Exploring in the City with Your Personal Guide: Design and User Study of T-Leap, a Telepresence System

Minori Manabe
The University of Tokyo
minori@star.rcast.u-tokyo.ac.jp

Daisuke Uriu
The University of Tokyo
uriu@star.rcast.u-tokyo.ac.jp

Takeshi Funatsu
The University of Tokyo
takeshi.funatsu@star.rcast.u-tokyo.ac.jp

Atsushi Izumihara
The University of Tokyo
izumihara@star.rcast.u-tokyo.ac.jp

Takeru Yazaki
The University of Tokyo
takeru.yazaki@star.rcast.u-tokyo.ac.jp

I-Hsin Chen
National Cheng Kung University
p36084027@gs.ncku.edu.tw

Yi-Ya Liao
National Taipei University of Technology
simon7531109@gmail.com

Kang-Yi Liu
National Taipei University of Technology
kouichi860612@gmail.com

Ju-Chun Ko
National Taipei University of Technology
dablog@gmail.com

Zendai Kashino
The University of Tokyo
kashino@star.rcast.u-tokyo.ac.jp

Atsushi Hiyama
The University of Tokyo
hiyama@star.rcast.u-tokyo.ac.jp

Masahiko Inami
The University of Tokyo
inami@star.rcast.u-tokyo.ac.jp

Figure 1: Overview of our field study using the T-Leap system in Taipei city. Three Nodes exploring outside (Left to Right: N-P5, 6, 7) may communicate with Viewer (The far right: V-P2), to collaboratively complete tasks.

ABSTRACT

This paper describes a field study conducted with our system, T-Leap, a telepresence system connecting one person (the Viewer), situated indoors, with multiple destinations (the Nodes), that roam outdoors. Here, each Node is a person wearing a module that includes a 360-degree camera and a microphone-speaker. Through our study, we demonstrate that T-Leap enables the Viewer to perform various interactions with the Nodes including being helped by them, collaborating with them, and guiding them. These interactions were demonstrated through three studies completing different tasks: 1) Nodes purchasing souvenirs for the Viewer, 2) Nodes finding objects in the park, and 3) Viewer guiding Nodes to purchase things. The studies were primarily conducted with Taiwanese locals and Japanese visitors in Taipei. Throughout the studies, we found that T-Leap worked especially well for mediating communication between a Viewer with local knowledge acting as a guide and several Nodes who were being guided. To conclude the paper, we broadly discuss our findings, the lessons we learned from our field study, and present recommendations for the future development of mobile and wearable telepresence systems.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in ubiquitous and mobile computing**; **Empirical studies in collaborative and social computing**.

MUM 2020, November 22–25, 2020, Essen, Germany

ACM ISBN 978-1-4503-8870-2/20/11…$15.00
https://doi.org/10.1145/3428361.3428382

KEYWORDS

Remote Communication, Telepresence, Research through Design

ACM Reference Format:
Minori Manabe, Daisuke Uriu, Takeshi Funatsu, Atsushi Izumihara, Takeru Yazaki, I-Hsin Chen, Yi-Ya Liao, Kang-Yi Liu, Ju-Chun Ko, Zendai Kashino, Atsushi Hiyama, and Masahiko Inami. 2020. 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), November 22–25, 2020, Essen, Germany*. ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3428361.3428382

1 INTRODUCTION

Research on remote communications have been conducted over the last thirty years in various research communities [2, 9, 11, 20, 29]. In the early age, researchers focused on using site-fixed devices such as remote meeting systems connecting specific rooms [2]. Responding to the needs of the business community, site-fixed remote conference systems have been developed for not only one-to-one but also many-to-many communications [2]. In comparison, research on mobile remote communication systems mainly focus on one-to-one communications (e.g., [9, 11]) with few researchers focusing on one-to-many or many-to-many remote communication systems [20].

In pursuit of cutting-edge technologies for remote communication systems, researchers have recently been working on developing novel telepresence devices with enhanced expressiveness. For example, some recent works have tried to make human-like robots to remotely replicate human expressive behavior [1, 26]. However, these devices are often expensive and complicated, and are not suited for practical use outside of the controlled environment of a research laboratory. Other researchers, in contrast, have preferred to develop simple devices for examining how mobile telepresence systems might work when used in a non-laboratory setting [17, 23]. Even these works, however, typically consider the one-to-one communication scenario, as noted above, despite the prevalence of real-world scenarios which could benefit from one-to-many or many-to-many mobile telepresence systems.

In this paper, we present a field study conducted to investigate the usage of a one-to-many mobile telepresence system in a practical setting, Figure 1. The telepresence system was designed to mediate communication between a single Viewer, situated indoors, and multiple Nodes, persons roaming the outdoors. The practical setting chosen for this field study was that of casual leisure. Specifically we focused on casual leisure activities that could be performed in a city context such as shopping and exploration. The objective of this study was to investigate how, and what kind of mobile telepresence system could enhance the user experience during these activities.

The telepresence system used in our field study, Tag-Leap (T-Leap), is a novel one whose design we also propose in this paper. Through T-Leap, the Viewer is able to *tag along* with any Node and *leap* between Nodes at will. This experience is enabled by the Node module provided to each Node. The Node module is a wearable device that consists of a remotely accessible 360-degree camera and microphone-speaker.

Our proposed system is a simple and minimal telepresence system which asymmetrically connects a single Viewer to multiple Nodes. The proposed system is simple in that it only makes use of commercially available products with minimal custom hardware. The system is asymmetric in that the Viewer receives significantly more information from the Nodes than the Node does from the Viewer. Additionally, the Viewer is provided with significantly more control over the communication than the Node is. The effectiveness of our proposed design was thoroughly examined through several scenarios conducted in the field in Taipei city, Taiwan.

In this paper, we first describe the T-Leap system and the experiments we conducted to investigate the usage of T-Leap in a practical setting. We then summarize the results of our experiments in terms of our findings during the scenarios. Finally, we provide a discussion outlining our thoughts on how less (limited functionality) can mean more (higher efficacy) for mobile remote telepresence systems and our recommendations for those seeking to future mobile and wearable telepresence systems.

2 RELATED WORK

This paper's primary contributions are to the field of telepresence/telexistence. More specifically, it contributes a novel wearable telepresence system and the results of a field study conducted to evaluate its use in a practical setting. The related works below, therefore, provide an overview of telepresence/telexistence, wearable telepresence/telexistence systems, and some recent trends in the development of novel devices for remote communication.

2.1 Telepresence and Telexistence

Research into telepresence and telexistence focuses on sending the presence of a person to remotely connected sites. Researchers in this area often focus on producing a richer sense of presence for remotely connected users than is achievable with conventional video communication systems. A wide range of methods for "sending" a sense of presence have been investigated and developed. For example, recent works have developed mobile robotic technologies for sending gestures [1, 26], facial expressions [1, 25, 28], mobility controls [4, 14, 28], and physical movements [21, 22], in addition to audio/video communications to improve the user's sense of presence at the remote location.

These mobile telepresence / telexistence robots, however, are often not suitable for use in uncontrolled settings. They can, for example, be physically disturbed or impeded by obstacles and require human assistance [4]. While this issue could be resolved by equipping mobile telepresence robots with highly flexible mobility, this is known to be very challenging. Furthermore, even if these robots could move flexibly and be commonly accepted in the future, navigational and robot control challenges remain. This is especially true in crowded situations where many people may be in the vicinity of the robot. In this case, the robot risks colliding with the surrounding people and the people become ' obstacles' for remote users [18].

2.2 Wearable Telepresence/Telexistence

One approach to resolving the abovementioned mobility issue is to place telepresence modules on people as they are already capable of highly flexible mobility. Several telepresence systems already take this approach to enable highly mobile remote presence. "Polly"

[11], for example, is a shoulder mounted telepresence device that enables remote users to control the direction of their view to explore the environment. "TEROOS" [9] is a remote avatar mounted on the shoulder whose head direction can be remotely controlled by a user. "MH-2" [26], designed as a telecommunicator, supports a remotely connected user by expressing gestures. On the other hand, "Gusty-Avatar" [25], a light-weight system whose functions are only indicating direction and timing, realizes the assimilation of multiple remotely connected persons.

These approaches are particularly popular for remote communications between users indoors and outdoors. Instead of being active outdoors with a potentially cumbersome robotic telepresence platform, users will carry the telepresence system with them. Examples of this kind of telepresence can be found in [17], where it was used for a collaborative treasure hunt, and [10] where it was used for remote shopping and walking around town. These research works showed that shared vision in the form of a video feed increased the sense of connectivity between the local and remote users. Furthermore, this kind of remote communication has been shown to induce entirely new interactions caused by the shared visual information. Examples include the remote user asking the local user for help, and requesting a better view of objects.

However, one of the critical difficulties in communication mediated by video is the video stability. The quality of experience for the remote users is strongly dependent on the video quality and stability [5]. For the video to remain stable and provide a good field of view to the remote user, the local users must carefully control a camera at all times. Without a local user adept in camera control, the experience can be highly stressful for both parties.

To address the above mentioned problem of camera views in wearable telepresence devices, several research works have proposed the use of 360-degree cameras that enable users to freely control the remote field of view without asking others at the site to change camera angles. An example of this can be found in [23], where a remote guide was able to effectively assist visitors' by remotely looking around inside the museum .

Other researchers have also shown that the viewpoint independence and interactivity afforded to viewers by using 360-degree cameras can bring a rich sense of presence and immersive experience to the remotely connected users [4, 16, 29]. For example, the "JackIn" projects [6–8] tackled the problem of giving remote users stabilized first-person views. Their work showed that this remote interaction induces proactive behaviors of the local user and parallel task transaction to enhance the collaboration. "MM360" [20] allows multiple users to jump around among multiple 360-degree moving images taken and streamed by each of them.

Other works have built upon the use of 360-degree cameras to propose novel telepresence systems that improve the communication experience. Some works focus on allowing the local user to see where the remote user is looking. For example, *Shamma et al.* focused on the awareness of remote participants in a remote conference scenario. They proposed a system design that indicated the viewing direction of remote participants with an LED array [19]. *Angelo et al.* showed eye-gaze sharing can improve remote collaboration [3].

Other researchers have focused on improving the remote viewer's experience with additional functionality. *Unver et al.*, for example,

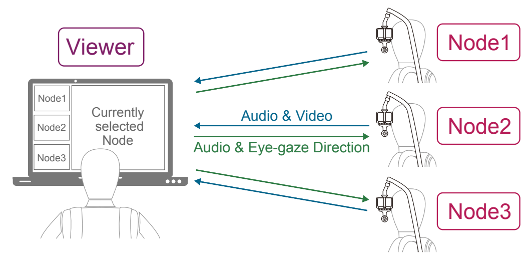

Figure 2: The Viewer is connected to multiple Nodes. The Viewer chooses one Node to connect.

indicated that allowing the remote viewer to change viewpoints made the viewing experience more enjoyable and voluntary in live video streaming scenarios [27]. "OmniGlobe," was an interface connecting remote spaces with a full spherical display [12]. It allowed a viewer to understand what was happening on all sides of the sphere with one glance by projecting a birds eye view of the spherical footage onto the top of the sphere. As shown in these research works, the ideal design of a remote communication system varies depending on the scenario. Furthermore, it is clear from the literature that there are many new interactions between local and remote persons that have yet to be explored. Our work presented herein demonstrates the new interactions which occurred during our field studies conducted using our T-Leap system, described in the next section.

3 METHODOLOGY

Methodologically, our research can be classified as first-person research [13] through design [30]. Namely, we, the authors, were the primary participants in the field study scenarios and we sought discoveries by designing the T-Leap system. We decided to conduct first-person research so that we could study behavior and interactions between well-acquainted individuals in a variety of scenarios. Furthermore, the first-person approach to research was taken to minimize issues encountered during the field study. As the field study described herein was the first experiment utilizing the T-Leap system, we expected to encounter unexpected troubles which would be difficult for a random participant to resolve. We acknowledge that taking the first-person approach means that our results likely include some positive bias. However, we believe that the insights gained through our research are still valid and worth considering by future works.

Our research through design approach began with an initial system design. This was followed by preliminary experiments which resulted in several adjustments to the design. We describe the final T-Leap system, as well as the adjustments made to the original system due to discoveries in the preliminary experiments, in Section 3.1. The final design was then used in a field study where the system was deployed and used by the participants to complete three scenarios. The field study is described in Section 3.2. The results of the field study (i.e., our findings) are detailed in Section 4.

3.1 The Tag-Leap (T-Leap) System

The Tag-Leap (T-Leap) system is a one-to-many asymmetric telepresence system that connects a single Viewer to multiple Nodes (one at a time). Figure 2 shows the system concept. The figure

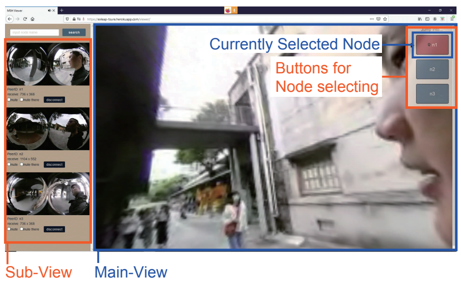

Figure 3: The UI presented to the Viewer on a laptop.

shows a single Viewer connecting to three Nodes wearing the Node (telepresence) module via a laptop interface. It is assumed that the Viewers are stationary indoors while the Nodes roam outdoors.

T-Leap is asymmetric two aspects: information flow and control. In terms of information flow, the system is asymmetric since the Viewer receives visual and audio information from the Nodes while the Nodes only receive audio information and a simple representation of viewing direction from the Viewer, Figure 2. We chose to create this asymmetry in information flow due to past research indicating that there is a natural asymmetry in information received by people, from their environment, when indoors and outdoors [1, 23]. The research showed that people outdoors receive significantly more information through their senses from their surroundings than people inside when these people are remotely connected. By creating an asymmetry in information flow that was skewed in the opposite direction (i.e., providing the Viewer with more information than the Node), we sought to provide the Viewer with a rich experience while not overstimulating and/or distracting the Node.

In terms of control, the system is asymmetric as the Viewer has full control over which Node to connect to at any time. The Node, in contrast, is only able to use a mobile messaging application, LINE, to contact the Viewer and request a connection. Simple, sticker-based communication via LINE was allowed after several preliminary experiments where the Node felt stressed about their inability to contact the Node. In one scenario conducted in Taipei, a Node, who could not speak Chinese, was queried by a local shop employee about the telepresence module they were wearing. This caused significant stress for the Node as they had no way of contacting the Viewer for help. In another scenario, a Node felt stress while waiting idly for the Viewer to contact them due to the lack of direction.

T-Leap is both simple and minimalistic in that it makes use of easily obtainable consumer products and has very few custom-made components. On the Viewer side, the only hardware required is a computer with an internet connection with the Viewer interface installed. The interface is described in further detail in Section 3.1.1. On the Node side, Nodes are provided with a Node module which is capable of recording 360-degree video and audio, and streaming the data to the Viewer interface. The Node module is described in further detail in Section 3.1.2.

3.1.1 The Viewer interface. The Viewer interface is an original software package that enables a remote user to switch between communicating with multiple Nodes. As seen in Figure 3, the Viewer interface consists of a Sub-View panel, a Main-View panel, and a set of Node selection buttons. The Sub-View panel shows the full 360-degree view obtained from each Node in dual fisheye format.

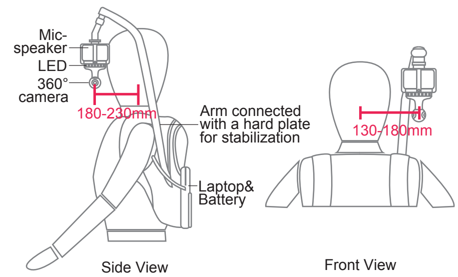

Figure 4: The design, components, and position of Node module

Using the Sub-View, the Viewer can get a sense of each and every Node's situation without having to switch to them. The Main-View panel shows a controllable first-person view from the currently selected Node. The Viewer is able to control the direction of this view using a mouse. Finally, the Node Selection buttons provide a simple way to change between Nodes. Clicking the appropriate button changes the first-person view displayed on the Main-View to that of the selected Node.

3.1.2 The Node module. The Node module, a wearable telepresence module, allows the Nodes to communicate with the Viewer. A schematic diagram of the module is shown in Figure 4. The Node module consists of a 360-degree camera (Insta360 Air), a video conference microphone/speaker (EMeet M0), a strip of LED tape controlled by an Arduino, and a tablet (Microsoft Surface Go). The tablet is bundled into a backpack to maximize portability and not impede the Node's motion. The tablet is used to stream video and sound from the camera and microphone to the Viewer and transmit audio from the Viewer to the speaker. The 360-degree camera, conference microphone/speaker, and LED strip are mounted on the end of an arm extending from the backpack. The arm is stabilized by a hard plate in the backpack such that it moves with the Node's upper body. The LED strip is arranged and controlled such that the lights indicate which direction the Viewer is looking using the Viewer interface. This function helps the Nodes to cooperate with the Viewer through the 360-degree cameras [19].

The arm is designed to hold the camera, microphone/speaker, and LED strip beside the Node's head. It is known that, while the height at which a 360-degree camera is placed does not significantly impact the viewer experience, the position on the body where the cameras is mounted has a significant impact on the remote viewer's experience [15]. For example, placing the camera inside a breast pocket evokes a feeling of being the avatar and induces active exploration behavior, while shoulder and overhead placements cause passive exploration behaviors due to a sense of being in a third-person viewpoint. In this work, we wanted the Viewer to feel as if they were walking together with the Node (i.e., that they were tagging along with the Node). A camera position that induced this feeling was found through preliminary experiments conducted in Tokyo. Through these preliminary experiments, we found that the dimensions indicated in Figure 4 allowed the Viewer to sense the Node's presence and feel as if they were walking beside the Node. After an appropriate camera position was determined, the speaker pickup threshold was also adjusted so that it would not pick up ambient noise in the outdoors.

3.2 The Field Study

The field study with the final T-Leap design was designed to demonstrate how the T-Leap system works and identify scenarios where T-Leap might be effective at facilitating remote cooperation. Prior to conducting our field study, we conjectured that the efficacy of T-Leap as a mobile telepresence system would be highlighted if there was a disparity in knowledge between the Viewer and Nodes. We identified geographical and linguistic knowledge as the primary factors which could strongly influence the user experience while performing the casual leisure activities we chose to focus on for this study (i.e., shopping and exploration). As such, we chose to conduct our field study in Taipei city, Taiwan, a location local to roughly half of the participants and foreign to the other half. Specifically, the Taiwanese authors were both familiar with the geography of the city and knew the local language while the Japanese authors had never been to the city and had no knowledge of the local language. Further knowledge discrepancies were created between the participants of the study and the local people by forcing participants to use English as their language of communication. The study was carried out in December 2019.

Three scenarios modeled after real casual leisure scenarios were conducted in the field study. In each scenario, all participants started in the university laboratory. Then, the Nodes left the laboratory to go to the location of the scenario while the Viewer remained in the laboratory. After the scenario, the Nodes returned to the laboratory. Each scenario lasted approximately one hour.

All scenarios were recorded in three ways: 1) by capturing the screen of the PC used by the Viewer, 2) by recording the Viewer with an iPhoneXR from a third person perspective, and 3) by on-site video recording with GoPro 8 cameras attached to each Node's front. The recordings span approximately three hours, and constitute the primary data obtained from this study. All conversations between Viewer and Nodes were transcribed from these sound and video recordings. Important scenes were selected from these transcriptions and analyzed by the authors. This study has been ethically approved by the University of Tokyo (No.19-40).

3.2.1 Participants. The field study was conducted with eight participants (P1-P8) consisting of the authors and their colleagues. P1, P2, and P3 spoke Japanese as their first language, knew no Mandarin, and had never been to Taipei. In contrast, P4, P5, P6, P7, and P8 were native to Taipei and spoke fluent Mandarin.

Hereafter, we will denote a participant who is acting as the Viewer as V-P# (e.g., when P1 is the Viewer, we denote them as V-P1). Similarly, we will denote a participant who is acting as a Node as N-P# (e.g., when P1 is a Node, we denote them as N-P1).

3.2.2 Scenario 1 (S1): A Demanding Viewer! In this scenario, the Viewer asks the Nodes to help in choosing and buying souvenirs at the Huasan 1914 Creative Park. The aim of this study was to observe the participants' behaviors when attempting to remotely buy souvenirs, a common situation in sightseeing trips. In this study, P2 was the Viewer (V-P2), and P5, P6, P7, were the Nodes (N-P5, N-P6, N-P7). V-P2 had never been to the Park, did not know the shops there, and had no concrete idea of the souvenirs they wanted. The Nodes knew the area and were able to communicate with the local people in Mandarin.

3.2.3 Scenario 2 (S2): Collaborative Exploration. In this scenario, we asked the Viewer and Nodes to collaboratively find three landmarks in Daan Park. This scenario aimed to replicate the experience of traveling to and exploring a place which is new to both the Viewer and Nodes. In this study, P6 was the Viewer (V-P6) and P1, P3, P7 were Nodes (N-P1, N-P3, N-P7). V-P6 spoke Mandarin but had never been to the park. N-P1 and N-P3 did not speak Mandarin and were also not familiar with the park. N-P7 spoke Mandarin but did not speak it for the duration of this scenario to simulate uniform language knowledge across the Nodes and promote conversation between the Viewer and the local people. N-P7 was also unfamiliar with the park. None of the participants knew where the objects were. The Nodes took a taxi to go to the park and return to the lab.

3.2.4 Scenario 3 (S3): The Professional Guide. In this scenario, we asked the Viewer to guide the Nodes to buy items they needed in an area around Bade Street where there were a many electronics shops. In this scenario, P5 was the Viewer (V-P5) and P1, P2, P4, were Nodes (N-P1, N-P2, N-P4). V-P5 knew the area very well as he frequently visited the location to shop in person. The Nodes, however, had never been to the area. At the start of the study, the Viewer asked the Nodes what they wanted to buy. The responses were: "a cable to connect HDMI with lightening" (N-P1), "a smartphone case" (N-P2), and "something weird for a Christmas present" (N-P4).

4 FINDINGS

There were three main findings we obtained from our studies. Firstly, we found that, the best overall performance was achieved when the Viewer was familiar with the local area. In this case, they were able to assist the Nodes through T-Leap and enabled the participants to efficiently complete the task at hand. Secondly, we found that spontaneous collaboration occurred between the Viewer and Nodes for tasks where neither side was at an advantage due to their greater geographical or linguistic knowledge. Thirdly, we found that, Nodes were able to remotely give the Viewer the experience of being on-site via T-Leap, but that this required relatively high quality video through a strong network connection if the Viewer was not familiar with the location.

Other findings of interest that are reported herein include how the local people appeared to interpret T-Leap, how the 360-degree camera was used and positioned effectively, and the variance in the cognitive load on the Viewer based on their level of local knowledge.

4.1 The Knowledgeable Viewer: an Effective Guide

The Viewer contributed most to the task at hand through T-Leap when they had local knowledge. In the following subsections, we discuss our findings on how the Viewer made use of their geographical and linguistic knowledge to remotely assist the Viewer.

4.1.1 When the Viewer had geographical knowledge. We observed several scenes throughout the scenarios in which the Viewer was able to support the Node(s) using their geographical knowledge (shops, roads, locations, etc.). For example, in S2, there was a scene where N-P1 did not know where to pick up a taxi to go to the Daan Park. V-P6 and P8 (who happened to be near V-P6 at the time), who

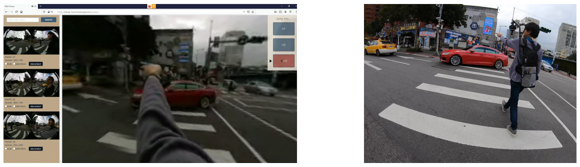

Figure 5: N-P1 finding a place to take a taxi while being guided by V-P6 in S2. (Left: Viewer's view, Right: A photo taken on-site)

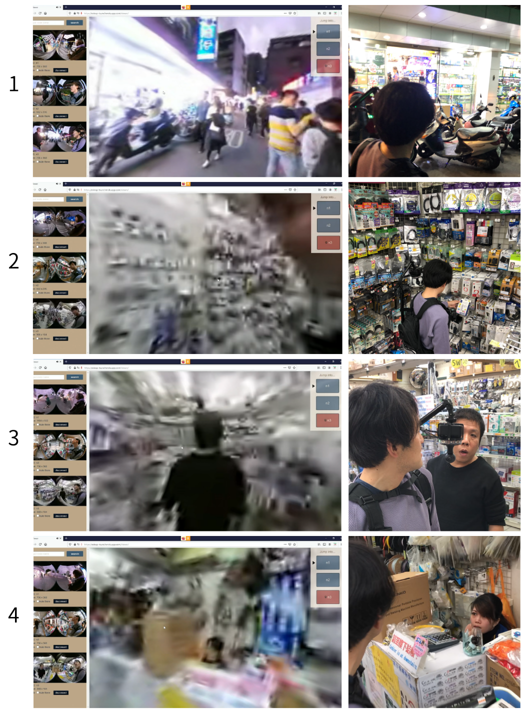

Figure 6: Sequence showing N-P1 seeking and purchasing a cable at a shop, guided by V-P5 in S3.

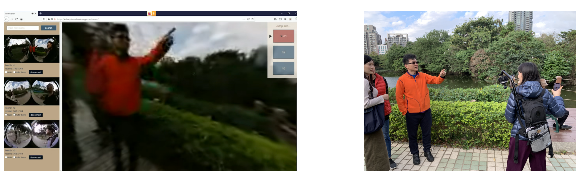

Figure 7: A local pedestrian telling V-P6 with N-P3 which way to go in Mandarin via T-Leap in S2.

knew the local area well, guided N-P1 to move to the best place to get a taxi. This scene is shown in Figure 5.

P1: Which road should I go to?
P8: Go straight to the MaiCoin (a signboard)!
P1: Where should I stop to take a taxi?
P6&P8: Cross this street. And the.......
P8: Just cross the street and turn left. Yes, yes, this way.

Similarly, in S3, V-P5 efficiently guided three Nodes to complete their shopping using his knowledge of the local area. Whenever V-P5 suggested a route to a Node, he mentally calculated how long it would take for the Node to reach their destination. This allowed him to help other Nodes in the meantime and check back on the Node around the time they reached their destination. V-P5's detailed geographical knowledge enabled him to provide a high degree of support to the Nodes who did not know the area. For example, N-P1 was able to quickly find a store and buy the cable he wanted within 11 minutes, despite not knowing the local geography or language.

Figure 6 shows a set of photos that sequentially describe this scene. 1) V-P5 guided N-P1 to a store he thought sold electric cables, and told N-P1 that the cables would be on the second floor. While N-P1 was going up the stairs, V-P5 switched the connection and guided the other Nodes. After about 2 minutes, he re-connected with N-P1. 2) N-P1 found there were many types of cables, and requested that V-P5 ask an employee where the cable he wanted was. 3) V-P5 asked a shop employee where the cable was via T-Leap in Mandarin. After the employee guided N-P1 to the correct cable, V-P5 switched connection to the other Nodes to guide them. V-P5 then re-connected with N-P1 when he arrived at the register on the first floor. 4) V-P5 provided his membership number in Taiwanese and purchased the cable with his membership discount.

In this case, V-P5 was able to recall the detailed layout of the shop from his frequent visits. Therefore, despite the poor video quality transmitted via T-Leap (as shown in Figure 6), V-P5 was able to successfully guide N-P1, and allowed N-P1 to move through the shop as if he were a regular customer.

4.1.2 *When the Viewer helped Nodes speak the local language.*

There were several scenes where the Viewer was able to support the Nodes using their local language skills. For example, when the Nodes were looking for landmarks in S2, N-P3 was able to reach one of the landmarks thanks to V-P6 talking with local pedestrians (Figure 7). The conversation is quoted below (speech in Mandarin is written in brackets).

P3: Excuse me.
P6: (... Are there bicycles by the water somewhere in the park?)
Ped: (Yes. There are.)
P6: (Where? Which way?)"
Ped: (How to say...), [pointing at a direction] this way.
P3: Thank you!

In the same scenario, V-P6 remotely asked a pedestrian group and a policeman where the bicycle landmark was to help N-P1 reach it. Thanks to this, N-P1 was able to arrive at the bicycle landmark taking a route different from the one N-P3 took.

P6: (Do you know where is something like a bicycle near the water in the park?)
Ped: (I don't know. Maybe you should ask the policeman)
P6&P1: (Ok, thank you)
P6: So we just ask the policeman. Yeah.
P6: [after N-P1 came to the policeman] (Hello, excuse me, I want to find a bicycle near the water in the park.)
Pol: (Go straight, and it's on the right hand side.)

In S2, the Viewer and Node collaborated to complete the task of reaching the landmarks as neither knew where they were. The Node found local persons in the park and the Viewer was able to ask them for direction in the local language via T-Leap. Besides this division of labor based on mobility and knowledge, we also observed spontaneous collaboration on tasks which did not require geographical or linguistic knowledge. In the next section, we will highlight these collaborations between the Viewer and Node.

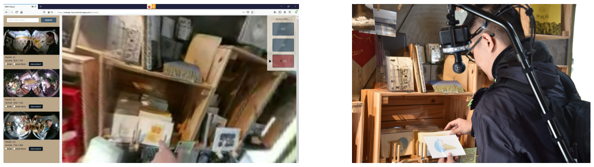

Figure 8: V-P2 and N-P6 collaboratively finding a postcard that V-P2 liked in S1.

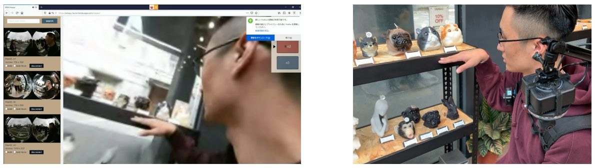

Figure 9: V-P2 and N-P7 collaboratively finding a candle V-P2 liked in S1.

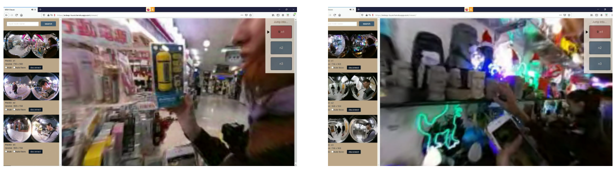

Figure 10: V-P5 advising N-P4 on what to buy for a gift in S3.

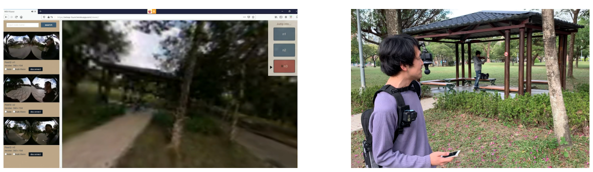

Figure 11: N-P1 letting V-P5 listen to a violin player in S2.

Figure 12: V-P6 jumping around inside a taxi with N-P2, N-P6, N-P7 in S2.

4.2 Spontaneous Collaboration on Location and Language Independent Tasks

Throughout the field study, we observed several scenes of spontaneous collaboration between the Viewer and Nodes, especially when they had no clear goal or mission. We suspect that such spontaneous collaboration was more easily induced in goal-less tasks as these typically did not require significant geographical or linguistic knowledge (i.e., there was no reliance relationship). In these collaborations, both the Viewer and Nodes flexibly modified their actions to collaboratively achieve their purpose.

In S1, for example, the Viewer and Nodes collaborated to find and buy souvenirs. While there was a general objective of buying souvenirs, there was no specific goal as to what souvenirs to buy.

After the Nodes arrived at the Huasan 1914 Creative Park, the Nodes and Viewer explored the shops. In one scene, V-P2 connected to N-P6 and entered a bookstore. N-P6 suggested that they buy a postcard there and V-P2 found one that he liked (Figure 8).

P6: Oh, you want a postcard? (... P2 and P6 looked at and talked about several postcards.)
P2: I like husky very much, so it's nice.
P6: Buy this?
P2: Yeah, I want this one. Husky one.

We observed another instance of spontaneous collaboration in S1. V-P2 went to a candle shop with N-P7 after exploring several places. They collaboratively found a candle that V-P2 liked (Figure 9). V-P2 had not made up his mind to buy a souvenir with a dog motif beforehand, but he noticed that he liked dogs more than cats while looking at several products with N-P7. The conversation below was held in the candle shop.

P2: Oh.. cute.
P7: This is a cat.
P2: Hum, cat face. (looking around) Can I choose some small one of the dog… I like the dog.

In S3, N-P4 and V-P5 collaboratively found and decided on a Christmas present to buy. N-P4 had not decided what to buy beforehand. N-P4 had only told V-P5 that she needed some weird Christmas presents for her friends. V-P5 was not sure what would be in the shops, but he gave suggestions while shopping, as if shopping together in person. (Right subfigure in Figure 10)

P5: What is that?
P4: Power bank
P5: Oh, Power bank yeah it seems good, it's like a ...
P4: Weird
P5: Yeah, it's quite weird (Left in Figure 10)
P5: I think this one looks better
P4: But the price is a bit high.

T-Leap enabled Nodes to share the experience of being outdoors with the Viewer as if they were at the location together. Besides the functional benefits (i.e., enabling the Viewer to guide Nodes), T-Leap provides the Nodes with the ability to share joyous experiences with the Viewer. We discuss this further in the next section.

4.3 The Remote Experience: Shared Joys and Challenges

We observed several scenes during the field study where a Node had serendipitous encounters and discoveries and was able to share them with the Viewer via T-Leap. These encounters and discoveries sometimes led towards the completion of the tasks at hand.

In S2, for example, N-P1 found a man playing the violin while exploring the park and contacted V-P6 via LINE in order to share the experience with him. V-P6 connected to N-P1 and enjoyed listening to it despite not being able to see the violinist due to the low video quality (See, Figure 11).

Experiencing a taxi ride via T-Leap was also a notably novel experience for the Viewer. V-P6 connected to N-P7 when all Nodes were in the same taxi (Figure 12) in S2. P2, who was observing V-P6 at the time, described the situation as follows:

P2: Ah… Between them is like a child.
P6: And sitting in the middle of them.
P7: Invisible guy [sitting between us]!

Inside the taxi, the Viewer was able to jump from one Node to another and see the other Nodes. The Viewer commented later that it felt like they were possessing the Nodes.

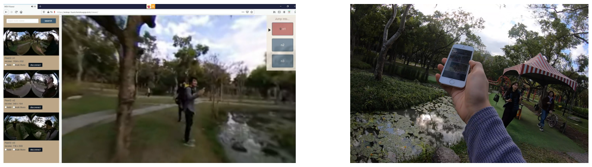

Figure 13: V-P6 and N-P1 discovering that N-P3 had already found the object in S2. V-P6 jumped into N-P3 soon after.

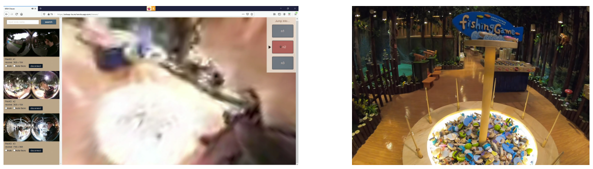

Figure 14: A blurry and bad view in S1.

In a similar case in S2, when V-P6 and N-P1 collaboratively found the bicycle landmark, N-P1 physically saw that N-P3 was already in front of the bicycle. At this point, N-P1 suggested that V-P6 connect to N-P3 (Figure 13). Then, V-P6 switched connection to N-P3 and observed N-P1 coming towards them.

The enjoyment of such serendipitous moments, however, seemed heavily reliant on the quality of video sent to the Viewer.

When the Viewer was someone who did not know the area where the Nodes were exploring, it was difficult for him/her to recognize what was happening. In S1, V-P2 did not know the place they went to (Huasan 1914 Creative Park) and the video sent to the Viewer was very low resolution, 368×184 px (Figure 14). As such, he was not very aware of the surroundings and often did not know what he was being shown by the Nodes.

However, the minimum resolution required for the Viewer to recognize things around the Node appeared to depend on the Viewer's knowledge about the Node's location. V-P5 in S3, for example, was sufficiently familiar with the shops the Nodes visited and easily recognized locations from the low resolution image (e.g., in Figure 6). As such, it is thought that the level of enjoyment the Viewer experiences from shared serendipitous encounters could remain high despite a low quality video stream if the Viewer is familiar with the physical location.

4.4 Additional Findings

In this subsection, we summarize other findings of interest that we gathered from our field study. These findings include how the local people appeared to interpret T-Leap, how the 360-degree camera was used and positioned effectively, and the variance in the cognitive load on the Viewer based on their level of local knowledge.

4.4.1 How local people interpreted T-Leap. Over the course of the field study, we had many opportunities to interact with the local people and observe how they interpreted what T-Leap was and how it worked. Some examples of local people that the participants met during the studies include shop employees (S1,3), taxi drivers (S2), and pedestrians in the park (S2). Surprisingly, all of these people easily understood that a remote person was connected via T-Leap and were able to converse with the Viewer.

In S1, a shop employee recognized T-Leap as an object *who* spoke in English but not the local language. In this encounter, V-P2 requested that N-P7 ask the employee about a candle after the scene shown in Figure 9. Initially, N-P7 and the employee were talking in Mandarin, but the employee changed their language to English after she noticed that V-P2 and N-P7 were talking in English. The conversation is quoted below.

P7: Let me ask the staff. (Excuse me, is this a set for two candles?)
Staff: (Yes, these are in one set.)
P7: (To P1) So both of them, just the one thousand. Two buy?
Staff: I think we get to have that one, the Shiba.

In S2, we observed a taxi driver growing to accept T-Leap, gradually changing his attitude towards the participants and T-Leap. This scene began after V-P6 told that Nodes that they were to go to Daan Park. The taxi driver they flagged was initially suspicious of T-Leap. Yet, after V-P6 explained what the system is, and what the Viewer and Nodes were doing, he was reassured and happily talked to V-P6. The entire conversation quoted below took place in Chinese (Tax:Taxi driver).

Tax: What are you guys doing? Interview? Don't video me.
P6: No. We won't do that. I'm at another place. They are foreigners and I am guiding them to get to some places.
Tax: Where do you want to go?
P6: The Daan Forest Park.
Tax: OK, good. OK, thank you. Amazing!

Again, in S2, there was an encounter with a pedestrian where, we interpreted that, a pedestrian did not notice that the Viewer would be able to see him via the camera on T-Leap. In this encounter, V-P6, connected to N-P3, asked pedestrians in Chinese if they knew where the bicycle landmark was (Figure 7). V-P6 wanted the pedestrian to show the way to reach the landmark, but the pedestrian was confused because they did not know how to give directions to V-P6. After hesitating for a while, he gave up trying to guide V-P6 using words and gestured to let N-P3 know the way to go.

In another case in S2, however, other pedestrians seemed to understand that the Viewer can move with the Node. As before, V-P6 asked pedestrians, in Mandarin, where the bicycle landmark was via N-P1's Node module. Since they did not know where it was, they suggested that V-P6 ask a nearby policeman for help.

We also observed a case in S2 where a pedestrian recognized T-Leap as a remote tour guide. Below is a quote from a pedestrian who saw N-P1 with the Node module:

Ped: (He wears a thing, a tour guide!)

Although limited in number, the T-Leap modules were, in every case, interpreted as an avatar remotely connected with a person.

4.4.2 Effective use and positioning of the 360-degree camera. Over the course of the field study, we observed some notable cases of the Viewer effectively utilizing the 360-degree view afforded by the 360-degree camera. For example, in S1, V-P2 successfully utilized the 360-degree sight from N-P6's camera to see a monument which he liked even though N-P6 did not notice it. V-P2 looked back at the monument twice using the 360-degree sight, even though N-P6 walked past it without noticing it. Similarly, in S3, V-P5 was able to determine where he was by independently looking around while the Nodes walked and explored. By making use of the 360-degree sight, V-P5 was able to smoothly guide the Nodes without having to ask them to wander to get his bearings.

Similar to these situations, there was a scene in which the Viewer obtained more information than Nodes who were on-site. In S3, N-P4 did not notice that two men behind her were looking at her, but V-P5, through his independent view of the surroundings, was able to notice that the two men, and others, were looking at her.

P5: There are two guys watching you. Hello!

P4: Oh my god!

P5: (to people around P4) See you!

The live video feed provided through the 360-degrees camera frequently allowed the Viewer to actively look around the Nodes' surroundings as if physically standing by the Nodes. From this, we believe that 360-degree cameras have the potential to balance the relationship between people indoors and outdoors where an asymmetry of information exists.

Through our study, we also reconfirmed that the position of the Node module (positions so the Viewer could see the face of the Node) was suitable for smooth communication between the Viewer and Nodes. The position was also useful for seeing both the landscape and Node's facial expression at the same time. We observed that several Viewers looked at the Node's face when chatting and rotated their view to see the landscape when interesting things were around. Also, the position (Figure4) was found to be appropriate in terms of safety as well. It was not so far out that it would hit other objects or people. These findings provide new knowledge about effective positioning of 360-degree cameras, supplementing discoveries described in previous works [15].

Camera stability, however, remained a problem. The video from the camera was sometimes unstable and caused the Viewers significant stress. Future work should consider improved methods of attachment to the body (e.g., an arm with built in stabilization).

4.4.3 Variance in cognitive load on the Viewer. In every scenario, the Viewer was very busy because he/she had to take care of three people at the same time. In S1, V-P2 said:

P2: I'm so busy because everyone calls me.

The frequency of being called was different for each Viewer and the difference was especially significant between S1 and S3. We suspect that this is due to the asymmetry in the level local knowledge between the Viewer and Nodes in S1 and S3. The Viewer had less knowledge than the Nodes in S1 and vice versa in S3. The Viewer was called 8 times in S1 but only 2 times in S3. Also, the duration of each connection was about 3 minutes on average in S1 and 1 minute in S3. From this, we can infer that the cognitive load on the Viewer is significantly higher when they do not know the environment the Nodes are in. This is due to the attention demanded by the Nodes and the need for the Viewer to interpret the incoming video to guess what is happening on the Node's side.

5 DISCUSSION

5.1 How Limited Functionality Works for Telepresence

T-Leap was designed for one-to-many remote communications connecting indoors with outdoors. Furthermore, it was designed with minimalism and simplicity in mind as we hoped to use it in a practical setting. In this section, we highlight how the simple features in T-Leap affect user actions during practical use and discuss how we should make use of this knowledge to develop future mobile telepresence systems. From our field study, we noticed that T-Leap, despite its simple design and limited functionality, successfully supported a high degree of communication between the Viewer and Nodes. This may suggest that "simple is best" may apply to telepresence systems as well. This could be taken as a novel design direction for telepresence and telexistence researchers who are competitively developing telepresence devices (e.g., [26]).

Despite the simple and purely functional design of the T-Leap Node module, there were many scenes throughout the field study that implied that persons surrounding the Node(s) felt and recognized the Viewer's presence. This is suggested by the fact that it was not only the Nodes(s) that communicated with the Viewer, but also some pedestrians as well. Though our system does not have the complex features proposed by some previous telepresence research, such as a robotic body [1, 26] or the face-view of the remotely connected person [1, 25, 28], the Nodes were able to guess the feelings of the remotely connected person from his/her tone of voice and actions (for example, in the cases shown in Figures 8, 9, 10). The most we did to assist the Node in interpreting the Viewer's actions was to adopt an LED on the Node module indicating the eye-gaze direction of the Viewer. This function is similar to previous work [19], but we did not observe any scenes in which it greatly benefited communications between the Viewer and Nodes.

Another simple feature which characterizes T-Leap is the leaping feature where the Viewer can only be connected with one Node at a time. This feature places some limitations on the types of communication that can occur between participants. For example, the Nodes cannot communicate with each other directly through T-Leap and must wait for the Viewer while they are connecting with another Node. Also, Nodes cannot know what the other Nodes and Viewer are talking about and can never know what the other Nodes are doing if they are separated. As a result, this feature is not suitable for cases in which Nodes are trying to cooperate to complete a common task. We observed this in S2 where the participants spent a significant amount of time searching for a non-existent landmark. In this case, time spent searching for the non-existent landmark was extended significantly due to information about the non-existence of the landmark not being shared with other Nodes immediately after discovery.

At the same time, however, this feature allowed the Viewer to carefully and precisely communicate with each Node. Due to this feature, the Viewer can speak to one Node at a time without interruption. If the Viewer is an adept guide (e.g., V-P5 in S3), every Node could very quickly complete their mission. We suggest that our one-to-many model (where only one channel is active at a time) could work well in professional guiding scenarios where the Viewer might be a travel guide, interpreter, or museum guide.

5.2 Implications for Designing Future Mobile and Wearable Telepresence Systems

5.2.1 Wearable 360-degree live streaming. With the current T-Leap, the quality of the video sent to the Viewer is highly limited and dependent on network conditions. If faster mobile network connections become available, T-Leap could be used as a wearable live-streaming system with which Node(s) could broadcast to Viewer(s). In this research, we assumed only one Viewer. Technically, however, it is possible to allow multiple Viewers to connect to a Node(s).

T-Leap, therefore, could be used as a new live streaming platform to present 360-degree video to multiple users. Since our T-Leap consists of low-cost devices that can easily be acquired by many people, the system itself could easily be commercialized.

Such a commercialized system could be used to encourage bi-directional communication between live streamers and their audience. In S3, as shown in Figure 10, V-P5 suggested that N-P4 take action and collaboratively found a Christmas gift to buy. In this case, N-P4 initially passed by the product but re-considered it after V-P5 gave the suggestion. Observing this case, we interpret that wearable 360-degree broadcasting can provide novel opportunities for persons (like YouTubers) who want to do live streaming and engage in bi-directional communication that allows the audience to gauge the streamer's reaction. This is difficult with a unidirectional camera as it requires significant camera work to be performed on-site and viewers are not afforded the freedom of viewing direction.

We found that the view direction freedom provided to Viewers by using a 360-degree camera allows the Viewer to feel a sense of agency which is independent of the Node. We observed a case where the Viewer felt as if he was on-site with the Node: V-P2 said, *"I am attached to those souvenirs even though it is the first time I have seen them,"* in an interview after S1. He felt as if he had bought a postcard himself with the Nodes in the field. This sense of agency might be favored by Viewers compared to the more traditional live streaming methods which give the viewers much less freedom of movement. The sense of presence in remote communication has been addressed in some research works [4, 16, 29]. During the field study, we also observed a case in which the Viewer felt a sense of self-efficacy as well as a sense of presence while using T-Leap.

A sense of presence, especially with multiple Viewers, could be enhanced by the use of eye-gaze visualization. Though the LED visualization of the Viewer's eye-gaze direction did not contribute to communication with the Nodes in our study, it might be more meaningful if multiple Viewers could join and talk with multiple Nodes at the same time. In this case, the Nodes may have difficulty figuring out which Viewer joined them from just their voices. This is similar to cases in which many users browse a live stream (for example on YouTube live) at the same time. If multiple LED arrays were attached to the T-Leap module or other systems, it could help Nodes to recognize the presence of multiple Viewers.

5.2.2 Remote one-to-many guidance with the 360-degree views. From V-P5's success at guiding the Nodes in S3, we believe that an online guiding service with the wearable mobile telepresence system could be highly proficient and inexpensive while also preserving privacy. Shibahara et al. [20] developed a system for sharing 360-camera views for many-to-many users who are outdoors. However, they did not examine its efficacy in specific scenarios. We explored more specific, real use cases and scenarios with T-Leap and found that the best practical use for T-Leap is guiding. Here, we describe some recommendations for future applications of telepresence to guiding.

Firstly, the one-to-many "connections" are required for efficient guide performance. In our study, by maintaining connections with multiple Nodes at the same time, the Viewer was able to flexibly and efficiently use their time. Namely, they did not have to wait while a Node did not need any help and was able to support the other Node(s) in the time they would have otherwise spent idly. While one-to-one remote communication with 360-degree camera has already been conducted [11, 23], we found that a proficient guide is able to guide multiple persons through one-to-many remote communications (e.g., shown in Figure 6). In addition, our easy-switching UI for the Viewer greatly contributed to the Viewer's ability to guide, because a proficient guide may switch the main-view in short spans (less than 3 minutes) as observed in S3.

Secondly, one-to-one "communication" is recommended. In T-Leap, we only allowed Viewer to talk with one Node at a time. In our study, this limitation provided valuable interactions between people in different places. While Tanikawa et al. constructed a system to guide many people at once [24], T-Leap is designed to always maintain one-to-one communication. This allows the Viewer to take care of each Node according to their scenarios. Therefore, it is possible for the Viewer to carry out person-specific remote guiding. By using one-to-one communication, a professional guide may flexibly get over not only linguistic barriers but also cultural barriers, as shown in Figure 5. Since this communication model has no Node to Node connections and each Node only has to communicate with the Viewer (i.e., the guide), the model inherently protects the Nodes' privacy whenever they are using the guiding service.

Finally, a proficient guide does not require a high quality view. Due to the network conditions, we sometimes encountered terribly low quality video streaming during our field study. However, as a talented guide, V-P5 in S3 successfully guided all Nodes to their destinations despite frequent appearances of low resolution video. Since he was very familiar with the area, he adapted to the bad network connection without any problems. By utilizing a 360-degree camera, a person who is greatly familiar with a place can easily recognize where it is or what it is, even from a low resolution view. As mentioned in a previous study [20], richly experienced guides do not need location information, such as that provided by GPS, to effectively guide people. Guides with T-Leap could perform as remote guides without GPS signals or even high resolution views.

6 CONCLUSION

In this paper, we proposed T-Leap, an integrated one-to-many wearable telepresence system that makes use of a 360-degree camera. With this system, we carried out a field study with three scenarios to observe participants using T-Leap to connect remotely (with the Viewer staying indoors, and the Nodes exploring outdoors) in situations where there exist linguistic and geographical barriers. We found that T-Leap is most effective at facilitating communication and cooperation when a Viewer with good local knowledge guides multiple Nodes to complete tasks or explore an area. We also formulated some recommendations for designing mobile, wearable, and one-to-many telepresence systems in general. T-Leap suggests a new design direction for telepresence and telexistence that leverages simple and minimal technology.

ACKNOWLEDGMENTS

This project is supported by TIS Inc., and JST ERATO Grant Number JPMJER1701, Japan.

REFERENCES

[1] Sigurdur O. Adalgeirsson and Cynthia Breazeal. 2010. MeBot: A Robotic Platform for Socially Embodied Presence. In *Proceedings of the 5th ACM/IEEE International*

Conference on Human-Robot Interaction (Osaka, Japan) *(HRI '10)*. IEEE Press, 15–22.

[2] Sara A Bly, Steve R Harrison, and Susan Irwin. 1993. Media spaces: bringing people together in a video, audio, and computing environment. *Commun. ACM* 36, 1 (1993), 28–46.

[3] Sarah D'Angelo and Darren Gergle. 2016. Gazed and Confused: Understanding and Designing Shared Gaze for Remote Collaboration. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 2492–2496. https://doi.org/10.1145/2858036.2858499

[4] Yasamin Heshmat, Brennan Jones, Xiaoxuan Xiong, Carman Neustaedter, Anthony Tang, Bernhard E. Riecke, and Lillian Yang. 2018. 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* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, Article Paper 359, 13 pages. https://doi.org/10.1145/3173574.3173933

[5] Brennan Jones, Anna Witcraft, Scott Bateman, Carman Neustaedter, and Anthony Tang. 2015. Mechanics of Camera Work in Mobile Video Collaboration. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 957–966. https://doi.org/10.1145/2702123.2702345

[6] Shunichi Kasahara, Shohei Nagai, and Jun Rekimoto. 2015. 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* (Brussels, Belgium) *(TVX '15)*. Association for Computing Machinery, New York, NY, USA, 33–42. https://doi.org/10.1145/2745197.2745202

[7] Shunichi Kasahara and Jun Rekimoto. 2014. JackIn: Integrating First-Person View with out-of-Body Vision Generation for Human-Human Augmentation. In *Proceedings of the 5th Augmented Human International Conference* (Kobe, Japan) *(AH '14)*. Association for Computing Machinery, New York, NY, USA, Article Article 46, 8 pages. https://doi.org/10.1145/2582051.2582097

[8] Shunichi Kasahara and Jun Rekimoto. 2015. 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* (Beijing, China) *(VRST '15)*. Association for Computing Machinery, New York, NY, USA, 217–225. https://doi.org/10.1145/2821592.2821608

[9] Tadakazu Kashiwabara, Hirotaka Osawa, Kazuhiko Shinozawa, and Michita Imai. 2012. TEROOS: A Wearable Avatar to Enhance Joint Activities. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. Association for Computing Machinery, New York, NY, USA, 2001–2004. https://doi.org/10.1145/2207676.2208345

[10] Seungwon Kim, Sasa Junuzovic, and Kori Inkpen. 2014. The Nomad and the Couch Potato: Enriching Mobile Shared Experiences with Contextual Information. In *Proceedings of the 18th International Conference on Supporting Group Work* (Sanibel Island, Florida, USA) *(GROUP '14)*. Association for Computing Machinery, New York, NY, USA, 167–177. https://doi.org/10.1145/2660398.2660409

[11] Sven Kratz, Don Kimber, Weiqing Su, Gwen Gordon, and Don Severns. 2014. Polly: "Being There" through the Parrot and a Guide. In *Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices & Services* (Toronto, ON, Canada) *(MobileHCI '14)*. Association for Computing Machinery, New York, NY, USA, 625–630. https://doi.org/10.1145/2628363.2628430

[12] Zhengqing Li, Shio Miyafuji, Erwin Wu, Hideaki Kuzuoka, Naomi Yamashita, and Hideki Koike. 2019. OmniGlobe: An Interactive I/O System For Symmetric 360-Degree Video Communication. In *Proceedings of the 2019 on Designing Interactive Systems Conference* (San Diego, CA, USA) *(DIS '19)*. Association for Computing Machinery, New York, NY, USA, 1427–1438. https://doi.org/10.1145/3322276.3322314

[13] Andrés Lucero, Audrey Desjardins, Carman Neustaedter, Kristina Höök, Marc Hassenzahl, and Marta E Cecchinato. 2019. A sample of one: First-person research methods in HCI. In *Companion Publication of the 2019 on Designing Interactive Systems Conference 2019 Companion*. 385–388.

[14] Eric Paulos and John Canny. 2001. Social Tele-Embodiment: Understanding Presence. *Autonomous Robots* 11, 1 (01 Jul 2001), 87–95. https://doi.org/10.1023/A:1011264330469

[15] 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* (Barcelona, Spain) *(SAP '19)*. Association for Computing Machinery, New York, NY, USA, Article Article 13, 10 pages. https://doi.org/10.1145/3343036.3343120

[16] Thammathip Piumsomboon, Gun A. Lee, Andrew Irlitti, Barrett Ens, Bruce H. Thomas, and Mark Billinghurst. 2019. On the Shoulder of the Giant: A Multi-Scale Mixed Reality Collaboration with 360 Video Sharing and Tangible Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, Article Paper 228, 17 pages. https://doi.org/10.1145/3290605.3300458

[17] Jason Procyk, Carman Neustaedter, Carolyn Pang, Anthony Tang, and Tejinder K. Judge. 2014. Exploring Video Streaming in Public Settings: Shared Geocaching over Distance Using Mobile Video Chat. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 2163–2172. https://doi.org/10.1145/2556288.2557198

[18] Irene Rae and Carman Neustaedter. 2017. Robotic Telepresence at Scale. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 313–324. https://doi.org/10.1145/3025453.3025855

[19] David A Shamma, Tony Dunnigan, Yulius Tjahjadi, and John Doherty. 2019. Visualizing Gaze Presence for 360 Cameras. (2019).

[20] N. Shibahara, R. Kondo, and M. Iwai. 2017. MM360: A GPS-assisted 360-degree video sharing system for participatory events. In *2017 IEEE International Conference on Big Data (Big Data)*. 4123–4127. https://doi.org/10.1109/BigData.2017.8258432

[21] Susumu Tachi. 2015. *Telexistence*. Springer International Publishing, Cham, 229–259. https://doi.org/10.1007/978-3-319-17043-5_13

[22] S. Tachi, K. Minamizawa, M. Furukawa, and C. L. Fernando. 2012. Telexistence — from 1980 to 2012. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 5440–5441. https://doi.org/10.1109/IROS.2012.6386296

[23] Anthony Tang, Omid Fakourfar, Carman Neustaedter, and Scott Bateman. 2017. Collaboration with 360° Videochat: Challenges and Opportunities. In *Proceedings of the 2017 Conference on Designing Interactive Systems* (Edinburgh, United Kingdom) *(DIS '17)*. Association for Computing Machinery, New York, NY, USA, 1327–1339. https://doi.org/10.1145/3064663.3064707

[24] Tomohiro Tanikawa, Makoto Ando, Kazuhiro Yoshida, Hideaki Kuzuoka, and Michitaka Hirose. 2004. Virtual gallery talk in museum exhibition. In *Proceedings of ICAT*, Vol. 2004. 369–376.

[25] H. Tobita. 2017. Gutsy-Avatar: Computational Assimilation for Advanced Communication and Collaboration. In *2017 First IEEE International Conference on Robotic Computing (IRC)*. 8–13. https://doi.org/10.1109/IRC.2017.82

[26] Y. Tsumaki, F. Ono, and T. Tsukuda. 2012. The 20-DOF miniature humanoid MH-2: A wearable communication system. In *2012 IEEE International Conference on Robotics and Automation*. 3930–3935. https://doi.org/10.1109/ICRA.2012.6224810

[27] Baris Unver, Sarah D'Angelo, Matthew Miller, John Tang, Gina Venolia, and Kori Inkpen. 2018. Hands-Free Remote Collaboration Over Video: Exploring Viewer and Streamer Reactions. 85–95. https://doi.org/10.1145/3279778.3279803

[28] Lillian Yang, Brennan Jones, Carman Neustaedter, and Samarth Singhal. 2018. Shopping Over Distance through a Telepresence Robot. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article Article 191 (Nov. 2018), 18 pages. https://doi.org/10.1145/3274460

[29] J. Young, T. Langlotz, M. Cook, S. Mills, and H. Regenbrecht. 2019. Immersive Telepresence and Remote Collaboration using Mobile and Wearable Devices. *IEEE Transactions on Visualization and Computer Graphics* 25, 5 (May 2019), 1908–1918. https://doi.org/10.1109/TVCG.2019.2898737

[30] John Zimmerman, Jodi Forlizzi, and Shelley Evenson. 2007. Research through design as a method for interaction design research in HCI. *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems* (2007), 493–502.