamer maintained complete control over the interaction, the viewer *still* has the ability to provide considerable input, and is not just an idle watcher.

Regarding the actual viewing device that should be included in a setup, 4 groups (the ones that included panoramic cameras) thought that the person at home should have access to a virtual reality platform, so they could have an immersive viewing experience. However, 2 of these also suggested that the viewer should have the ability to choose their own device, even offering the ability to swap between an immersive device and a flat screen at will. In this way, our participants noted that a viewer's individual preferences should be considered; as long as they are able to see and hear the environment, talk to their partner, and offer valuable non-verbal feedback, then the viewer should be satisfied. We now shift to the final subset of features, which generally ensure that the interaction runs smoothly.

Additional Considerations: Infrastructure Assumed to Work Natively

The last major dimension of our feature codebook is that for "additional considerations", which includes battery charging capabilities, stable network connection, and flight parameters for a drone. In Activity 4, none of the groups included a drone, so the flight parameters are not necessary. Further, none of the groups stated that the streamer should carry a wireless aircard - network connectivity is assumed. The only remaining feature in this category is battery charging or replacements, which 4 groups specified is one minimum requirement for a telepresence setup.

All of these groups specified inclusion of a portable power bank, to which the various devices could connect and recharge. Though these power banks were thought to be contained in a personal storage item or pocket, they could be taken out and connected to the devices that need them - but one group thought that it would be prudent to incorporate an extension cable which could be fed through the user's shirt, so that charging could occur without having to expose the devices to public eye. This technique would allow the streamer to charge various devices even during use. In this way, there would be no down-time of the interaction.

Discussion

The results of our study provide new insight on designing a more balanced interpersonal telepresence system that incorporates both advances in previous literature and a user-centric approach. The following section provides details on the implications of our findings.

Streamer and Viewer Values Align More than Diverge (RQ1)

One of the most unexpected findings through our work was that both stakeholder groups made significant consideration for the other, despite being given a design prompt that primed them to be greedy for their own role. We found that the participants' set of values (in the context of interpersonal telepresence) were mostly aligned, with little divergence. As such, although previous work has shown how interpersonal telepresence developers tend to design a system to benefit the Viewer [198], end users of such systems actually desire a prototype that supports the needs of both major stakeholders. After all, if the Viewers are receiving an experience from a loved one, it is natural for them to make sure that their partner's social well-being is guaranteed. Even in cases where the two users are strangers, we need to make explicit consideration for the Streamers as we ask them to use a system that ultimately benefits a Viewer.

The two most emergent themes within our codebook pertained to ensuring the Streamer had a physically comfortable rig, and ensuring that their experience using it was socially acceptable to others. Previous literature in this space does seem to support the *physical* comfort of Streamers by introducing various wearable devices (e.g. camera mounts and harnesses [87, 153, 205]), but these wearables are ultimately too bulky or too noticeable. As noted in previous field studies, if the designed wearable devices are not socially acceptable, then they need to be hidden, in some way, from the public eye [115]. Participants from both primary stakeholder roles in our study confirm that this is a high priority, and future designs should heed this warning that conspicuous devices may be met with negative feedback.

Although the stakeholder values generally overlapped, we did find two major themes of which our participants had differing opinions. First, the Viewer participants expressed a desire to be able to interact with the remote environment, or to at least have a sense of “being there”; a system that supports this value would allow the Viewer to have autonomy and would let them offload some of the work the Streamer would otherwise need to perform for them. Unlike with mobile robotic telepresence, where the Viewer has an actuated platform that they can directly manipulate [125], this interpersonal style of telepresence finds the Viewer relying on the Streamer to navigate an environment and interact with it for them. Some previous literature posited the use of robotic avatars that the Streamer could wear, ranging from devices that declared a Viewer was watching (e.g. TEROOS [106]) to one that allowed the Viewer to provide physical non-verbal communication (e.g., the 20DOF Humanoid [253]) to one that supports direct manipulation of the environment through actuated robot arms (e.g., Fusion [221]). We find, however, that these wearable avatars are not yet socially acceptable, and thus are not ready for inclusion. More appropriate features to support a Viewer’s sense of embodiment might include graphic overlays, such that the two users could see each other’s faces. Using AR/VR technology, the Viewer’s face (or more) could be displayed within the view of the Streamer’s gaze, in 3D space [91, 92]. The benefit of utilizing AR/VR in this regard is that this aspect of the interaction would not be noticeable by other people, and thus the Streamer's sense of social well-being could remain in-tact.

The other major value with Streamer/Viewer conflict was the type of shared experience that a given telepresence system should convey. Whereas the Streamer participants wanted their partner to see *exactly* what they wanted to show — and thus have the experience be more reflective of a first-person experience through their eyes — the Viewers wanted to have the ability to explore the remote environment at will, and related to the above, they also wanted to feel like they were "actually there" with their friend. As we have previously demonstrated in Chapter 4, simple design choices such as the location of a wearable camera have a profound impact on the sense of story [195], and in a telepresence context, this can be the difference between a first-person view and side-by-side experience. For instance, Manabe et al. concluded that an over-the-shoulder 360° camera placement could let the Viewer feel like they were actually with the Streamer [149], and Misawa used a camera attached to the face of the Streamer, in an effort to let the Viewer feel as if they *were* the Streamer [162]. Naturally, however, these prototypes might not support a positive Streamer self-image.

Ultimately, both sets of participants had relatively similar values. They felt empathetic to their counterparts in the other role, and thus wanted to ensure that both Viewers and Streamers would have a positive shared experience. However, although the participants had aligned values, there were significant deviations regarding how the values were embedded in interpersonal telepresence features. In the next section we discuss these divergences.

Design Strategies to Foster Stakeholder Values (RQ2)

Our 2 stakeholder groups held similar values, but we note that the final design strategies to support these values, in some places, diverged significantly. In this section, we discuss the similarities and differences of these design strategies.

Communication is Key: Supporting Voice Chat and More, in a Natural Way

Supporting a conversation between the Streamer and Viewer was a feature that every group in our study wanted, and nine of ten sub-groups selected earbuds that would connect to the user's personal smartphone; one group chose noise-cancelling headphones instead. Using ubiquitous devices, voice chat is easily supported in any interpersonal telepresence prototype. However, as noted by previous work, other forms of natural communication are often lost when the two users are geographically distributed [115, 243]. Most of our groups were able to also identify this detriment and thus designed features to help solve the problem — but *how* to solve the problem was where our groups differed.

Few of our Streamer groups wanted to use advanced technology in their prototype, instead opting to use minimalist directional indicators such as LED lights, which would give visual feedback that a streamer could use reference to understand where their partner was looking or wanted to look. This differed from our Viewer participants, who wanted to include new and exciting pieces of equipment such as AR Glasses and robotic arms. This is a reflection of the findings found in our literature review (Chapter 3) in that the "bells and whistles" are often developed, but they are not necessarily appreciated by the end users [198]. One design consideration emerging from this discrepancy is to use familiar devices *when possible*, but if a novel device is developed to support new affordances, then its form factor should be minimal in nature.

Making Streaming Life Easier by Reducing the Workload

Most of our participants wanted to reduce the Streamer's physical burden when using an interpersonal telepresence prototype, but there was some divergence in the features to manage this value as well. More accurately, the Viewer participants wanted to use devices that would support the user throughout the interaction, whereas the Streamer participants decided to simply reduce the number of items. It seems that these participants view the *absence* of features as the only true way to reduce the physical toll — after all, helper items such as backpacks *implicitly require effort to use*.

In the framework developed by Rae et al., **Initiation** was one of the major dimensions, comprised of themes such as “amount of planning” and “setup costs” [205]. Our participants identified that reducing the workload for the Streamer would not just entail low setup costs, but also low *tear-down* costs — for instance, if the Streamer needs to take a bathroom break, the setup needs to be quickly collapsible. We view physical storage as a necessity for interpersonal telepresence for this reason — but following our participants’ thoughts, we expect that the size of a bag or other container should scale with the size of the live-stream equipment.

A Struggle to Find the “Where” for Wearable Cameras

Although not all participants in our study expressed the Togetherness value, the ones that did clearly diverged depending on their group; and, their desired togetherness trait also seems to predict other values. The Viewers expressed the desire to have the side-by-side feeling, and although these individuals barely expressed the value of Natural Sight, they all wanted to have a form of autonomy. In essence, these individuals are asking to feel as if they were “actually there,” by being able to see their partner and look around the remote environment on their own accord. In contrast, the Streamers expressed the desire to have a shared experience with their partner, where both users could see and hear the exact same thing. These individuals also expressed the desire to give their partner a Natural View, but their designs ultimately reflected a first-person camera feed from *their* eyes.

This begs the question: what exactly is a *Natural View*? To our Streamer participants, it means a view point that could represent their eye-sight — this is indeed natural to the Streamer, but it is not necessarily natural to the Viewer. As we identified in Chapter 4, camera placement influences the perceived sense of "story" [195], and as we identified in Chapter 5, men and women differ in terms of preferred camera height [196]. So, placing the camera around the Streamer's eyes may actually be a detriment — it could be too tall or too short; and, placing the camera around the eyes also prevents the Viewer from seeing their partner's face. Previous work has identified that different wearable camera placements have an effect on the wearer's self-image [279], and our participants indeed want the camera to be worn in an inconspicuous manner; we thus suggest inclusion of multiple hidden cameras spread across the Streamer's body. One could be around the user's shoulder, allowing the Viewer to see their partner's face, and another could be placed around the Streamer's eyes or on their head. In this manner, the Viewer could swap between cameras in order to see what they want *and* to see what their partner wants.

Designing a Balanced Prototype Based on Stakeholder Agreements (RQ3)

Using the participants' values and the final design activity, we were able to draft an interpersonal telepresence prototype that we expect would satisfy both major stakeholders; see Figure 6.7. In this illustration, we used images that were not found in our "Bag of Stuff", yet were reflective of participant feedback. The top features that were agreed upon include synchronous voice chat, panoramic video, visual augmentation (AR), battery charging accessories, and physical storage.

The first major feature, which all groups desired across all stages of the study, was synchronous voice. Here, to foster this feature, we include the wireless ear buds that most of our groups selected. The Apple AirPods Pro include a microphone in each earbud, and a "Transparency" feature that allows the Streamer to hear their environment if they desire. These earbuds can by paired with the user's smartphone, and in this way we are including the current ubiquitous nature of phone calls.

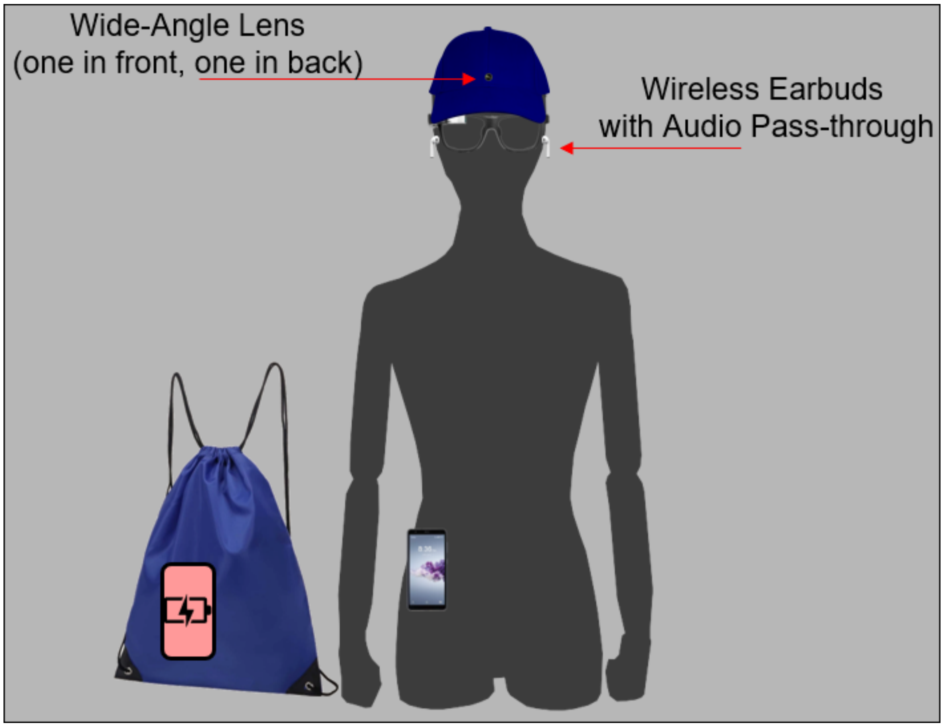


Figure 6.7: Final prototype design based on the top compromised features from our study; note that colors and patterns are arbitrary, and can be changed to match Streamer preferences.

The next major feature, as agreed upon by most of our groups, was panoramic video. Here, to support the Streamer's self-image, we choose to embed tiny, wide field-of-view cameras inside a hat; one lens will be on the front of the hat, and the other will be on the back. Some hats have certain patterns, logos, or words that would help to hide the lenses, making them look like they are part of the hat. In this picture, we use a hat to help illustrate that a small lens can be embedded inside it.

The third feature agreed upon by our groups was visual augmentation, such that the Viewer can communicate through non-verbal means. The beauty of AR Glasses in this context is that additional software features can be incorporated to the system without changing the outward appearance of the prototype. The software capabilities that our participants brainstormed can all be included in this way; these include a map feature, such that the Viewer can help navigate the Streamer; a face-to-face video chat window; directional indicators so the Streamer knows where the Viewer is looking; and virtual annotations, such that the Viewer can circle, point at, or otherwise highlight interesting pieces of the environment.

The fourth feature that was agreed upon by most groups was a portable battery charger. In our design, we place a power bank inside the user's bag, but we note the possibility of placing one inside the Streamer's pocket, and feeding a wire through their clothing. This practice could be too cumbersome, so ultimately the end-user should be able to choose how to use the device; but, we find that the power bank is a must-have, instead of a wall charger. The portable power bank will allow the users to recharge the device in use, instead of having to remove the device and plug it into a stationary outlet.

Lastly, we add a draw-string bag to our final prototype. The majority of our groups wanted physical storage of some sort, but none of them wanted to use an entire backpack — the bag should be only as big as it needs to be. Inside the bag, we plan on placing the portable power bank, but it also needs to be large enough to fit the rest of the devices (the AR Glasses and the 360° camera hat).

Through this design, we are able to support the needs of both major stakeholders. We are able to provide a viewpoint which enables the Viewer to explore the remote environment at will, while allowing the Streamer to feel socially comfortable; we took an approach that embeds technology into pieces of clothing, without sacrificing capability; and we provide the Viewer with ways to interact with the Streamer beyond simply watching a video feed and talking with their partner. Through these means, we can create a safe and comfortable telepresence experience that truly fosters the sense of togetherness that people are craving.

Applying Value-Sensitive Design in a Novel Domain

We make a methodological contribution to novel interaction research by leveraging co-design to draft a hardware-based system using a "Bag of Stuff" with "paper dolls" approach [84, 117], which is typically performed more within social computing/HCI research and with software-based systems. To our knowledge, ours is the first study in the interpersonal telepresence domain that uses co-design to let stakeholders identify and select desirable features. Further, we utilized the VSD methodology to help us identify the human values that should be embedded in a telepresence prototype, which (to our knowledge) has not yet been applied to this domain. This is particularly useful because of the vast range of interpersonal telepresence prototypes that are being created without a complete user evaluation. Given the two distinct stakeholders with values that naturally conflict and needed to be reconciled, co-design and VSD were well-suited for our study. We were able to identify compromises between these stakeholders without making it a zero-sum game.

Limitations and Future Work

We recognize that our study is not without its limitations and discuss future work that could help to reconcile possible shortcomings. First, though our procedure allowed us to collect data from a diverse set of people and foster thought-provoking discussions, the final designs are representative of the group, and not the individual. As such, we were unable to capture dissent within a given group, and more importantly any corrections to the design that those dissenting individuals would make. Naturally, to receive real feedback for improvement, we conducted a follow-up study in which we deploy a prototype based on the results of this study (see Chapter 7).

We also note that our results do not generalize to all interaction scenarios. In our study, we specifically asked our participants to create their designs with an outdoor theme park location in mind.

We found many participant comments to revolve around this criterion, and some of the features (such as Mickey Mouse ears or sweatbands) will not translate outside of this context. While we can envision other wearables to replace these context-sensitive features, we cannot speak to the social acceptability of the future devices.

Next, though previous work has shown how the Streamers of interpersonal telepresence are under-considered, our study was less sensitive to the Viewers — we did not ask participants to design a Viewer prototype to compliment the Streamer's system. While we believe our work is necessary to swing the pendulum back towards the center, we must remind ourselves that our goal is to identify ways to support *both* major stakeholders.

Lastly, we note that our design recommendations involve use of a live-streaming camera while hiding the device from public eye. We must ask ourselves if this solution, though practical, is *ethical*. Our work falls short of identifying the needs of a third major stakeholder for any public interaction — the third-party bystanders. Previous work has shown that even simple technology probes with private voice chat can be confusing or even offensive to those collocated with the Streamer [202, 205], and, to our knowledge, no work has been performed to directly identify how to balance their expectations within this context. Future work should directly explore third-party members' direct expectations for the growth of this interaction paradigm.

Conclusion

In this chapter, we found with our co-design sessions that people want to make considerations for both major stakeholders, even though they were instructed to take a greedy approach for their own, single role. The empirical findings here compliment the results of our literature review (Chapter 3), in that the Streamer's technology should be as unobtrusive as possible; and, harmonizing with our previous chapters, the participants also valued Viewer autonomy, such that they could look around on their own. However, this reiterates natural value tensions caused by a desire for high capability with reduced form factor.

Thus, in the next chapter, we describe work in which we built an unobtrusive interpersonal telepresence prototype that meets the needs of both stakeholders. We worked to embed as many human values as possible within our design, as identified through this chapter's co-design sessions, and, as the VSD framework suggests, provide iterative investigation into this problem space.

CHAPTER 7: DEPLOYING AN UNOBTRUSIVE TELEPRESENCE PROTOTYPE BASED ON CO-DESIGN REQUIREMENTS

Introduction

Having identified the human values that should be embedded into an interpersonal telepresence system, and having considered the features that can support these values, we aimed at developing a prototype that would satisfy these conditions. In this chapter, we describe a technological investigation in which we discuss the steps taken to create such a prototype, and describe a study in which participant pairs completed an artwork scavenger hunt with our system and a control (see Figure 7.1). Before conducting this study, we recalled lessons learned in one of our pilot studies; see Appendix A for our first study in this area.

This chapter is divided into multiple sections. First, we describe the networking, software, and

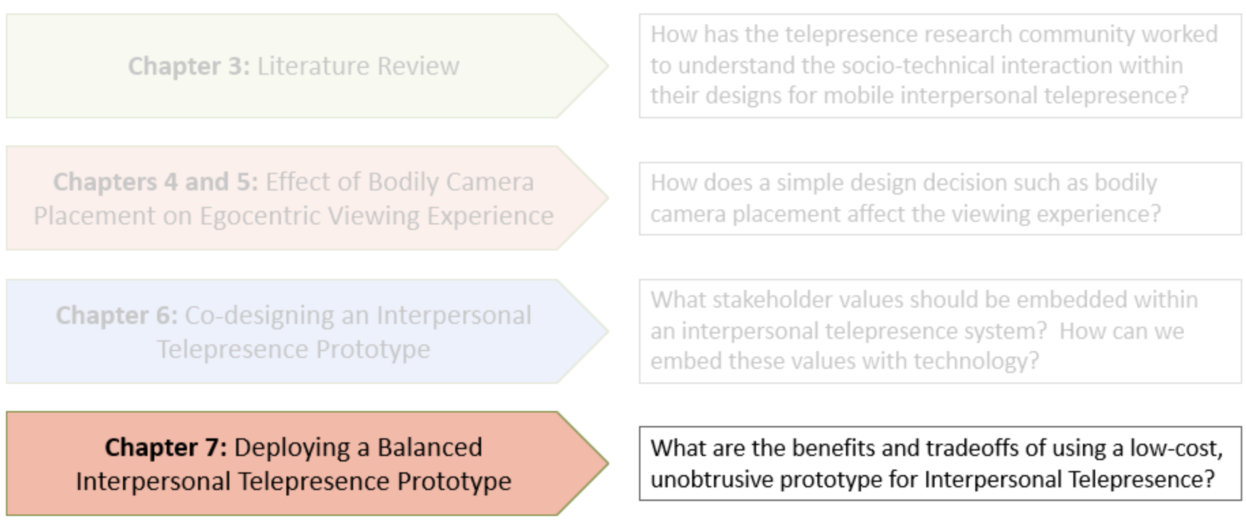


Figure 7.1: In this chapter, we perform a technological investigation in which we deploy an interpersonal telepresence prototype that embeds the human values identified in our co-design study.

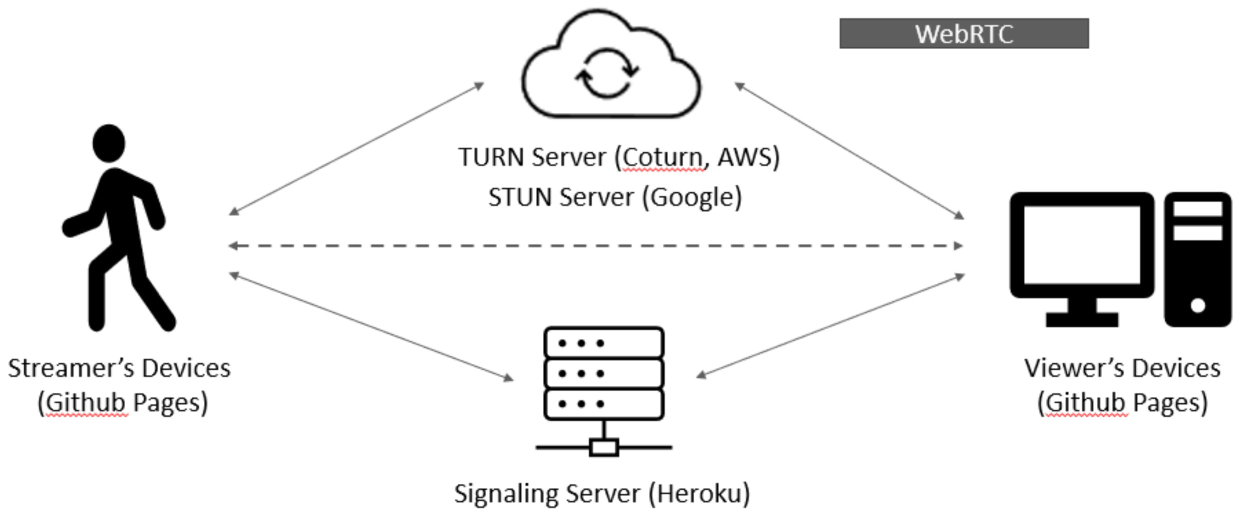


Figure 7.2: Using WebRTC, we developed a custom peer-to-peer interpersonal telepresence system that relies on multiple servers, including STUN, TURN, and signaling servers.

hardware infrastructure that comprised our new prototype, followed by a description of the methods we used to evaluate this prototype against a control. Next, we describe the results of our user study, and lastly discuss the significance of our findings.

An Unobtrusive Interpersonal Telepresence Prototype

In this section, we describe our design and development of an new prototype to support interpersonal telepresence. We discuss the various hardware and software requirements, our design iterations, and final implementations. See Figure 7.2 for illustration of the system architecture.

Networking & Web Hosting Architecture

First we discuss the networking and web hosting architecture to support our telepresence platform. We aimed to build an application that could be accessed by any user with an internet connection, and, per our identified requirements from Chapter 6, we aimed to be able to stream multiple cameras at the same time. We elected to use WebRTC[1] to construct our application. WebRTC is an attractive solution because with only a simple understanding of the framework, a developer can rapidly create a custom peer-to-peer application using only high-level functions.

However, there are some networking and server requirements that must be resolved to get a WebRTC application to work. Our application relies on three servers and a secure web hosting platform that serves our client-side application. It is beyond this dissertation to detail the protocols behind the various servers, but, as they are essential to the application, we briefly describe the purpose for each.

The STUN Server

The purpose of a Session Traversal Utilities for NAT (STUN) server is to tell a device what its public IP address is; it is useful when a computer is behind a firewall and does not know what ports can be used to send and receive UDP packets, which are used for audio and video streams. For our application, we simply harness public STUN servers hosted by Google. There exists (at the time of writing) a list of public STUN servers that can be used for any WebRTC application (see footnote[2]). Some listed servers may not be trustworthy, but we used the publicly available STUN servers provided by Google to support our application; we did not need to setup our own.

[1] http://webrtc.org

[2] https://gist.github.com/mondain/b0ec1cf5f60ae726202e

The TURN Server

STUN servers alone do not work for every possible scenario, however. If every device is on the same network, perhaps that would be sufficient; but rarely will that be the case. For our planned study, we expected to use cellular data via a 5G hotspot, and a STUN server would be insufficient. This is where the Traversal Using Relay NAT (TURN) server is used, because it helps clients identify network paths that will let them send and receive audio and video streams. At the time of writing, we have not been able to identify a list of public TURN servers; we deployed our own, by leveraging existing open-sourced software.

We requisitioned an Amazon Web Services (AWS) Elastic Compute Cloud server[3] and deployed a Coturn[4] server on our instance. We followed instructions described by Kosta Malsev (see reference [148]) and we were able to get our TURN server online in just a couple of hours. Our AWS account was assessed approximately $9 USD per month to keep this server online.

The Signaling Server

The final server required for our WebRTC application to function is the signaling server. This helped us manage users logging into our web application, and it handled transmission of various messages between users. WebRTC has increased in popularity over the past few years, and there are some tutorials that help setup a basic application. We followed one such tutorial[5] to help us get started with a Node.js server, but, we needed to change the code according to our needs. We modified the code to handle various message types, and deployed the server as a free Heroku[6]

---

[3] https://aws.amazon.com/ec2/

[4] https://github.com/coturn/coturn

[5] https://www.tutorialspoint.com/webrtc/webrtc_quick_guide.htm

[6] https://www.heroku.com

application.

After deploying and testing the Heroku application, we noticed over time that the server would stop working. This was due to Heroku automatically putting the server into a “sleep” state; though it would work again when a user requested the application, it would not remember which users were logged in. This meant that, if there was a lapse of interactivity between users for some period of time, the functionality would stop working altogether. To prevent the server from automatically “sleeping”, we registered our application with Kaffeine[7]. This free tool automatically pings an application to keep it online.

The Secure Hosting Platform

With the various servers in place, we simply needed a platform that could host our client application, including html, css, and javascript. There are many hosting sites available for use, but we elected to use GitHub Pages[8]. In addition to providing a simple and accessible platform for us to host our web content via the *https* protocol, which is required for WebRTC to function, it ties directly to the project repository. We provide a link to our personal repository so that our application can be used or modified for future projects. The root GitHub repository is found at h2bo.github.io, but the directory for this project is found at h2bo.github.io/PeerConn/.

With the networking requirements solved and the various servers in place, we just needed to develop our front-end application and Streamer hardware. In the next section, we first describe the front-end software that was deployed; we follow this with a description of the hardware implementation.

[7] https://kaffeine.herokuapp.com

[8] https://pages.github.com

Front-End Software Architecture

In this section, we describe in detail the main front-end pieces of the application. We wanted our app to be usable by anyone, in any location, while allowing us to record the interaction between the two users of the prototype. We thus implemented 4 main pieces: 1.) a Streamer module; 2.) a Viewer module; 3.) a Researcher module; and 4.) a Controller module.

Streamer Module

The Streamer module handled the identification and collection of the devices that would be streamed by the user; it also specified how to handle feedback from the other modules. The first main piece of this module is the collection and organization of audio and video that would be transmitted to the other users. Since we wanted to handle streaming multiple cameras — one for each major direction — we needed a way to let the user tell the system which camera belonged to which direction. The first step in using this module was to turn on the cameras using WebRTC functionality, and then select with drop-down menus the direction (front, back, left, right) that each camera faced. We added a logical check to ensure that none of the directions were repeated. See Figure 7.3 for

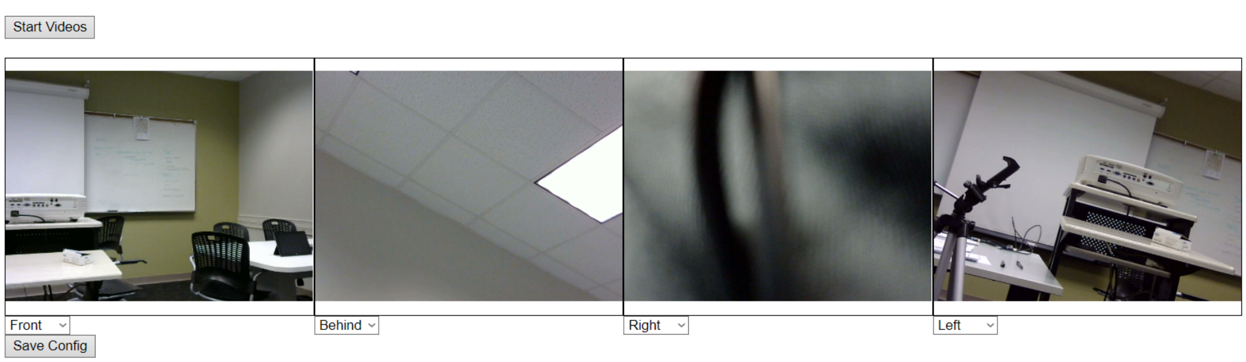


Figure 7.3: A screen capture of the Streamer's client application. The user simply needed to determine which camera pointed in which direction using drop-down menus.

illustration.

After specifying the direction associated with each camera, the user would then click the Login button, registering the user with the signaling server. For simplicity, we predetermined the user's login name; the participants did not need to specify any credentials. When the server detected a successful login, it messaged the Streamer module with this event. The module then automatically began calling the Viewer module using another predetermined login name. In this way, the only setup the Streamer needed to perform was identifying which camera faced which direction; but, in future implementations, the Streamer should also provide their own login name and choose which Viewer to call.

Viewer Module

The Viewer module provided a graphical user interface to choose and view the desired directional camera feeds. First, the user clicked a button to login to the page. As with the Streamer module, we automatically defined the user's name for our study. After logging in and receiving a request from the Streamer, the GUI is displayed. The selected feed was maximized on the screen, and a series of buttons were displayed in a list along the side of the screen; these buttons allow the user to send a message back to the Streamer module which, upon receiving the message, automatically replaced the current camera feed with the one the Viewer selected. One additional button was displayed on this page, allowing the user to take a picture and save it on their local machine. See Figure 7.4 for illustration.

In addition to receiving the audio and video feed from the Streamer, the Viewer module detected the user's microphone and transmitted audio back to the Streamer. This allowed users to have bidirectional voice communication. Finally, if the Viewer's connection was disrupted, or if they accidentally closed the tab, the signaling server would automatically detect this event and inform

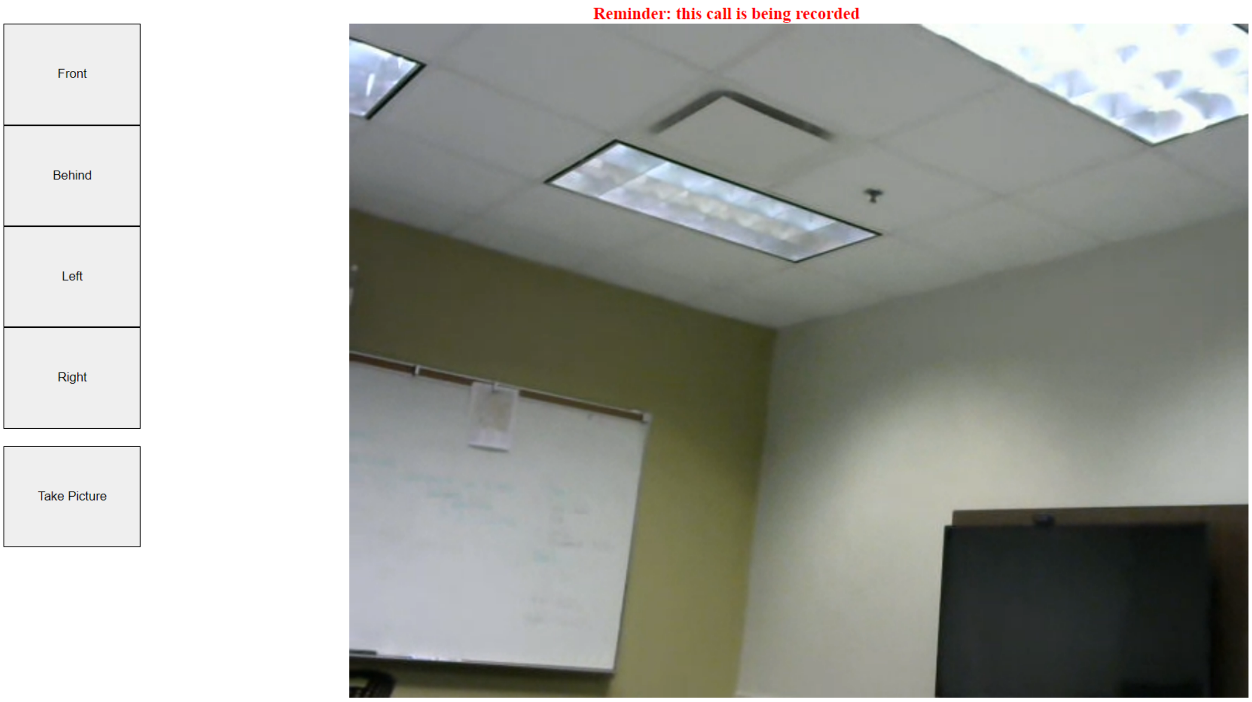


Figure 7.4: A screen capture of the Viewer's client application. Buttons allowed the Viewer to choose the camera they wanted to view, as well as take a picture of the current camera view.

the Streamer module; in such an event, the Streamer module would resume calling the Viewer until they logged in again, to resume the call.

Controller Module

An additional module was developed such that the Streamer user would have an avenue of controlling the stream in more sensitive contexts. This module simply allowed the user to access the website from their smartphone device and send two messages to the Streamer' module via button presses: Pause or Resume.

The Pause button sent a message to the Streamer module which removed the video track from the Stream. Audio could still be transmitted and received, but not video. The Resume button replaced the previously selected video track back into the stream. We added this functionality for the event that a Streamer would need to pause the interaction without completely hanging up, such as entering a restroom.

Researcher Module

The final module was a Researcher module which we planned to use to collect and record the verbal communication between users, and the selected video feed that the Viewer saw. In this way, we could record the entire interaction that would take place between our participants, which is useful for when researchers could not be collocated with participants. This module was unused for our study, but future researchers can leverage it as needed.

Next, we discuss our hardware selections and the iterations of our prototype that were tested before we reached our final design.

Hardware Design Iterations

Per our co-design requirements, in accordance with our VSD approach, our main goal was to embed as many human values into the Streamer's hardware setup as possible; this included making the setup as unobtrusive, inconspicuous, and inexpensive as possible, while still providing the viewer with opportunity to explore the environment. In this section, we discuss our hardware selections and the tested iterations of our prototype before reaching our final design. In total, we attempted four different designs. We discuss each in their own section below. Each of these designs used four of the same cameras — Logitech C270 Webcams[9] — which we stripped from

[9]https://www.logitech.com/en-us/products/webcams/c270-hd-webcam.960-000694.html

the plastic casings and sewed into the fabric of the various clothing and accessories. The cameras had 60° field of view and were USB plug-and-play. We selected this camera because it was easily detectable by computers and it was relatively inexpensive (at the time of the work, $40 USD). We did not write our own drivers for this device.

A Hidden Sports Hat Camera Rig

Per our co-design feedback, we first tried to embed camera lenses inside a simple sports hat. We used an all-black cap because the color matched the camera lenses, and we felt that this would help prevent the cameras from being detected. The hat was stretch-fit and had a sweatband along the inside of the cap, which we leveraged to prevent the cameras' circuitry from making contact with the user's skin. We cut holes in the hat's fabric just big enough for the lens to fit through, and after positioning / rotating the lenses to be upright, we anchored them into place by sewing black thread around the board. Four cameras were used to display the front, back, left, and right of the streamer; see Figure 7.5 for illustration.

We stress-tested our design before deploying it with a user. We ran our application for 90 consecutive minutes with all four cameras turned on, and the prototype worked successfully. However, we noticed that the front camera was tilted backwards, and the lens thus aimed too far up to be useful. In addition, wearing the hat for a prolonged period of time was somewhat painful. The electronics heated up to the point where a burning sensation was felt on the tester's head, and the electronics pressed into the skin, even though the sweatband was buffered between the two. A large red mark was present on the testers forehead at the conclusion of the test.

We thus opted to try a different style of hat; we wanted an accessory that was less rigid than the sports cap, and we wanted to be able to place the cameras further away from the skin, to prevent pain and discomfort. We decided to try a boonie hat for these reasons.

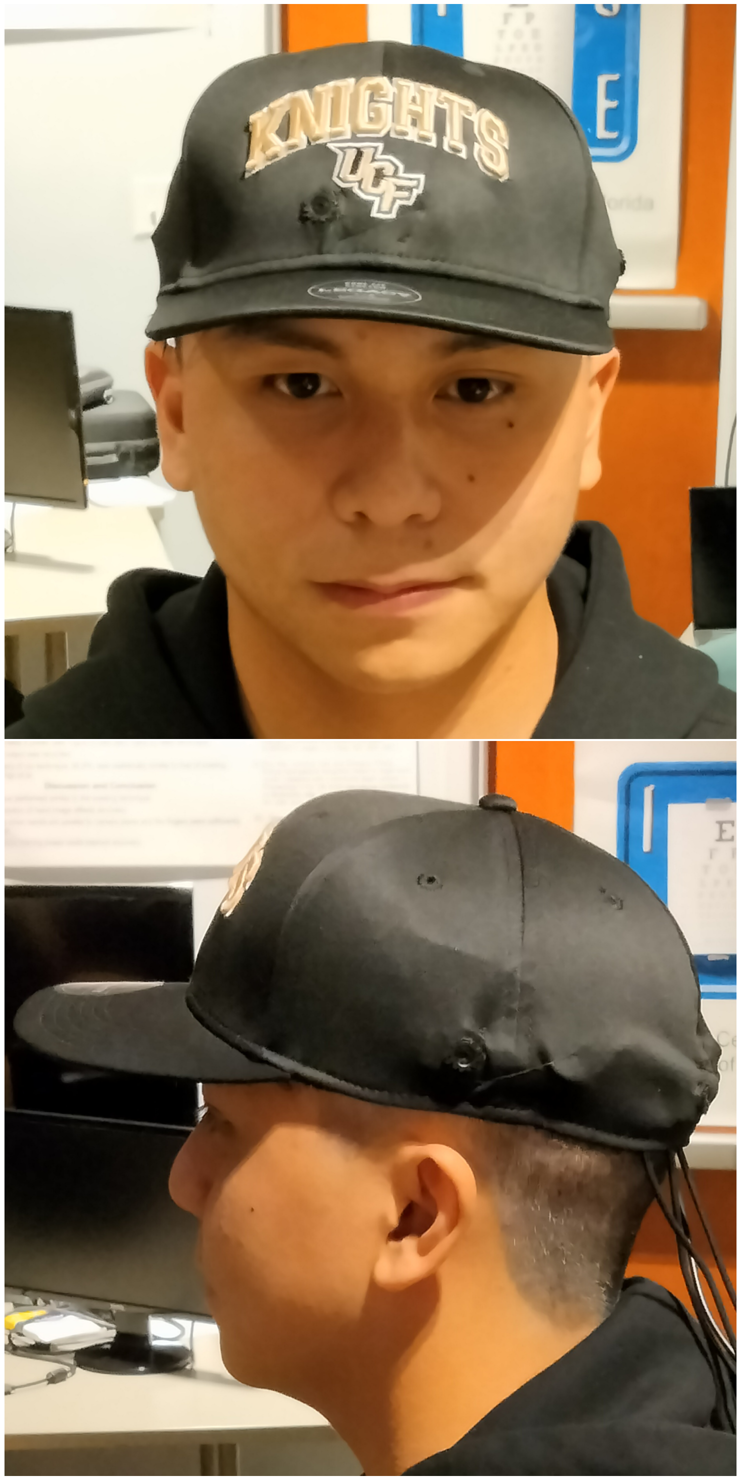


Figure 7.5: The hidden sports cap camera rig. Cameras were embedded on the front, back, left, and right of the cap.

A Hidden Boonie Hat Camera Rig

We found a simple, unisex, olive-colored boonie hat that had a breathable mesh on the top, with an elastic, adjustable drawstring in the back. For the previous reasons, we suspected this hat would be superior to the sports cap. Instead of embedding cameras between the hat fabric and the sweatband, we instead opted to anchor the cameras on the outside layer of the hat; to hide them, we sewed black ribbon around the entire hat, so that it looked natural and would shroud the electronics from view. Similar to before, we cut small holes in the ribbon for the lenses to fit through, and the same four camera directions were supported. See Figure 7.6 for illustration.

We stress-tested this design and found that the discomfort was indeed alleviated; there was no reported pain, and the heat was not felt. However, we did notice a problem that would challenge any head-mounted camera rigs. As the cameras were tied to the head movements, natural head nods would pitch the front and back cameras, and roll the side cameras. This meant that the view for the side cameras would often appear tilted if the user moved their head, such as for looking at stimuli below or above their eye line. Another problem we encountered was that the camera cables were visible, and with four cameras mounted on the hat, there was significant visibility that might cause a user to feel self-conscious.

We took these problems into consideration and reasoned that it might be more appropriate to decouple the cameras from the head, and position them closer to the torso. In the next section, we describe a Backpack Rig with which we attempt to overcome these issues.

A Hidden Backpack Camera Rig

Similar to the previous iterations, we wanted to provide a complete view of the environment with multiple cameras, but we wanted to decouple the peripheral cameras off of the head. We used a

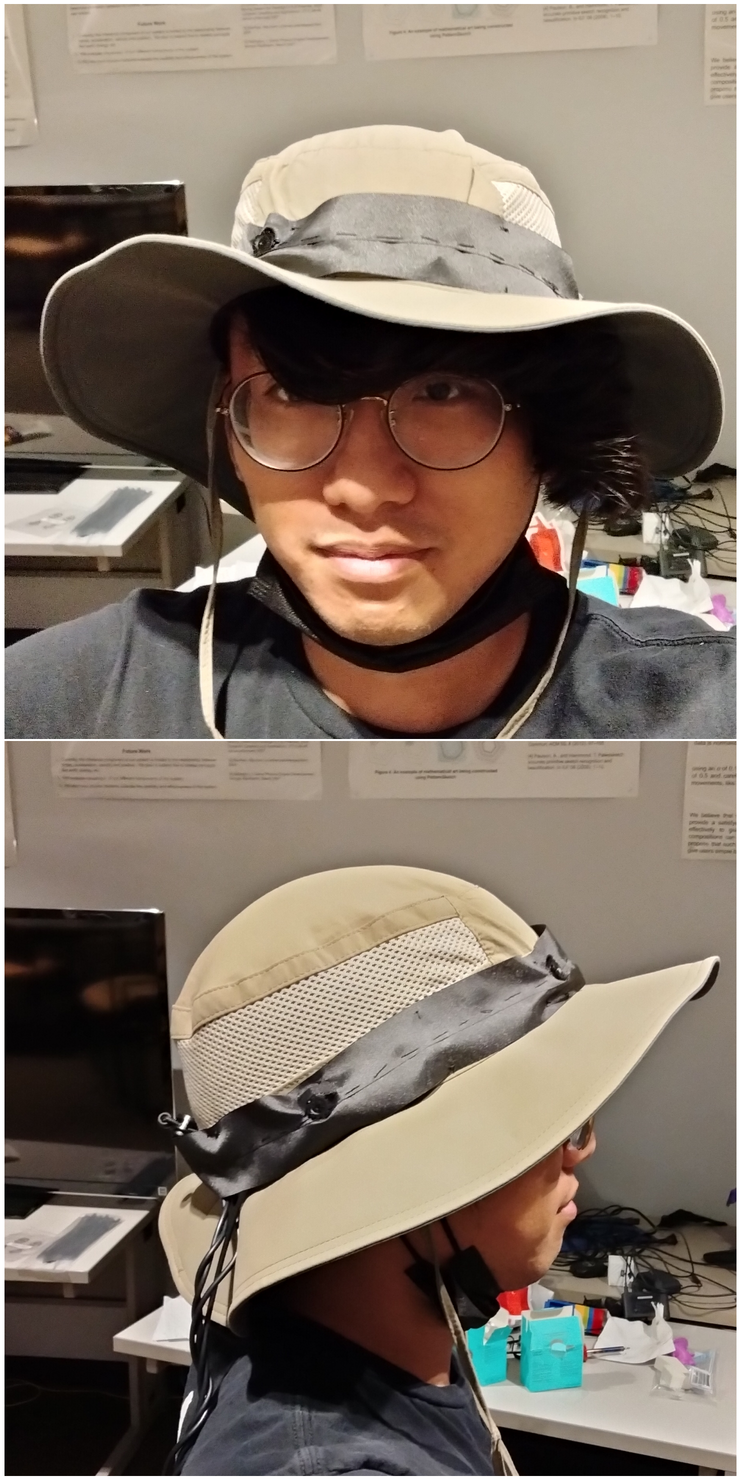


Figure 7.6: The hidden boonie hat camera rig. Cameras were positioned on the front, back, left, and right of the hat, and veiled with black ribbon.

black backpack and modified it by securing the cameras with thread on the outside of the left and right pockets, as well as a flap that pointed behind the wearer. Lastly, we secured the front-facing camera to one of the shoulder straps. Similar to our boonie hat rig, we used fabric to shroud the electronics from view; we also cut holes in the bag to feed the wires to the computer. See Figure 7.7 for illustration.

As anticipated, the cameras were much more stable, and walking gait did not significantly impact the pitch or roll of the cameras. The only issue we found was that the shoulder strap camera was not easily positioned to "look forward", as individual chest sizes affected the strap's orientation. For instance, the camera would often tilt upwards, or off to one side, depending on who wore the bag. Further, as there was no longer a camera anchored to the forehead, it proved difficult to identify where exactly the streamer would be looking.

For these reasons, we decided to iterate our design one final time. We combined the boonie hat and the backpack to form our final prototype.

Final Prototype: Boonie Hat + Backpack Camera Rig

For our final prototype, we used the forward-facing camera attached to the boonie hat, and the left, right, and behind cameras from the backpack. In this way, a viewer could understand which direction the streamer would look, and also have additional viewpoints to help understand the remote surroundings. We removed the other 3 cameras from the hat, such that only one cable was visible. All of the cameras attached to a USB-C hub, which plugged into a laptop that was stored in the backpack. See Figure 7.8 for illustration.

Analyzing the prototypes that were represented in Chapter 3, we find that our final design positions itself between the JackIn Head system, which is a head-mounted multi-lens 360° camera

Figure 7.7: The hidden backpack camera rig. Cameras were positioned on the front strap, back, left, and right of the bag, and veiled with black material.

Figure 7.8: The final prototype used in this work. We leveraged the front-facing camera of the hat, and the back, left, and right cameras of the backpack.

(see [100]), and an eyeglasses-mounted device which was used by Procyk et al. [202]. Next, we describe a user study in which we deployed our prototype for pairs of participants, garnering their feedback toward this system and that of a control.

Methods

In this section, we describe a user study in which we deployed our prototype and had participants compare it with Skype.

Study Design

We conducted a 2x1 within-subjects experiment, and the only independent variable was *Device*. The two conditions were Skype and our custom camera rig. We randomly assigned condition order and applied counter-balancing. Our study had IRB approval.

Apparatus

For the custom camera rig condition, the streamer was given the backpack and the developed prototype. We used a laptop with an Intel i7 processor clocked at 2.3GHz, 8GB of RAM, and the Windows 10 operating system. The laptop had a USB-C port, and to connect the cameras to the laptop, a compatible 4-port USB hub was used. For internet connectivity, we used a OnePlus Nord N200 phone to function as a mobile 5G hotspot. To support audio communication, we used Jabra Elite 3 earbuds that connected to the laptop via BlueTooth; they included an embedded microphone, and we enabled the pass-through audio mode such that the user could hear their surroundings. The viewer participants viewed the stream on a Microsoft Surface tablet that had an Intel i5 processor clocked at 1.7GHz, 4GB of RAM, and the Windows 8 operating system.

For the Skype condition, we asked that the streaming participant would use their own phone; but in the event that this was not possible, we let them use our OnePlus Nord N200. The viewer participant viewed the stream on the same Surface tablet for both conditions. Both participants simply used the basic features of Skype, which supports audio communication and video streaming from a laptop's webcam or smartphone's built-in camera.

Procedure

The study occurred on UCF's main campus in Orlando, Florida. The campus is circular in shape, with a major road enveloping the majority of the buildings (see Appendix D for a map). Our participant dyads met us at the Barbara Ying Center building, which is located near the southern-most side of campus, and is separated from the main campus by that major road. After completing the informed consent process, we issued a demographics questionnaire. Following, we explained that one participant would need to act as a streamer for both study conditions, while the other would need to act as the viewer. We let the participants decide how they wanted to assume their respective roles.

After they determined the roles, we collected photo IDs from both participants, which we held until the study was complete. We then applied the first condition in a randomized, counter-balanced order. For the Skype condition, we had the streamer participant call the researcher's Surface tablet, and we received the call using the researcher's Skype account. For the custom camera rig condition, we showed the participants how the devices operated, and then had the streamer don the hat, backpack, and wireless ear buds.

We gave the streamer a map of campus, and sent them to complete the task, described below.

When the streamer left the room, we gave the viewer a map which was marked with the artwork locations. In addition, we supplied the viewer with a simple website that displayed the artwork that was needed for the current task iteration. In this way, the viewer knew which artwork was needed, and where exactly they resided, while the streamer only knew the general layout of campus. We recorded the interaction with a smartphone camera, pointing it at the viewer and their screen.

After the task was completed, or after 60 minutes elapsed, the streamer was recalled back to the starting location. When the streamer arrived, we issued a questionnaire to both users that contained many Likert scale items regarding the experience; each individual had a different set of questions. We allowed the streamer to have a break, and when they were ready, we repeated the procedure with the second condition.

When the participants completed the experiment, we conducted a semi-structured interview to garner their detailed feedback about the conditions. The study lasted approximately three hours, and at the end of the session, each participant was paid $30 USD in cash.

Scavenger Hunt Task

For our study, the participants needed to work together to complete an artwork scavenger hunt. Around the UCF campus, there are many art installations and paintings, both outside and indoors. We selected a subset of these artworks to include in our scavenger hunt task, favoring the outdoor art to prevent network connectivity issues; in total, we selected 14 pieces of art scattered around campus, but divided these into two subsets. These two subsets are found in Appendix D.

We did not polarize the artwork into two halves of campus, instead choosing to have the locations overlap; the presented art subsets were counterbalanced along with condition. This introduces a learning effect, but it also helps us understand how users behave with varying levels of familiarity.

To complete the scavenger hunt, the participants needed to work together to guide the streamer to each piece of art, and the viewer was required to take a picture when the art was in view of the camera, in a way that closely matched the example picture provided to them. For both conditions, the viewer had an option to take a screen capture. For the custom camera rig condition, the picture needed to be taken from the front camera. Our task mirrors that of Tang et al., who used a 360° camera mounted on a pole worn by the Streamer, who needed to take pictures of art from certain angles described by the Viewer [243]. One difference is that our Viewers were tasked with taking pictures of the art instead of the Streamer.

Dependent Variables

Our participants were given questionnaires after each condition to measure a variety of constructs. All items were measured on a 7-point Likert scale. In accordance with the findings from our literature review (Chapter 3), we wanted to ensure that we captured the perceived sense of social presence from *both* stakeholders, instead of just the Viewer. We also wanted to measure how self-conscious our devices made our Streamer participants feel. Lastly, we wanted to understand benefits and harms associated with each device. We thus applied a variety of constructs in questionnaires that were given after each condition was complete.

The viewer was given questions to measure the following constructs: Co-Presence, Psychological Involvement, and Behavioral Engagement operationalized through the Networked Minds Measure of Social Presence questionnaire [22]; Intention, Usefulness, and Ease of Use, operationalized through the Technology Acceptance Model [261]; Emotional Expression, Engagement & Play, Presence-in-Absence, Social Support, and Threat to Privacy, operationalized through the Affective Benefits and Costs questionnaire [276]; Spatial Presence, operationalized through the Presence questionnaire [271]; Quality of View (see Chapter 4); and social connectedness (custom items).

The streamer was given the same questions to measure Co-Presence, Psychological Involvement, and Behavioral Engagement; Intention, Usefulness, and Ease of Use; Emotional Expression, Engagement & Play, Presence-in-Absence, Social Support, and Threat to Privacy; and Social Connectedness. In addition, these users were asked questions regarding their level of Self-Consciousness. There does not exist an instrument to measure the self-consciousness of a user as a state; we thus leveraged a trait-based survey from Scheier and Carver [224], and participant quotes from previous field studies to generate a new instrument. All instruments can be found in Appendix D.

Semi-Structured Interview Questions

After both conditions were completed, we finished the session by administering a semi-structured interview with both participants simultaneously. We used the following questions as a guide for our semi-structured interview, but dug deeper on user feedback to help uncover meaning and more details. See Table 7.1 for a list of questions that guided our interview.

Table 7.1: Semi-structured interview questions used in our study.

(Streamer) Which was easier to use and why?
(Viewer) How could we have improved your experience?
Why did you choose the roles that you chose?
Between the two systems, which did you like better and why?
Which made you felt more socially connected and why?
Which made you feel more self-conscious and why?
Which was more useful and engaging with the environment and why?
If you could redesign the Streamer capabilities, how would you do that?
Were there any other considerations we should have taken?
What type of situation would you use this system and why?
When would it not be useful?
For what context do you think Skype would work better?
For what context do you think the system would work better?

Table 7.2: Participant Demographics. We recruited eight unique participant dyads, who chose their own roles.

ID	Relationship	Role	Age	Gender	O	C	E	A	N
1	Married	Streamer	27	F	3.8	2.0	3.0	3.0	3.3
		Viewer	30	M	3.0	4.0	3.0	3.8	3.5
2	Friends	Streamer	22	M	3.0	3.3	2.5	4.3	3.5
		Viewer	30	M	5.0	3.8	3.5	4.3	2.0
3	Friends	Streamer	25	M	4.3	4.5	3.5	4.0	3.3
		Viewer	22	M	3.3	3.8	2.5	4.0	2.3
4	Boyfriend/Girlfriend	Streamer	23	M	4.3	4.0	3.3	4.0	2.0
		Viewer	22	F	4.3	2.8	4.3	3.5	3.3
5	Friends	Streamer	26	M	3.5	3.0	1.8	3.5	3.0
		Viewer	20	M	4.0	3.0	2.8	4.3	2.3
6	Best Friends	Streamer	19	F	5.0	2.8	2.5	3.8	4.0
		Viewer	20	F	4.8	3.8	3.0	4.0	1.5
7	Classmates	Streamer	24	M	3.5	4.0	4.0	4.5	2.5
		Viewer	23	M	3.8	2.3	2.3	3.5	2.5
8	Friends	Streamer	19	M	3.8	4.0	2.8	3.3	1.8
		Viewer	19	M	3.5	3.3	3.3	4.0	2.8

Subjects

We recruited 10 participant pairs, but one did not appear for their scheduled time, and one other was canceled due to thunderstorms. Our final participant pool consisted of 8 user dyads. The demographic data for these users can be found in Table 7.2.

We asked our participants their age, gender, and relationship with their partner, and how experienced they were with virtual reality technology, conducting live-streams, and watching live-streams. We also administered the Mini-IPIP [42], an inventory that measures the Big 5 personality traits with a smaller set of questions.

Data Analysis Approach

Since we were unable to recruit a large enough sample to run meaningful statistical analyses, we do not report statistical tests in this work. However, we do report the descriptive statistics for each questionnaire construct to support the findings from our qualitative analysis.

With the audio and video recordings, we planned on conducting a thematic analysis using a grounded approach. We first used open coding and subsequently iterated our codes until themes emerged. We report the major themes that emerged from our dataset. We combine the session ID with the participant role to provide their quotes; for instance, S1 corresponds to the Streamer from session 1, and V8 corresponds to the Viewer from session 8.

Results

In this section we present results of our study. First we present the descriptive statistics of the questionnaire data, followed by the emergent themes from our qualitative analysis.

Descriptive Statistics

Descriptive statistics for the questionnaire items can be found in Table 7.3. Though we do not have enough data to run statistical tests, we do see the data hint towards differences and comparability within the more compelling constructs. In the Prototype conditions, our Viewers reported a much stronger sense of co-presence than the Streamers, and this sense of co-presence was stronger for both stakeholders in the Skype conditions. While our Viewers seemed to strongly desire the Prototype due to it providing better viewing angles than Skype, the Streamers were not so enthusiastic. The Streamers, on average, felt more self-conscious using the Prototype, but we find that the scores

Table 7.3: Descriptive Statistics of the questionnaire indices

Construct	Condition	Streamer		Viewer	
		M	SD	M	SD
Co-Presence	Prototype	3.9	0.4	4.9	0.6
	Skype	4.3	0.7	5.1	0.8
Psychological Involvement	Prototype	4.9	0.6	5.6	0.3
	Skype	4.8	0.9	5.2	0.5
Behavioral Engagement	Prototype	5.8	0.7	6.0	0.9
	Skype	5.5	0.7	6.3	0.3
Intent to Use	Prototype	4.7	2.0	5.7	1.6
	Skype	5.0	1.8	4.4	1.8
Usefulness	Prototype	5.0	1.5	5.3	1.2
	Skype	5.2	1.9	4.5	1.0
Ease of Use	Prototype	5.3	1.7	6.3	1.1
	Skype	5.1	1.7	5.6	0.7
Emotional Expression	Prototype	5.0	1.0	4.3	1.6
	Skype	5.1	1.2	4.1	0.8
Engagement & Play	Prototype	5.3	0.9	5.5	1.3
	Skype	4.9	1.5	4.7	1.5
Presence-in-Absence	Prototype	4.5	1.2	4.8	1.2
	Skype	4.5	1.2	4.3	1.0
Social Support	Prototype	4.4	0.9	4.3	1.7
	Skype	4.7	1.3	4.0	1.4
Threat to Privacy	Prototype	2.8	1.3	2.2	1.5
	Skype	2.8	1.4	2.6	1.9
Social Connectedness	Prototype	4.5	1.0	4.9	1.0
	Skype	4.9	1.0	4.4	0.9
Self-Consciousness	Prototype	3.4	1.5	—	—
	Skype	2.3	0.9	—	—
Spatial Presence	Prototype	—	—	4.9	0.5
	Skype	—	—	4.3	0.4
Quality of View	Prototype	—	—	4.7	1.7
	Skype	—	—	5.1	1.8

for both conditions were not indicative of a severe problem in this regard. As we discuss in the next section, our participants still liked the idea of an unobtrusive multi-camera rig for many reasons.

Qualitative Analysis

In this section, we present the qualitative findings that emerged from our dataset. We first discuss device preferences, and provide reasoning behind these choices. Next, we describe how well our Streamer participants' social needs were supported per condition. Lastly, in accordance with the identified value tension of autonomy, we describe how our Viewer participants perceived the varying ability to explore the remote environment.

User Autonomy Supported: Participants Preferred Prototype

Six of the eight Streamers preferred to use the custom prototype to perform their role, and six of eight Viewers preferred to use the prototype to receive the stream. The primary reason Streamers preferred the prototype was because their hands were freed (4/8):

> "Definitely the [prototype], because its hands-free. Meanwhile when using Skype, I had to hold my phone most of the time. Sometimes I'd put it to the side, but I only had one hand free, compared to the system where I was totally free" -S5

This was particularly the case for three Streamers who rode their bike or skateboard around campus. Other reasons for selecting the prototype included providing a more socially acceptable paradigm (1/8) and a more socially connected experience (1/8). The two Streamers who preferred Skype indicated that it was more lightweight (1/8) and that there were fewer adjustment issues (1/8).

The primary reason Viewers selected their preferred device was quality of view. For the six of eight participants who preferred the prototype, they enjoyed being able to select the various viewing angles because it offered something they could do autonomously, while providing ability to explore the remote environment:

> "From my perspective, I prefer the system. First you can see a lot of point of views, you get the right angle, left angle, front and back; it is really useful if you want to identify some stuff around you." -V5

However, the participants who selected Skype as their favorite also pointed to quality of view as the major point for their decision. One noted that their streaming partner did not do a good job aligning the front camera with what they wanted to see, while the other noted that the brim of the hat got in the way of the camera; this group participated on a particularly windy day *and* the streamer used a bike, so the brim flipped upward often.

> "I might say [I preferred the] prototype, but as he was going on cycle, I couldn't see the front camera because the hat flipped up. If he walked, I might recommend the [prototype]. But for what we just did, I would prefer the Skype." -V7

Our participants described their preferences in multiple ways, but they generally selected the prototype as their preferred device because it supported their autonomy — the Streamers' hands were kept free, and the Viewers were allowed to explore on their own. They also showed us through their actions and behaviors different reasons why each system has their advantages and disadvantages. In the following sections we dig deeper into the major themes that emerged in our dataset.

Image Misalignment Causes Frustration and Self-Consciousness

One of the advantages Skype had over our prototype was that the Streamer could sometimes have visual feedback of what they were pointing the camera towards. Some participants chose to use the front / selfie camera to complete the task, but others used the phone's back camera — this meant that, when they wanted to show the art to their partner, they could see exactly what their partner was seeing. This feature was missing in our prototype system, and it caused frustration because the partners had significant difficulty completing the task, as evidenced by our participants' in situ communications:

> "Move towards the right a little... two or three steps forward... now move your head towards left... more left... a little left... move your head towards left. Rotate your head. A little right... OK that's it." -V3, Prototype

All of our participant pairs encountered this problem, though only 3 of the pairs verbalized this in our interview process:

> "With the [prototype], I was frustrated a little bit. She said 'Put the hat down!' and I'm like, 'I am!'" -S4

In addition to adding frustration, this misalignment problem also caused some participants to feel more self-conscious. Since aligning the cameras for the perfect photograph of an art piece sometimes took upwards of ten seconds, and the camera angle needed constant correction, this brought on a sense of awkwardness:

> "[People were thinking:] 'she's still talking to herself; do we call someone? She's

> just like standing in the grass right now... looking around weirdly... just circling sculptures.' Literally I was just there staring creepily with a backpack." -S6

This sense of self-consciousness was thus induced because of a *lack* of features; however, as we show in the next section, a simple design decision like the style of hat was enough to make many Streamers feel even more conscious about their looks.

Streamers Desire Prototype Customization

One of the most pronounced themes from our dataset was that every individual Streamer wanted something different to meet their needs. While the backpack and the three peripheral cameras were met with high praise, the boonie hat was met with an array of disapproval. Only one Streamer thought that this hat was a good choice. The others voiced their own opinions of what hat (or other accessory) should be used instead, including a sports cap, *no* hat, a jacket-embedded placement, a backpack-strap, shirt pocket, harness, and eyeglass cam. In addition, participants noted that any given accessory can not possibly be appropriate for every given scenario:

> "I think it should be customized, so that if it's rainy season, you could wear a different hat, and if it's sunny, normal hats could do a really great job. Sports cap, so it doesn't feel weird... and the boonie hat feels weird" -S3

> "You should have like 3 different options: casual, professional, and sport" - V6

We even found some discrepancy within the individual participant pairs:

> Researcher: "Is there a better piece of headwear that you'd prefer?" S4: "A fedora." V4: [rolls eyes and makes a face] "Who wears Fedoras?!"

To further investigate the problem of self-consciousness in public, we polled our streamer participants during the exit interview if they encountered any instances where bystanders stared at them or made them feel uncomfortable. Most of the time, participants responded that they did not particularly feel like they were being watched, but, using Skype brought more attention to the Streamer than the prototype:

> "People definitely noticed. The responses to that notice was different though. I would say people were a lot more subtle staring at me in the prototype, but whereas when I was recording [with Skype], people tried to photobomb me... Some people tried ducking under, they were all apologetic, I felt bad." -S6

These types of interactions were rare with our prototype, so it does seem that we achieved what we were aiming for — a less noticeable interaction style; yet as most of our participants indicated, our singular prototype did not meet their needs completely. Most of them wanted our device to have a different form, and the boonie hat that we used did contribute to their self-consciousness. This, in conjunction with the visible wire that connected the camera to the USB hub, was the main cause for our participant's level of self-consciousness.

Freedom of Exploration Provides More Engaging Interaction

Asking our participants about which device helped foster a sense of togetherness, we found that both were able to provide this feeling in their own way. Skype's innate video chat features allowed both users to see and hear each other, which was the primary way our users were able to feel co-present. Our custom prototype did not allow the users to see their partners' face, but, they still had a sense of co-presence, though in a different way than Skype:

> “I would actually say that [with the prototype] it felt a little more like a mission, like a task really we’re completing together, whereas [Skype] was more like him telling me a place and then I go there” -S8

> “When I’m personally looking at a computer screen and I see that person, and that’s where the audio is coming from, my brain is like, OK, there’s a level of this that isn’t sincere or real because this a video call; but if I don’t see the video, then it’s almost more like you are in my head, so it feels more personal. My brain can’t automatically associate “this is a call”. -S6

Other participants noted that one of the weaknesses of our prototype was that users were not able to see each other. Though the Viewer could see their partner’s surroundings, they were not able to see the reactions and facial expressions of their partner, which is something that users desire.

While our prototype did not enhance the sense of co-presence among participant pairs, it did provide a better viewing experience, and helped to keep the users engaged with each other. During navigation between the different pieces of art, we constantly observed Viewers pressing the buttons to explore the peripheral cameras, which, by allowing them to see more of the environment, provided them with more opportunities to comment on the remote surroundings and keep conversations going. In contrast, when Skype was employed, we observed many instances where the Streamer needed to put the smartphone down due to their arm tiring, or to concentrate on navigation — this was particularly true for the individuals riding their bike or skateboard, who needed to focus on steering. When the phone was put into their pocket or otherwise placed into a position that did not offer exploration, we observed the Viewer participants turning on their own personal smartphones to check text messages, social media posts, or other distractions that did not relate to the study. Here, the participants wanted to fill the gap between task locations, because there was nothing else to see or do.

General Discussion

In this section, we provide a general discussion of our findings and position them among previous literature.

Prototype Enhancements Required

Though the majority of our participants verbally expressed their preference for our prototype, we acknowledge that there are still improvements that can be made. Both roles encountered frustration when the participants needed to align the camera to take a picture, and this can be alleviated in part by including additional features. Visual feedback was requested for future iterations, and some participants suggested the incorporation of inconspicuous Augmented Reality devices such as Google Glasses to tackle the problem. Such a device could include a simple window to display a graphical overlay of the active camera such that the Streamer would know what their partner sees at any given time. Further, these low-profile devices would support the value of social acceptability. Another alternative is the addition of a module which the Streamer could use on their own smartphone device to preview what they are streaming.

One weakness of augmented reality glasses is cost; currently, the inclusion of such a device would significantly increase the price of our prototype. As such, other avenues should be considered. Although our camera rig is a pseudo-360° device, it does not capture the entire environment. Wider lenses would help approach this true panoramic view, which would also provide the Streamer with a buffer, so that they could focus less on achieving a completely aligned image. This would in turn help to alleviate the behavior-induced sense of self-consciousness. Previous work using true 360° cameras mount them on a pole [243] or embed lenses in a rigid bracket [105], which does not meet our needs; but, future designs might find it possible to apply wider lenses and a stitching algorithm, such that the Viewer will miss less of the environment.

We also find that our participants desire a way see the face of their partner. Using an augmented reality display, a graphic overlay could toggle to show their partner; likewise, using a smartphone, the prototype could include this feature in the same manner as basic video chat. This feature would boost the users' sense of togetherness while interacting, and as it is optional, it would allow users to choose how they want to interact with their partner, particularly if the social landscape changes around them. Though it is simple to capture the facial reactions of the Viewer, it is more difficult to do so for Streamer's face; after all, that would require yet another camera far enough away from their body — but, this seemingly contradicts the purpose of the prototype, which is to provide an unobtrusive streaming setup.

One Size Does NOT Fit All

One unexpected result was our participants having very strong opinions towards the fashion of the hat. Though the boonie hat was unisex and seemingly appropriate for an outdoor excursion, we found that many participants simply would not wear that style of hat, or any hat in general. One of the limitations of our study is that we provided a singular prototype to our users, expecting that most would feel comfortable with it. The opposite was true; while the backpack and peripheral cameras were met with acclaim, the hat and primary camera were perceived negatively. This likely was a direct influence on the scores of our questionnaires. In this respect, we failed to create the prototype that was designed by our co-design participants (see Chapter 6). In our next iterations, we plan on developing wearable attachments that can be easily added to any given hat, or other accessories that can be used to anchor a front camera; in this way, any user could select the style of device that best suits them.

Knowing What "Over There" Means

In previous literature, one of the common themes is that users often do not know where their partner is looking. For instance, with a dismounted 360° camera in the loop, Viewers may make reference to specific stimuli, but gestures such as pointing and head nods are unseen because Streamers do not have an accessory to display this feedback [205, 243]. Additionally, as these 360° cameras are not coupled with the head, the Viewer is unable to immediately understand which direction their partner is looking. There are two different forward headings — one for the head, and one for the body; and if the Viewer and Streamer have different ways of communicating and interpreting directional feedback, the problem is exacerbated even further. If a Viewer would say "look to your left", the Streamer might wonder, "which left do they mean — left of the body, or left of the head?"

Our prototype helped, in a way, to alleviate these issues, as the front camera allowed the Viewer to know exactly where the Streamer was looking at any given time. This is not to say it is impossible for users to discern headings when viewing a 360° cameras; Lee et al. [134] combined one with a regular camera to allow the Viewer to see the complete remote surroundings, while also having a window that showed exactly where the Streamer was looking. Our prototype works similar, in a much less obtrusive manner. We find that Viewers using a prototype such as ours will have their desire for autonomy supported, and they will also be able to quickly identify the forward heading of their partner by simply switching to the front camera coupled to the users' head. By decoupling the peripheral cameras from the head and attaching them to the user's torso, we were able to achieve better understanding between users while still providing a relatively inconspicuous setup.

Most of the works in this space that aim at enhancing the viewing experience typically use one camera, whether it is a 360° device or not, but we find that the inclusion of multiple cameras helps to solve this problem. For instance, Kim et al. [115] used multiple smartphones, to enhance regular video chat with a second view of the remote environment — one was handheld, and the other was coupled with the torso. Likewise, Young et al. [278] deployed a coupled smartphone and 360° camera. Using an application on the smartphone, each user could see where the other was looking within a shared view of the environment. Our prototype follows these designs, in a way, by providing multiple cameras — one coupled with the head and others coupled with the torso — and we suggest the use of such designs to help the Viewer know where their partner is looking.

Limitations and Future Work

We acknowledge our limitations. Foremost is the bias in the responses we have received. While we wanted to understand how various user personalities affected behavior, the participants who volunteered for our study are likely more extroverted or less neurotic than the ones who would not volunteer; in other words, we have an inherit sampling bias. We also do not have enough data to report statistical findings, including correlation tests. We thus were unable to measure how user personality influences their decision-making processes. Further, the task duration was relatively lengthy, and as it was a within-subjects design, there was certainly a physical toll exacted from our Streamers. Some participants might have felt more comfortable with our hands-free system due to their personal strength (or lack thereof). We did not measure physical attributes of our users, but we do suspect that the weight of our prototype influenced responses; for instance, one participant ran around campus in a display of physical aptitude, while others walked and required a substantial break before the second leg of the trial.

Apart from difficulties in study recruitment and execution, we acknowledge that our results cannot generalize to every possible use scenario. Our study was conducted outside, in Florida, during sunny weather; we cannot say for certain that our devices would be appropriate indoors, or in cold or rainy weather, or in other parts of the world. We would expect that a prototype such as ours would be met with much hesitation during more formal settings, or in private spaces. As such, additional work should be conducted with alternative venues and use cases in mind.

We also note that our prototype, though based on values expressed in a co-design setting, did not completely meet the needs of the Streamers; due to varying sense of fashion, our singular accessory was met with some apprehension. We envision an array of telepresence devices that would allow any given individual to pick the devices that meets their own personal needs. Future work in this space should be conducted at the intersection of human-computer interaction and fashion, towards customization and accommodation of multiple scenarios.

Conclusion

In this chapter, we constructed a new interpersonal telepresence prototype that was based on human values identified by our co-design sessions, which in turn was influenced by previous literature in this space. By embedding cameras inside clothing and accessories, we were able to achieve a low-profile camera rig that still offered more opportunity to explore a remote environment than the baseline of video chat. However, there is still much work to be done towards supporting Streamer individual differences.

We find that our prototype is less obtrusive than a smartphone, but it is not usable in every situation. Our work is specific to shared, outdoor experiences in an informal setting, but individual differences directing fashion choices need to be considered in future iterations of devices such as ours. In the next chapter, we highlight the overarching themes and findings of this dissertation, providing design recommendations and avenues for future investigation.

CHAPTER 8: CONCLUSION AND FUTURE RESEARCH

In this dissertation, we describe multiple investigations into human values that should be supported by interpersonal telepresence designs; see Figure 8.1. We opened with a survey of the relevant literature in this space, highlighting the state-of-the-art and the significant problems that stand in the way of its ubiquity. We noted that there was an overabundance of works approaching the topic from a technical perspective, but a lack of work that focuses on the needs of the end users — specifically, we found that some streaming users felt socially unacceptable and self-conscious when using devices, including simple video chat applications. To help rectify this situation, we aimed at taking a human-centered approach to create guidelines for design. Our two VR-based studies showed that even simple design decisions in this space such as where to place the camera on one's body can cause a value tension between the stakeholders of interpersonal telepresence.

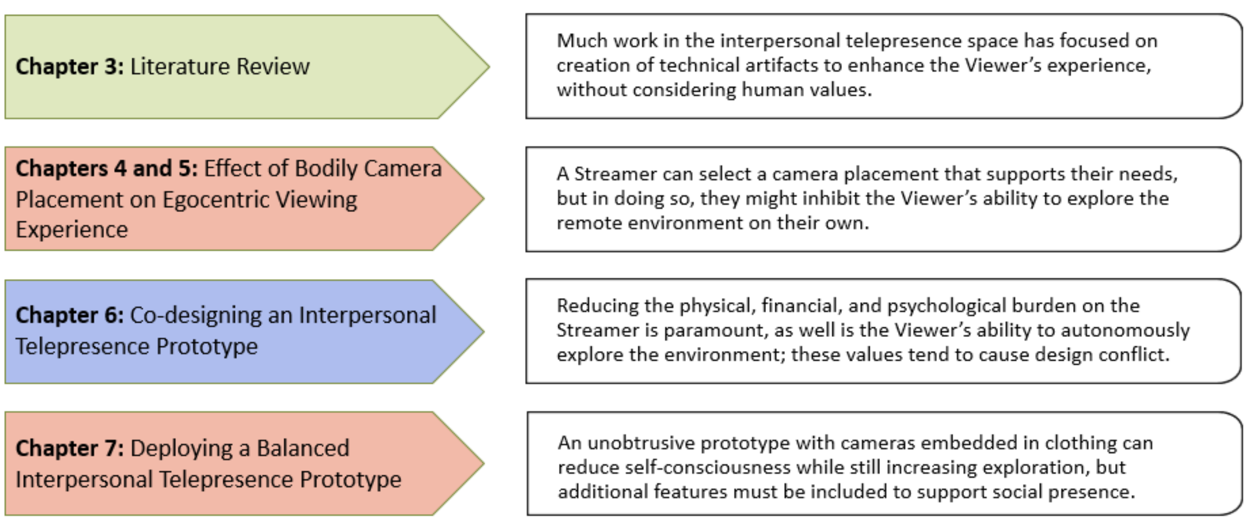


Figure 8.1: Using the VSD framework, we provided a series on contributions to the telepresence field. We reviewed the literature in this space, demonstrated how simple design decisions can cause value conflict, identified human values that should be embedded into future technologies, and evaluated a prototype that embeds these human values.

Further, where many efforts suggest the use of a device or technique that fits the "average" user, we find that individual differences must be supported in this space if it is truly going to become ubiquitous.

With our co-design study, in which prospective live-streaming users were asked to provide their input to interpersonal telepresence prototype design, we were able to identify human values that should be embedded into a telepresence system, including a reduced social and physical burden on the streamer; however, to accommodate these values, a system should support multiple interaction styles, such that the user can choose *how* to use it. Lastly, we developed a prototype based on user feedback and deployed it in the midst bystanders; we found that users each had their own idea of what the system should look like, indicating that any given device cannot meet the needs of every end user.

While each of the studies described in this dissertation had their own set of discussion points, we found that there were recurring themes that echoed across each: a heavy desire for customization, and the desperate need for an ethical discussion. In this concluding chapter, we discuss our plans to tackle these two themes such that we can truly balance the user experience for mobile, one-to-one interpersonal telepresence.

Customization is Key

This dissertation shows in every chapter how we need to take a human-centered approach to designing telepresence systems. While findings of the past propose application of a "one-size-fits-all" solution, we show that, in this domain, we need to give power to the actual users of the technologies we are designing. It is not enough to design a prototype that *we* believe is suitable for the masses; rather, it has become clear that we need to give power to the people who want to use these.

Thus, in the future, we plan on developing a toolkit that shows users how to develop their own systems. Following previous maker culture work, our toolkit will provide a suite of materials, tools, and instructions, such that any given user could create a telepresence system that would work for them (see [88, 236]). For instance, Jelen et al. [89] developed a toolkit to support elderly citizens' learning of newer fabrication technologies, and Kazemitabaar et al. [109] developed one that allowed children to create interesting wearable devices while learning skills that bolster their STEM education. Hartikainen et al. [74] describes that maker culture can be used to enrich the lives of vulnerable and marginalized populations. Research has shown how maker culture and development of new toolkits can help propagate technical skills to populations who traditionally do not have access to that type of education, and we expect that we can leverage such toolkits or create a new one, such that prospective users of interpersonal telepresence can develop a device that would meet their personal needs.

Another boon of helping a user construct their own prototype is that they might appreciate it more than ones that are made by others; a documented phenomenon, the "IKEA effect" shows how humans appreciate and prefer things that they personally build [185]. Previous work in maker culture space has referenced this psychological effect, demonstrating that even simple, low-quality prototypes deserve their makers' appreciation [89, 107]. We hypothesize that constructing one's own telepresence devices would also help alleviate their sense of self-consciousness because they would use a device that they created; and, if two users build a complete interpersonal telepresence prototype together, it might also increase their connectedness while using it, since it is a device and/or program that they worked on together. Thus, we find it critical to let users have the power to design and build a telepresence system that works for them. We plan on initializing a research program in which we investigate the creation of a physical hardware toolkit, such that any user could construct their own wearable telepresence prototype; and, as part of this effort, we plan to continue development of our software platform such that the users' hardware could easily be used.

This includes the modularization of our platform, so that users can develop and customize their own viewing experiences.

Towards Supporting the Third Stakeholder: Bystanders

One of the limitations of our work — and that of this research space in general — is a lack of attention given to the bystanders who might be collocated with a Streamer. As noted by previous researchers, the literature in this space to address this concern is essentially nonexistent [205]. The research community is still pondering how to balance the user experience of Viewers and Streamers with the social well-being of these Bystanders. In the United States, federal law protects the right to record others in public — the legal view is that there is no reasonable right to privacy in common spaces [124, 239]. However, we must still ask how, as researchers, we can support these users, to ensure that all parties are satisfied. One way to do this is by including them in design decisions; similar to how we conducted our co-design sessions in Chapter 6 to give power to the Streamers and Viewers, we can give a voice to this stakeholder as well.

As users in the studies of our previous two chapters pointed out, telepresence devices can be used in nefarious ways. Is it an issue if the bystanders never know they're being recorded? Do we have a moral obligation to consider their privacy, if our devices will only be used in public? How can we ensure that people only use these devices in appropriate ways? What should we plan on doing if a user breaches the social contract? Who bears the responsibility for ensuring safety of all users? These questions remain unanswered in this dissertation, but these types of questions can guide future work aimed at providing a balance between all stakeholders. In addition to conducting this type of work, we plan on organizing a workshop or panel in which a diverse set of HCI researchers can discuss these ethical concerns. By approaching this problem from multiple angles, we can triangulate best practices and recommendations for future iterations of telepresence devices. While we do not anticipate the generation of a rigid set of guidelines, we do suspect that we can uncover considerations that can be supplied to future researchers.

Final Thoughts

In this final paragraph, somewhat strong opinions are presented; thus, I switch to first person to ensure these thoughts only represent my own. After finishing the literature review of the interpersonal telepresence space (Chapter 3), I couldn't help but wonder — with all of the various prototypes described in the literature, why don't we see them deployed in society? Especially during times of COVID-19, I would expect that loved ones would want to share remote experiences together, yet a platform to support such an interaction did not emerge. I suspect it is because these devices and applications are simply not attainable, and in some cases, they are not desirable. In my opinion, we as a research community have been missing the mark with our efforts. While there is substantial value in technological advancements, I feel that it is now more appropriate to focus on fostering human-centered values with what we have. That is why I plan on open-sourcing my software platform used in Chapter 7; anyone who could benefit from this work should have it available to them. I also truly plan on conducting the aforementioned future work. It is time to put power in the hands of the end users so they can create and share wonderful experiences with their loved ones. With this dissertation, I hope to influence the next wave of telepresence designers; we have an opportunity to do good for the world — let's not let it go to waste.

APPENDIX A: PILOT STUDY: TOWARDS BALANCING VIEWER AND STREAMER NEEDS DURING INTERPERSONAL TELEPRESENCE

The following is a pilot study that we conducted for interpersonal telepresence in 2017. This is based on a work that has been previously published:

Pfeil, K.P., Kapalo, K.A., Koh, S.L., Wisniewski, P., LaViola, J.J. (2021). Exploring Human-to-Human Telepresence and the Use of Vibro-Tactile Commands to Guide Human Streamers. In: Chen, J.Y.C., Fragomeni, G. (eds) Virtual, Augmented and Mixed Reality. HCII 2021. Lecture Notes in Computer Science(), vol 12770. Springer, Cham. https://doi.org/10.1007/978-3-030-77599-5_15

Introduction

Telepresence is the ability to perceive and/or interact with a remote environment, as if actually there. Originally conceived by Minsky in 1980, it has been hypothesized that telecommunications can provide an avenue for a worker to perform a task from across the globe using a natural, egocentric point of view [159]. While this ultimate telepresence experience has not yet been fully realized, we have seen immense technological advances in that direction. For instance, the combination of 360° cameras and virtual reality (VR) head-mounted displays (HMDs) allow an individual to explore a remote environment with a greater sense of *immersion*, or a sense of being enveloped in an interaction (see [235, 271]), than ever before. To facilitate mobile exploration, the robotics field has brought forth wheeled platforms which can be piloted through a simple graphical user interface (GUI) [125]. However, these robots are typically constrained to a particular, pre-planned environment such as conference venues [182], hospitals [45], and sidewalks [78], as they still have problems in navigating difficult terrain.

To circumvent this difficulty, and to provide a more interpersonal experience, some researchers have hypothesized that the robot can be replaced with another human (called Streamer). This concept of interpersonal telepresence has recently emerged, and the HCI community is examining its benefits and detriments. However, we have still seen the implementation of interpersonal telepresence in the wake of the recent COVID-19 pandemic, which has forced worldwide communities to practice social distancing. For instance, the Faroe Islands conducted telepresence experiences by having a Streamer walk about the islands while remote users directed them where to go [3]. Additionally, some researchers suggest that we might even see this concept provide a new avenue for job creation; one person could perform tasks while being directed by a remote user. Misawa and Rekimoto describe a "physical body marketplace" [162], where the Streamer's role is equivalent to ride-sharing drivers (e.g. Uber and Lyft).

It is this style of interaction that thus influences our research interests. While numerous previous works have focused on the user experience for the Viewer (the person watching the video stream), we note that the user experience for the Streamer has been woefully understudied. One consideration is the Viewer may not have the ability to pick their partner, or perhaps they will share a Streamer's service with other Viewers simultaneously. As such, verbal guidance might not be desirable, or even possible. To help tackle this problem, we draw from Teleguidance literature, in which researchers evaluated vibro-tactile devices (in this work, we use the shortened term VibTac). These devices use feedback in the form of vibrations to convey information (such as directional commands), and they have been shown to succeed for guidance and navigation as primary or auxiliary communication channels [37, 108, 155, 157]. We suspect that they are suitable for use in interpersonal telepresence as well.

Therefore, in this work, we explore this concept with a lab-based study with which we gather feedback from 16 participants regarding a telepresence setting using VibTac devices. We describe a prototypical interactive system that allows a Streamer to give a live-streamed tour to a Viewer, who in turn can provide navigational commands in the form of tactile cues through a VibTac belt or hat. Additionally, we surveyed 30 other observers to help understand their how the third-party perceives this style of interaction. Our primary research questions include the following:

- RQ1: What types of VibTac devices are preferred by Streamers?
- RQ2: How do third-party members perceive interpersonal telepresence interaction?
- RQ3: What scenarios would benefit from interpersonal telepresence?

We found that our participants did not have a particular preference between our VibTac devices, though the navigational commands were more strongly felt with our developed hat. Additionally, our participants responded that, although they were comfortable acting as a Streamer, they were not particularly energized to be one. Lastly, observers were somewhat comfortable being collocated with a Streamer, although they felt a greater sense of trust in the Streamer than they did in the remote Viewer.

Related Work

Telepresence Streamers have been leveraged in a variety of projects, but we have yet to understand how humans truly regard this role. In this section, we review the relevant literature at the intersection of interpersonal telepresence and VibTac interfaces, and discuss our unique contributions to the telepresence community.

Interpersonal Telepresence

Interpersonal Telepresence is a fairly new concept, as the technology and infrastructure to support this kind of interaction has just recently emerged; but, we have seen many projects that in part help to realize this concept. The JackIn system by Kasahara et al. shows how the use of omni-directional video cameras worn on the head by Streamers can give Viewers an immersive avenue to explore remote environments [100, 102, 105]. Here, the Viewer typically utilizes a VR HMD, granting the ability to explore the environment through natural head rotations. Pfeil et al. investigated how high a camera could be placed, and found that camera height is not a significant factor to consider for telepresence design [195, 196]; thus, this interaction could be enjoyed by broad audiences. Even a simple live-streaming device, such as a mobile phone, can provide an adequate view, as with the Polly system which is mounted on a Streamer's shoulder [116, 122, 123]. Here, though the Viewer does not wear a VR HMD, they can change the viewpoint by controlling a gimbal that holds the camera, through a GUI. Between these two projects, the Viewer has either a first-person or third-person point of view in relation to the Streamer. The latter allows for a more personable connection between the two users, but this may or may not be desirable, as the Streamer could be a friend or family member, but they could be a stranger.

Some researchers have explored *who* can be a Streamer. Ishak et al. deployed a technology probe to the university classroom [87]. Here, the authors found that students would be interested in having their friends act as live Streamers, but there were reservations in having strangers be these proxies. This concern was in part addressed through works by Misawa and Rekimoto, where Streamers wore a tablet on their head, to display the face of the Viewer [161, 162]. Here, the Streamer did not interact with the environment, except as instructed by the Viewer's audible instructions. In this way, there was an illusion that the Viewer was actually in the remote environment. The authors suggest that the optimal Streamer would be someone known to the Viewer, and someone who has similar physical traits, so to enhance the illusion.

One of the problems in the interpersonal telepresence literature is the common finding that the Streamers feel socially awkward when other people see them using this technology [115, 198, 205]. In all of the above examples, the devices are clearly identifiable by third-party users, resulting in this awkward feeling. It is thus important to find a way to balance user experience for both members. Baishya and Neustaedter envisioned a way for long distance couples to increase their togetherness using always-on video chat [14]. Here, a smartphone was placed in the shirt pocket of both users, with the camera facing out. In this way, it was inconspicuous to third party observers, although the opportunity to explore the remote environment is stymied because of where the camera is placed. As such, there is work that must be performed to reach an acceptable balance of user experience between all parties involved.

In our research, we extend prior work by letting a Viewer command a Streamer through the use of VibTac devices. Our work is similar to that of Misawa and Rekimoto, in that the Streamer is asked to follow specific commands given by the Viewer [161, 162], and our technology probe is not unlike that of the Faroe Islands remote tourism application [3]. However, our prototype devices were designed to be inconspicuous, by embedding them into clothing. Our work contributes to the telepresence literature by providing an understanding of Streamer's perception of interpersonal telepresence devices, and to understand how comfortable they are with the general interaction style.

Vibro-Tactile Interfaces

VibTac interfaces have been studied in the past, commonly in the form of belts [155, 238, 252, 257]; but there have also been others in the form of vests [156], helmets [174], head bands [29], caps [108], and even socks and shoes [157]. They have been integrated into user interfaces to support the visually impaired; for instance, McDaniel et al. described their belt prototype in an effort to convey where other people were, in both direction and proximity [155], and Wang et al. developed a belt that provided directional commands to help users navigate around obstacles without a cane [265]. In these implementations, computer vision techniques are used to identify detect people and obstacles, and to calculate a route around them.

However, algorithms do not operate at 100% accuracy. It is sometimes recommended to have a human-in-the-loop (HITL) interaction, so to leverage human expertise and correct problems when algorithms make mistakes. Scheggi et al. proposed an HITL system in which a remote expert could view the surroundings of a visually impaired person, and provide real-time feedback in the form of vibrations [223]. Our work is inspired by this prior work, to help understand if VibTac devices can be used for general telepresence use. There is not a wide range of consumer-grade VibTac wearables, so we developed two simple devices; one is a belt similar to Tsukada's ActiveBelt [252], and the other is a hat based on a reduction of Kaul's HapticHead device [108]. Literature does not seem to point to optimal tactor configurations, so iterative development helped direct our design of these devices. In this research, we elicit Streamer feedback to the VibTac belt and hat, to understand which provides the clearest set of instructions.

Prototype System Design

We developed a custom live-streaming application in Unity3D; Unity's multiplayer service allowed us to easily transmit data over the network. To provide visuals, we used the Ricoh Theta S 360° camera, which supports live streaming. The camera was mounted to a backpack rig, giving an overhead view of the Streamer, similar to the work by Tang et al. [243]. We modified FFPLAY[1] code and used it as a DLL in Unity. A script decoded the live, equirectangluar video frames, turned them into RGB values, and mapped them into a spherical image. We were able to achieve a streaming rate of 30 frames per second with approximately 1 second of latency. Using a virtual reality head-mounted display (HMD), or by simply click-and-dragging with a mouse, a Viewer could manipulate the viewpoint to any desired angle. See Figure A.1 for a visual of the hardware worn by the Streamer; this consisted of a backpack rig, the pole-mounted camera, a VibTac device, and a laptop to send and receive data over the network. The total weight was approximately 20lbs

[1] https://ffmpeg.org

(9.1kg).

The VibTac devices we developed were a belt and a hat (see left section of Figure A.1). Each utilized four coin vibration motors that were sewn into the clothing and interfaced with an Arduino Uno, which was programmed to read serial port data. The Uno was plugged into the Streamer's computer via USB, and Unity's multiplayer service allowed us to easily transmit commands. When activating the motors, we used 3.3V (output of the digital pins), as we found 5V to burn the motors. The motors did not make direct contact with the body, and were veiled behind a layer of fabric. This is in accordance with the findings by Kaul and Rohs, that direct contact could cause marks on the skin of the user [108]. Our navigational commands were similar to those used by the Faroe Islands remote tourism [3]; we simply needed our devices to convey four commands - *Go Forward*, *Go Backward*, *Turn Left*, and *Turn Right*.

For the belt, the mounted motors were sewn into an elastic band that used a belt buckle for fastening. Our final design called for motors contacting the belly, the back, and each hip. Activating the belly motor indicated *Go Forward*, activating the right hip motor meant *Turn Right*, and so forth. For the hat, the motors were sewn into the fabric - our final design consisted of two placed near the temples, and two placed near the sides of the neck. Activating the front motors indicated *Go Forward*, activating the two right motors meant *Turn Right*, and so forth.

When a Viewer pressed a navigation button, a command was sent to the Streamer's laptop, activating the appropriate motors. Latency from button press to motor activation was less than 1 second. For our study, we utilized two Viewer conditions. The first condition used the HTC Vive VR headset, where the built-in head tracking allowed the user to change viewpoints while watching the 360° camera feed. The second condition used a flat laptop screen, and by click-and-dragging a mouse pointer, the omnidirectional video feed could be manipulated in real-time.

Figure A.1: **Left**: The developed tactile belt (top) and hat (bottom). Each utilized a 4-tactor design in order to provide feedback to the Streamer. The motors were controlled by an Arduino Uno. **Right**: A user wearing the backpack rig, which consisted of a 360 camera for livestreaming, a vibro-tactile device for feedback, and a laptop.

User Study

We conducted a lab-based study at the University of Central Florida with IRB approval. The study aimed at garnering user feedback from Streamers, Viewers, and other third-party members who were not part of the interaction.

Study Design

We conducted a 2x2 mixed-design study; participants were split into 2 groups - one group consisted of the Streamers, in which the VibTac device selection was a within-subjects variable. The other group consisted of Viewers, in which the Viewing Mode was a within-subjects variable. Per group, condition order was counter-balanced. In addition, we conducted a short survey with non-participants. Our dependent variables aimed at measuring the perceived value of this kind of interpersonal telepresence, effectiveness of VibTac devices as a primary communication channel, and perception of the Viewer-Streamer relationship.

Subjects and Apparatus

We recruited eight participants to act as a Viewer. These participants consisted of 4 males and 4 females, and their age ranged between 18 and 26 (M = 21.8; SD = 3.20). On a 5-point Likert scale where 1 means Never and 5 means Always, the participants very rarely watched 360° videos (M = 1.75; SD = 0.46). They again very rarely used VR (M = 1.89; SD = 0.64). They very rarely drove robots (M = 1.89; SD = 0.64), and they rarely played first-person video games (M = 2.25; SD = 1.04).

We recruited eight participants to act as a Streamer. These participants consisted of 6 males and 2 females, and their age ranged between 18 and 38 (M = 24.4; SD = 6.30). Only 1 of our participants had experience with a wearable device (such as an Apple Watch or Fitbit). On a 5-point Likert scale where 1 means Never and 5 means Always, the participants exercised somewhat often (M = 3.5; SD = 0.76). They sometimes felt phantom vibrations regarding their cell phone (M = 2.88, SD = 0.83).

We surveyed an additional 30 university students (17 female) to understand their thoughts regarding the concept of interpersonal telepresence. All of our participants were recruited from the University of Central Florida.

Procedure

Viewers and Streamers

Our recruited participants reported to our lab, where they reviewed an informed consent form. After agreeing to participate, they were given a demographics survey, followed by a brief description of the telepresence setup. We explained how the VibTac interfaces worked, and how the controls were akin to a computer game (WASD keys). After introducing the hardware, our participants moved to their assigned position.

At any given time, a participant was accompanied by an experimenter, and we did not have interaction between participants. During Streamer trials, an experimenter provided the navigational commands. During Viewer trials, an experimenter served as the Streamer. As such, each participant only interacted with authors.

The Streamer was guided through a building on our university campus. The building layout consists of multiple hallways and corridors, which allowed us to take advantage of all VibTac commands. When controlling a Streamer participant, we held down the commands such that the vibration effect was constant. The *Stop Moving* command was issued when no vibrations were activated. We did this to ensure our Streamers had ample interaction time with the VibTac interfaces.

The navigational route was randomized and kept secret from the Streamer. As such, they had to rely on the VibTac devices to complete the trial; Streamers were asked to walk at a comfortable pace, and not run. After a Streamer completed a route, they changed VibTac devices and was then navigated through a different route. After a Viewer participant completed a route, they changed viewing modes and then performed navigation through a different route. In total, a Streamer participant was navigated twice; similarly, a Viewer participant conducted navigation twice. After each condition, a questionnaire was administered to elicit feedback regarding that particular condition. After both trials were complete, a final questionnaire was administered to garner feedback in terms of preference. Between participants, we cleaned the HMD and hat instruments with rubbing alcohol to ensure sanitation. The study took approximately 30 minutes to complete, and participants were paid $5 for their time. This study was conducted prior to the COVID-19 pandemic.

Survey Respondents

Survey respondents were approached in public and asked if they would like to provide their comments regarding interpersonal telepresence. One author wore the backpack rig, and we described the concept. Feedback was received in the form of a questionnaire and verbal communication. Respondents were not compensated.

Soliciting User Feedback

We administered quantitative measures to help understand user feedback, in terms of enjoyment and comfort, as well as social and economic aspects; see Table A.1 for a list of questions asked to our participants and respondents. All closed-ended questions were given on a 7-pt Likert scale where 1 meant "Strongly Disagree" and 7 meant "Strongly Agree." Open-ended questions were given as free-response prompts.

For the Viewer participants, we also asked simple questions to see if there was a preference in Viewing Mode, but we were more interested in their thoughts regarding the concept of controlling

Table A.1: Questions administered to participants. The Group column indicates which set of participants received the question: V = Viewers; S = Streamers; R = Survey Respondents. Asterisks (*) denote questions were given for multiple conditions.

Question	Group
I felt comfortable knowing I was controlling another human.	V
I liked having control over the Streamer.	V
I trusted the Streamer to go where I wanted them to go.	V
I would feel comfortable navigating the Streamer into a private place, such as a restroom.	V
I felt comfortable knowing I was being controlled by another human.	S
I liked being controlled by the Viewer.	S
I trusted the Viewer to give me good directions.	S
I would feel comfortable being navigated into a private place, such as a restroom.	S
It was easy to understand the Viewer's directions.	*S
I could easily feel the *Move Forward* command.	*S
I could easily feel the *Move Backward* command.	*S
I could easily feel the *Turn Left* command.	*S
I could easily feel the *Turn Right* command.	*S
Assuming the Streamer did a good job, I would consider paying for a live virtual tour.	V, R
How many dollars (USD) would you pay for a live virtual tour?	V, R
Assuming I would be adequately compensated, I would consider being a Streamer as a job.	S, R
How many dollars (USD) would you expect to receive, to be a Streamer?	S, R
I would trust the Streamer to do the right thing.	R
I would trust the Viewer to do the right thing.	R
I think it would be fun to be a Streamer.	R
I think it would be fun to be a Viewer.	R
I would be comfortable knowing a Streamer is near me in a public place.	R
I would be comfortable knowing a Streamer is near me in a private place.	R
What did you like about the system?	V, S
What did you dislike about the system?	V, S
What changes would you make to improve the overall experience?	V, S
What scenarios would you use this system for?	V, S, R

another human. For our Streamer participants, we were also interested in understanding VibTac device preference. We asked questions to understand how well the devices conveyed instructions and how comfortable they were.

Data Analysis Approach

As we did not have enough power to run statistical tests within the Streamer and Viewer groups, we report our results using descriptive statistics. To determine the overall effectiveness of each VibTac device, we averaged the 4 directional command responses into an index, per device. For the responses garnered from the final questionnaire, we report descriptive statistics and also discuss the positive / negative aspects as indicated by our users.

Results

In this section we report the results of our study, split by the different roles. We begin with the Viewer participants, followed by the Streamer participants, and finish with the survey respondents.

Viewer Considerations

Overall, our Viewers liked using the system and they had no major discomforts with either Viewing Mode. Six participants were most comfortable with the flat screen, while only two were more comfortable with the HMD. However, five indicated that they preferred the HMD to the flat screen.

Our Viewer participants felt comfortable controlling the Streamer (M = 5.5, SD = 2.1) and typically liked the interaction (M = 5.4, SD = 2.1). Only one participant responded negatively, as they were afraid that they would cause harm to the Streamer: *"Controlling people scared me half to death!*

The building is new to me and I was extremely scared I was going to run the Streamer off a ledge." This participant was the only one to express consideration for the Streamer.

Generally, our participants had great trust that the Streamer would follow the commands accurately (M = 6.6, SD = 0.5). However, the participants were very apprehensive about the idea of sending a Streamer into a private place; see Figure A.4, with only one participant feeling comfortable regarding that situation. Considering the concept as a whole, participants responded that they would be likely to hire a Streamer for a tour (M = 5.5, SD = 1.6). The range of value was very broad, however; the smallest sum of money participants were willing to pay was $5, and the most was $25 (see Figure A.3).

Our open-ended questions revealed that the most negative features of our telepresence setup was the latency; half of our participants complained about lag. From button press to motor activation, approximately 1 second elapsed; but, there was additional overhead time from motor activation to Streamer response, followed by another second of video lag. As such, there were many times where the Streamer walked too far down a hallway, or missed a turn. In addition to better visuals, some of our participants indicated that they would prefer to hear and speak to the Streamer. Usage scenarios for interpersonal telepresence included campus orientation for new or prospective students, playing augmented reality games, virtual tourism, and remote shopping.

Streamer Considerations

Our participants generally had no issue with either VibTac device. Comfort was comparable between the Belt (M = 5.4, SD = 1.3) and the Hat (M = 5.4, SD = 1.2); likewise, users felt only slightly embarrassed when wearing either device (Belt: M = 2.3, SD = 1.5; Hat: M = 2.6, SD = 1.8). Regarding the quality of VibTac commands, there was only a slight difference in the ability to detect them (Belt: M = 5.2, SD = 1.5; Hat: M = 5.6, SD = 1.0), and regardless of device, par-

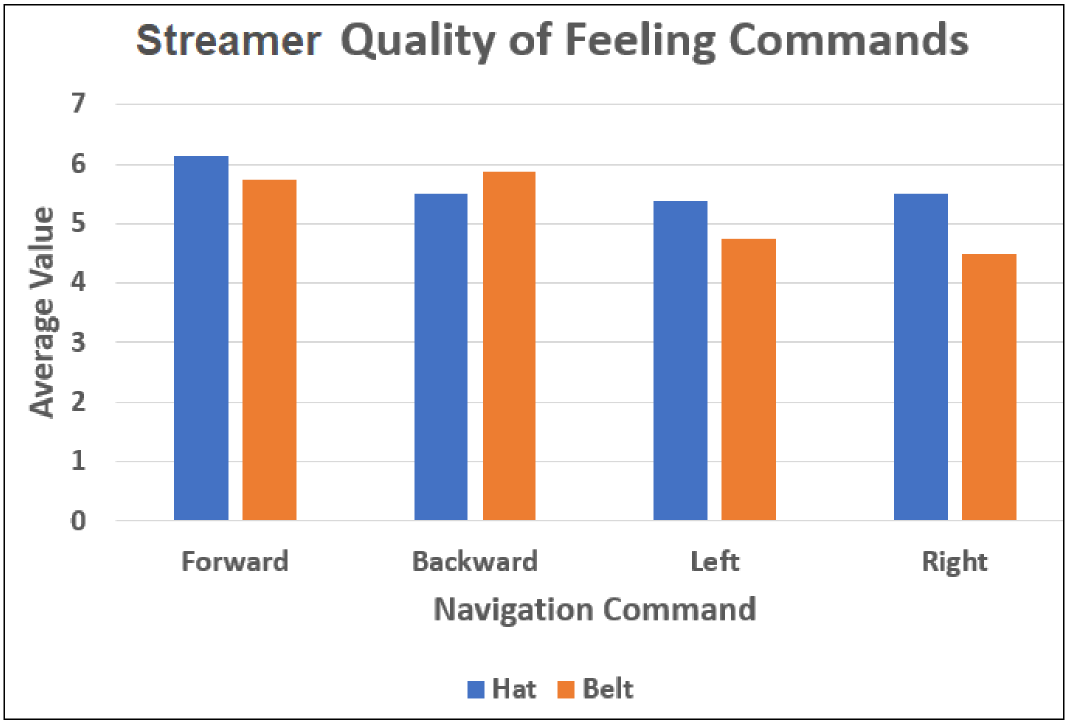


Figure A.2: The commands from both VibTac devices were perceived well, but the Hat device was slightly more conducive to feeling the vibrations.

ticipants felt that they understood where they were being directed to go (Belt: M = 5.3, SD = 1.7; Hat: M = 5.8, SD = 1.0). Our participants indicated that they generally felt the belt vibrations with ease, but the hat vibrations were more distinct (see Figure A.2). As such, device preference was not in favor of one over the other; three users preferred the belt, and five preferred the hat.

Participants indicated that they enjoyed the novelty of our system, and we suspect that this novelty may have impacted participant Likert-scale responses. However, open-ended questions did help to reveal weaknesses of our devices. Five of our participants indicated that while the vibrations were enough to complete the walking task, they could have been better. For example, the commands from the belt were perceived as having different strengths. One user informed us that they needed to put their hands on the belt in order to recognize the vibrations. On the other hand, while perception

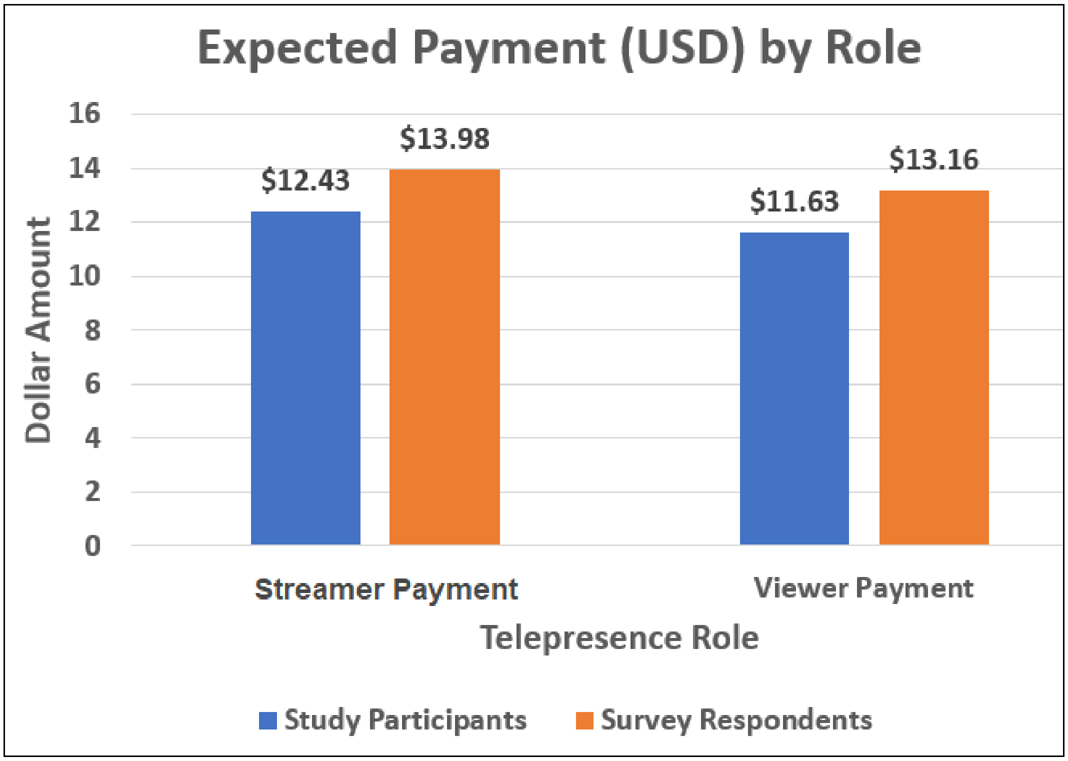


Figure A.3: Expected payment (USD) per telepresence role. Participants and survey respondents both indicated that the Streamer would garner more compensation than Viewers should pay.

towards the hat was comparable to the belt, users indicated that the vibrations caused a tickling sensation, or were more distracting. This is due to the motors being close to the users' forehead and ears, so the vibrations were more distinct and audible. One user suggested that they were able to hear the motors vibrating near their ears, but this helped them to determine which direction was requested.

Survey Respondents

Our survey respondents, who were third-party bystanders, had mixed feelings regarding this concept. While they had general trust in a Streamer (M = 5.3, SD = 1.1), they were much less trusting

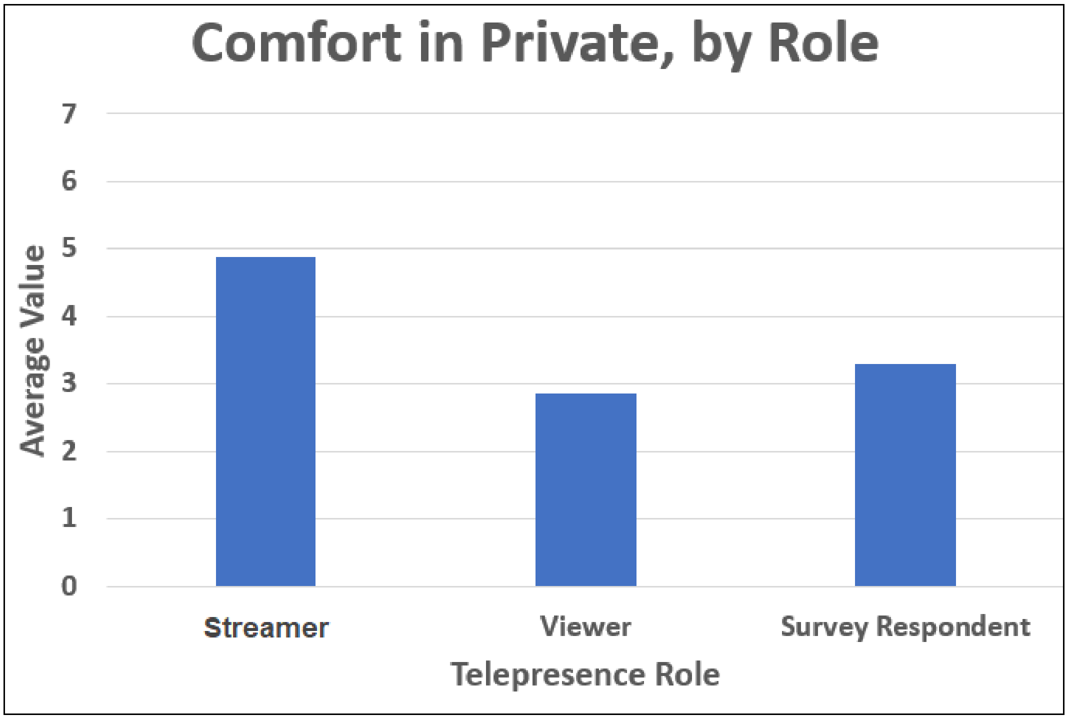


Figure A.4: Average response to comfort of telepresence in a private setting. Interestingly, the Streamers had little issue with the concept, whereas the Viewers and survey respondents were much less enthused.

in the Viewer (M = 4.4, SD = 1.7). A Wilcoxon Signed Rank Test revealed a significant difference in who they would trust ($Z = -2.964, p < .005$). This is because they are able to see and perhaps identify the local user, but they have no indication as to who is on the other end of the interaction. As such, they were somewhat comfortable with the idea of a Streamer being nearby in a public place (M = 4.9, SD = 1.3), but were far less enthused with the idea of a Streamer being collocated in a private area (M = 3.3, SD = 1.6). A Wilcoxon Signed Rank Test reveals a significant difference in the comfort of these scenarios ($Z = -4.168, p < .001$). Although they did not see a telepresence scenario taking place, the respondents thought it would be somewhat fun to be a viewer (M = 5.8, SD = 1.4), and less fun to be a Streamer (M = 5.2, SD = 1.6). A Wilcoxon Signed Rank Test revealed a significant difference between these sentiments ($Z = -2.056, p < .05$). Interestingly, however, they responded with slight interest in being a Streamer for payment (M = 5.0, SD = 1.8), and were less interested in paying for a live tour (M = 4.4, SD = 1.6). A Wilcoxon Signed Rank Test revealed a significant difference of interest for these roles ($Z = -2.024, p < .05$). The range of expected dollars-per-hour payment for being a Streamer was broad; the smallest rate was $7, and the largest was $30 (M = 13.98, SD = 5.80). The dollars-per-hour range for being a viewer was from $4 to $20 (M = 13.15, SD = 7.15). Thus, the respondents felt that a Streamer would warrant more compensation than a viewer would need to pay. See Figure A.3 for illustration. The respondents had many scenarios in mind for interpersonal telepresence, including tele-health physical therapy, remote exploration of live events like concerts, housing and campus tours, vlogging, tourism, military missions where talking is prohibited, and giving directions.

Discussion

In this section, we unpack the findings which emerged from our study, to address our research questions. We begin with how all parties view each other; next we discuss the perceived value of the interaction; third we describe the potential scenarios pondered by our users; and last we discuss VibTac device efficiency.

RQ1: Towards Better Telepresence VibTac Devices

Although our Streamer participants responded (with bias) in favor of the VibTac devices, we learned from them that there is room for improvement. First, there are benefits and drawbacks for each device. The belt was not as conducive as the hat in terms of interpreting the commands. McDaniel et al. found users to easily localize the vibrations emitted from their 7-tactor belt [155], but our participants had a measure of uncertainty in simply feeling the vibrations. Although ours and their tactors each vibrated on 3V, theirs were felt more strongly. While we cannot confirm, we believe that this difference stems from the distinct procedures between our studies. While their participants were standing still, ours were engaged in a walking task. As such, our participants had more distractions as they navigated their route. To overcome this problem, we plan on reiterating our belt with the use of motors which can withstand greater voltage ratings.

The hat allowed our participants to recognize the commands more strongly than the belt, but it was not perceived to be perfect. First, the constant vibration of the motors against the cranium caused a tickling sensation to some, a result also found by Kaul and Rohs with HapticHead [108]. They suggest to decrease the frequency in which the motors activate, and we believe that would have caused more comfort with our users. We purposefully kept the vibrations constant during interaction to elicit user responses, but in a next iteration we would transition hat vibrations towards more gestural commands.

Although neither of our VibTac devices were "perfect" according to our participants, they were still able to convey commands properly and discretely. We do not anticipate for one VibTac wearable to emerge as an optimal device, and believe that it is more appropriate to provide a range of options for a Streamer to use.

RQ2: Third-Party More Trusting of Streamer than of Viewer

Our survey respondents were interestingly more trusting of a Streamer than of a remote Viewer. Although the Streamer will be the individual who is wearing the hardware, respondents see remote Viewers as the potential wrong-doers; they are the ones consuming the video stream and potentially making recordings. This is exemplified with the general discomfort of the idea of being collocated with a Streamer in private. Still, respondents were not particularly thrilled with the idea of being collocated with a Streamer in public. Prior research has found a negative attitude towards live streaming, especially without consent [140]; but in countries where public video recording is protected, e.g. in the United States, there is no real obligation to inform people that they might be in a stream [124, 239]. With technological advances, some researchers have begun asking questions regarding streaming ethics [52]; future research should target interpersonal telepresence to help understand how Streamers and their remote Viewers are perceived in the scope of ethics and legality.

Imbalance of Streamer/Viewer Payment

While monetary exchange is not necessary for friends and family to use telepresence, it is implied that it would help to bring about more general use. Our survey respondents indicated an imbalance regarding monetary payment and compensation between the viewer and Streamer, in that they would expect a Streamer to receive more money than they would be willing to pay. This is an issue that would prevent a "physical body marketplace," as conceived by Misawa and Rekimoto [162], from becoming reality, at least for interactions that involve 1 Streamer and 1 viewer. If this is to become a mainstream interaction style, it is clear that there needs to be more incentive to participate, in both ends, as suggested by prior work [19, 63]. Future research can help to identify avenues which will provide these incentives; as an example, some researchers are focusing on ways to increase immersion through multisensory stimulation, including touch [119, 188] and smell [17, 172], which could lead to a more enjoyable experience and thus a stronger desire to participate. Additionally, it may be the case where a 1-to-1 interaction might not be suitable for interpersonal telepresence. Instead, by adopting a 1-to-many paradigm, where a single Streamer could give tours to multiple viewers simultaneously, the cost per viewer could decrease. Although there are additional challenges to be met in this type of interaction, it is possible to achieve. For example, the Faroe Islands remote tourism experience gives control to a single viewer at a time [3], but as such, control time is limited. We look forward to future studies and design ideas regarding how to improve upon the Streamer-viewer experience, to make it desirable by all parties.

RQ3: Expected Interaction Scenarios

Our participants and survey respondents were able to conceptualize a broad range of scenarios in which interpersonal telepresence may prosper, including more personal cases such as physical therapy or playing augmented reality games with a friend as a Streamer. More intimate scenarios such as giving directions and physical therapy can be found in prior research [32, 203], and popular social media sites such as FaceBook, YouTube, and Twitch.tv provide platforms for larger audience engagement such as vlogging of travels and activities [9]. Popular live streams (especially those for games, found on Twitch) typically have a specific goal or direction, but some streamers do poll their viewers to provide more personalized content. Viewers watch these streams for multiple reasons [79, 233], including to live vicariously through the experiences of another person [10], but with interpersonal telepresence, Viewers have the opportunity to engage in an even more personal experience. We suggest that the creation of a platform specific to one-on-one telepresence would help create more engaging experiences which are currently unavailable.

VibTac Devices Not Optimal for Primary Communication Mode

The novelty of VibTac-based navigation gave rise to positive feedback regarding our belt and hat. We did hear some suggestions for improving the devices, including a way to make them more inconspicuous (e.g. replacing the wires with a Bluetooth connection); but, our short-term study did reveal some disinterest with solely relying on VibTac as a primary mode of navigation. Participants on both sides of the interaction expressed a desire for audible communication with their interaction partner. Additionally, our users were on the cusp of becoming annoyed with constant vibrations. As such, we would recommend telepresence designers to consider adding VibTac as an auxiliary mode of communication, as well as exploring additional modes not studied here. As wearable VibTac devices have been broadly researched with positive findings [108, 155–157, 174], there is opportunity to let Streamer users pick their own as an option.

Limitations and Future Work

Although our interaction prototype was met with positive feedback, we did also find apprehensions. As such, there is still much work to be done on many fronts. Our results are from a test where the participants and authors met before conducting the experiment. We do not assume our results to generalize to situations where both parties are absolute strangers. As such, field tests in real-world scenarios are needed to further our understanding; but, ethical considerations must be made to ensure the rights of all parties (viewers, streamers, and third-party) are not infringed upon. Regarding Streamer control, it remains to be seen how users react to extended lengths of interaction exposure. Longitudinal studies should also be conducted to help identify problems which arise over time. Further, our study did not utilize pre-validated instruments, and instead offer insight through custom measures. In our future work, we will identify proper instruments to measure perception towards VibTac devices.

Our prototype was relatively bulky, and iterative ergonomic enhancements can be made to ensure Streamer comfort. Additionally, although we chose to study two of the most prevalent haptic devices found in previous literature, there are others which need to be thoroughly studied. Lastly, legal and ethical considerations must be investigated. In spirit, our study assumed that the interaction would be performed as intended. However, it is possible for a Streamer to be directed to questionable locations, or perform questionable actions (such as commit a crime). Though we do not want to instigate such a scenario, the telepresence community must ask what the proper response would be. We hope that our work contributes to the discussion for this type of interaction which is rapidly approaching the mainstream.

Conclusion

We have presented our prototype in which a person can direct another through vibro-tactile commands. Our results indicate that there is a real opportunity to provide novel and desirable interaction, but more work is needed to make that a widespread reality. We envision this type of system becoming a prominent avenue in social media, allowing Viewers to have a more direct form of experience through the use of a Streamer. We look forward to seeing how this type of technology will engage users of all backgrounds, in order to explore the world around them.

APPENDIX B: SAP 2019 QUESTIONNAIRES

Demographics Survey Participant: _______

Age: _______________

Major: __________________________________

Gender:

Male Female

Academic Standing:

Freshman Sophomore Junior Senior Graduate N/A

How often do you play First Person Shooter video games?

Never Rarely Sometimes Frequently Always

How often do you watch 360-degree videos?

Never Rarely Sometimes Frequently Always

How often do you use Virtual Reality (e.g. HTC Vive, Google Cardboard, Oculus Rift, etc.)?

Never Rarely Sometimes Frequently Always

In-Between Survey   Participant: _______

The camera placement allowed me to see everything I needed to see

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

Nothing blocked my view to the point where I became disoriented

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

My field of view was clear, so I could perform the task

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

The height of the camera felt natural to me

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

The camera height let me view the environment with ease

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

I liked the height of the camera placement

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

All things considered, please give a score to that camera placement

1	2	3	4	5	6	7
Terrible			Neutral			Excellent

*The camera placement **allowed me to see what I wanted** to see*

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

In-Between Survey Participant: _______

Was the view shorter, taller, or equal to your natural height?

Shorter Natural Taller

What did you like about that camera placement?

What did you dislike about that camera placement?

APPENDIX C: CO-DESIGN "BAG OF STUFF" ARTIFACTS

Table C.1: Camera artifacts within our Bag of Stuff. Each entry lists an image, short item description, and (if applicable) an exemplar citation that used the device.

(a) Mobile Tablet (iPad Air) [250]

(b) Smartphone Camera (Android/iPhone) [87]

(c) Single-lens 360° Camera (360Fly)

(d) Action Camera (GoPro)

(e) Ear-mounted Camera (Looxcie) [202]

(f) Dual-lens 360° Camera (Ricoh Theta) [26]

(g) Multi-lens 360° camera (from [176])

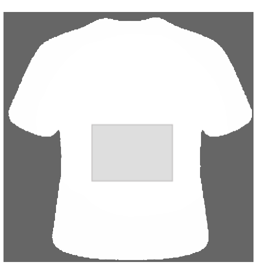
Table C.2: Harnesses and Mounts artifacts within our Bag of Stuff. Each entry lists an image, short item description, and (if applicable) an exemplar citation that used the device.

(a) Shoulder Strap [205]

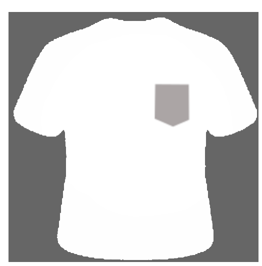
(b) Backpack [243]

(c) Gimbal Stabilizer [123]

(d) Telescopic Pole [243]

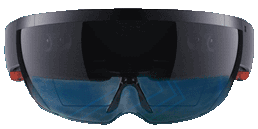


(e) Shirt with Window [250]



(f) Shirt with Pocket [14]

Table C.3: Feedback Accessories within our Bag of Stuff. Each entry lists an image, short item description, and (if applicable) an exemplar citation that used the device.



(a) AR Goggles (Microsoft Hololens) [184]

(b) AR Glasses (Google Glass) [26]

(c) Noise-Cancelling Headphones

(d) Wireless Earbuds (Apple AirPods Pro) [202]

(e) Strip of LED Lights [149]

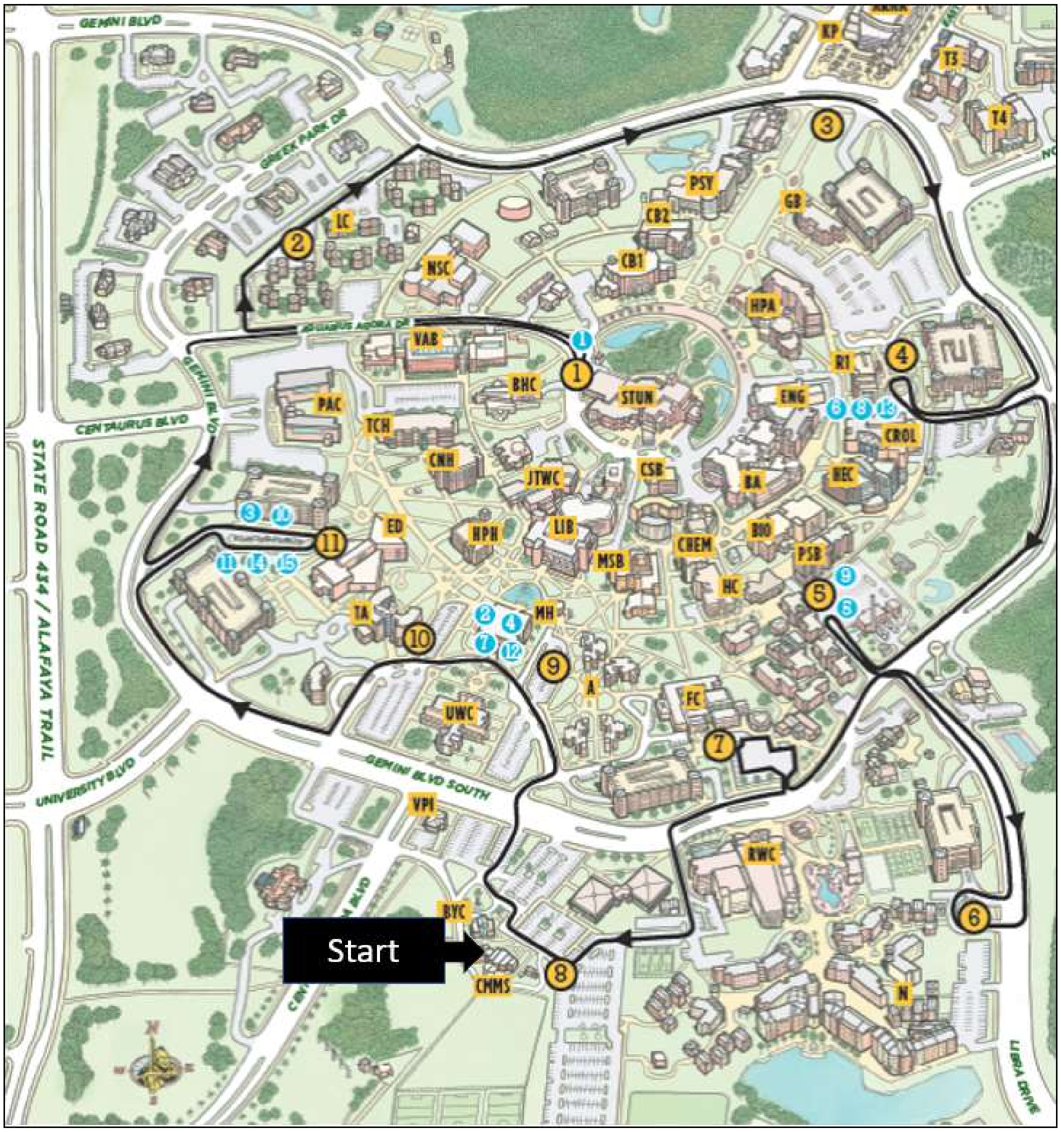
(f) Ma Petite Cherie (from [164])

(a) Robot Arms (Robotis) [253]

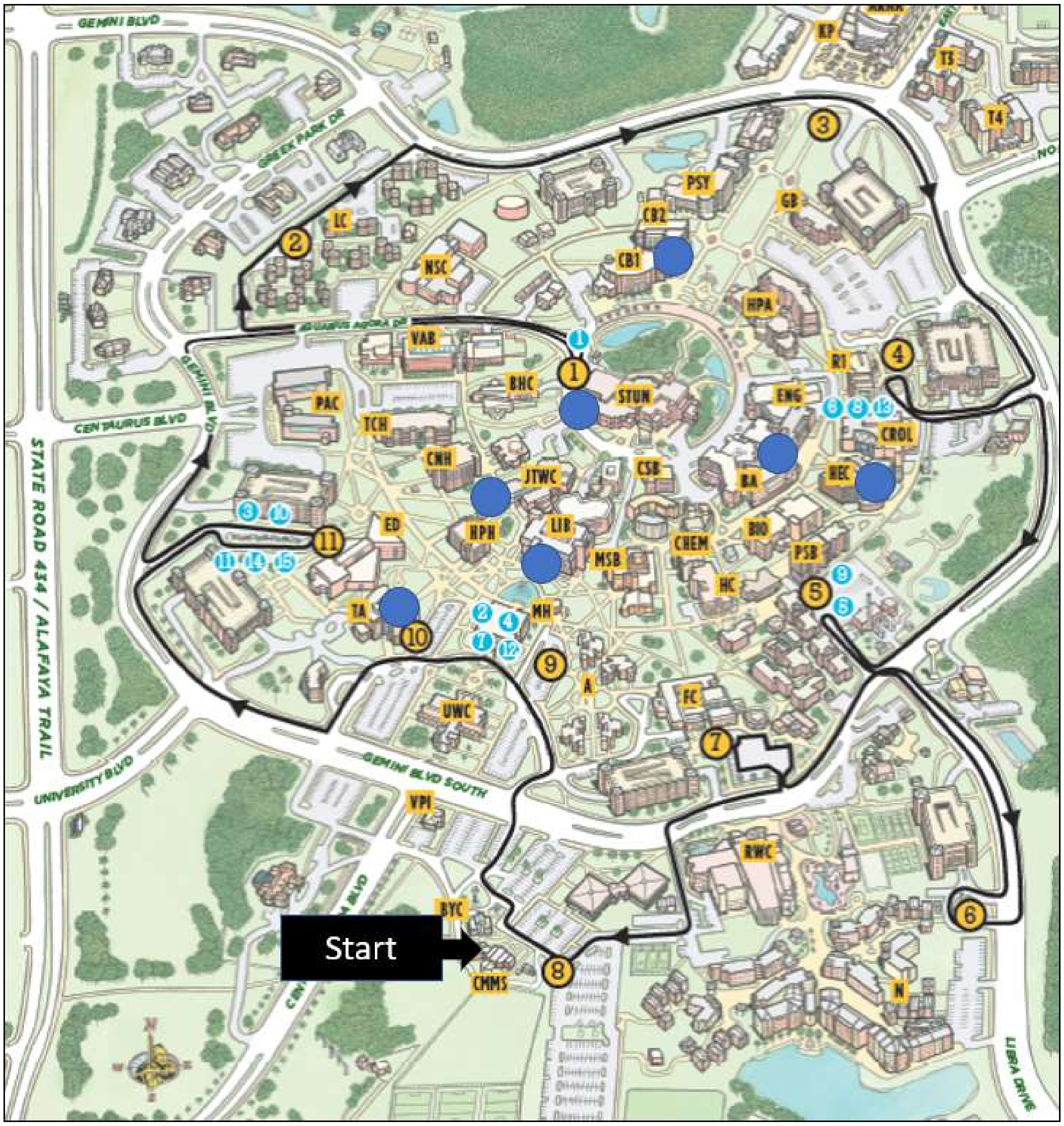
(b) Wearable Speakers (Bluedio) [153]

(c) TEROOS (from [106])

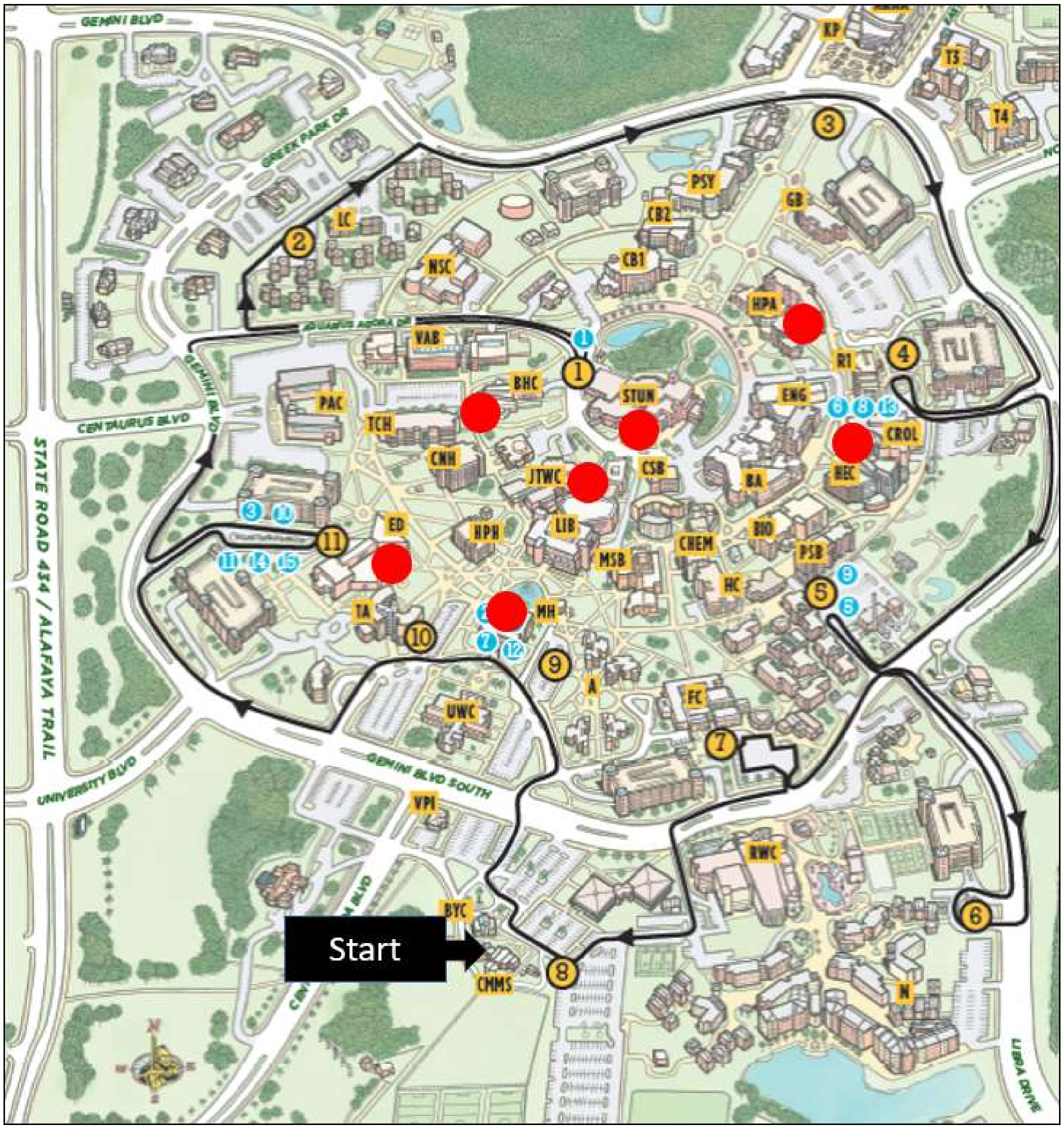
APPENDIX D: FIELD STUDY INSTRUMENTS



BA	Business Administration	**ED**	Education Complex	**PAC**	Performing Arts Center
BHC	Burnett Honors College	**ENG**	Engineering 1 and 2	**PSB**	Physical Science Bldg
BIO	Biology Building	**FC**	Ferrell Commons	**PSY**	Psychology
BYC	Barbara Ying	**HC**	Health Center	**R1**	Research 1 Bldg
CB1	Classroom Building 1	**HEC**	Harris Engineering Bldg	**STUN**	Student Union
CB2	Classroom Building 2	**HPA**	Health & Public Affairs	**TA**	Teaching Academy
CHEM	Chemistry Building	**HPH**	Howard Phillips Hall	**TCH**	Trevor Colbourn Hall
CMMS	Barbara Ying	**JTWC**	John T. Washington Ctr	**UWC**	University Writing Ctr
CNH	Colbourn Hall	**LIB**	Library	**VAB**	Visual Arts Building
CROL	CREOL Building	**MH**	Millican Hall		
CSB	College of Sciences	**MSB**	Math Science Building		



BA Business Administration
BHC Burnett Honors College
BIO Biology Building
BYC Barbara Ying
CB1 Classroom Building 1
CB2 Classroom Building 2
CHEM Chemistry Building
CMMS Barbara Ying
CNH Colbourn Hall
CROL CREOL Building
CSB College of Sciences
ED Education Complex
ENG Engineering 1 and 2
FC Ferrell Commons
HC Health Center
HEC Harris Engineering Bldg
HPA Health & Public Affairs
HPH Howard Phillips Hall
JTWC John T. Washington Ctr
LIB Library
MH Millican Hall
MSB Math Science Building
PAC Performing Arts Center
PSB Physical Science Bldg
PSY Psychology
R1 Research 1 Bldg
STUN Student Union
TA Teaching Academy
TCH Trevor Colbourn Hall
UWC University Writing Ctr
VAB Visual Arts Building



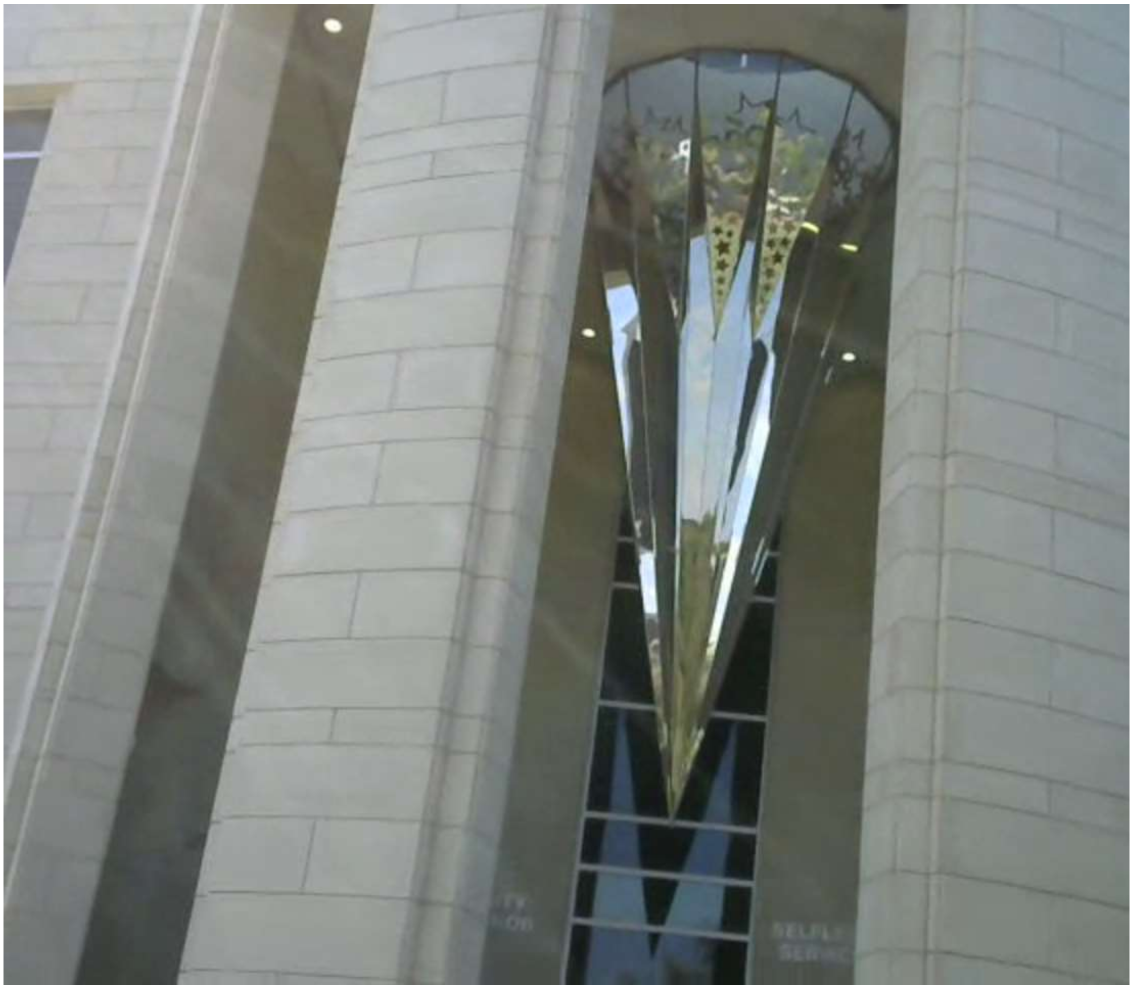
BA	Business Administration	**ED**	Education Complex	**PAC**	Performing Arts Center
BHC	Burnett Honors College	**ENG**	Engineering 1 and 2	**PSB**	Physical Science Bldg
BIO	Biology Building	**FC**	Ferrell Commons	**PSY**	Psychology
BYC	Barbara Ying	**HC**	Health Center	**R1**	Research 1 Bldg
CB1	Classroom Building 1	**HEC**	Harris Engineering Bldg	**STUN**	Student Union
CB2	Classroom Building 2	**HPA**	Health & Public Affairs	**TA**	Teaching Academy
CHEM	Chemistry Building	**HPH**	Howard Phillips Hall	**TCH**	Trevor Colbourn Hall
CMMS	Barbara Ying	**JTWC**	John T. Washington Ctr	**UWC**	University Writing Ctr
CNH	Colbourn Hall	**LIB**	Library	**VAB**	Visual Arts Building
CROL	CREOL Building	**MH**	Millican Hall		
CSB	College of Sciences	**MSB**	Math Science Building		


Engineering 2 Building (ENG)

Harris Engineering Center (HEC)


John T. Washington Center (JTWC)

Library (LIB)



Classroom 2 / ROTC Building (CB2)


Student Union (STUN)


Teaching Academy (TA)

Burnett Honors College (BHC)

CREOL Building (CROL)


Education Complex (ED)


Millican Hall (MH)

Health and Public Affairs (HPA)

John T. Washington Center (JTWC)


Student Union (STUN)

Streamer Survey Participant: _______

Age: ___________________ Role: _______________

Gender: __________________________

Relationship with your partner: ____

I am the life of the party

1	2	3	4	5
Strongly Disagree		Neutral		Strongly Agree

I sympathize with others' feelings

1	2	3	4	5
Strongly Disagree		Neutral		Strongly Agree

I get chores done right away

1	2	3	4	5
Strongly Disagree		Neutral		Strongly Agree

I have frequent mood swings

1	2	3	4	5
Strongly Disagree		Neutral		Strongly Agree

I have a vivid imagination

1	2	3	4	5
Strongly Disagree		Neutral		Strongly Agree

I don't talk a lot

1	2	3	4	5
Strongly Disagree		Neutral		Strongly Agree

I am not interested in other peoples' problems

1	2	3	4	5
Strongly Disagree		Neutral		Strongly Agree

Streamer Survey Participant: _______

I often forget to put things back in their proper place

1	2	3	4	5
Strongly Disagree		Neutral		Strongly Agree

I am relaxed most of the time

1	2	3	4	5
Strongly Disagree		Neutral		Strongly Agree

I am not interested in abstract ideas

1	2	3	4	5
Strongly Disagree		Neutral		Strongly Agree

I talk to a lot of different people at parties

1	2	3	4	5
Strongly Disagree		Neutral		Strongly Agree

I feel others' emotions

1	2	3	4	5
Strongly Disagree		Neutral		Strongly Agree

I like order

1	2	3	4	5
Strongly Disagree		Neutral		Strongly Agree

I get upset easily

1	2	3	4	5
Strongly Disagree		Neutral		Strongly Agree

I have difficulty understanding abstract ideas

1	2	3	4	5
Strongly Disagree		Neutral		Strongly Agree

I keep in the background

1	2	3	4	5
Strongly Disagree		Neutral		Strongly Agree

Streamer Survey Participant: _______

I am not really interested in others

1	2	3	4	5
Strongly Disagree		Neutral		Strongly Agree

I make a mess of things

1	2	3	4	5
Strongly Disagree		Neutral		Strongly Agree

I seldom feel blue

1	2	3	4	5
Strongly Disagree		Neutral		Strongly Agree

I do not have a good imagination

1	2	3	4	5
Strongly Disagree		Neutral		Strongly Agree

I use virtual reality technology…

1	2	3	4	5
Never	Rarely	Sometimes	Very Often	Always

I perform live-streaming…

1	2	3	4	5
Never	Rarely	Sometimes	Very Often	Always

I watch live-streams…

1	2	3	4	5
Never	Rarely	Sometimes	Very Often	Always

Streamer Survey Participant: _______

For each of the following, please mark exactly one response.

I felt concerned about my style of doing things

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

I felt concerned about the way I presented myself

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

I felt self-conscious about the way I looked

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

I felt worried about making a good impression

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

I felt concerned about what others thought of me

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

I felt aware of my appearance

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

I felt self-conscious using the system in public

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

I felt like people were looking at me

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

I felt like people gave me strange looks

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

I felt like people stared at me

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

I felt like people noticed me

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

Streamer Survey Participant: _______

I felt stupid using the system

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

I felt awkward using the system

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

I felt like I received unwanted attention

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

I often felt as if I was all alone

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

I think the other individual often felt alone

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

I hardly noticed another individual

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

The other individual didn't notice me

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

I was often aware of others in the environment

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

Others were often aware of me

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

I sometimes pretended to pay attention to the other individual

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

The other individual sometimes pretended to pay attention to me

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

Streamer Survey Participant: _______

The other individual paid close attention to me						
1 Strongly Disagree	2	3	4 Neutral	5	6	7 Strongly Agree

I paid close attention to the other individual						
1 Strongly Disagree	2	3	4 Neutral	5	6	7 Strongly Agree

My partner was easily distracted when other things were going on around us						
1 Strongly Disagree	2	3	4 Neutral	5	6	7 Strongly Agree

I was easily distracted when other things were going on around me						
1 Strongly Disagree	2	3	4 Neutral	5	6	7 Strongly Agree

The other individual tended to ignore me						
1 Strongly Disagree	2	3	4 Neutral	5	6	7 Strongly Agree

I tended to ignore the other individual						
1 Strongly Disagree	2	3	4 Neutral	5	6	7 Strongly Agree

When I was happy, the other was happy						
1 Strongly Disagree	2	3	4 Neutral	5	6	7 Strongly Agree

When the other was happy, I was happy						
1 Strongly Disagree	2	3	4 Neutral	5	6	7 Strongly Agree

The other individual was influenced by my moods						
1 Strongly Disagree	2	3	4 Neutral	5	6	7 Strongly Agree

I was influenced by my partner’s moods						
1 Strongly Disagree	2	3	4 Neutral	5	6	7 Strongly Agree

The other’s mood did NOT affect my mood/emotional-state						
1 Strongly Disagree	2	3	4 Neutral	5	6	7 Strongly Agree

Streamer Survey Participant: _______

My mood did NOT affect the other’s mood/emotional-state

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

My opinions were clear to the other

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

The opinions of the other were clear

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

My thoughts were clear to the other

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

The other individual’s thoughts were clear to me

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

The other understood what I meant

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

I understood what the other meant

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

My actions were dependent on the other’s actions

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

The other’s actions were dependent on my actions

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

My behavior was in direct response to the other’s behavior

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

The behavior of the other was in direct response to my behavior

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

Streamer Survey Participant: _______

What the other did affected what I did

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

What I did affected what the other did

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

My partner did not help me very much

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

I did not help the other very much

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

My partner worked with me to complete the task

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

I worked with the other individual to complete the task

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

The other could not act without me

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

I could not act without the other

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

Assuming I have access to the system, I intend to use it

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

Given that I have access to the system, I predict that I would use it

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

Using the system improves my performance

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

Streamer Survey Participant: _______

Using the system increases my productivity

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

Using the system enhances my effectiveness

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

I find the system to be useful

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

My interaction with the system is clear and understandable

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

Interacting with the system does not require a lot of my mental effort

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

I find the system to be easy to use

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

I find it easy to get the system to do what I want to do

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

I felt connected with my partner when using the system

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

My relationship with my partner is enhanced when using the system

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

The system satisfies my need for social interaction

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

Use of this system is a good substitute for face-to-face interaction with my partner

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

Streamer Survey Participant: _______

Using the system deepens my relationship with my partner

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

Communicating with my partner using the system helps me tell how my partner is feeling

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

Communicating with my partner using the system helps me let my partner know how I am feeling

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

Communicating with my partner using the system helps me see how much my partner cares about me

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

I feel that contact with my using the system is engaging for my partner

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

I am excited about using the system with my partner

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

I have fun with my partner using the system

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

Communicating with my partner using the system helps me feel closer to my partner

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

When done communicating, I keet thinking about something my partner shared thru the system

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

Communicating with my partner using the system helps me feel more connected to my partner

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

Communicating with my partner using the system helps me provide my partner with social support

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

Streamer Survey Participant: _______

My partner makes me feel special in our contact using the system

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

I worry that my partner might learn something using the system that I want to keep secret

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

I worry about my privacy while my partner and I are using the system together

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

I worry that others may overhear or see something that my partner and I share using the system

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

I worry that I am violating my partner's privacy during our contact using the system

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

Viewer Survey Participant: _______

For each of the following, please mark exactly <u>one</u> response.

I hardly noticed another individual
1 2 3 4 5 6 7
Strongly Disagree Neutral Strongly Agree

The other individual didn’t notice me
1 2 3 4 5 6 7
Strongly Disagree Neutral Strongly Agree

I was often aware of others in the environment
1 2 3 4 5 6 7
Strongly Disagree Neutral Strongly Agree

Others were often aware of me
1 2 3 4 5 6 7
Strongly Disagree Neutral Strongly Agree

I think the other individual often felt alone
1 2 3 4 5 6 7
Strongly Disagree Neutral Strongly Agree

I often felt as if I was all alone
1 2 3 4 5 6 7
Strongly Disagree Neutral Strongly Agree

I sometimes pretended to pay attention to the other individual
1 2 3 4 5 6 7
Strongly Disagree Neutral Strongly Agree

The other individual sometimes pretended to pay attention to me
1 2 3 4 5 6 7
Strongly Disagree Neutral Strongly Agree

The other individual paid close attention to me
1 2 3 4 5 6 7
Strongly Disagree Neutral Strongly Agree

I paid close attention to the other individual
1 2 3 4 5 6 7
Strongly Disagree Neutral Strongly Agree

My partner was easily distracted when other things were going on around us
1 2 3 4 5 6 7
Strongly Disagree Neutral Strongly Agree

Viewer Survey   Participant: _______

I was easily distracted when other things were going on around me

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

The other individual tended to ignore me

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

I tended to ignore the other individual

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

When I was happy, the other was happy

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

When the other was happy, I was happy

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

The other individual was influenced by my moods

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

I was influenced by my partner’s moods

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

The other’s mood did NOT affect my mood/emotional-state

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

My mood did NOT affect the other’s mood/emotional-state

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

My opinions were clear to the other

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

The opinions of the other were clear

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

Viewer Survey Participant: _______

My thoughts were clear to the other

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

The other individual's thoughts were clear to me

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

The other understood what I meant

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

I understood what the other meant

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

My actions were dependent on the other's actions

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

The other's actions were dependent on my actions

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

My behavior was in direct response to the other's behavior

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

The behavior of the other was in direct response to my behavior

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

What the other did affected what I did

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

What I did affected what the other did

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

My partner did not help me very much

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

Viewer Survey Participant: _______

I did not help the other very much

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

My partner worked with me to complete the task

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

I worked with the other individual to complete the task

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

The other could not act without me

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

I could not act without the other

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

The view allowed me to see everything I needed to see

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

Nothing blocked my view to the point where I became disoriented

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

My field of view was clear, so I could perform the task

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

How much were you able to control events?

1	2	3	4	5	6	7
Not Much			Neutral			Very Much

How natural did your interactions with the environment seem?

1	2	3	4	5	6	7
Not Natural			Neutral			Very Natural

How completely were all of your senses engaged?

1	2	3	4	5	6	7
Not at All			Neutral			Completely

Viewer Survey Participant: _______

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

1	2	3	4	5	6	7
Not at All			Neutral			Completely

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

1	2	3	4	5	6	7
Not at All			Neutral			Completely

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

1	2	3	4	5	6	7
Not Natural			Neutral			Very Natural

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

1	2	3	4	5	6	7
Not Aware			Neutral			Very Aware

How aware were you of your display and control devices?

1	2	3	4	5	6	7
Not Aware			Neutral			Very Aware

How compelling was your sense of objects moving through space?

1	2	3	4	5	6	7
Not Compelling			Neutral			Very Compelling

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

1	2	3	4	5	6	7
Not Consistent			Neutral			Very Consistent

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

1	2	3	4	5	6	7
Not at All			Neutral			Completely

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

1	2	3	4	5	6	7
Not at All			Neutral			Completely

How well could you identify sounds?

1	2	3	4	5	6	7
Not Well			Neutral			Very Well

How well could you localize sounds?

1	2	3	4	5	6	7
Not Well			Neutral			Very Well

Viewer Survey Participant: _______

How closely were you able to examine objects?

1	2	3	4	5	6	7
Not at All			Neutral			Very Closely

How well could you examine objects from multiple viewpoints?

1	2	3	4	5	6	7
Not Well			Neutral			Very Well

To what degree did you feel confused or disoriented at the beginning of breaks or at the end of the experimental session?

1	2	3	4	5	6	7
Not Confused			Neutral			Very Confused

How involved were you in the experience?

1	2	3	4	5	6	7
Not Involved			Neutral			Very Involved

How much delay did you experience between your actions and expected outcomes?

1	2	3	4	5	6	7
Not Much Delay			Neutral			Much Delay

How quickly did you adjust to the experience?

1	2	3	4	5	6	7
Very Slowly			Neutral			Very Quickly

How much did the visual display quality interfere or distract you from performing assigned tasks or required activities?

1	2	3	4	5	6	7
Not Much			Neutral			Very Much

How much did the control devices interfere with the performance of assigned tasks or with other activities?

1	2	3	4	5	6	7
Not Much			Neutral			Very Much

How well could you concentrate on the assigned tasks or required activities rather than on the mechanisms used to perform those tasks or activities?

1	2	3	4	5	6	7
Not Well			Neutral			Very Well

Did you learn new techniques that enabled you to improve your performance?

1	2	3	4	5	6	7
Not At All			Neutral			Completely

Viewer Survey Participant: _______

Were you involved in the experimental task to the extent that you lost track of time?

1	2	3	4	5	6	7
Not Involved			Neutral			Very Involved

Assuming I have access to the system, I intend to use it

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

Given that I have access to the system, I predict that I would use it

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

Using the system improves my performance

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

Using the system increases my productivity

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

Using the system enhances my effectiveness

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

I find the system to be useful

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

My interaction with the system is clear and understandable

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

Interacting with the system does not require a lot of my mental effort

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

I find the system to be easy to use

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

I find it easy to get the system to do what I want to do

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

Viewer Survey Participant: _______

I felt connected with my partner when using the system

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

My relationship with my partner is enhanced when using the system

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

The system satisfies my need for social interaction

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

Use of this system is a good substitute for face-to-face interaction with my partner

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

Using the system deepens my relationship with my partner

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

Communicating with my partner using the system helps me tell how my partner is feeling

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

Communicating with my partner using the system helps me let my partner know how I am feeling

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

Communicating with my partner using the system helps me see how much my partner cares about me

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

I feel that contact with my using the system is engaging for my partner

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

I am excited about using the system with my partner

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

I have fun with my partner using the system

1	2	3	4	5	6	7
Strongly Disagree			Neutral			Strongly Agree

Viewer Survey Participant: _______

Communicating with my partner using the system helps me feel closer to my partner

1 2 3 4 5 6 7

Strongly Disagree Neutral Strongly Agree

When done communicating, I keet thinking about something my partner shared thru the system

1 2 3 4 5 6 7

Strongly Disagree Neutral Strongly Agree

Communicating with my partner using the system helps me feel more connected to my partner

1 2 3 4 5 6 7

Strongly Disagree Neutral Strongly Agree

Communicating with my partner using the system helps me provide my partner with social support

1 2 3 4 5 6 7

Strongly Disagree Neutral Strongly Agree

My partner makes me feel special in our contact using the system

1 2 3 4 5 6 7

Strongly Disagree Neutral Strongly Agree

I worry that my partner might learn something using the system that I want to keep secret

1 2 3 4 5 6 7

Strongly Disagree Neutral Strongly Agree

I worry about my privacy while my partner and I are using the system together

1 2 3 4 5 6 7

Strongly Disagree Neutral Strongly Agree

I worry that others may overhear or see something that my partner and I share using the system

1 2 3 4 5 6 7

Strongly Disagree Neutral Strongly Agree

I worry that I am violating my partner’s privacy during our contact using the system

1 2 3 4 5 6 7

Strongly Disagree Neutral Strongly Agree

APPENDIX E: IRB APPROVAL LETTERS

University of Central Florida Institutional Review Board
Office of Research & Commercialization
12201 Research Parkway, Suite 501
Orlando, Florida 32826-3246
Telephone: 407-823-2901 or 407-882-2276
www.research.ucf.edu/compliance/irb.html

Approval of Human Research

From: **UCF Institutional Review Board #1**
FWA00000351, IRB00001138

To: **Kevin Pfeil** and Co-PI: **Joseph J Laviola II**

Date: **October 16, 2018**

Dear Researcher:

On 10/16/2018 the IRB approved the following human participant research until 10/15/2019 inclusive:

Type of Review:	IRB Continuing Review Application Form Expedited Review
Project Title:	Human-human telepresence camera placement
Investigator:	Kevin Pfeil
IRB Number:	SBE-18-13836
Funding Agency:	
Grant Title:	
Research ID:	N/A

The scientific merit of the research was considered during the IRB review. The Continuing Review Application must be submitted 30 days prior to the expiration date for studies that were previously expedited, and 60 days prior to the expiration date for research that was previously reviewed at a convened meeting. Do not make changes to the study (i.e., protocol, methodology, consent form, personnel, site, etc.) before obtaining IRB approval. A Modification Form **cannot** be used to extend the approval period of a study. All forms may be completed and submitted online at https://iris.research.ucf.edu .

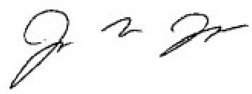
If continuing review approval is not granted before the expiration date of 10/15/2019, approval of this research expires on that date. When you have completed your research, please submit a Study Closure request in iRIS so that IRB records will be accurate.

Use of the approved, stamped consent document(s) is required. The new form supersedes all previous versions, which are now invalid for further use. Only approved investigators (or other approved key study personnel) may solicit consent for research participation. Participants or their representatives must receive a copy of the consent form(s).

All data, including signed consent forms if applicable, must be retained and secured per protocol for a minimum of five years (six if HIPAA applies) past the completion of this research. Any links to the identification of participants should be maintained and secured per protocol. Additional requirements may be imposed by your funding agency, your department, or other entities. Access to data is limited to authorized individuals listed as key study personnel.

In the conduct of this research, you are responsible to follow the requirements of the Investigator Manual.

This letter is signed by:



Signature applied by Jessica Jacques on 10/16/2018 10:00:19 AM EDT

Designated Reviewer

UNIVERSITY OF CENTRAL FLORIDA

Institutional Review Board
FWA00000351
IRB00001138, IRB00012110
Office of Research
12201 Research Parkway
Orlando, FL 32826-3246

APPROVAL

April 5, 2021

Dear Kevin Pfeil:

On 4/5/2021, the IRB reviewed the following submission:

Type of Review:	Initial Study
Title:	Human-to-Human Telepresence
Investigator:	Kevin Pfeil
IRB ID:	STUDY00002064
Funding:	None
Grant ID:	None
IND, IDE, or HDE:	None
Documents Reviewed:	• study3ApprovalForm_signed.pdf, Category: Faculty Research Approval; • BagOfStuff, Category: Other; • Big 5 personality inventory, Category: Other; • COVID_Safety_Plan_FieldStudy.pdf, Category: Other; • fieldStudy, Category: Consent Form; • FieldStudy_SemiStructuredInterview.docx, Category: Survey / Questionnaire; • PD_Session_Slides.pptx, Category: Other; • PD_StudyScript, Category: Other; • PDStudy, Category: Consent Form; • protocol.docx, Category: IRB Protocol; • Qualtrics Recruitment Form, Category: Survey / Questionnaire; • Qualtrics_Final_Survey.docx, Category: Survey / Questionnaire; • questionnaire_fieldStudy_streamer.docx, Category: Survey / Questionnaire; • questionnaire_fieldStudy_viewer.docx, Category: Survey / Questionnaire; • recruitment_fieldStudy.docx, Category: Recruitment Materials; • recruitment_PDStudy.docx, Category: Recruitment Materials;

The IRB approved the protocol on 4/5/2021.

In conducting this protocol, you are required to follow the requirements listed in the Investigator Manual (HRP-103), which can be found by navigating to the IRB Library within the IRB system. Guidance on submitting Modifications and a Continuing Review or Administrative Check-in are detailed in the manual. When you have completed your research, please submit a Study Closure request so that IRB records will be accurate.

If you have any questions, please contact the UCF IRB at 407-823-2901 or irb@ucf.edu. Please include your project title and IRB number in all correspondence with this office.

Sincerely,

Gillian Bernal
Designated Reviewer

University of Central Florida Institutional Review Board
Office of Research & Commercialization
12201 Research Parkway, Suite 501
Orlando, Florida 32826-3246
Telephone: 407-823-2901 or 407-882-2276
www.research.ucf.edu/compliance/irb.html

Approval of Human Research

From: **UCF Institutional Review Board #1**
FWA00000351, IRB00001138

To: **Kevin Pfeil** and Co-PIs **Joseph J LaViola II and Katelynn A Kapalo**

Date: **October 18, 2017**

Dear Researcher:

On 10/18/2017 the IRB approved the following human participant research until 10/17/2018 inclusive:

Type of Review:	UCF Initial Review Submission Form Expedited Review Category #6 & 7 This approval includes a Waiver of Written Documentation of Consent
Project Title:	Navigation and Direction Using Vibro-Tactile Interfaces
Investigator:	Kevin Pfeil
IRB Number:	SBE-17-13387
Funding Agency:	
Grant Title:	
Research ID:	N/A

The scientific merit of the research was considered during the IRB review. The Continuing Review Application must be submitted 30days prior to the expiration date for studies that were previously expedited, and 60 days prior to the expiration date for research that was previously reviewed at a convened meeting. Do not make changes to the study (i.e., protocol, methodology, consent form, personnel, site, etc.) before obtaining IRB approval. A Modification Form **cannot** be used to extend the approval period of a study. All forms may be completed and submitted online at https://iris.research.ucf.edu .

If continuing review approval is not granted before the expiration date of 10/17/2018, approval of this research expires on that date. When you have completed your research, please submit a Study Closure request in iRIS so that IRB records will be accurate.

Use of the approved, stamped consent document(s) is required. The new form supersedes all previous versions, which are now invalid for further use. Only approved investigators (or other approved key study personnel) may solicit consent for research participation. Participants or their representatives must receive a copy of the consent form(s).

All data, including signed consent forms if applicable, must be retained and secured per protocol for a minimum of five years (six if HIPAA applies) past the completion of this research. Any links to the identification of participants should be maintained and secured per protocol. Additional requirements may be imposed by your funding agency, your department, or other entities. Access to data is limited to authorized individuals listed as key study personnel.

In the conduct of this research, you are responsible to follow the requirements of the Investigator Manual.

On behalf of Sophia Dziegielewski, Ph.D., L.C.S.W., UCF IRB Chair, this letter is signed by:

Signature applied by Jennifer Neal-Jimenez on 10/18/2017 02:37:56 PM EDT

Designated Reviewer

APPENDIX F: COPYRIGHT RELEASE AGREEMENTS

ACM Copyright and Audio/Video Release

Title of the Work: An Analysis of User Perception Regarding Body-Worn 360° Camera Placements and Heights
Submission ID:5
Author/Presenter(s): Kevin P. Pfeil; Pamela J. Wisniewski; Joseph J. Laviola Jr.
Type of material: Full Paper

Publication and/or Conference Name: SAP '19: ACM Symposium on Applied Perception 2019 Proceedings

I. **Copyright Transfer, Reserved Rights and Permitted Uses**

* Your Copyright Transfer is conditional upon you agreeing to the terms set out below.

Copyright to the Work and to any supplemental files integral to the Work which are submitted with it for review and publication such as an extended proof, a PowerPoint outline, or appendices that may exceed a printed page limit, (including without limitation, the right to publish the Work in whole or in part in any and all forms of media, now or hereafter known) is hereby transferred to the ACM (for Government work, to the extent transferable) effective as of the date of this agreement, on the understanding that the Work has been accepted for publication by ACM.

Reserved Rights and Permitted Uses

(a) All rights and permissions the author has not granted to ACM are reserved to the Owner, including all other proprietary rights such as patent or trademark rights.

(b) Furthermore, notwithstanding the exclusive rights the Owner has granted to ACM, Owner shall have the right to do the following:

(i) Reuse any portion of the Work, without fee, in any future works written or edited by the Author, including books, lectures and presentations in any and all media.

(ii) Create a "Major Revision" which is wholly owned by the author

(iii) Post the Accepted Version of the Work on (1) the Author's home page, (2) the Owner's institutional repository, (3) any repository legally mandated by an agency funding the research on which the Work is based, and (4) any non-commercial repository or aggregation that does not duplicate ACM tables of contents, i.e., whose patterns of links do not substantially duplicate an ACM-copyrighted volume or issue. Non-commercial repositories are here understood as repositories owned by non-profit organizations that do not charge a fee for accessing deposited articles and that do not sell advertising or otherwise profit from serving articles.

(iv) Post an "Author-Izer" link enabling free downloads of the Version of Record in the ACM Digital Library on (1) the Author's home page or (2) the Owner's institutional repository;

(v) Prior to commencement of the ACM peer review process, post the version of the Work as submitted to ACM ("Submitted Version" or any earlier versions) to non-peer reviewed servers;

(vi) Make free distributions of the final published Version of Record internally to the Owner's employees, if applicable;

(vii) Make free distributions of the published Version of Record for Classroom and Personal Use;

(viii) Bundle the Work in any of Owner's software distributions; and

(ix) Use any Auxiliary Material independent from the Work. (x) If your paper is withdrawn before it is published in the ACM Digital Library, the rights revert back to the author(s).

When preparing your paper for submission using the ACM TeX templates, the rights and permissions information and the bibliographic strip must appear on the lower left hand portion of the first page.

The new ACM Consolidated TeX template Version 1.3 and above automatically creates and positions these text blocks for you based on the code snippet which is system-generated based on your rights management choice and this particular conference.

NOTE: For authors using the ACM Microsoft Word Master Article Template and Publication Workflow, The ACM Publishing System (TAPS) will add the rights statement to your papers for you. Please check with your conference contact for information regarding submitting your source file(s) for processing.

Please copy and paste \setcopyright{acmcopyright} before \begin{document} and please copy and paste the following code snippet into your TeX file between \begin{document} and \maketitle, either after or before CCS codes.

```
\copyrightyear{2019}
\acmYear{2019}
\acmConference[SAP '19]{ACM Symposium on Applied Perception 2019}{September 19--20, 2019}{Barcelona, Spain}
\acmBooktitle{ACM Symposium on Applied Perception 2019 (SAP '19), September 19--20, 2019, Barcelona, Spain}
\acmPrice{15.00}
\acmDOI{10.1145/3343036.3343120}
\acmISBN{978-1-4503-6890-2/19/09}
```

If you are using the ACM Microsoft Word template, or still using an older version of the ACM TeX template, or the current versions of the ACM SIGCHI, SIGGRAPH, or SIGPLAN TeX templates, you must copy and paste the following text block into your document as per the instructions provided with the templates you are using:

SAP '19, September 19–20, 2019, Barcelona, Spain

ACM ISBN 978-1-4503-6890-2/19/09…$15.00
https://doi.org/10.1145/3343036.3343120

NOTE: Make sure to include your article's DOI as part of the bibstrip data; DOIs will be registered and become active shortly after publication in the ACM Digital Library. Once you have your camera ready copy ready, please send your source files and PDF to your event contact for processing.

☑ A. Assent to Assignment. I hereby represent and warrant that I am the sole owner (or authorized agent of the copyright owner(s)), with the exception of third party materials detailed in section III below. I have obtained permission for any third-party material included in the Work.

☐ B. Declaration for Government Work. I am an employee of the National Government of my country and my Government claims rights to this work, or it is not copyrightable (Government work is classified as Public Domain in U.S. only)

Are any of the co-authors, employees or contractors of a National Government? ○ Yes ● No

---

II. Permission For Conference Recording and Distribution

* Your Audio/Video Release is conditional upon you agreeing to the terms set out below.

I hereby grant permission for ACM to include my name, likeness, presentation and comments in any and all forms, for the Conference and/or Publication.

I further grant permission for ACM to record and/or transcribe and reproduce my presentation as part of the ACM Digital Library, and to distribute the same for sale in complete or partial form as part of an ACM product on CD-ROM, DVD, webcast, USB device, streaming video or any other media format now or hereafter known.

I understand that my presentation will not be sold separately as a stand-alone product without my direct consent. Accordingly, I give ACM the right to use my image, voice, pronouncements, likeness, and my name, and any biographical material submitted by me, in connection with the Conference and/or Publication, whether used in excerpts or in full, for distribution described above and for any associated advertising or exhibition.

Do you agree to the above Audio/Video Release? ● Yes ○ No

III. Auxiliary Material

Do you have any Auxiliary Materials? ○ Yes ● No

IV. Third Party Materials

In the event that any materials used in my presentation or Auxiliary Materials contain the work of third-party individuals or organizations (including copyrighted music or movie excerpts or anything not owned by me), I understand that it is my responsibility to secure any necessary permissions and/or licenses for print and/or digital publication, and cite or attach them below.

● We/I have not used third-party material.
○ We/I have used third-party materials and have necessary permissions.

V. Artistic Images
If your paper includes images that were created for any purpose other than this paper and to which you or your employer claim copyright, you must complete Part V and be sure to include a notice of copyright with each such image in the paper.
● We/I do not have any artistic images.
○ We/I have any artistic images.

VI. Representations, Warranties and Covenants

The undersigned hereby represents, warrants and covenants as follows:

(a) Owner is the sole owner or authorized agent of Owner(s) of the Work;

(b) The undersigned is authorized to enter into this Agreement and grant the rights included in this license to ACM;

(c) The Work is original and does not infringe the rights of any third party; all permissions for use of third-party materials consistent in scope and duration with the rights granted to ACM have been obtained, copies of such permissions have been provided to ACM, and the Work as submitted to ACM clearly and accurately indicates the credit to the proprietors of any such third-party materials (including any applicable copyright notice), or will be revised to indicate such credit;

(d) The Work has not been published except for informal postings on non-peer reviewed servers, and Owner covenants to use best efforts to place ACM DOI pointers on any such prior postings;

(e) The Auxiliary Materials, if any, contain no malicious code, virus, trojan horse or other software routines or hardware components designed to permit unauthorized access or to disable, erase or otherwise harm any computer systems or software; and

(f) The Artistic Images, if any, are clearly and accurately noted as such (including any applicable copyright notice) in the Submitted Version.

☑ I agree to the Representations, Warranties and Covenants

DATE: **08/16/2019** sent to kevin.pfeil@knights.ucf.edu at **10:08:43**

ACM Copyright and Audio/Video Release

Title of the Work: The Effects of Gender and the Presence of Third-Party Humans on Telepresence Camera Height Preferences
Submission ID:4949125
Author/Presenter(s): Kevin Pfeil,Pamela Wisniewski,Joseph Laviola
Type of material: Full Paper

Publication and/or Conference Name: SAP '20: ACM Symposium on Applied Perception 2020 Proceedings

I. Copyright Transfer, Reserved Rights and Permitted Uses

* Your Copyright Transfer is conditional upon you agreeing to the terms set out below.

Copyright to the Work and to any supplemental files integral to the Work which are submitted with it for review and publication such as an extended proof, a PowerPoint outline, or appendices that may exceed a printed page limit, (including without limitation, the right to publish the Work in whole or in part in any and all forms of media, now or hereafter known) is hereby transferred to the ACM (for Government work, to the extent transferable) effective as of the date of this agreement, on the understanding that the Work has been accepted for publication by ACM.

Reserved Rights and Permitted Uses

(a) All rights and permissions the author has not granted to ACM are reserved to the Owner, including all other proprietary rights such as patent or trademark rights.

(b) Furthermore, notwithstanding the exclusive rights the Owner has granted to ACM, Owner shall have the right to do the following:

(i) Reuse any portion of the Work, without fee, in any future works written or edited by the Author, including books, lectures and presentations in any and all media.

(ii) Create a "Major Revision" which is wholly owned by the author

(iii) Post the Accepted Version of the Work on (1) the Author's home page, (2) the Owner's institutional repository, (3) any repository legally mandated by an agency funding the research on which the Work is based, and (4) any non-commercial repository or aggregation that does not duplicate ACM tables of contents, i.e., whose patterns of links do not substantially duplicate an ACM-copyrighted volume or issue. Non-commercial repositories are here understood as repositories owned by non-profit organizations that do not charge a fee for accessing deposited articles and that do not sell advertising or otherwise profit from serving articles.

(iv) Post an "Author-Izer" link enabling free downloads of the Version of Record in the ACM Digital Library on (1) the Author's home page or (2) the Owner's institutional repository;

(v) Prior to commencement of the ACM peer review process, post the version of the Work as submitted to ACM ("Submitted Version" or any earlier versions) to non-peer reviewed servers;

(vi) Make free distributions of the final published Version of Record internally to the Owner's employees, if applicable;

(vii) Make free distributions of the published Version of Record for Classroom and Personal Use;

(viii) Bundle the Work in any of Owner's software distributions; and

(ix) Use any Auxiliary Material independent from the Work. (x) If your paper is withdrawn before it is published in the ACM Digital Library, the rights revert back to the author(s).

When preparing your paper for submission using the ACM TeX templates, the rights and permissions information and the bibliographic strip must appear on the lower left hand portion of the first page.

The new ACM Consolidated TeX template Version 1.3 and above automatically creates and positions these text blocks for you based on the code snippet which is system-generated based on your rights management choice and this particular conference.

NOTE: For authors using the ACM Microsoft Word Master Article Template and Publication Workflow, The ACM Publishing System (TAPS) will add the rights statement to your papers for you. Please check with your conference contact for information regarding submitting your source file(s) for processing.

Please put the following LaTeX commands in the preamble of your document - i.e., before \begin{document}:

```
\copyrightyear{2020}
\acmYear{2020}
\setcopyright{acmcopyright}\acmConference[SAP '20]{ACM Symposium on Applied Perception 2020}{September 12--13, 2020}{Virtual Event, USA}
\acmBooktitle{ACM Symposium on Applied Perception 2020 (SAP '20), September 12--13, 2020, Virtual Event, USA}
\acmPrice{15.00}
\acmDOI{10.1145/3385955.3407924}
\acmISBN{978-1-4503-7618-1/20/07}
```

NOTE: For authors using the ACM Microsoft Word Master Article Template and Publication Workflow, The ACM Publishing System (TAPS) will add the rights statement to your papers for you. Please check with your conference contact for information regarding submitting your source file(s) for processing.

If you are using the ACM Interim Microsoft Word template, or still using or older versions of the ACM SIGCHI template, you must copy and paste the following text block into your document as per the instructions provided with the templates you are using:

ACM ISBN 978-1-4503-7618-1/20/07...$15.00
https://doi.org/10.1145/3385955.3407924

NOTE: Make sure to include your article's DOI as part of the bibstrip data; DOIs will be registered and become active shortly after publication in the ACM Digital Library. Once you have your camera ready copy ready, please send your source files and PDF to your event contact for processing.

- [x] A. Assent to Assignment. I hereby represent and warrant that I am the sole owner (or authorized agent of the copyright owner(s)), with the exception of third party materials detailed in section III below. I have obtained permission for any third-party material included in the Work.

- [ ] B. Declaration for Government Work. I am an employee of the National Government of my country and my Government claims rights to this work, or it is not copyrightable (Government work is classified as Public Domain in U.S. only)

II. Permission For Conference Recording and Distribution

* Your Audio/Video Release is conditional upon you agreeing to the terms set out below.

I hereby grant permission for ACM to include my name, likeness, presentation and comments in any and all forms, for the Conference and/or Publication.

I further grant permission for ACM to record and/or transcribe and reproduce my presentation as part of the ACM Digital Library, and to distribute the same for sale in complete or partial form as part of an ACM product on CD-ROM, DVD, webcast, USB device, streaming video or any other media format now or hereafter known.

I understand that my presentation will not be sold separately as a stand-alone product without my direct consent. Accordingly, I give ACM the right to use my image, voice, pronouncements, likeness, and my name, and any biographical material submitted by me, in connection with the Conference and/or Publication, whether used in excerpts or in full, for distribution described above and for any associated advertising or exhibition.

Do you agree to the above Audio/Video Release? (●) Yes ( ) No

III. Auxiliary Material

Do you have any Auxiliary Materials? ( ) Yes (●) No

IV. Third Party Materials

In the event that any materials used in my presentation or Auxiliary Materials contain the work of third-party individuals or organizations (including copyrighted music or movie excerpts or anything not owned by me), I understand that it is my responsibility to secure any necessary permissions and/or licenses for print and/or digital publication, and cite or attach them below.

(●) We/I have not used third-party material.
( ) We/I have used third-party materials and have necessary permissions.

V. Artistic Images
If your paper includes images that were created for any purpose other than this paper and to which you or your employer claim copyright, you must complete Part V and be sure to include a notice of copyright with each such image in the paper.
(•) We/I do not have any artistic images.
( ) We/I have any artistic images.

VI. Representations, Warranties and Covenants

The undersigned hereby represents, warrants and covenants as follows:

(a) Owner is the sole owner or authorized agent of Owner(s) of the Work;

(b) The undersigned is authorized to enter into this Agreement and grant the rights included in this license to ACM;

(c) The Work is original and does not infringe the rights of any third party; all permissions for use of third-party materials consistent in scope and duration with the rights granted to ACM have been obtained, copies of such permissions have been provided to ACM, and the Work as submitted to ACM clearly and accurately indicates the credit to the proprietors of any such third-party materials (including any applicable copyright notice), or will be revised to indicate such credit;

(d) The Work has not been published except for informal postings on non-peer reviewed servers, and Owner covenants to use best efforts to place ACM DOI pointers on any such prior postings;

(e) The Auxiliary Materials, if any, contain no malicious code, virus, trojan horse or other software routines or hardware components designed to permit unauthorized access or to disable, erase or otherwise harm any computer systems or software; and

(f) The Artistic Images, if any, are clearly and accurately noted as such (including any applicable copyright notice) in the Submitted Version.

[✔] I agree to the Representations, Warranties and Covenants

Funding Agents

1. Research, Development and Engineering Command award number(s): W911QX13C0052
2. National Science Foundation award number(s): IIS-1638060,IIS-1917728

DATE: **06/25/2020** sent to kevin.pfeil@knights.ucf.edu at **19:06:32**

ACM Copyright Form and Audio/Video Release

Title of the Work: Bridging the Socio-Technical Gaps in Body-worn Interpersonal Live-Streaming Telepresence through a Critical Review of the Literature

Author/Presenter(s): Kevin Pfeil;Neeraj Chatlani;Joseph LaViola;Pamela J. Wisniewski

Type of material: full paper

Publication: PACM on Human-Computer Interaction

I. Copyright Transfer, Reserved Rights and Permitted Uses

* Your Copyright Transfer is conditional upon you agreeing to the terms set out below.

Copyright to the Work and to any supplemental files integral to the Work which are submitted with it for review and publication such as an extended proof, a PowerPoint outline, or appendices that may exceed a printed page limit, (including without limitation, the right to publish the Work in whole or in part in any and all forms of media, now or hereafter known) is hereby transferred to the ACM (for Government work, to the extent transferable) effective as of the date of this agreement, on the understanding that the Work has been accepted for publication by ACM.

Reserved Rights and Permitted Uses

(a) All rights and permissions the author has not granted to ACM are reserved to the Owner, including all other proprietary rights such as patent or trademark rights.

(b) Furthermore, notwithstanding the exclusive rights the Owner has granted to ACM, Owner shall have the right to do the following:

(i) Reuse any portion of the Work, without fee, in any future works written or edited by the Author, including books, lectures and presentations in any and all media.

(ii) Create a "Major Revision" which is wholly owned by the author

(iii) Post the Accepted Version of the Work on (1) the Author's home page, (2) the Owner's institutional repository, or (3) any repository legally mandated by an agency funding the research on which the Work is based.

(iv) Post an "Author-Izer" link enabling free downloads of the Version of Record in the ACM Digital Library on (1) the Author's home page or (2) the Owner's institutional repository;

(v) Prior to commencement of the ACM peer review process, post the version of the Work as submitted to ACM ("Submitted Version" or any earlier versions) to non-peer reviewed servers;

(vi) Make free distributions of the final published Version of Record internally to the Owner's employees, if applicable;

(vii) Make free distributions of the published Version of Record for Classroom and Personal Use;

(viii) Bundle the Work in any of Owner's software distributions; and

(ix) Use any Auxiliary Material independent from the Work.

(x)If your paper is withdrawn before it is published in the ACM Digital Library, the rights revert back to the author(s).

When preparing your paper for submission using the ACM templates, you will need to include the rights management and bibstrip text blocks below to the lower left hand portion of the first page. As this text will provide rights information for your paper, please make sure that this text is displayed and positioned correctly when you submit your manuscript for publication.

Authors should understand that consistent with ACM's policy of encouraging dissemination of information, each work published by ACM appears with a copyright and the following notice:

If you are using Authorized ACM TeX templates, the following code will generate the proper statements based on your rights choices. Please copy and paste it into your TeX file between \begin{document} and \maketitle, either after or before CCS codes.

```
\setcopyright{acmcopyright}
\acmJournal{PACMHCI}
\acmYear{2021} \acmVolume{5} \acmNumber{CSCW1} \acmArticle{120}
\acmMonth{4} \acmPrice{15.00}\acmDOI{10.1145/3449194}
```

If you are using Word, copy and paste these words in the space provided at the bottom of your first page:

https://doi.org/10.1145/3449194

NOTE: DOIs will be registered and become active shortly after publication in the ACM Digital Library

☑ A. Assent to Assignment. I hereby represent and warrant that I am the sole owner (or authorized agent of the copyright owner(s)), with the exception of third party materials detailed in section III below. I have obtained permission for any third-party material included in the Work.

☐ B. Declaration for Government Work. I am an employee of the National Government of my country and my Government claims rights to this work, or it is not copyrightable (Government work is classified as Public Domain in U.S. only)

Are any of the co-authors, employees or contractors of a National Government?
○ Yes ◉ No
Country:

II. PERMISSION FOR CONFERENCE TAPING AND DISTRIBUTION (Check A and, if applicable, B)
A. Audio /Video Release

I hereby grant permission for ACM to include my name, likeness, presentation and comments in any and all forms, for the Conference and/or Publication.

I further grant permission for ACM to record and/or transcribe and reproduce my presentation as part of the ACM Digital Library, and to distribute the same for sale in complete or partial form as part of an ACM product on CD-ROM, DVD, webcast, USB device, streaming video or any other media format now or hereafter known.

I understand that my presentation will not be sold separately by itself as a stand-alone product without my direct consent. Accordingly, I give ACM the right to use my image, voice, pronouncements, likeness, and my name, and any biographical material submitted by me, in connection with the Conference and/or Publication, whether used in excerpts or in full, for distribution described above and for any associated advertising or exhibition.

Do you agree to the above Audio/Video Release? ◉ Yes ○ No

III. Auxiliary Materials, not integral to the Work

* Your Auxiliary Materials Release is conditional upon you agreeing to the terms set out below.

[Defined as additional files, including software and executables that are not submitted for review and publication as an integral part of the Work but are supplied by the author as useful resources for the reader.]

I hereby grant ACM permission to serve files containing my Auxiliary Material from the ACM Digital Library. I hereby represent and warrant that any of my Auxiliary Materials do not knowingly and surreptitiously incorporate malicious code, virus, trojan horse or other software routines or hardware components designed to permit unauthorized access or to disable, erase or otherwise harm any computer systems or software.

☐ I agree to the above Auxiliary Materials permission statement.

☐ This software is knowingly designed to illustrate technique(s) intended to defeat a system's security. The code has been explicitly documented to state this fact.

IV. Third Party Materials

In the event that any materials used in my presentation or Auxiliary Materials contain the work of third-party individuals or organizations (including copyrighted music or movie excerpts or anything not owned by me), I understand that it is my responsibility to secure any necessary permissions and/or licenses for print and/or digital publication, and cite or attach them below.

- (●) We/I have not used third-party material.
- ( ) We/I have used third-party materials and have necessary permissions.

V. Artistic Images

If your paper includes images that were created for any purpose other than this paper and to which you or your employer claim copyright, you must complete Part IV and be sure to include a notice of copyright with each such image in the paper.

- (●) We/I do not have any artistic images.
- ( ) We/I have have any artistic images.

VI. Representations, Warranties and Covenants

The undersigned hereby represents, warrants and covenants as follows:

(a) Owner is the sole owner or authorized agent of Owner(s) of the Work;

(b) The undersigned is authorized to enter into this Agreement and grant the rights included in this license to ACM;

(c) The Work is original and does not infringe the rights of any third party; all permissions for use of third-party materials consistent in scope and duration with the rights granted to ACM have been obtained, copies of such permissions have been provided to ACM, and the Work as submitted to ACM clearly and accurately indicates the credit to the proprietors of any such third-party materials (including any applicable copyright notice), or will be revised to indicate such credit;

(d) The Work has not been published except for informal postings on non-peer reviewed servers, and Owner covenants to use best efforts to place ACM DOI pointers on any such prior postings;

(e) The Auxiliary Materials, if any, contain no malicious code, virus, trojan horse or other software routines or hardware components designed to permit unauthorized access or to disable, erase or otherwise harm any computer systems or software; and

(f) The Artistic Images, if any, are clearly and accurately noted as such (including any applicable copyright notice) in the Submitted Version.

- [x] I agree to the Representations, Warranties and Covenants

Funding Agents

1. National Science Foundation IIS-1917728
2. NSF (National Science Foundation) CHS-1844881

DATE: **02/04/2021** sent to kevin.pfeil@knights.ucf.edu at **14:02:56**

Consent to Publish
Lecture Notes in Computer Science

Title of the Book or Conference Name: HCI INTERNATIONAL 2021

Volume Editor(s) Name(s): --

Title of the Contribution: Exploring Human-to-Human Telepresence and the Use of Vibro-Tactile Commands to Guide Human Streamers

Author(s) Full Name(s): Kevin Pfeil, Katelynn Kapalo, Seng Lee Koh, Pamela Wisniewski, Joseph LaViola Jr.

Corresponding Author's Name, Affiliation Address, and Email:

Kevin Pfeil, kevin.pfeil@knights.ucf.edu
University of Central Florida: 4000 Central Florida Blvd, Orlando, FL 32816

When Author is more than one person the expression "Author" as used in this agreement will apply collectively unless otherwise indicated.

The Publisher intends to publish the Work under the imprint **Springer**. The Work may be published in the book series **Lecture Notes in Computer Science (LNCS, LNAI or LNBI)**.

§ 1 Rights Granted

Author hereby grants and assigns to **Springer Nature Switzerland AG, Gewerbestrasse 11, 6330 Cham, Switzerland** (hereinafter called **Publisher**) the exclusive, sole, permanent, world-wide, transferable, sub-licensable and unlimited right to reproduce, publish, distribute, transmit, make available or otherwise communicate to the public, translate, publicly perform, archive, store, lease or lend and sell the Contribution or parts thereof individually or together with other works in any language, in all revisions and versions (including soft cover, book club and collected editions, anthologies, advance printing, reprints or print to order, microfilm editions, audiograms and videograms), in all forms and media of expression including in electronic form (including offline and online use, push or pull technologies, use in databases and data networks (e.g. the Internet) for display, print and storing on any and all stationary or portable end-user devices, e.g. text readers, audio, video or interactive devices, and for use in multimedia or interactive versions as well as for the display or transmission of the Contribution or parts thereof in data networks or search engines, and posting the Contribution on social media accounts closely related to the Work), in whole, in part or in abridged form, in each case as now known or developed in the future, including the right to grant further time-limited or permanent rights. Publisher especially has the right to permit others to use individual illustrations, tables or text quotations and may use the Contribution for advertising purposes. For the purposes of use in electronic forms, Publisher may adjust the Contribution to the respective form of use and include links (e.g. frames or inline-links) or otherwise combine it with other works and/or remove links or combinations with other works provided in the Contribution. For the avoidance of doubt, all provisions of this contract apply regardless of whether the Contribution and/or the Work itself constitutes a database under applicable copyright laws or not.
The copyright in the Contribution shall be vested in the name of Publisher. Author has asserted his/her right(s) to be identified as the originator of this Contribution in all editions and versions of the Work and parts thereof, published in all forms and media. Publisher may take, either in its own name or in that of Author, any necessary steps to protect the rights granted under this Agreement against infringement by third parties. It will have a copyright notice inserted into all editions of the Work and on the Contribution according to the provisions of the Universal Copyright Convention (UCC).
The parties acknowledge that there may be no basis for claim of copyright in the United States to a Contribution prepared by an officer or employee of the United States government as part of that person's official duties. If the Contribution was performed under a United States government contract, but Author is not a United States government employee, Publisher grants the United States government royalty-free permission to reproduce all or part of the Contribution and to authorise others to do so for United States government purposes. If the Contribution was prepared or published by or under the direction or control of the Crown (i.e., the constitutional monarch of the Commonwealth realm) or any Crown government department, the copyright in the Contribution shall, subject to any

28.05.2019 09:36

agreement with Author, belong to the Crown. If Author is an officer or employee of the United States government or of the Crown, reference will be made to this status on the signature page.

§ 2 Rights Retained by Author

Author retains, in addition to uses permitted by law, the right to communicate the content of the Contribution to other research colleagues, to share the Contribution with them in manuscript form, to perform or present the Contribution or to use the content for non-commercial internal and educational purposes, provided the original source of publication is cited according to the current citation standards in any printed or electronic materials. Author retains the right to republish the Contribution in any collection consisting solely of Author's own works without charge, subject to ensuring that the publication of the Publisher is properly credited and that the relevant copyright notice is repeated verbatim. Author may self-archive an author-created version of his/her Contribution on his/her own website and/or the repository of Author's department or faculty. Author may also deposit this version on his/her funder's or funder's designated repository at the funder's request or as a result of a legal obligation. He/she may not use the Publisher's PDF version, which is posted on the Publisher's platforms, for the purpose of self-archiving or deposit. Furthermore, Author may only post his/her own version, provided acknowledgment is given to the original source of publication and a link is inserted to the published article on the Publisher's website. The link must be provided by inserting the DOI number of the article in the following sentence: "The final authenticated version is available online at https://doi.org/[insert DOI]." The DOI (Digital Object Identifier) can be found at the bottom of the first page of the published paper.
Prior versions of the Contribution published on non-commercial pre-print servers like ArXiv/CoRR and HAL can remain on these servers and/or can be updated with Author's accepted version. The final published version (in pdf or html/xml format) cannot be used for this purpose. Acknowledgment needs to be given to the final publication and a link must be inserted to the published Contribution on the Publisher's website, by inserting the DOI number of the article in the following sentence: "The final authenticated publication is available online at https://doi.org/[insert DOI]".
Author retains the right to use his/her Contribution for his/her further scientific career by including the final published paper in his/her dissertation or doctoral thesis provided acknowledgment is given to the original source of publication. Author also retains the right to use, without having to pay a fee and without having to inform the Publisher, parts of the Contribution (e.g. illustrations) for inclusion in future work. Authors may publish an extended version of their proceedings paper as a journal article provided the following principles are adhered to: a) the extended version includes at least 30% new material, b) the original publication is cited, and c) it includes an explicit statement about the increment (e.g., new results, better description of materials, etc.).

§ 3 Warranties

Author agrees, at the request of Publisher, to execute all documents and do all things reasonably required by Publisher in order to confer to Publisher all rights intended to be granted under this Agreement. Author warrants that the Contribution is original except for such excerpts from copyrighted works (including illustrations, tables, animations and text quotations) as may be included with the permission of the copyright holder thereof, in which case(s) Author is required to obtain written permission to the extent necessary and to indicate the precise sources of the excerpts in the manuscript. Author is also requested to store the signed permission forms and to make them available to Publisher if required.
Author warrants that Author is entitled to grant the rights in accordance with Clause 1 "Rights Granted", that Author has not assigned such rights to third parties, that the Contribution has not heretofore been published in whole or in part, that the Contribution contains no libellous or defamatory statements and does not infringe on any copyright, trademark, patent, statutory right or proprietary right of others, including rights obtained through licences. Author agrees to amend the Contribution to remove any potential obscenity, defamation, libel, malicious falsehood or otherwise unlawful part(s) identified at any time. Any such removal or alteration shall not affect the warranty given by Author in this Agreement.

§ 4 Delivery of Contribution and Publication

Author agrees to deliver to the responsible Volume Editor (for conferences, usually one of the Program Chairs), on a date to be agreed upon, the manuscript created according to the Publisher's Instructions for Authors. Publisher will undertake the reproduction and distribution of the Contribution at its own expense and risk. After submission of the Consent to Publish form signed by the Corresponding Author, changes of authorship, or in the order of the authors listed, will not be accepted by the Publisher.

§ 5 Author's Discount for Books

Author is entitled to purchase for his/her personal use (if ordered directly from Publisher) the Work or other books published by Publisher at a discount of 40% off the list price for as long as there is a contractual arrangement between Author and Publisher and subject to applicable book price regulation.
Resale of such copies is not permitted.

§ 6 Governing Law and Jurisdiction

If any difference shall arise between Author and Publisher concerning the meaning of this Agreement or the rights and liabilities of the parties, the parties shall engage in good faith discussions to attempt to seek a mutually satisfactory resolution of the dispute. This agreement shall be governed by, and shall be construed in accordance with, the laws of Switzerland. The courts of Zug, Switzerland shall have the exclusive jurisdiction.

Corresponding Author signs for and accepts responsibility for releasing this material on behalf of any and all Co-Authors.

Signature of Corresponding Author: **Date:**

Kevin Pfeil 02/04/2021

☐ I'm an employee of the US Government and transfer the rights to the extent transferable (Title 17 §105 U.S.C. applies)

☐ I'm an employee of the Crown and copyright on the Contribution belongs to the Crown

For internal use only:
Legal Entity Number: 1128 Springer Nature Switzerland AG
Springer-C-CTP-07/2018

28.05.2019 09:36

LIST OF REFERENCES

[1] Wearable & on-body mounts, 2019. URL https://www.bhphotovideo.com/c/buy/body-mounts/ci/26980/N/3673516282.

[2] Gopro camera accessories, 2019. URL http://shop.gopro.com/mounts-accessories.

[3] Remote tourism - visit faroe islands, 2020. URL https://visitfaroeislands.com/remote-tourism/.

[4] Certificated remote pilots including commercial operators, 2020. URL https://www.faa.gov/uas/commercial_operators/.

[5] Get Your Mass Gatherings or Large Community Events Ready, 2020. URL https://www.cdc.gov/coronavirus/2019-ncov/community/large-events/mass-gatherings-ready-for-covid-19.html.

[6] Covid: New lockdowns for england and scotland ahead of 'hardest weeks', Jan 2021. URL https://www.bbc.com/news/uk-55538937.

[7] M. S. Ackerman. The Intellectual Challenge of CSCW: The Gap Between Social Requirements and Technical Feasibility. *HumanComputer Interaction*, 15(2-3):179–203, Sept. 2000. ISSN 0737-0024, 1532-7051. doi: 10.1207/S15327051HCI1523_5. URL https://www.tandfonline.com/doi/full/10.1207/S15327051HCI1523_5.

[8] L. Alem, F. Tecchia, and W. Huang. HandsOnVideo: Towards a Gesture based Mobile AR System for Remote Collaboration. In L. Alem and W. Huang, editors, *Recent Trends of Mobile Collaborative Augmented Reality Systems*, pages 135–148. Springer New York, New York, NY, 2011. ISBN 978-1-4419-9844-6 978-1-4419-9845-3. doi:

10.1007/978-1-4419-9845-3_11. URL http://link.springer.com/10.1007/978-1-4419-9845-3_11.

[9] A. Alohali, K. Kunze, and R. Earle. Run with me: designing storytelling tools for runners. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, pages 5–8. ACM, 2016.

[10] K. E. Anderson. Getting acquainted with social networks and apps: streaming video games on twitch. tv. *Library Hi Tech News*, 35(9):7–10, 2018.

[11] S. Aseeri, K. Paraiso, and V. Interrante. Investigating the influence of virtual human entourage elements on distance judgments in virtual architectural interiors. *Front. Robot. AI 6: 44. doi: 10.3389/frobt*, 2019.

[12] N. S. Asjad, H. Adams, R. Paris, and B. Bodenheimer. Perception of height in virtual reality: a study of climbing stairs. In *Proceedings of the 15th ACM Symposium on Applied Perception*, page 4. ACM, 2018.

[13] S. S. Ayyagari, K. Gupta, M. Tait, and M. Billinghurst. CoSense: Creating Shared Emotional Experiences. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '15*, pages 2007–2012, Seoul, Republic of Korea, 2015. ACM Press. ISBN 978-1-4503-3146-3. doi: 10.1145/2702613.2732839. URL http://dl.acm.org/citation.cfm?doid=2702613.2732839.

[14] U. Baishya and C. Neustaedter. In Your Eyes: Anytime, Anywhere Video and Audio Streaming for Couples. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing - CSCW '17*, pages 84–97, Portland, Oregon, USA, 2017. ACM Press. ISBN 978-1-4503-4335-0. doi: 10.1145/2998181.2998200. URL http://dl.acm.org/citation.cfm?doid=2998181.2998200.

[15] T. Bajarin. Are 360-degree cameras the next big thing in video?, 2015. URL https://www.pcmag.com/article2/0,2817,2492098,00.asp.

[16] D. Banakou, R. Groten, and M. Slater. Illusory ownership of a virtual child body causes overestimation of object sizes and implicit attitude changes. *Proceedings of the National Academy of Sciences*, 110(31):12846–12851, 2013.

[17] O. Baus and S. Bouchard. Exposure to an unpleasant odour increases the sense of presence in virtual reality. *Virtual Reality*, 21(2):59–74, 2017.

[18] D. C. Beidel, B. C. Frueh, S. M. Neer, C. A. Bowers, B. Trachik, T. W. Uhde, and A. Grubaugh. Trauma management therapy with virtual-reality augmented exposure therapy for combat-related ptsd: A randomized controlled trial. *Journal of anxiety disorders*, 2017.

[19] V. Bellotti, A. Ambard, D. Turner, C. Gossmann, K. Demkova, and J. M. Carroll. A muddle of models of motivation for using peer-to-peer economy systems. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 1085–1094. ACM, 2015.

[20] C. Benke, L. K. Autenrieth, E. Asselmann, and C. A. Pan-Farr. Lockdown, quarantine measures, and social distancing: Associations with depression, anxiety and distress at the beginning of the covid-19 pandemic among adults from germany. *Psychiatry Research*, 293:113462, 2020. ISSN 0165-1781. doi: https://doi.org/10.1016/j.psychres.2020.113462. URL https://www.sciencedirect.com/science/article/pii/S0165178120331231.

[21] M. Billinghurst, A. Nassani, and C. Reichherzer. Social panoramas: using wearable computers to share experiences. In *SIGGRAPH Asia 2014 Mobile Graphics and Interactive Ap-*

plications on - SA '14, pages 1–1, Shenzhen, China, 2014. ACM Press. ISBN 978-1-4503-1891-4. doi: 10.1145/2669062.2669084. URL `http://dl.acm.org/citation.cfm?doid=2669062.2669084`.

[22] F. Biocca, C. Harms, and J. Gregg. The networked minds measure of social presence: Pilot test of the factor structure and concurrent validity. In *4th annual international workshop on presence, Philadelphia, PA*, pages 1–9, 2001.

[23] D. Bowman, E. Kruijff, J. J. LaViola Jr, and I. P. Poupyrev. *3D User interfaces: theory and practice, CourseSmart eTextbook*. Addison-Wesley, 2004.

[24] J. Brooks, S. Nagels, and P. Lopes. Trigeminal-based Temperature Illusions. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2020.

[25] M. Cai and J. Tanaka. Trip Together: A Remote Pair Sightseeing System Supporting Gestural Communication. In *Proceedings of the 5th International Conference on Human Agent Interaction - HAI '17*, pages 317–324, Bielefeld, Germany, 2017. ACM Press. ISBN 978-1-4503-5113-3. doi: 10.1145/3125739.3125762. URL `http://dl.acm.org/citation.cfm?doid=3125739.3125762`.

[26] M. Cai, S. Masuko, and J. Tanaka. Gesture-based Mobile Communication System Providing Side-by-side Shopping Feeling. In *Proceedings of the 23rd International Conference on Intelligent User Interfaces Companion - IUI 18*, pages 1–2, Tokyo, Japan, 2018. ACM Press. ISBN 978-1-4503-5571-1. doi: 10.1145/3180308.3180310. URL `http://dl.acm.org/citation.cfm?doid=3180308.3180310`.

[27] M. Cai, S. Masuko, and J. Tanaka. Shopping Together: A Remote Co-shopping System Utilizing Spatial Gesture Interaction. In M. Kurosu, editor, *Human-Computer Interaction. Interaction Technologies*, volume 10903, pages 219–232. Springer International Publishing, Cham, 2018. ISBN 978-3-319-91249-3 978-3-319-91250-9. doi:

10.1007/978-3-319-91250-9_17. URL http://link.springer.com/10.1007/978-3-319-91250-9_17.

[28] Y. Cao, X. Qian, T. Wang, R. Lee, K. Huo, and K. Ramani. An Exploratory Study of Augmented Reality Presence for Tutoring Machine Tasks. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–13, Honolulu HI USA, Apr. 2020. ACM. ISBN 978-1-4503-6708-0. doi: 10.1145/3313831.3376688. URL https://dl.acm.org/doi/10.1145/3313831.3376688.

[29] A. Cassinelli, C. Reynolds, and M. Ishikawa. Augmenting spatial awareness with haptic radar. In *Wearable Computers, 2006 10th IEEE International Symposium on*, pages 61–64. IEEE, 2006.

[30] A. Cavelaars, A. Kunst, J. Geurts, R. Crialesi, L. Grötvedt, U. Helmert, E. Lahelma, O. Lundberg, A. Mielck, N. K. Rasmussen, et al. Persistent variations in average height between countries and between socio-economic groups: an overview of 10 european countries. *Annals of human biology*, 27(4):407–421, 2000.

[31] C.-T. Chang, S. Takahashi, and J. Tanaka. WithYou - a communication system to provide out together feeling. In *Proceedings of the International Working Conference on Advanced Visual Interfaces - AVI '12*, page 320, Capri Island, Italy, 2012. ACM Press. ISBN 978-1-4503-1287-5. doi: 10.1145/2254556.2254617. URL http://dl.acm.org/citation.cfm?doid=2254556.2254617.

[32] B. Chaudary, I. Paajala, E. Keino, and P. Pulli. Tele-guidance based navigation system for the visually impaired and blind persons. In *eHealth 360*, pages 9–16. Springer, 2017.

[33] E. P. H. Choi, B. P. H. Hui, and E. Y. F. Wan. Depression and anxiety in hong kong during covid-19. *International journal of environmental research and public health*, 17(10):3740, 2020.

[34] J. Christian. Doctor tells man he's going to die via telepresence robot, 2019. URL https://futurism.com/doctor-death-telepresence-robot.

[35] A. Chua, A. Forghani, and C. Neustaedter. Shared Bicycling Over Distance. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '17*, pages 455–455, Denver, Colorado, USA, 2017. ACM Press. ISBN 978-1-4503-4656-6. doi: 10.1145/3027063.3049776. URL http://dl.acm.org/citation.cfm?doid=3027063.3049776.

[36] K. Corti and A. Gillespie. Revisiting Milgrams Cyranoid Method: Experimenting With Hybrid Human Agents. *The Journal of Social Psychology*, 155(1):30–56, Jan. 2015. ISSN 0022-4545, 1940-1183. doi: 10.1080/00224545.2014.959885. URL http://www.tandfonline.com/doi/abs/10.1080/00224545.2014.959885.

[37] A. Cosgun, E. A. Sisbot, and H. I. Christensen. Guidance for human navigation using a vibro-tactile belt interface and robot-like motion planning. In *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, pages 6350–6355. IEEE, 2014.

[38] L. J. Cronbach. Coefficient alpha and the internal structure of tests. *psychometrika*, 16(3): 297–334, 1951.

[39] G. Cui. Evaluating online social presence: An overview of social presence assessment. *Journal of Educational Technology Development and Exchange (JETDE)*, 6(1):3, 2013.

[40] L. de Greef, M. E. Morris, and K. Inkpen. TeleTourist: Immersive Telepresence Tourism for Mobility-Restricted Participants. In *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion - CSCW '16 Companion*, pages 273–276, San Francisco, California, USA, 2016. ACM Press. ISBN 978-1-4503-3950-6. doi: 10.1145/2818052.2869082. URL http://dl.acm.org/citation.cfm?doid=2818052.2869082.

[41] W. H. DeLone and E. R. McLean. Information Systems Success: The Quest for the Dependent Variable. *Information Systems Research*, 3(1):60–95, 1992. ISSN 10477047, 15265536. URL http://www.jstor.org/stable/23010781. Publisher: INFORMS.

[42] M. B. Donnellan, F. L. Oswald, B. M. Baird, and R. E. Lucas. The mini-ipip scales: tiny-yet-effective measures of the big five factors of personality. *Psychological assessment*, 18 (2):192, 2006.

[43] B. A. Duncan and R. R. Murphy. Effects of Speed, Cyclicity, and Dimensionality on Distancing, Time, and Preference in Human-Aerial Vehicle Interactions. *ACM Transactions on Interactive Intelligent Systems*, 7(3):1–27, Sept. 2017. ISSN 21606455. doi: 10.1145/2983927. URL http://dl.acm.org/citation.cfm?doid=3143523.2983927.

[44] F. El Jamiy and R. Marsh. Survey on depth perception in head mounted displays: distance estimation in virtual reality, augmented reality, and mixed reality. *IET Image Processing*, 13 (5):707–712, 2019. doi: 10.1049/iet-ipr.2018.5920.

[45] L. M. Ellison, M. Nguyen, M. D. Fabrizio, A. Soh, S. Permpongkosol, and L. R. Kavoussi. Postoperative Robotic Telerounding: A Multicenter Randomized Assessment of Patient Outcomes and Satisfaction. *Archives of Surgery*, 142(12):1177–1181, 12 2007. ISSN 0004-0010. doi: 10.1001/archsurg.142.12.1177. URL https://doi.org/10.1001/archsurg.142.12.1177.

[46] D. Ergrel. 6 tips for learning how to shoot 360 videos, Jun 2016. URL https://ijnet.org/en/blog/6-tips-learning-how-shoot-360-videos.

[47] C. K. Ettman, S. M. Abdalla, G. H. Cohen, L. Sampson, P. M. Vivier, and S. Galea. Preva-

lence of depression symptoms in us adults before and during the covid-19 pandemic. *JAMA network open*, 3(9):e2019686–e2019686, 2020.

[48] B. Everson, K. A. Mackintosh, M. A. McNarry, C. Todd, and G. Stratton. Can wearable cameras be used to validate school-aged childrens lifestyle behaviours? *Children*, 6(2):20, 2019.

[49] Facebook. Better practices for 360, Apr 2016. URL `https://www.facebook.com/facebookmedia/blog/better-practices-for-360`.

[50] K. Fagan. This 24-year-old is hitchhiking across America and live-streaming the whole thing on Amazon's Twitch for his thousands of followers. URL `https://www.businessinsider.com/trevor-daneliuk-hitchhikes-across-america-on-twitch-2018-8`.

[51] J. A. Fails, M. L. Guha, and A. Druin. Methods and techniques for involving children in the design of new technology for children. *Found. Trends Hum.-Comput. Interact.*, 6(2):85166, Dec. 2013. ISSN 1551-3955. doi: 10.1561/1100000018. URL `https://doi.org/10.1561/1100000018`.

[52] C. Faklaris, F. Cafaro, S. A. Hook, A. Blevins, M. O'Haver, and N. Singhal. Legal and ethical implications of mobile live-streaming video apps. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, pages 722–729. ACM, 2016.

[53] F. Faul, E. Erdfelder, A.-G. Lang, and A. Buchner. G* power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, 39(2):175–191, 2007.

[54] G. Fitzpatrick. A personal letter to my CHI colleagues, 2019. URL

http://www.changingacademiclife.com/gerifitz/2019/2/19/a-personal-letter.

[55] E. Frankenberg and N. R. Jones. Self-rated health and mortality: does the relationship extend to a low income setting? *Journal of health and social behavior*, 45(4):441–452, 2004.

[56] B. Friedman, P. H. Kahn, A. Borning, and A. Huldtgren. *Value Sensitive Design and Information Systems*, pages 55–95. Springer Netherlands, Dordrecht, 2013. ISBN 978-94-007-7844-3. doi: 10.1007/978-94-007-7844-3_4. URL https://doi.org/10.1007/978-94-007-7844-3_4.

[57] C. D. Fryar, Q. Gu, C. L. Ogden, and K. M. Flegal. Anthropometric reference data for children and adults; united states, 2011-2014. 2016.

[58] H. C. Gagnon, L. Buck, T. N. Smith, G. Narasimham, J. Stefanucci, S. H. Creem-Regehr, and B. Bodenheimer. Far distance estimation in mixed reality. In *ACM Symposium on Applied Perception 2020*, SAP '20, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450376181. doi: 10.1145/3385955.3407933. URL https://doi.org/10.1145/3385955.3407933.

[59] S. Gauglitz, B. Nuernberger, M. Turk, and T. Hllerer. World-stabilized annotations and virtual scene navigation for remote collaboration. In *Proceedings of the 27th annual ACM symposium on User interface software and technology - UIST '14*, pages 449–459, Honolulu, Hawaii, USA, 2014. ACM Press. ISBN 978-1-4503-3069-5. doi: 10.1145/2642918.2647372. URL http://dl.acm.org/citation.cfm?doid=2642918.2647372.

[60] M. Geuss, J. Stefanucci, S. Creem-Regehr, and W. B. Thompson. Can i pass? using affordances to measure perceived size in virtual environments. In *Proceedings of the 7th*

Symposium on Applied Perception in Graphics and Visualization, APGV '10, page 6164, New York, NY, USA, 2010. Association for Computing Machinery. ISBN 9781450302487. doi: 10.1145/1836248.1836259. URL https://doi.org/10.1145/1836248.1836259.

[61] K. Gilbert, 2018. URL https://www.etsy.com/listing/520519597/darth-vader-death-starry-night-vincent.

[62] W. Glauser. Doctors among early adopters of Google Glass. *Canadian Medical Association.Journal*, 185(16):1385, Nov. 2013. URL https://search.proquest.com/docview/1476500582?accountid=10003. ISBN: 08203946.

[63] M. Glöss, M. McGregor, and B. Brown. Designing for labour: uber and the on-demand mobile workforce. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, pages 1632–1643. ACM, 2016.

[64] O. A. Godin, V. G. Irisov, and M. I. Charnotskii. Passive acoustic measurements of wind velocity and sound speed in air. *The Journal of the Acoustical Society of America*, 135 (2):EL68–EL74, Feb. 2014. ISSN 0001-4966. doi: 10.1121/1.4862885. URL http://asa.scitation.org/doi/10.1121/1.4862885.

[65] L. R. Goldberg. An alternative" description of personality": the big-five factor structure. *Journal of personality and social psychology*, 59(6):1216, 1990. Publisher: American Psychological Association.

[66] J. Graham. Why 360 video is the next big thing in tech, 2016. URL https://www.usatoday.com/story/tech/2016/01/08/why-360-video-next-big-thing-tech/78499508/.

[67] T. Y. Grechkin, T. D. Nguyen, J. M. Plumert, J. F. Cremer, and J. K. Kearney. How does

presentation method and measurement protocol affect distance estimation in real and virtual environments? *ACM Trans. Appl. Percept.*, 7(4), July 2010. ISSN 1544-3558. doi: 10.1145/1823738.1823744. URL https://doi.org/10.1145/1823738.1823744.

[68] C. N. Gunawardena and F. J. Zittle. Social presence as a predictor of satisfaction within a computer-mediated conferencing environment. *American journal of distance education*, 11 (3):8–26, 1997. Publisher: Taylor & Francis.

[69] J. Guo, Y. T.-Y. Hou, H. Mueller, K. Tang, and S. R. Fussell. As If I Am There: A New Video Chat Interface Design for Richer Contextual Awareness. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems - CHI EA '19*, pages 1–6, Glasgow, Scotland Uk, 2019. ACM Press. ISBN 978-1-4503-5971-9. doi: 10.1145/3290607.3312759. URL http://dl.acm.org/citation.cfm?doid=3290607.3312759.

[70] M. Gwilliam. Japanese Twitch streamer reveals the problem with IRL streaming in Tokyo, 2020. URL https://www.dexerto.com/entertainment/japanese-twitch-streamer-reveals-the-problem-with-irl-streaming-tokyo-

[71] O. L. Haimson and J. C. Tang. What Makes Live Events Engaging on Facebook Live, Periscope, and Snapchat. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 48–60, Denver Colorado USA, May 2017. ACM. ISBN 978-1-4503-4655-9. doi: 10.1145/3025453.3025642. URL https://dl.acm.org/doi/10.1145/3025453.3025642.

[72] W. A. Hamilton, J. Tang, G. Venolia, K. Inkpen, J. Zillner, and D. Huang. Rivulet: Exploring Participation in Live Events through Multi-Stream Experiences. In *Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video -*

TVX '16, pages 31–42, Chicago, Illinois, USA, 2016. ACM Press. ISBN 978-1-4503-4067-0. doi: 10.1145/2932206.2932211. URL http://dl.acm.org/citation.cfm?doid=2932206.2932211.

[73] D. W. Harper Jr and others. Comparing tourists crime victimization. *Annals of Tourism Research*, 28(4):1053–1056, 2001. Publisher: Pergamon Press.

[74] H. Hartikainen, M. Cortés Orduña, M. Käsmä, I. Sánchez Milara, and L. Ventä-Olkkonen. Make4change: Empowering unemployed youth through digital fabrication. In *FabLearn Europe/MakeEd 2021-An International Conference on Computing, Design and Making in Education*, pages 1–5, 2021.

[75] M. Hassenzahl, S. Heidecker, K. Eckoldt, S. Diefenbach, and U. Hillmann. All You Need is Love: Current Strategies of Mediating Intimate Relationships through Technology. *ACM Transactions on Computer-Human Interaction*, 19(4):1–19, Dec. 2012. ISSN 10730516. doi: 10.1145/2395131.2395137. URL http://dl.acm.org/citation.cfm?doid=2395131.2395137.

[76] R. A. Hayes, C. T. Carr, and D. Y. Wohn. One Click, Many Meanings: Interpreting Paralinguistic Digital Affordances in Social Media. *Journal of Broadcasting & Electronic Media*, 60(1):171–187, Jan. 2016. ISSN 0883-8151, 1550-6878. doi: 10.1080/08838151.2015.1127248. URL http://www.tandfonline.com/doi/full/10.1080/08838151.2015.1127248.

[77] C. Heeter. Being there: The subjective experience of presence. *Presence: Teleoperators & Virtual Environments*, 1(2):262–271, 1992.

[78] Y. Heshmat, B. Jones, X. Xiong, C. Neustaedter, A. Tang, B. E. Riecke, and L. Yang. Geocaching with a Beam: Shared Outdoor Activities through a Telepresence Robot with 360

Degree Viewing. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, pages 1–13, Montreal QC, Canada, 2018. ACM Press. ISBN 978-1-4503-5620-6. doi: 10.1145/3173574.3173933. URL http://dl.acm.org/citation.cfm?doid=3173574.3173933.

[79] Z. Hilvert-Bruce, J. T. Neill, M. Sjöblom, and J. Hamari. Social motivations of live-streaming viewer engagement on twitch. *Computers in Human Behavior*, 84:58–67, 2018.

[80] S. Hodgkinsoi and N. Tilley. Travel-to-crime: homing in on the victim. *International Review of Victimology*, 14(3):281–298, 2007. Publisher: SAGE Publications Sage UK: London, England.

[81] S. Holm. A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, pages 65–70, 1979.

[82] W. Huang, L. Alem, and F. Tecchia. HandsIn3D: Supporting Remote Guidance with Immersive Virtual Environments. In D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, P. Kotz, G. Marsden, G. Lindgaard, J. Wesson, and M. Winckler, editors, *Human-Computer Interaction INTERACT 2013*, volume 8117, pages 70–77. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013. ISBN 978-3-642-40482-5 978-3-642-40483-2. doi: 10.1007/978-3-642-40483-2_5. URL http://link.springer.com/10.1007/978-3-642-40483-2_5. Series Title: Lecture Notes in Computer Science.

[83] Huazhong Li, Yongsheng Liang, Tao He, and Yi Li. Real-time shared control of space robots teleoperation without time delay. In *2012 24th Chinese Control and Decision Conference (CCDC)*, pages 3637–3642, May 2012. doi: 10.1109/CCDC.2012.6244582.

[84] S. Hussain and E. B.-N. Sanders. Fusion of horizons: Co-designing with cambodian children who have prosthetic legs, using generative design tools. *CoDesign*, 8(1):43–79, 2012.

[85] P. Hyland, M. Shevlin, O. McBride, J. Murphy, T. Karatzias, R. P. Bentall, A. Martinez, and F. Vallières. Anxiety and depression in the republic of ireland during the covid-19 pandemic. *Acta Psychiatrica Scandinavica*, 142(3):249–256, 2020.

[86] K. Inkpen, B. Taylor, S. Junuzovic, J. Tang, and G. Venolia. Experiences2Go: sharing kids' activities outside the home with remote family members. In *Proceedings of the 2013 conference on Computer supported cooperative work - CSCW '13*, page 1329, San Antonio, Texas, USA, 2013. ACM Press. ISBN 978-1-4503-1331-5. doi: 10.1145/2441776.2441926. URL `http://dl.acm.org/citation.cfm?doid=2441776.2441926`.

[87] C. Ishak, C. Neustaedter, D. Hawkins, J. Procyk, and M. Massimi. Human Proxies for Remote University Classroom Attendance. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*, pages 931–943, Santa Clara, California, USA, 2016. ACM Press. ISBN 978-1-4503-3362-7. doi: 10.1145/2858036.2858184. URL `http://dl.acm.org/citation.cfm?doid=2858036.2858184`.

[88] P. Jarusriboonchai, J. L. Meissner, N. B. Hansen, and B. Schouten. Thinking outside the (tool) box: empowering people with toolkits. In *Proceedings of the 10th Nordic Conference on Human-Computer Interaction*, pages 980–983, 2018.

[89] B. Jelen, A. Freeman, M. Narayanan, K. M. Sanders, J. Clawson, and K. A. Siek. Craftec: Engaging older adults in making through a craft-based toolkit system. In *Proceedings of the Thirteenth International Conference on Tangible, Embedded, and Embodied Interaction*, pages 577–587, 2019.

[90] B. Jones, A. Witcraft, S. Bateman, C. Neustaedter, and A. Tang. Mechanics of Camera Work in Mobile Video Collaboration. In *Proceedings of the 33rd Annual ACM Conference*

on Human Factors in Computing Systems - CHI '15, pages 957–966, Seoul, Republic of Korea, 2015. ACM Press. ISBN 978-1-4503-3145-6. doi: 10.1145/2702123.2702345. URL http://dl.acm.org/citation.cfm?doid=2702123.2702345.

[91] B. Jones, Y. Zhang, P. N. Y. Wong, and S. Rintel. Vroom: Virtual robot overlay for online meetings. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI EA '20, page 110, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450368193. doi: 10.1145/3334480.3382820. URL https://doi.org/10.1145/3334480.3382820.

[92] B. Jones, Y. Zhang, P. N. Y. Wong, and S. Rintel. Belonging there: Vroom-ing into the uncanny valley of xr telepresence. *Proc. ACM Hum.-Comput. Interact.*, 5(CSCW1), Apr. 2021. doi: 10.1145/3449133. URL https://doi.org/10.1145/3449133.

[93] J. A. Jones, J. E. Swan, G. Singh, E. Kolstad, and S. R. Ellis. The effects of virtual reality, augmented reality, and motion parallax on egocentric depth perception. In *Proceedings of the 5th Symposium on Applied Perception in Graphics and Visualization*, APGV '08, page 914, New York, NY, USA, 2008. Association for Computing Machinery. ISBN 9781595939814. doi: 10.1145/1394281.1394283. URL https://doi.org/10.1145/1394281.1394283.

[94] J. A. Jones, J. E. Swan, G. Singh, and S. R. Ellis. Peripheral visual information and its effect on distance judgments in virtual and augmented environments. In *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception in Graphics and Visualization*, APGV '11, page 2936, New York, NY, USA, 2011. Association for Computing Machinery. ISBN 9781450308892. doi: 10.1145/2077451.2077457. URL https://doi.org/10.1145/2077451.2077457.

[95] J. A. Jones, J. E. Swan II, and M. Bolas. Peripheral stimulation and its effect on perceived

spatial scale in virtual environments. *IEEE Transactions on Visualization and Computer Graphics*, 19(4):701–710, 2013. doi: 10.1109/TVCG.2013.37.

[96] T. K. Judge, C. Neustaedter, and A. F. Kurtz. The family window: the design and evaluation of a domestic media space. In *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10*, page 2361, Atlanta, Georgia, USA, 2010. ACM Press. ISBN 978-1-60558-929-9. doi: 10.1145/1753326.1753682. URL `http://portal.acm.org/citation.cfm?doid=1753326.1753682`.

[97] T. K. Judge, C. Neustaedter, S. Harrison, and A. Blose. Family portals: connecting families through a multifamily media space. In *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11*, page 1205, Vancouver, BC, Canada, 2011. ACM Press. ISBN 978-1-4503-0228-9. doi: 10.1145/1978942.1979122. URL `http://dl.acm.org/citation.cfm?doid=1978942.1979122`.

[98] B. Kang, I. Hwang, J. Lee, S. Lee, T. Lee, Y. Chang, and M. K. Lee. My Being to Your Place, Your Being to My Place: Co-present Robotic Avatars Create Illusion of Living Together. In *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services - MobiSys '18*, pages 54–67, Munich, Germany, 2018. ACM Press. ISBN 978-1-4503-5720-3. doi: 10.1145/3210240.3210348. URL `http://dl.acm.org/citation.cfm?doid=3210240.3210348`.

[99] E. Karaaslan, U. Bagci, and F. N. Catbas. Artificial Intelligence Assisted Infrastructure Assessment using Mixed Reality Systems. *Transportation Research Record: Journal of the Transportation Research Board*, 2673(12):413–424, Dec. 2019. ISSN 0361-1981, 2169-4052. doi: 10.1177/0361198119839988. URL `http://journals.sagepub.com/doi/10.1177/0361198119839988`.

[100] S. Kasahara and J. Rekimoto. JackIn: integrating first-person view with out-of-body vi-

sion generation for human-human augmentation. In *Proceedings of the 5th Augmented Human International Conference on - AH '14*, pages 1–8, Kobe, Japan, 2014. ACM Press. ISBN 978-1-4503-2761-9. doi: 10.1145/2582051.2582097. URL http://dl.acm.org/citation.cfm?doid=2582051.2582097.

[101] S. Kasahara and J. Rekimoto. JackIn head: an immersive human-human telepresence system. In *SIGGRAPH Asia 2015 Emerging Technologies on - SA '15*, pages 1–3, Kobe, Japan, 2015. ACM Press. ISBN 978-1-4503-3925-4. doi: 10.1145/2818466.2818486. URL http://dl.acm.org/citation.cfm?doid=2818466.2818486.

[102] S. Kasahara and J. Rekimoto. JackIn head: immersive visual telepresence system with omnidirectional wearable camera for remote collaboration. In *Proceedings of the 21st ACM Symposium on Virtual Reality Software and Technology - VRST '15*, pages 217–225, Beijing, China, 2015. ACM Press. ISBN 978-1-4503-3990-2. doi: 10.1145/2821592.2821608. URL http://dl.acm.org/citation.cfm?doid=2821592.2821608.

[103] S. Kasahara, S. Nagai, and J. Rekimoto. LiveSphere: immersive experience sharing with 360 degrees head-mounted cameras. In *Proceedings of the adjunct publication of the 27th annual ACM symposium on User interface software and technology - UIST'14 Adjunct*, pages 61–62, Honolulu, Hawaii, USA, 2014. ACM Press. ISBN 978-1-4503-3068-8. doi: 10.1145/2658779.2659114. URL http://dl.acm.org/citation.cfm?doid=2658779.2659114.

[104] S. Kasahara, S. Nagai, and J. Rekimoto. First Person Omnidirectional Video: System Design and Implications for Immersive Experience. In *Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video - TVX '15*, pages 33–42, Brussels, Belgium, 2015. ACM Press. ISBN 978-1-4503-3526-3. doi: 10.1145/

2745197.2745202. URL http://dl.acm.org/citation.cfm?doid=2745197.2745202.

[105] S. Kasahara, S. Nagai, and J. Rekimoto. JackIn Head: Immersive Visual Telepresence System with Omnidirectional Wearable Camera. *IEEE Transactions on Visualization and Computer Graphics*, 23(3):1222–1234, Mar. 2017. ISSN 1077-2626. doi: 10.1109/TVCG.2016.2642947. URL http://ieeexplore.ieee.org/document/7792698/.

[106] T. Kashiwabara, H. Osawa, K. Shinozawa, and M. Imai. TEROOS: a wearable avatar to enhance joint activities. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12*, page 2001, Austin, Texas, USA, 2012. ACM Press. ISBN 978-1-4503-1015-4. doi: 10.1145/2207676.2208345. URL http://dl.acm.org/citation.cfm?doid=2207676.2208345.

[107] E.-S. Katterfeldt, N. Dittert, and H. Schelhowe. Eduwear: smart textiles as ways of relating computing technology to everyday life. In *Proceedings of the 8th International Conference on Interaction Design and Children*, pages 9–17, 2009.

[108] O. B. Kaul and M. Rohs. Haptichead: A spherical vibrotactile grid around the head for 3d guidance in virtual and augmented reality. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 3729–3740. ACM, 2017.

[109] M. Kazemitabaar, J. McPeak, A. Jiao, L. He, T. Outing, and J. E. Froehlich. Makerwear: A tangible approach to interactive wearable creation for children. In *Proceedings of the 2017 chi conference on human factors in computing systems*, pages 133–145, 2017.

[110] J. W. Kelly, L. A. Cherep, and Z. D. Siegel. Perceived space in the htc vive. *ACM Trans. Appl. Percept.*, 15(1):2:1–2:16, July 2017. ISSN 1544-3558. doi: 10.1145/3106155. URL http://doi.acm.org.ezproxy.net.ucf.edu/10.1145/3106155.

[111] P. Kelly, A. R. Doherty, A. Hamilton, A. Matthews, A. M. Batterham, M. Nelson, C. Foster, and G. Cowburn. Evaluating the feasibility of measuring travel to school using a wearable camera. *American journal of preventive medicine*, 43(5):546–550, 2012.

[112] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology*, 3(3):203–220, 1993. Publisher: Taylor & Francis.

[113] T. Keskinen, V. Mäkelä, P. Kallioniemi, J. Hakulinen, J. Karhu, K. Ronkainen, J. Mäkelä, and M. Turunen. The effect of camera height, actor behavior, and viewer position on the user experience of 360 videos. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2019.

[114] S. Kim, G. A. Lee, N. Sakata, A. Dunser, E. Vartiainen, and M. Billinghurst. Study of augmented gesture communication cues and view sharing in remote collaboration. In *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 261–262, Adelaide, Australia, Oct. 2013. IEEE. ISBN 978-1-4799-2869-9. doi: 10.1109/ISMAR.2013.6671795. URL `http://ieeexplore.ieee.org/document/6671795/`.

[115] S. Kim, S. Junuzovic, and K. Inkpen. The Nomad and the Couch Potato: Enriching Mobile Shared Experiences with Contextual Information. In *Proceedings of the 18th International Conference on Supporting Group Work - GROUP '14*, pages 167–177, Sanibel Island, Florida, USA, 2014. ACM Press. ISBN 978-1-4503-3043-5. doi: 10.1145/2660398.2660409. URL `http://dl.acm.org/citation.cfm?doid=2660398.2660409`.

[116] D. Kimber, P. Proppe, S. Kratz, J. Vaughan, B. Liew, D. Severns, and W. Su. Polly: Telepresence from a Guides Shoulder. In L. Agapito, M. M. Bronstein, and C. Rother, editors,

Computer Vision - ECCV 2014 Workshops, volume 8927, pages 509–523. Springer International Publishing, Cham, 2015. ISBN 978-3-319-16198-3 978-3-319-16199-0. doi: 10.1007/978-3-319-16199-0_36. URL http://link.springer.com/10.1007/978-3-319-16199-0_36.

[117] S. Kinch, M. Pakanen, K. Heiselberg, C. Dindler, A.-M. Iversen, and P. G. Krogh. An exploratory study of using speculative artefacts in co-design. *CoDesign*, pages 1–19, 2022.

[118] S. Koceski and N. Koceska. Evaluation of an Assistive Telepresence Robot for Elderly Healthcare. *Journal of Medical Systems*, 40(5):121, May 2016. ISSN 0148-5598, 1573-689X. doi: 10.1007/s10916-016-0481-x. URL http://link.springer.com/10.1007/s10916-016-0481-x.

[119] D. Kontaris, D. Harrison, E.-E. Patsoule, S. Zhuang, and A. Slade. Feelybean: communicating touch over distance. In *CHI'12 Extended Abstracts on Human Factors in Computing Systems*, pages 1273–1278. ACM, 2012.

[120] J. Kopf. 360 video stabilization. *ACM Transactions on Graphics (TOG)*, 35(6):195, 2016.

[121] L. Kovcs, T. Haidegger, and I. Rudas. Surgery from a distanceApplication of intelligent control for telemedicine. In *2013 IEEE 11th International Symposium on Applied Machine Intelligence and Informatics (SAMI)*, pages 125–129, Jan. 2013. doi: 10.1109/SAMI.2013.6480959.

[122] S. Kratz, D. Kimber, W. Su, G. Gordon, and D. Severns. Polly: "being there" through the parrot and a guide. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services - MobileHCI '14*, pages 625–630, Toronto, ON, Canada, 2014. ACM Press. ISBN 978-1-4503-3004-6. doi: 10.1145/2628363.2628430. URL http://dl.acm.org/citation.cfm?doid=2628363.2628430.

[123] S. Kratz, D. Avrahami, D. Kimber, J. Vaughan, P. Proppe, and D. Severns. Polly Wanna Show You: Examining Viewpoint-Conveyance Techniques for a Shoulder-Worn Telepresence System. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct - MobileHCI '15*, pages 567–575, Copenhagen, Denmark, 2015. ACM Press. ISBN 978-1-4503-3653-6. doi: 10.1145/2786567.2787134. URL http://dl.acm.org/citation.cfm?doid=2786567.2787134.

[124] S. F. Kreimer. Pervasive image capture and the First Amendment: Memory, discourse, and the right to record. *U. Pa. L. Rev.*, 159:335, 2010. Publisher: HeinOnline.

[125] A. Kristoffersson, S. Coradeschi, and A. Loutfi. A Review of Mobile Robotic Telepresence. *Advances in Human-Computer Interaction*, 2013:1–17, 2013. ISSN 1687-5893, 1687-5907. doi: 10.1155/2013/902316. URL http://www.hindawi.com/journals/ahci/2013/902316/.

[126] S. A. Kuhl, W. B. Thompson, and S. H. Creem-Regehr. Hmd calibration and its effects on distance judgments. *ACM Transactions on Applied Perception (TAP)*, 6(3):19, 2009.

[127] B. R. Kunz, L. Wouters, D. Smith, W. B. Thompson, and S. H. Creem-Regehr. Revisiting the effect of quality of graphics on distance judgments in virtual environments: A comparison of verbal reports and blind walking. *Attention, Perception, & Psychophysics*, 71(6):1284–1293, 2009.

[128] M. Kytö, A. Mäkinen, T. Tossavainen, and P. T. Oittinen. Stereoscopic depth perception in video see-through augmented reality within action space. *Journal of Electronic Imaging*, 23 (1):011006, 2014.

[129] E. Langbehn, G. Bruder, and F. Steinicke. Scale matters! analysis of dominant scale esti-

mation in the presence of conflicting cues in multi-scale collaborative virtual environments. In *2016 IEEE Symposium on 3D User Interfaces (3DUI)*, pages 211–220. IEEE, 2016.

[130] C. Lavigne. 360 video production tactics: What weve learned so far, Apr 2016. URL `https://wistia.com/learn/production/360-video-shooting-techniques`.

[131] J. J. LaViola, E. Kruijff, R. P. McMahan, D. A. Bowman, and I. Poupyrev. *3D user interfaces: theory and practice*. Addison-Wesley usability and HCI series. Addison-Wesley, Boston, second edition edition, 2017. ISBN 978-0-13-403432-4. OCLC: ocn935986831.

[132] J. Lazar, J. H. Feng, and H. Hochheiser. *Research Methods in Human-Computer Interaction*. 2017. ISBN 978-0-12-809343-6. URL `https://public.ebookcentral.proquest.com/choice/publicfullrecord.aspx?p=4851896`. OCLC: 1076644193.

[133] G. A. Lee, T. Teo, S. Kim, and M. Billinghurst. Mixed reality collaboration through sharing a live panorama. In *SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications*, page 14. ACM, 2017.

[134] G. A. Lee, T. Teo, S. Kim, and M. Billinghurst. Sharedsphere: MR collaboration through shared live panorama. In *SIGGRAPH Asia 2017 Emerging Technologies on - SA '17*, pages 1–2, Bangkok, Thailand, 2017. ACM Press. ISBN 978-1-4503-5404-2. doi: 10.1145/3132818.3132827. URL `http://dl.acm.org/citation.cfm?doid=3132818.3132827`.

[135] G. A. Lee, T. Teo, S. Kim, and M. Billinghurst. A User Study on MR Remote Collaboration Using Live 360 Video. In *2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 153–164, Munich, Germany, Oct. 2018. IEEE. ISBN 978-1-5386-

7459-8. doi: 10.1109/ISMAR.2018.00051. URL `https://ieeexplore.ieee.org/document/8613761/`.

[136] D. Lester. Achieving Human Presence in Space Exploration. *Presence*, 22(4):345–349, Nov. 2013. ISSN 1054-7460. doi: 10.1162/PRES_a_00160.

[137] D. F. Lester, K. V. Hodges, and R. C. Anderson. Exploration telepresence: A strategy for optimizing scientific research at remote space destinations. *Science Robotics*, 2(7):eaan4383, June 2017. ISSN 2470-9476. doi: 10.1126/scirobotics.aan4383. URL `http://robotics.sciencemag.org/lookup/doi/10.1126/scirobotics.aan4383`.

[138] M. Leyrer, S. A. Linkenauger, H. H. Bülthoff, U. Kloos, and B. Mohler. The influence of eye height and avatars on egocentric distance estimates in immersive virtual environments. In *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception in Graphics and Visualization*, pages 67–74. ACM, 2011.

[139] H. Li, C. Ji, X. Zong, and Y. Zhang. Height and weight standardized growth charts for chinese children and adolescents aged 0 to 18 years. *Zhonghua er ke za zhi= Chinese journal of pediatrics*, 47(7):487–492, 2009.

[140] Y. Li, Y. Kou, J. S. Lee, and A. Kobsa. Tell me before you stream me: Managing information disclosure in video game live streaming. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW):107, 2018.

[141] C. F. Liew and T. Yairi. Companion unmanned aerial vehicles: A survey. *arXiv preprint arXiv:2001.04637*, 2020.

[142] M. A. Livingston, Z. Ai, J. E. Swan, and H. S. Smallman. Indoor vs. outdoor depth percep-

tion for mobile augmented reality. In *2009 IEEE Virtual Reality Conference*, pages 55–62, Lafayette, LA, 2009. IEEE. doi: 10.1109/VR.2009.4810999.

[143] J. M. Loomis and J. W. Philbeck. Measuring spatial perception with spatial updating and action. In *Carnegie Symposium on Cognition, 2006, Pittsburgh, PA, US*, pages 1–43, Pittsburgh, PA, US, 2008. Psychology Press, Psychology Press.

[144] Z. Lu, M. Annett, M. Fan, and D. Wigdor. "I feel it is my responsibility to stream": Streaming and Engaging with Intangible Cultural Heritage through Livestreaming. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, pages 1–14, Glasgow, Scotland Uk, 2019. ACM Press. ISBN 978-1-4503-5970-2. doi: 10.1145/3290605.3300459. URL `http://dl.acm.org/citation.cfm?doid=3290605.3300459`.

[145] Z. Lu, M. Annett, and D. Wigdor. Vicariously Experiencing it all Without Going Outside: A Study of Outdoor Livestreaming in China. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):1–28, Nov. 2019. ISSN 2573-0142, 2573-0142. doi: 10.1145/3359127. URL `https://dl.acm.org/doi/10.1145/3359127`.

[146] A. M. Lund. Measuring usability with the use questionnaire12. *Usability interface*, 8(2): 3–6, 2001.

[147] T. Maeda, H. Ando, H. Iizuka, T. Yonemura, D. Kondo, and M. Niwa. Parasitic Humanoid: the wearable robotics as a behavioral assist interface like oneness between horse and rider. In *Proceedings of the 2nd Augmented Human International Conference on - AH '11*, pages 1–8, Tokyo, Japan, 2011. ACM Press. ISBN 978-1-4503-0426-9. doi: 10.1145/1959826.1959844. URL `http://portal.acm.org/citation.cfm?doid=1959826.1959844`.

[148] K. Malsev. Set up a turn server on aws in 15 minutes, Jan 2021. URL https://kostya-malsev.medium.com/set-up-a-turn-server-on-aws-in-15-minutes-25beb145bc77.

[149] M. Manabe, D. Uriu, T. Funatsu, A. Izumihara, T. Yazaki, I.-H. Chen, Y.-Y. Liao, K.-Y. Liu, J.-C. Ko, Z. Kashino, A. Hiyama, and M. Inami. Exploring in the city with your personal guide:design and user study of t-leap, a telepresence system. In *19th International Conference on Mobile and Ubiquitous Multimedia*, MUM 2020, page 96106, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450388702. doi: 10.1145/3428361.3428382. URL https://doi.org/10.1145/3428361.3428382.

[150] J. Marn-Morales, J. L. Higuera-Trujillo, C. De-Juan-Ripoll, C. Llinares, J. Guixeres, S. Iarra, and M. Alcaiz. Navigation comparison between a real and a virtual museum: Time-dependent differences using a head mounted display. *Interacting with Computers*, 31(2):208–220, 07 2019. ISSN 0953-5438. doi: 10.1093/iwc/iwz018. URL https://doi.org/10.1093/iwc/iwz018.

[151] S. Masnadi, K. P. Pfeil, J.-V. T. Sera-Josef, and J. J. LaViola. Field of view effect on distance perception in virtual reality. In *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pages 542–543. IEEE, 2021.

[152] A. Matsuda and J. Rekimoto. Scalablebody: A telepresence robot supporting socially acceptable interactions and human augmentation through vertical actuation. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, pages 103–105. ACM, 2016.

[153] A. Matsuda, K. Nozawa, and J. Rekimoto. JackIn Neck: A Neckband Wearable Telepresence System Designed for High Comfortability. In *Proceedings of the 2018 ACM International Conference on Interactive Surfaces and Spaces - ISS '18*, pages 415–418, Tokyo,

Japan, 2018. ACM Press. ISBN 978-1-4503-5694-7. doi: 10.1145/3279778.3279917. URL http://dl.acm.org/citation.cfm?doid=3279778.3279917.

[154] A. Matsuda, K. Nozawa, K. Takata, A. Izumihara, and J. Rekimoto. HapticPointer: A Neck-worn Device that Presents Direction by Vibrotactile Feedback for Remote Collaboration Tasks. In *Proceedings of the Augmented Humans International Conference*, pages 1–10, Kaiserslautern Germany, Mar. 2020. ACM. ISBN 978-1-4503-7603-7. doi: 10.1145/3384657.3384777. URL https://dl.acm.org/doi/10.1145/3384657.3384777.

[155] T. McDaniel, S. Krishna, V. Balasubramanian, D. Colbry, and S. Panchanathan. Using a haptic belt to convey non-verbal communication cues during social interactions to individuals who are blind. In *Haptic Audio visual Environments and Games, 2008. HAVE 2008. IEEE International Workshop on*, pages 13–18. IEEE, 2008.

[156] B. McGrath, A. Estrada, M. Braithwaite, A. Raj, and A. Rupert. Tactile situation awareness system flight demonstration. Technical report, ARMY AEROMEDICAL RESEARCH LAB FORT RUCKER AL, 2004.

[157] A. Meier, D. J. Matthies, B. Urban, and R. Wettach. Exploring vibrotactile feedback on the body and foot for the purpose of pedestrian navigation. In *Proceedings of the 2nd international Workshop on Sensor-based Activity Recognition and Interaction*, page 11. ACM, 2015.

[158] R. W. Miller. What's 'zoom fatigue'? here's why video calls can be so exhausting, 2020. URL https://www.usatoday.com/story/news/nation/2020/04/23/zoom-fatigue-video-calls-coronavirus-can-make-us-tired-anxious/3010478001/.

[159] M. Minsky. Telepresence. omni magazine. *New York, Jun*, 1980.

[160] K. Misawa and J. Rekimoto. ChameleonMask: a human-surrogate system with a telepresence face. In *SIGGRAPH Asia 2015 Emerging Technologies on - SA '15*, pages 1–3, Kobe, Japan, 2015. ACM Press. ISBN 978-1-4503-3925-4. doi: 10.1145/2818466.2818473. URL http://dl.acm.org/citation.cfm?doid=2818466.2818473.

[161] K. Misawa and J. Rekimoto. ChameleonMask: Embodied Physical and Social Telepresence using Human Surrogates. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '15*, pages 401–411, Seoul, Republic of Korea, 2015. ACM Press. ISBN 978-1-4503-3146-3. doi: 10.1145/2702613.2732506. URL http://dl.acm.org/citation.cfm?doid=2702613.2732506.

[162] K. Misawa and J. Rekimoto. Wearing another's personality: a human-surrogate system with a telepresence face. In *Proceedings of the 2015 ACM International Symposium on Wearable Computers - ISWC '15*, pages 125–132, Osaka, Japan, 2015. ACM Press. ISBN 978-1-4503-3578-2. doi: 10.1145/2802083.2808392. URL http://dl.acm.org/citation.cfm?doid=2802083.2808392.

[163] K. Misawa and J. Rekimoto. Who am I Touching?: User Study of Remote Handshaking with a Telepresence Face. In *Proceedings of the Fourth International Conference on Human Agent Interaction - HAI '16*, pages 163–170, Biopolis, Singapore, 2016. ACM Press. ISBN 978-1-4503-4508-8. doi: 10.1145/2974804.2974821. URL http://dl.acm.org/citation.cfm?doid=2974804.2974821.

[164] K. Misawa, Y. Ishiguro, and J. Rekimoto. Ma petite chrie: what are you looking at?: a small telepresence system to support remote collaborative work for intimate communication. In *Proceedings of the 3rd Augmented Human International Conference on - AH '12*, pages 1–5, Megève, France, 2012. ACM Press. ISBN 978-1-4503-1077-2. doi: 10.1145/

2160125.2160142. URL http://dl.acm.org/citation.cfm?doid=2160125.2160142.

[165] R. Mitchell, A. Gillespie, and B. O'Neill. Cyranic contraptions: using personality surrogates to explore ontologically and socially dynamic contexts. In *Procedings of the Second Conference on Creativity and Innovation in Design - DESIRE '11*, page 199, Eindhoven, Netherlands, 2011. ACM Press. ISBN 978-1-4503-0754-3. doi: 10.1145/2079216.2079246. URL http://dl.acm.org/citation.cfm?doid=2079216.2079246.

[166] B. J. Mohler, S. H. Creem-Regehr, and W. B. Thompson. The influence of feedback on egocentric distance judgments in real and virtual environments. In *Proceedings of the 3rd symposium on Applied perception in graphics and visualization*, pages 9–14. ACM, 2006.

[167] B. J. Mohler, S. H. Creem-Regehr, W. B. Thompson, and H. H. Blthoff. The effect of viewing a self-avatar on distance judgments in an hmd-based virtual environment. *Presence: Teleoperators and Virtual Environments*, 19(3):230–242, 2010. doi: 10.1162/pres.19.3.230. URL https://doi.org/10.1162/pres.19.3.230.

[168] P. Mohr, S. Mori, T. Langlotz, B. H. Thomas, D. Schmalstieg, and D. Kalkofen. Mixed Reality Light Fields for Interactive Remote Assistance. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–12, Honolulu HI USA, Apr. 2020. ACM. ISBN 978-1-4503-6708-0. doi: 10.1145/3313831.3376289. URL https://dl.acm.org/doi/10.1145/3313831.3376289.

[169] S. Mosadeghi, M. W. Reid, B. Martinez, B. T. Rosen, and B. M. R. Spiegel. Feasibility of an immersive virtual reality intervention for hospitalized patients: an observational cohort study. *JMIR mental health*, 3(2), 2016.

[170] A. E. Mostafa, K. Inkpen, J. C. Tang, G. Venolia, and W. A. Hamilton. SocialStreamViewer: Guiding the Viewer Experience of Multiple Streams of an Event. In *Proceedings of the*

19th International Conference on Supporting Group Work - GROUP '16, pages 287–291, Sanibel Island, Florida, USA, 2016. ACM Press. ISBN 978-1-4503-4276-6. doi: 10.1145/2957276.2957286. URL http://dl.acm.org/citation.cfm?doid=2957276.2957286.

[171] J. Muller, T. Langlotz, and H. Regenbrecht. PanoVC: Pervasive telepresence using mobile phones. In *2016 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pages 1–10, Sydney, NSW, Mar. 2016. IEEE. ISBN 978-1-4673-8779-8. doi: 10.1109/PERCOM.2016.7456508. URL http://ieeexplore.ieee.org/document/7456508/.

[172] B. G. Munyan, S. M. Neer, D. C. Beidel, and F. Jentsch. Olfactory stimuli increase presence during simulated exposure. In *International Conference on Virtual, Augmented and Mixed Reality*, pages 164–172. Springer, 2016.

[173] I. B. Myers. The Myers-Briggs Type Indicator: Manual (1962). 1962. Publisher: Consulting Psychologists Press.

[174] K. Myles and J. T. Kalb. Guidelines for head tactile communication. Technical report, Army Research Lab Aberdeen Proving Ground Md Human Research And Engineering Directorate, 2010.

[175] S. Nagai, S. Kasahara, and J. Rekimoto. Directional communication using spatial sound in human-telepresence. In *Proceedings of the 6th Augmented Human International Conference on - AH '15*, pages 159–160, Singapore, Singapore, 2015. ACM Press. ISBN 978-1-4503-3349-8. doi: 10.1145/2735711.2735818. URL http://dl.acm.org/citation.cfm?doid=2735711.2735818.

[176] S. Nagai, S. Kasahara, and J. Rekimoto. LiveSphere: Sharing the Surrounding Visual Environment for Immersive Experience in Remote Collaboration. In *Proceedings of the Ninth*

International Conference on Tangible, Embedded, and Embodied Interaction - TEI '14, pages 113–116, Stanford, California, USA, 2015. ACM Press. ISBN 978-1-4503-3305-4. doi: 10.1145/2677199.2680549. URL http://dl.acm.org/citation.cfm?doid=2677199.2680549.

[177] NBA. Nba vr and mr, 2019. URL https://www.nba.com/xr.

[178] C. Neustaedter and S. Greenberg. Intimacy in long-distance relationships over video chat. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12*, page 753, Austin, Texas, USA, 2012. ACM Press. ISBN 978-1-4503-1015-4. doi: 10.1145/2207676.2207785. URL http://dl.acm.org/citation.cfm?doid=2207676.2207785.

[179] C. Neustaedter and T. K. Judge. Peek-A-Boo: the design of a mobile family media space. In *Proceedings of the 12th ACM international conference adjunct papers on Ubiquitous computing - Ubicomp '10*, page 449, Copenhagen, Denmark, 2010. ACM Press. ISBN 978-1-4503-0283-8. doi: 10.1145/1864431.1864482. URL http://portal.acm.org/citation.cfm?doid=1864431.1864482.

[180] C. Neustaedter and A. Tang. Explorations of Remote Attendance at CHI, 2019. URL https://chi2019.acm.org/2019/02/15/explorations-of-remote-attendance-at-chi/.

[181] C. Neustaedter, C. Pang, A. Forghani, E. Oduor, S. Hillman, T. K. Judge, M. Massimi, and S. Greenberg. Sharing Domestic Life through Long-Term Video Connections. *ACM Transactions on Computer-Human Interaction*, 22(1):1–29, Feb. 2015. ISSN 10730516. doi: 10.1145/2696869. URL http://dl.acm.org/citation.cfm?doid=2737795.2696869.

[182] C. Neustaedter, G. Venolia, J. Procyk, and D. Hawkins. To Beam or Not to Beam: A Study of Remote Telepresence Attendance at an Academic Conference. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing - CSCW '16*, pages 417–430, San Francisco, California, USA, 2016. ACM Press. ISBN 978-1-4503-3592-8. doi: 10.1145/2818048.2819922. URL `http://dl.acm.org/citation.cfm?doid=2818048.2819922`.

[183] C. Neustaedter, J. Procyk, A. Chua, A. Forghani, and C. Pang. Mobile Video Conferencing for Sharing Outdoor Leisure Activities Over Distance. *HumanComputer Interaction*, 35(2):103–142, Mar. 2020. ISSN 0737-0024, 1532-7051. doi: 10.1080/07370024.2017.1314186. URL `https://www.tandfonline.com/doi/full/10.1080/07370024.2017.1314186`.

[184] M. Norman, G. Lee, R. T. Smith, and M. Billinghurst. A mixed presence collaborative mixed reality system. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 1106–1107, 2019. doi: 10.1109/VR.2019.8797966.

[185] M. I. Norton, D. Mochon, and D. Ariely. The ikea effect: When labor leads to love. *Journal of consumer psychology*, 22(3):453–460, 2012.

[186] M. Obaid, F. Kistler, G. Kasparaviciute, A. E. Yanta, and M. Fjeld. How would you gesture navigate a drone?: a user-centered approach to control a drone. In *Proceedings of the 20th International Academic Mindtrek Conference on - AcademicMindtrek '16*, pages 113–121, Tampere, Finland, 2016. ACM Press. ISBN 978-1-4503-4367-1. doi: 10.1145/2994310.2994348. URL `http://dl.acm.org/citation.cfm?doid=2994310.2994348`.

[187] C. S. Oh, J. N. Bailenson, and G. F. Welch. A Systematic Review of Social Presence: Definition, Antecedents, and Implications. *Frontiers in Robotics and AI*, 5:114, Oct. 2018. ISSN

2296-9144. doi: 10.3389/frobt.2018.00114. URL https://www.frontiersin.org/article/10.3389/frobt.2018.00114/full.

[188] T. Pallarino, A. Free, K. Mutuc, and S. Yarosh. Feeling distance: An investigation of mediated social touch prototypes. In *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion*, pages 361–364. ACM, 2016.

[189] R. Pan and C. Neustaedter. Streamer.Space: A Toolkit for Prototyping Context-Aware Mobile Video Streaming Apps. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '17*, pages 1947–1954, Denver, Colorado, USA, 2017. ACM Press. ISBN 978-1-4503-4656-6. doi: 10.1145/3027063.3053083. URL http://dl.acm.org/citation.cfm?doid=3027063.3053083.

[190] R. Pan, S. Singhal, B. E. Riecke, E. Cramer, and C. Neustaedter. "MyEyes": The Design and Evaluation of First Person View Video Streaming for Long-Distance Couples. In *Proceedings of the 2017 Conference on Designing Interactive Systems - DIS '17*, pages 135–146, Edinburgh, United Kingdom, 2017. ACM Press. ISBN 978-1-4503-4922-2. doi: 10.1145/3064663.3064671. URL http://dl.acm.org/citation.cfm?doid=3064663.3064671.

[191] K. Paraiso and V. Interrante. Can virtual human entourage elements facilitate accurate distance judgments in vr? In *Virtual Reality and Augmented Reality*, pages 119–133. Springer International Publishing, 2017. ISBN 978-3-319-72323-5.

[192] T. Pejsa, 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*, CSCW '16, pages 1716–1725, New York, NY, USA, 2016. ACM. ISBN 978-1-4503-3592-8.

doi: 10.1145/2818048.2819965. URL http://doi.acm.org/10.1145/2818048.2819965. event-place: San Francisco, California, USA.

[193] H. R. M. Pelikan, A. Cheatle, M. F. Jung, and S. J. Jackson. Operating at a Distance - How a Teleoperated Surgical Robot Reconfigures Teamwork in the Operating Room. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW):1–28, Nov. 2018. ISSN 25730142. doi: 10.1145/3274407. URL http://dl.acm.org/citation.cfm?doid=3290265.3274407.

[194] K. Pfeil, S. L. Koh, and J. LaViola. Exploring 3d gesture metaphors for interaction with unmanned aerial vehicles. In *Proceedings of the 2013 international conference on Intelligent user interfaces - IUI '13*, page 257, Santa Monica, California, USA, 2013. ACM Press. ISBN 978-1-4503-1965-2. doi: 10.1145/2449396.2449429. URL http://dl.acm.org/citation.cfm?doid=2449396.2449429.

[195] K. Pfeil, P. Wisniewski, and J. LaViola Jr. An analysis of user perception regarding body-worn 360° camera placements and heights. In *ACM Symposium on Applied Perception 2019*, SAP '19, pages 13:1–13:10, New York, NY, USA, 2019. ACM. ISBN 978-1-4503-6890-2. doi: 10.1145/3343036.3343120. URL http://doi.acm.org/10.1145/3343036.3343120.

[196] K. Pfeil, P. J. Wisniewski, and J. J. Laviola Jr. The effects of gender and the presence of third-party humans on telepresence camera height preferences. In *ACM Symposium on Applied Perception 2020*, SAP '20, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450376181. doi: 10.1145/3385955.3407924. URL https://doi.org/10.1145/3385955.3407924.

[197] K. Pfeil, S. Masnadi, J. Belga, J.-V. T. Sera-Josef, and J. LaViola. Distance perception with a video see-through head-mounted display. In *Proceedings of the 2021 CHI Conference on*

Human Factors in Computing Systems, CHI '21, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450380966. doi: 10.1145/3411764.3445223. URL https://doi.org/10.1145/3411764.3445223.

[198] K. P. Pfeil, N. Chatlani, J. J. LaViola, and P. Wisniewski. Bridging the socio-technical gaps in body-worn interpersonal live-streaming telepresence through a critical review of the literature. *Proc. ACM Hum.-Comput. Interact.*, 5(CSCW1), Apr. 2021. doi: 10.1145/3449194. URL https://doi.org/10.1145/3449194.

[199] J. M. Plumert, J. K. Kearney, J. F. Cremer, and K. Recker. Distance perception in real and virtual environments. *ACM Trans. Appl. Percept.*, 2(3):216233, July 2005. ISSN 1544-3558. doi: 10.1145/1077399.1077402. URL https://doi.org/10.1145/1077399.1077402.

[200] E. Price. How to shoot amazing pictures with a 360-degree camera, Jun 2017. URL https://lifehacker.com/how-to-shoot-amazing-pictures-with-a-360-degree-camera-1795928246.

[201] E. Price. 10 samsung gear 360 tips and tricks, May 2017. URL https://www.lifewire.com/tips-for-samsung-gear-360-4121408.

[202] J. Procyk, C. Neustaedter, C. Pang, A. Tang, and T. K. Judge. Exploring video streaming in public settings: shared geocaching over distance using mobile video chat. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14*, pages 2163–2172, Toronto, Ontario, Canada, 2014. ACM Press. ISBN 978-1-4503-2473-1. doi: 10.1145/2556288.2557198. URL http://dl.acm.org/citation.cfm?doid=2556288.2557198.

[203] P. Pulli, Z. Asghar, M. Siitonen, R. Niskala, E. Leinonen, A. Pitkänen, J. Hyry, J. Lehtonen, V. Kramar, and M. Korhonen. Mobile augmented teleguidance-based safety navigation

concept for senior citizens. In *2nd. International Conference on Applied and Theoretical Information Systems Research (2nd. ATISR2012)*, pages 1–9, 2012.

[204] I. Rae, L. Takayama, and B. Mutlu. The influence of height in robot-mediated communication. In *Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction*, pages 1–8. IEEE Press, 2013.

[205] I. Rae, G. Venolia, J. C. Tang, and D. Molnar. A Framework for Understanding and Designing Telepresence. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing - CSCW '15*, pages 1552–1566, Vancouver, BC, Canada, 2015. ACM Press. ISBN 978-1-4503-2922-4. doi: 10.1145/2675133.2675141. URL http://dl.acm.org/citation.cfm?doid=2675133.2675141.

[206] E. Ragan, C. Wilkes, D. A. Bowman, and T. Hollerer. Simulation of augmented reality systems in purely virtual environments. In *2009 IEEE Virtual Reality Conference*, pages 287–288. IEEE, 2009.

[207] E. D. Ragan, C. Wilkes, Y. Cao, and D. A. Bowman. The effects of virtual character animation on spatial judgments. In *2012 IEEE Virtual Reality Workshops (VRW)*, pages 141–142, March 2012. doi: 10.1109/VR.2012.6180921.

[208] RealGM. Nba players, 2019. URL https://basketball.realgm.com/nba/players.

[209] S. Reeves, C. Greiffenhagen, M. Flintham, S. Benford, M. Adams, J. Row Farr, and N. Tandavantij. I'd Hide You: Performing Live Broadcasting in Public. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*, pages 2573–2582, Seoul, Republic of Korea, 2015. ACM Press. ISBN 978-1-4503-3145-6. doi: 10.1145/2702123.2702257. URL http://dl.acm.org/citation.cfm?doid=2702123.2702257.

[210] J. Rekimoto, K. Uragaki, and K. Yamada. Behind-the-mask: A Face-through Head-mounted Display. In *Proceedings of the 2018 International Conference on Advanced Visual Interfaces*, pages 1–5, 2018.

[211] R. S. Renner, B. M. Velichkovsky, and J. R. Helmert. The perception of egocentric distances in virtual environments - a review. *ACM Comput. Surv.*, 46(2), Dec. 2013. ISSN 0360-0300. doi: 10.1145/2543581.2543590. URL https://doi.org/10.1145/2543581.2543590.

[212] G. Riva, F. Davide, and W. A. IJsselsteijn. Being there: The experience of presence in mediated environments. *Being there: Concepts, effects and measurement of user presence in synthetic environments*, 5, 2003.

[213] R. M. B. Rivera, C. B. Arbona, A. García-Palacios, S. Q. Castellano, and J. B. López. Treating emotional problems with virtual and augmented reality. *The handbook of the psychology of communication technology*, 32:548, 2015.

[214] R. Rocca. Here are the fines people could face for violating ontarios coronavirus shutdown laws, Dec 2020. URL https://globalnews.ca/news/7542586/ontario-coronavirus-lockdown-fines/.

[215] S. Rothe, B. Kegeles, M. Allary, and H. Hußmann. The impact of camera height in cinematic virtual reality. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*, page 124. ACM, 2018.

[216] S. Rothe, B. Kegeles, and H. Hussmann. Camera heights in cinematic virtual reality: How viewers perceive mismatches between camera and eye height. In *Proceedings of the 2019 ACM International Conference on Interactive Experiences for TV and Online Video*, TVX '19, pages 25–34, New York, NY, USA, 2019. ACM. ISBN 978-1-4503-6017-3.

doi: 10.1145/3317697.3323362. URL http://doi.acm.org/10.1145/3317697.3323362.

[217] A. Rugg and B. Burroughs. Periscope, live-streaming and mobile video culture. *Geoblocking and global video culture*, 18:64–73, 2016.

[218] C. S. Sahm, S. H. Creem-Regehr, W. B. Thompson, and P. Willemsen. Throwing versus walking as indicators of distance perception in similar real and virtual environments. *ACM Trans. Appl. Percept.*, 2(1):3545, Jan. 2005. ISSN 1544-3558. doi: 10.1145/1048687.1048690. URL https://doi.org/10.1145/1048687.1048690.

[219] M. M. Samson, A. Crowe, P. De Vreede, J. A. Dessens, S. A. Duursma, and H. J. Verhaar. Differences in gait parameters at a preferred walking speed in healthy subjects due to age, height and body weight. *Aging clinical and experimental research*, 13(1):16–21, 2001.

[220] Samsung. How to film the best 360 videos?, Nov 2016. URL https://www.samsung.com/ae/discover/how-to/how-to-film-the-best-360-videos/.

[221] M. Saraiji, T. Sasaki, R. Matsumura, K. Minamizawa, and M. Inami. Fusion: full body surrogacy for collaborative communication. In *ACM SIGGRAPH 2018 Emerging Technologies*, page 7. ACM, 2018.

[222] P. Sarconi. How to shoot a 360 video, Feb 2017. URL https://www.wired.com/2017/02/shoot-360-video/.

[223] S. Scheggi, A. Talarico, and D. Prattichizzo. A remote guidance system for blind and visually impaired people via vibrotactile haptic feedback. In *22nd Mediterranean Conference on Control and Automation*, pages 20–23. IEEE, 2014.

[224] M. F. Scheier and C. S. Carver. The self-consciousness scale: A revised version for use with general populations 1. *Journal of Applied Social Psychology*, 15(8):687–699, 1985.

[225] J. Schlueter and E. Winer. Expert-assisted field maintenance using augmented reality. In *Interservice/Industry Training, Simulation, and Education (I/ITSEC)*, 2017.

[226] T. Schubert, F. Friedmann, and H. Regenbrecht. The experience of presence: Factor analytic insights. *Presence: Teleoperators & Virtual Environments*, 10(3):266–281, 2001. Publisher: MIT Press.

[227] M. J. Schuemie, P. Van Der Straaten, M. Krijn, and C. A. Van Der Mast. Research on presence in virtual reality: A survey. *CyberPsychology & Behavior*, 4(2):183–201, 2001.

[228] N. Shade, 2018. URL `https://assetstore.unity.com/packages/3d/environments/showroom-environment-73740`.

[229] H. Shakeri and C. Neustaedter. Teledrone: Shared outdoor exploration using telepresence drones. In *Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing*, CSCW '19, page 367371, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450366922. doi: 10.1145/3311957.3359475. URL `https://doi.org/10.1145/3311957.3359475`.

[230] J. Short, E. Williams, and B. Christie. *The social psychology of telecommunications*. John Wiley & Sons, 1976.

[231] S. Singanamalla, W. Thies, and C. Scott. Avatar: Enabling Immersive Collaboration via Live Mobile Video. In *Proceedings of the 3rd International Workshop on Multimedia Alternate Realities - AltMM'18*, pages 9–14, Seoul, Republic of Korea, 2018. ACM Press. ISBN 978-1-4503-5979-5. doi: 10.1145/3268998.3269003. URL `http://dl.acm.org/citation.cfm?doid=3268998.3269003`.

[232] S. Singhal and C. Neustaedter. BeWithMe: An Immersive Telepresence System for Distance Separated Couples. In *Companion of the 2017 ACM Conference on Computer Sup-*

ported Cooperative Work and Social Computing - CSCW '17 Companion, pages 307–310, Portland, Oregon, USA, 2017. ACM Press. ISBN 978-1-4503-4688-7. doi: 10.1145/3022198.3026310. URL http://dl.acm.org/citation.cfm?doid=3022198.3026310.

[233] M. Sjöblom and J. Hamari. Why do people watch others play video games? an empirical study on the motivations of twitch users. *Computers in Human Behavior*, 75:985–996, 2017.

[234] M. Slater, M. Usoh, and A. Steed. Depth of presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, 3(2):130–144, 1994. Publisher: MIT Press.

[235] M. Slater, V. Linakis, M. Usoh, and R. Kooper. Immersion, presence and performance in virtual environments: An experiment with tri-dimensional chess. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, VRST 96, page 163172, New York, NY, USA, 1996. Association for Computing Machinery. ISBN 0897918258. doi: 10.1145/3304181.3304216. URL https://doi.org/10.1145/3304181.3304216.

[236] M. Smyth, I. Helgason, F. Kresin, M. Balestrini, A. B. Unteidig, S. Lawson, M. Gaved, N. Taylor, J. Auger, L. K. Hansen, et al. Maker movements, do-it-yourself cultures and participatory design: Implications for hci research. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–7, 2018.

[237] J. Stark. 12 tips for shooting and editing 360-degree video for journalism, Dec 2017. URL https://medium.com/@StanfordJournalism/12-tips-for-shooting-and-editing-360-degree-video-for-journalism-4972c

[238] H. Steltenpohl and A. Bouwer. Vibrobelt: tactile navigation support for cyclists. In *Proceedings of the 2013 international conference on Intelligent user interfaces*, pages 417–426. ACM, 2013.

[239] D. R. . Stewart and J. Littau. Up, Periscope: Mobile Streaming Video Technologies, Privacy in Public, and the Right to Record. *Journalism & Mass Communication Quarterly*, 93(2): 312–331, June 2016. ISSN 1077-6990, 2161-430X. doi: 10.1177/1077699016637106. URL http://journals.sagepub.com/doi/10.1177/1077699016637106.

[240] B. Stoll, S. Reig, L. He, I. Kaplan, M. F. Jung, and S. R. Fussell. Wait, Can You Move the Robot?: Examining Telepresence Robot Use in Collaborative Teams. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, HRI '18, pages 14–22, New York, NY, USA, 2018. ACM. ISBN 978-1-4503-4953-6. doi: 10.1145/3171221.3171243. URL http://doi.acm.org/10.1145/3171221.3171243. event-place: Chicago, IL, USA.

[241] D. Stout. Social media statistics 2018: What you need to know, 2018. URL https://dustn.tv/social-media-statistics/.

[242] J. E. Swan, A. Jones, E. Kolstad, M. A. Livingston, and H. S. Smallman. Egocentric depth judgments in optical, see-through augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, 13(3):429–442, 2007.

[243] A. Tang, O. Fakourfar, C. Neustaedter, and S. Bateman. Collaboration with 360 Videochat: Challenges and Opportunities. In *Proceedings of the 2017 Conference on Designing Interactive Systems - DIS '17*, pages 1327–1339, Edinburgh, United Kingdom, 2017. ACM Press. ISBN 978-1-4503-4922-2. doi: 10.1145/3064663.3064707. URL http://dl.acm.org/citation.cfm?doid=3064663.3064707.

[244] J. Tang, G. Venolia, K. Inkpen, C. Parker, R. Gruen, and A. Pelton. Crowdcasting: Remotely Participating in Live Events Through Multiple Live Streams. *Proceedings of the ACM on Human-Computer Interaction*, 1(CSCW):1–18, Dec. 2017. ISSN 25730142. doi:

10.1145/3134733. URL http://dl.acm.org/citation.cfm?doid=3171581.3134733.

[245] P. Team. Periscope, by the numbers, Aug 2015. URL https://medium.com/periscope/periscope-by-the-numbers-6b23dc6a1704.

[246] F. Tecchia, L. Alem, and W. Huang. 3D helping hands: a gesture based MR system for remote collaboration. In *Proceedings of the 11th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry - VRCAI '12*, page 323, Singapore, Singapore, 2012. ACM Press. ISBN 978-1-4503-1825-9. doi: 10.1145/2407516.2407590. URL http://dl.acm.org/citation.cfm?doid=2407516.2407590.

[247] A. Tewari, J. Peabody, R. Sarle, G. Balakrishnan, A. Hemal, A. Shrivastava, and M. Menon. Technique of da vinci robot-assisted anatomic radical prostatectomy. *Urology*, 60(4):569–572, Oct. 2002. ISSN 00904295. doi: 10.1016/S0090-4295(02)01852-6. URL https://linkinghub.elsevier.com/retrieve/pii/S0090429502018526.

[248] W. B. Thompson, P. Willemsen, A. A. Gooch, S. H. Creem-Regehr, J. M. Loomis, and A. C. Beall. Does the quality of the computer graphics matter when judging distances in visually immersive environments? *Presence: Teleoperators and Virtual Environments*, 13 (5):560–571, 2004. doi: 10.1162/1054746042545292.

[249] B. Thoravi Kumaravel, F. Anderson, G. Fitzmaurice, B. Hartmann, and T. Grossman. Loki: Facilitating Remote Instruction of Physical Tasks Using Bi-Directional Mixed-Reality Telepresence. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, pages 161–174, New Orleans LA USA, Oct. 2019. ACM. ISBN 978-1-4503-6816-2. doi: 10.1145/3332165.3347872. URL https://dl.acm.org/doi/10.1145/3332165.3347872.

[250] H. Tobita. Gutsy-Avatar: Computational Assimilation for Advanced Communication and Collaboration. In *2017 First IEEE International Conference on Robotic Computing (IRC)*, pages 8–13, Taichung, Taiwan, Apr. 2017. IEEE. ISBN 978-1-5090-6724-4. doi: 10.1109/IRC.2017.82. URL `http://ieeexplore.ieee.org/document/7926508/`.

[251] K. Tsui, A. Norton, D. Brooks, H. Yanco, and D. Kontak. Designing telepresence robot systems for use by people with special needs. In *Int. Symposium on Quality of Life Technologies: Intelligent Systems for Better Living*, 2011.

[252] K. Tsukada and M. Yasumura. Activebelt: Belt-type wearable tactile display for directional navigation. In *UbiComp*, volume 3205, pages 384–399. Springer, 2004.

[253] Y. Tsumaki, F. Ono, and T. Tsukuda. The 20-DOF miniature humanoid MH-2: A wearable communication system. In *2012 IEEE International Conference on Robotics and Automation*, pages 3930–3935, St Paul, MN, USA, May 2012. IEEE. ISBN 978-1-4673-1405-3 978-1-4673-1403-9 978-1-4673-1578-4 978-1-4673-1404-6. doi: 10.1109/ICRA.2012.6224810. URL `http://ieeexplore.ieee.org/document/6224810/`.

[254] Twitch. 2017 year in review, 2017. URL `https://www.twitch.tv/year/2017/factsheet.jpg`.

[255] F. Uenuma. Video Chat Is Helping Us Stay Connected in Lockdown. But the Tech Was Once a Spectacular Flop'. URL `https://time.com/5834516/video-chat-zoom-history/`.

[256] J. Van Baren. Measuring presence: A guide to current measurement approaches. *Deliverable of the OmniPres project IST-2001-39237*, 2004.

[257] J. B. Van Erp, H. A. Van Veen, C. Jansen, and T. Dobbins. Waypoint navigation with

a vibrotactile waist belt. *ACM Transactions on Applied Perception (TAP)*, 2(2):106–117, 2005.

[258] T. van Oosterhout and A. Visser. A visual method for robot proxemics measurements. In *Proceedings of Metrics for Human-Robot Interaction: A Workshop at the Third ACM/IEEE International Conference on Human-Robot Interaction (HRI 2008). Citeseer*, pages 61–68, 2008.

[259] K. Vaziri, P. Liu, S. Aseeri, and V. Interrante. Impact of visual and experiential realism on distance perception in vr using a custom video see-through system. In *Proceedings of the ACM Symposium on Applied Perception*, SAP '17, New York, NY, USA, 2017. Association for Computing Machinery. ISBN 9781450351485. doi: 10.1145/3119881.3119892. URL https://doi.org/10.1145/3119881.3119892.

[260] K. Venkaiah, K. Damayanti, M. Nayak, and K. Vijayaraghavan. Diet and nutritional status of rural adolescents in india. *European journal of clinical nutrition*, 56(11):1119, 2002.

[261] V. Venkatesh and F. D. Davis. A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management science*, 46(2):186–204, 2000.

[262] G. Venolia, J. C. Tang, K. Inkpen, and B. Unver. Wish you were here: being together through composite video and digital keepsakes. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 1–11, Barcelona Spain, Sept. 2018. ACM. ISBN 978-1-4503-5898-9. doi: 10.1145/3229434.3229476. URL https://dl.acm.org/doi/10.1145/3229434.3229476.

[263] S. Vu, M. J. Rissanen, N. Pang, and S. Foo. Towards evaluating social telepresence in mobile context. In *Proceedings of the 11th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry - VRCAI '12*, page 75, Singapore,

Singapore, 2012. ACM Press. ISBN 978-1-4503-1825-9. doi: 10.1145/2407516.2407539. URL http://dl.acm.org/citation.cfm?doid=2407516.2407539.

[264] G. Walsh, E. Foss, J. Yip, and A. Druin. *FACIT PD: A Framework for Analysis and Creation of Intergenerational Techniques for Participatory Design*, page 28932902. Association for Computing Machinery, New York, NY, USA, 2013. ISBN 9781450318990. URL https://doi.org/10.1145/2470654.2481400.

[265] H.-C. Wang, R. K. Katzschmann, S. Teng, B. Araki, L. Giarré, and D. Rus. Enabling independent navigation for visually impaired people through a wearable vision-based feedback system. In *2017 IEEE international conference on robotics and automation (ICRA)*, pages 6533–6540. IEEE, 2017.

[266] V. Wanick, G. Xavier, and E. Ekmekcioglu. Virtual Transcendence Experiences: Exploring Technical and Design Challenges in Multi-Sensory Environments. In *Proceedings of the 10th International Workshop on Immersive Mixed and Virtual Environment Systems - MMVE '18*, pages 7–12, Amsterdam, Netherlands, 2018. ACM Press. ISBN 978-1-4503-5771-5. doi: 10.1145/3210438.3210444. URL http://dl.acm.org/citation.cfm?doid=3210438.3210444.

[267] S. Webel, U. Bockholt, T. Engelke, N. Gavish, M. Olbrich, and C. Preusche. An augmented reality training platform for assembly and maintenance skills. *Robotics and Autonomous Systems*, 61(4):398–403, Apr. 2013. ISSN 09218890. doi: 10.1016/j.robot.2012.09.013. URL https://linkinghub.elsevier.com/retrieve/pii/S0921889012001674.

[268] N. Weibel, D. Gasques, J. Johnson, T. Sharkey, Z. R. Xu, X. Zhang, E. Zavala, M. Yip, and K. Davis. ARTEMIS: Mixed-Reality Environment for Immersive Surgical Telementoring. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Com-*

puting Systems, pages 1–4, Honolulu HI USA, Apr. 2020. ACM. ISBN 978-1-4503-6819-3. doi: 10.1145/3334480.3383169. URL `https://dl.acm.org/doi/10.1145/3334480.3383169`.

[269] A. Wilberz, D. Leschtschow, C. Trepkowski, J. Maiero, E. Kruijff, and B. Riecke. FaceHaptics: Robot Arm based Versatile Facial Haptics for Immersive Environments. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–14, Honolulu HI USA, Apr. 2020. ACM. ISBN 978-1-4503-6708-0. doi: 10.1145/3313831.3376481. URL `https://dl.acm.org/doi/10.1145/3313831.3376481`.

[270] P. Willemsen, M. B. Colton, S. H. Creem-Regehr, and W. B. Thompson. The effects of head-mounted display mechanical properties and field of view on distance judgments in virtual environments. *ACM Trans. Appl. Percept.*, 6(2), Mar. 2009. ISSN 1544-3558. doi: 10.1145/1498700.1498702. URL `https://doi.org/10.1145/1498700.1498702`.

[271] B. G. Witmer and M. J. Singer. Measuring presence in virtual environments: A presence questionnaire. *Presence*, 7(3):225–240, 1998. Publisher: MIT Press.

[272] J. O. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 143–146. ACM, 2011.

[273] X. Wu, R. Thomas, E. Drobina, T. Mitzner, and J. Beer. An evaluation of a telepresence robot: User testing among older adults with mobility impairment. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, pages 325–326. ACM, 2017.

[274] L. Yang, B. Jones, C. Neustaedter, and S. Singhal. Shopping Over Distance through a Telepresence Robot. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW):

1–18, Nov. 2018. ISSN 25730142. doi: 10.1145/3274460. URL http://dl.acm.org/citation.cfm?doid=3290265.3274460.

[275] S. Yarosh, K. M. Inkpen, and A. B. Brush. Video playdate: toward free play across distance. In *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10*, page 1251, Atlanta, Georgia, USA, 2010. ACM Press. ISBN 978-1-60558-929-9. doi: 10.1145/1753326.1753514. URL http://portal.acm.org/citation.cfm?doid=1753326.1753514.

[276] S. Yarosh, P. Markopoulos, and G. D. Abowd. Towards a questionnaire for measuring affective benefits and costs of communication technologies. CSCW '14, page 8496, New York, NY, USA, 2014. Association for Computing Machinery. ISBN 9781450325400. doi: 10.1145/2531602.2531634. URL https://doi.org/10.1145/2531602.2531634.

[277] J. Young, T. Langlotz, M. Cook, S. Mills, and H. Regenbrecht. Immersive Telepresence and Remote Collaboration using Mobile and Wearable Devices. *IEEE Transactions on Visualization and Computer Graphics*, 25(5):1908–1918, May 2019. ISSN 1077-2626, 1941-0506, 2160-9306. doi: 10.1109/TVCG.2019.2898737. URL https://ieeexplore.ieee.org/document/8642375/.

[278] J. Young, T. Langlotz, S. Mills, and H. Regenbrecht. Mobileportation: Nomadic Telepresence for Mobile Devices. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(2):1–16, June 2020. ISSN 2474-9567, 2474-9567. doi: 10.1145/3397331. URL https://dl.acm.org/doi/10.1145/3397331.

[279] C. Zeagler. Where to wear it: Functional, technical, and social considerations in on-body location for wearable technology 20 years of designing for wearability. In *Proceedings of the 2017 ACM International Symposium on Wearable Computers*, ISWC '17, page 150157, New York, NY, USA, 2017. Association for Computing Machinery.

ISBN 9781450351881. doi: 10.1145/3123021.3123042. URL https://doi.org/10.1145/3123021.3123042.

[280] N. Zenati-Henda, A. Bellarbi, S. Benbelkacem, and M. Belhocine. Augmented reality system based on hand gestures for remote maintenance. In *2014 International Conference on Multimedia Computing and Systems (ICMCS)*, pages 5–8, Marrakech, Morocco, Apr. 2014. IEEE. ISBN 978-1-4799-3824-7 978-1-4799-3823-0. doi: 10.1109/ICMCS.2014.6911258. URL http://ieeexplore.ieee.org/document/6911258/.