International Conference on Artificial Reality and Telexistence
Eurographics Symposium on Virtual Environments (2017)
R. Lindeman, G. Bruder, and D. Iwai (Editors)

Sharing Gaze for Remote Instruction

Sathya Barathan[1], Gun A. Lee[2,1], Mark Billinghurst[2,1] and Robert W. Lindeman[1]

[1]The Human Interface Technology Laboratory New Zealand, University of Canterbury, New Zealand
[2]School of Information Technology & Mathematical Sciences, University of South Australia, Australia

Abstract

In this paper, we report on how sharing gaze cues can assist remote instruction. A person wearing a head-mounted display and camera can share his or her view with a remote collaborator and get assistance on completing a real-world task. This configuration has been extensively studied in the past, but there has been little research on how the addition of sharing gaze cues might affect the collaboration. This paper reports on a user study exploring how sharing the gaze of a remote expert affects the quality of collaboration over a head-worn video conferencing link. The results showed that the users performed faster when the local workers were aware of their remote collaborator's gaze, and the remote experts were in favour of shared gaze cues because of the ease-of-use and improved communication.

CCS Concepts

•***Human-centered computing*** → ***Mixed / augmented reality;*** *Computer supported cooperative work;*

1. Introduction

Enabled by applications such as Skype (http://skype.com), Periscope (http://periscope.tv) or FaceTime (http://apple.com/facetime), remote collaboration using live video sharing is becoming a common activity. Most of these applications focus on viewing the face of a remote person in a so called "talking head video". However, using a wearable computer with a head-mounted display (HMD) and head worn camera (HWC), a person can live stream their view [SKJC95]. The remote expert could also use a mouse to provide virtual pointing or annotations that can be shared back with the local worker, overlaid on their live video view. This configuration has been used for many applications such as remote collaboration for technical support activities [BHKS98] or shared problem solving [KLSB14].

Using an HMD and camera, Kraut et al. [KFS03] found that sharing visual information was one of the most useful resources for collaboration on physical tasks. Fussell et al. [FKS00] studied the effects of shared visual context on collaborative work and also found a positive effect of using shared visual cues. Similarly, researchers found that head-worn remote collaborative systems had a positive impact on maintenance duties on bicycles and aircraft [KMS96] [SKJC95]. However, in these cases the remote expert was only able to place a virtual pointer in the local user's view. In our research, we are interested in the effects of sharing gaze cues from the remote collaborator to a local worker. This would allow the local worker to know exactly where the remote collaborator is looking while helping them with their task.

Eye tracking [Duc07] can be used to identify the gaze of a person and where their attention is directed. In our research, eye tracking was used to show the local user where their remote collaborator was looking in the remotely shared view. Earlier studies have looked at shared gaze in desktop [CHN*10] or immersive virtual environments [NCD*10], but our research brings the shared gaze cue into the real world by combining video conferencing, wearable computing, and Augmented Reality (AR) technologies. There is earlier work by others focused on sharing the gaze of the local worker [GLB16] [MKB16], but less investigation is done on sharing the gaze of the remote expert. The primary goal of our research is to compare different communication cues (voice, mouse + voice, gaze + voice) from a remote expert in a wearable video conferencing system.

Compared to earlier work, our system makes the following novel contributions:

- It is one of the earliest examples of a wearable system for remote collaboration that shares gaze of the remote expert to the local user wearing a HMD+HWC system.
- It is one of the earliest user studies exploring the effect of sharing gaze cues of the remote expert on wearable remote collaboration.

In the remainder of the paper we first review related work, then describe the system we developed, and the user evaluation method. Next, we report on and discuss experimental results to draw conclusions.

2. Related Work

Our research is related to earlier work in wearable computing, remote collaboration, and gaze interaction. In this section we review

DOI: 10.2312/egve.20171356

work in each of these areas and emphasize the novel direction that we are taking.

There has been a significant amount of earlier research on using HMDs and HWCs for remote collaboration. Starting with Shared-View [IKTM90] a number of projects have shown that enabling a remote expert to see what a local worker is doing and being able to provide visual feedback through pointing or gesture significantly improves remote collaboration on physical tasks [BKS99] [FKS00] [KFS03]. Most recently AR techniques have been developed to precisely align virtual feedback cues with the real world [GNTH14] [KLSB14]. However, a limitation of these systems is that neither of the users knows exactly where the other person is looking; the remote expert may be focusing their attention on a different part of the live video view than the local worker, and so may have collaboration problems.

In face-to-face collaboration, gaze is an important cue for showing attention and mutual understanding. Similarly, gaze is valuable in remote collaboration and can be used to create a shared focus on appropriate objects, described as joint attention [MD95]. For example, Brennan et al. [BCD*08] experimented with eye tracking and vocal communication while users identified targets together in a desktop collaborative virtual environment. They concluded that the partner's gaze was vital to completing the task.

There has been some previous work on remote collaboration with gaze cues, although most of it has been done in a desktop setting. Typical of this is the work of Muller et al. [MHPV13], which involved two remote users solving a puzzle application on a GUI screen. Their experiment required participants to carry out a collaborative task in four different conditions: (1) gaze, (2) voice, (3) gaze and voice, or (4) mouse and voice. They found that participants received their instructions more clearly from partners when gaze information was used in addition to speech. These results have also been replicated in simulated bomb disposal [LMDG16] and remote puzzle-solving tasks [SB04]. Overall, it seems that gaze provides an excellent cue for inferring user attention and improving remote collaboration in a desktop interface.

Several researchers have also explored adding head-worn eye tracking to the local worker to enable gaze input in collaboration on a physical task. Fussell et al. [FSK03] attached an eye tracker to an HWC to enable a local worker to send gaze cues to the remote expert. Initially they found that sharing eye gaze information produced no performance improvement. However, in a follow-up study they found that the focus of attention can be predicted from monitoring eye gaze, and eye tracking provides a benefit [OOF*08]. In both cases only the local worker wore an eye-tracker and there was no way for the remote expert to give visual feedback.

Recently, Gupta et al. [GLB16] developed one of the first systems that combined an HMD with an HWC, and an eye tracker, so that the remote expert could send back virtual pointing cues. They conducted a user study comparing collaboration with and without eye gaze support, and with and without showing pointing cues. They found that the use of eye tracking and a pointer together significantly improved the co-presence felt between the users and overall task performance. In this paper, we want to explore the effect of sharing the remote expert's gaze with the local worker.

The closest directly related work is that of Higuch et al. [HYS16] which explores the effect of sharing the remote expert's eye gaze with the local worker. This work explored a number of different display conditions, including showing remote expert's gaze cues with spatial projection or on an HMD, but they did not compare sharing eye-gaze to mouse based remote pointing, as we are doing. Also, their work focused on performance measures, while we investigate more social presence aspects of the interface.

In summary, we can see that there has been significant work on using HMDs and HWCs for remote collaboration, showing the benefit of sharing remote view in collaboration on physical tasks. Desktop experiments have confirmed the value of sharing gaze for remote collaboration, but there have been very few examples of sharing gaze in an HMD/HWC system. While there is work that explores sending the remote expert's gaze back to the local worker, our work complements this by comparing different pointing technologies and investigating the social presence aspects.

3. Prototype Remote Collaboration System

To investigate the effect of sharing gaze over head-worn video conferencing, we developed a prototype system (see Figure 1) that integrates eye tracking on a desktop monitor where one user (remote expert) can see the video feed from another user (local worker) wearing an HMD with a world-facing HWC. The image shown on the remote expert's desktop is also shown on the local user's HMD so that both the participants see the same view. The remote expert and local worker are out of sight from one another, and so rely on voice and visual cues to communicate. Apart from the eye tracking system, similar configurations were used in previous research on remote collaboration [GNTH14] [GLB16] [KLSB14].

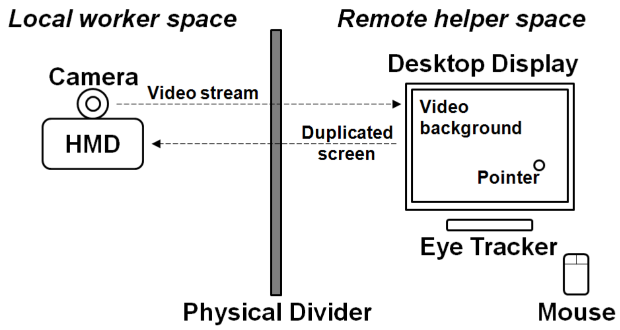

Figure 1: *System configuration of the prototype system.*

A virtual pointer was overlaid on the video feed captured with HMC. The pointer used was a semi-transparent red dot that indicated the point of interest of the remote expert to the local worker (see Figure 2), and it followed either the mouse motion or the eye tracking of the remote expert. The virtual pointer was shown on both remote expert's desktop display and local worker's HMD.

The prototype system used a laptop for the remote expert with an Intel Core i7 2.60 GHz processor, 16GB RAM, running Windows 10, and a 17.3 inch widescreen (16:9) with 1920x1080 resolution. The system used the Eye Tribe eye tracker (http://

Figure 2: *A pointer (red dot) overlaid on background video feed captured with HWC.*

theeyetribe.com; 75Hz update rate, latency < 16 milliseconds, 0.5-1° accuracy). The local worker wore a Vuzix Wrap 1200DX-VR HMD (https://www.vuzix.com; WVGA 852 x 480 pixel resolution and 35° diagonal field of view) with a Logitech HD Pro Webcam C920 attached on the front as a HWC to create a video see-through configuration. The software was implemented with Processing (http://processing.org) with the *Video Capture* library to view the camera feed and the *PGraphics* module to overlay the mouse pointer and gaze cues.

4. User Study Design

In order to explore the effect of sharing gaze cues, we conducted a user study. This was in a within-subjects design with one independent variable, the types of cues shared from the remote expert to the local worker. The three experimental conditions were:

- Voice (V): The remote expert could see the video from the local worker's camera but could only talk to communicate. This was the baseline condition.
- Mouse (M): In addition to the baseline condition with voice, the remote expert could use mouse input to point on the shared video.
- Gaze (G): In addition to the baseline condition with voice, the remote expert's gaze was shown on the shared video.

All three conditions included verbal communication. For each of these conditions, we collected task completion time as the objective performance measure, and used questionnaires and interviews for subjective feedback.

4.1. Procedure and Setup

At the beginning of the experiment, the researcher introduced the prototype system and experimental procedure to the participants, and asked them to fill-out a pre-experiment questionnaire that collected demographic information. For each experimental condition, the researcher explained the hardware configuration that the participants would be using, and the participants had some practice trials followed by an experimental trial. The remote experts were seated in front of a laptop and given instruction sheets for guiding their partners. Wearing the HMD with HWC, the local workers stood at the centre of a projection-based immersive visualisation system (see Figure 3) with three screens, each measuring 2.45 x 1.8 metres, with 1024 x 768 resolution. The projection screens showed projected control panels, simulating a physical task space that the local worker needs to operate. The HMD was adjusted to fit the local worker's vision and they were asked to move around to make sure all panels of the wall were accessible to perform the actions instructed by the remote expert. Depending on the condition, a virtual pointer was overlaid onto the video feed captured with HWC, shown on HMD and the remote expert's desktop monitor. Once the participants were comfortable, the lights were dimmed to ensure maximum visibility of the projection screens.

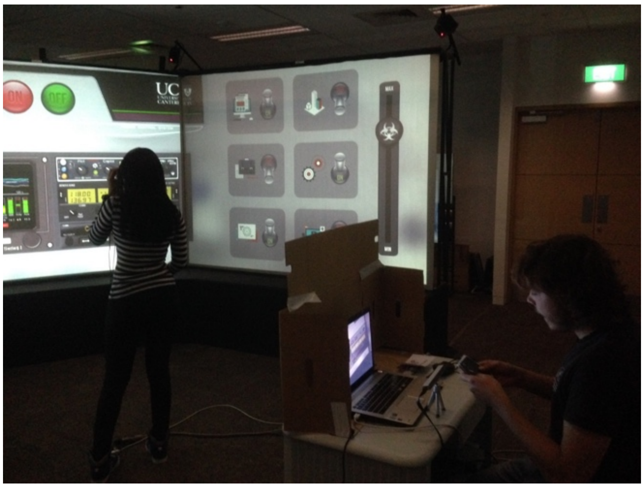

Figure 3: *Experimental setup.*

In each condition, the participant pairs (expert and worker) had to complete five instructional tasks. After completing each condition, the participants answered questionnaires about usability and social presence level. The order of the conditions was varied between participant groups using a balanced Latin Square design to counter balance ordering effects.

Practice trials were given at the beginning of each condition so that the participants were familiar with using the interface. In the Gaze condition, the eye tracker was first calibrated with a 12-point calibration process. This was repeated several times to ensure accurate calibration before the trial started. After finishing all the conditions, the participants filled out a post-experiment questionnaire that included ranking each condition and feedback on interface improvements. This was followed by a debriefing interview.

4.2. Experimental Task

The participants were required to complete a set of five tasks under each of the three conditions as fast as possible. The task for the local worker was to follow the instructions given by the remote expert, who was handed a series of instructions to operate icons shown on projection screen (see Figure 4). Instructions were printed on A5 size paper as shown in Figure 5. Each instruction included a figure of an icon (e.g., a button) with a brief description of an action (e.g.,

Figure 4: *Controls on the projection screen in the local worker's task space.*

"Press the button"). The icons were designed to be easily distinguishable from each other, yet hard to describe verbally to prevent the communication between the users from solely relying on verbal communication. The instruction sets were randomized between conditions.

The participants were informed that the tasks would be timed from when they began to perform the first instruction. The remote expert had to ask their partner to move around the projection screens to identify the icons related to the target task. Once they identified the icon, the remote expert then instructed the local worker to carry out the respective task. The instructions given for the local workers were to press or toggle buttons, turn switches on or off, move sliders, and turn knobs. The local workers were asked to act out what was instructed, although the mock-up controls were not interactive.

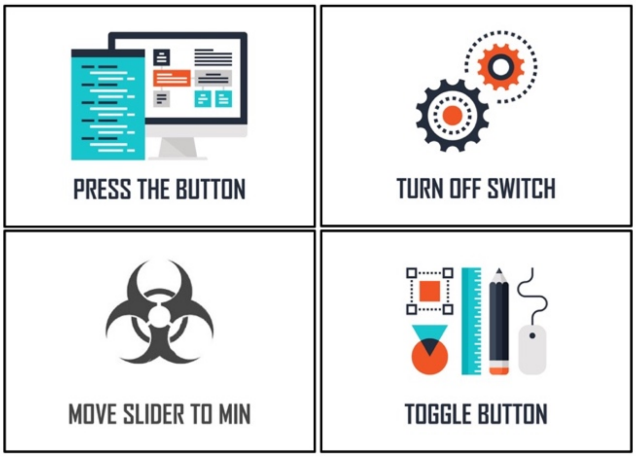

Figure 5: *Sample instructions given to the remote expert.*

4.3. Measurements

A set of qualitative and quantitative data was collected, including the task completion time and number of errors made while performing the task. The usability of the interface was measured using the System Usability Scale (SUS) [Bro96] and feedback on the collaborative experience was collected using the Social Presence survey [HB04], which included questions about co-presence, attention allocation, perceived message understanding and perceived affective understanding of their partners. In the post-experiment questionnaire, users were asked to rank the three conditions based on their preference and also provide written feedback on what they liked and disliked about the interfaces.

5. Results

Fifteen pairs of participants who knew each other were recruited for the user study. They were 55% men and 45% women with ages ranging from 24-30 years old. Most (95%) of them used video conferencing at least a few times per week. While both remote and local users answered the subjective questionnaires unfortunately the record of the local worker participants was lost, hence we only report the results from the remote expert participants' answers. In complement, we report on the local workers' interview results to show their perspective.

Overall, we found the following results:

- Participants performed significantly faster with visual cues added to the voice communication. The gaze cue helped reducing the task completion time significantly more than the mouse cue.
- Adding visual cues over voice communication improved the system usability, yet there was no significant difference between the gaze and mouse conditions.
- There were no significant differences between the conditions in terms of social presence from the remote expert's perspective.
- Ranking results showed the remote expert participants preferred the gaze cue over the other two conditions. The gaze cue was especially helpful in letting the participants feel connected and feel co-present with their partners.

5.1. Task Completion Time

The task completion time (measured in seconds) was recorded for each of the conditions (see Figure 6). Overall, participants were able to complete the task faster using Gaze ($M = 101.34$ sec., $S.D. = 22.18$) over Mouse ($M = 134.8$, $S.D. = 38.5$) and Voice

($M = 178.0$, $S.D. = 47.9$) conditions. The data was normally distributed based on the Shapiro-Wilk test (Voice: $W = 0.970$, $p = .864$; Mouse: $W = 0.972$, $p = .880$; Gaze: $W = 0.950$, $p = .526$). A repeated measures one-way ANOVA ($\alpha = .05$) found a statistically significant difference between the conditions ($F(2,28) = 17.410$, $p < .001$). A post-hoc analysis with the Bonferroni method revealed that there was a statistically significant differences between all three conditions, Voice vs. Mouse ($p = .015$), Voice vs. Gaze ($p < .001$) and Mouse vs. Gaze ($p = .013$).

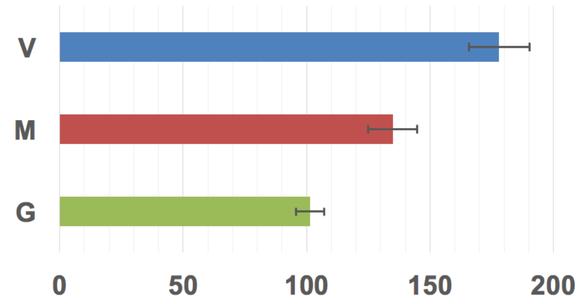

Figure 6: *Average task completion time (in seconds; whiskers: S.E.).*

5.2. System Usability Scale

The remote expert participants felt that the Gaze and Mouse conditions were significantly more usable than the Voice condition. Figure 7 shows the results of average SUS scores collected from the remote experts. The median SUS scores for Voice, Mouse, and Gaze conditions were 62.5 (Inter-Quartile Range: [45-75]), 85 [67.5-87.5] and 82.5 [72.5-90], respectively. A Friedman test ($\alpha = .05$) found a statistically significant difference in average SUS scores between the conditions ($\chi^2(2) = 12.441$, $p = .002$). Post hoc analysis with Wilcoxon signed-rank tests was conducted with a Bonferroni correction applied ($\alpha = .017$). There were no significant differences between the Gaze and Mouse conditions ($Z = -0.787$, $p = .431$). However, there was a statistically significant difference for the Voice condition as compared to the Mouse ($Z = -2.559$, $p = .010$) or Gaze ($Z = -2.502$, $p = .012$) conditions. Participants felt that the Voice condition was significantly less usable than the Mouse or Gaze conditions.

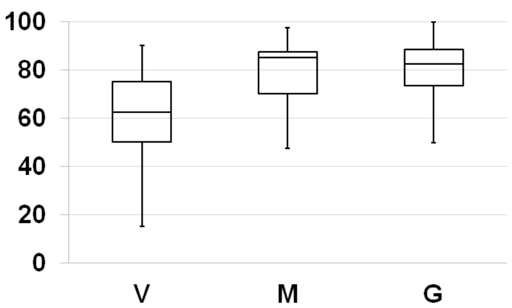

Figure 7: *Average System Usability Scale results.*

5.3. Social Presence

The remote expert participants rated all three conditions almost the same in Social Presence questionnaire (see Figure 8). A Friedman test ($\alpha = .05$) found no significant difference between the conditions ($\chi^2(2) = 3.448$, $p = .178$).

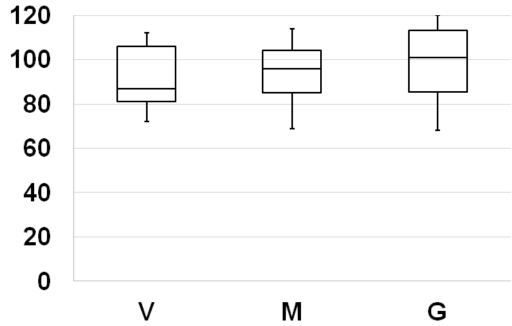

Figure 8: *Average Social Presence results.*

5.4. Ranking

At the end of the experiment, the participants were requested to rank the conditions from 1 (best) to 3 (worst) in various categories described in Table 1.

Table 1: *Ranking category.*

C#	Ranking Category
C1	At helping you complete the task.
C2	At making you feel connected with your partner.
C3	At helping you stay focused on the task
C4	At making you feel that you were present with you partner.
C5	For you (or the partner) to know that the partner (or you) needed assistance.
C6	At helping you understand the partner's message.

Overall the Gaze condition was ranked as the best by the remote expert participants in most categories, while the Voice condition was the worst (see Figure 9). We used Friedman tests ($\alpha = .05$) for analysing the ranking results, followed by post hoc tests using Wilcoxon Signed Rank tests with Bonferroni correction ($\alpha = .0167$). The results showed that the remote participant ranked the three conditions significantly different in all categories except for C5, knowing needing assistance (C1: $\chi^2(2) = 17.733$, $p < .001$, C2: $\chi^2(2) = 8.93$, $p = .011$, C3: $\chi^2(2) = 8.53$, $p = .014$, C4: $\chi^2(2) = 10.13$, $p = .01$, C5: $\chi^2(2) = 2.53$, $p = .28$, C6: $\chi^2(2) = 14.80$, $p = .01$).

Post hoc tests revealed the Gaze condition was ranked significantly higher (better) than the Voice condition in all five categories (C1 C4 and C6). The Gaze condition was also favoured over the Mouse condition in helping the participants feel connected during the collaboration process (C2: $Z = -2.40$, $p = .016$) and making them feel present with their partners (C4: $Z = -2.45$, $p = .010$). On the other hand, in terms of task related categories (C1: completing

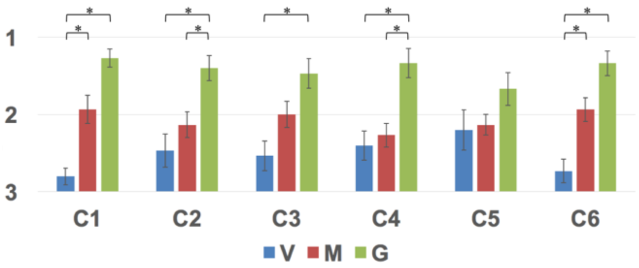

Figure 9: *Average ranking of each condition (higher the bar the better; whiskers: S.E.; *: significant different).*

the task, and C6: understanding partner), the Gaze condition was ranked as not being significantly different from the Mouse condition (C1: $Z = -2.055$, $p = .040$, C6: $Z = -1.93$, $p = .053$).

5.5. Qualitative Feedback

After completing all three conditions the participants were requested to provide some feedback on their most and least favourite conditions and give comments on what could be improved. The top answers have been summarised below.

Q1: What did you like the most in the condition you ranked best?

For the Mouse condition, participants felt that the mouse pointer was clear and simple to use and communicate the instructions. They also felt that being able to see the environment and physically guide the actions was easy. One of the participants who ranked the mouse condition the highest, said *"Mouse condition taps well into familiar behaviours, gave all of the possible information easily."*

For participants who felt that Gaze was best, they wrote that the instructions were spontaneous and easy to convey to the local workers. The advantage of having their partners visually look where the remote experts were looking also made Gaze a preferred choice. Finally, the fact that there was less vocal communication made the collaboration process using the gaze cues easy to use. Quoting a few of the participants who preferred Gaze; *"The ease with which my partner was able to pick up on my cues based on where I was looking," "Gaze condition was easy to communicate where I was looking once I had located it," "Gaze condition didn't need to give as much detail because I could just look at it."*

The Voice Only condition was not ranked best by any of the participants.

Q2: What did you dislike the most in the condition you ranked least?

Many people ranked the Voice condition the worst, for a number of reasons. Some people felt that just using voice made it hard to provide a context to the local users and difficult to explain the location of the target element. It also took an extra effort to instruct partners when the remote experts were not sure of a description for the action button.

To quote some of the participants who thought the Voice condition were the least favourite; *"I had to describe using only voice in simple language as my partner is a native foreign language speaker," "It could get confusing using only voice cause sometimes you don't know what the thing on the front is called," "It was hard to communicate where exactly I wanted to focus using my words."*

Participants who ranked the Mouse condition last said that since the local workers were constantly moving to identify the target element in the three screen projected control panel, it was difficult for the remote experts to move their mouse along with their partner's movement. To quote a few of the participants feedback; *"The system seemed difficult to use because when my partner moved it was in the wrong place," "There was an extra step and effort in dragging the mouse."*

Participants who ranked Gaze last said that the local users performed actions as soon as the remote experts fixed their gaze at a particular point even before the instructions were made. They also felt that gaze movements were not accurate in places where there were elements close together. This made it distracting for their partners. To quote a few of the participants; *"I had to focus too much on the mechanics rather than the content of the task," "Moving the eyes are distracting the partner a little bit."*

Q3: What could be improved in the three conditions?

Participants felt that the stability of the eye tracking could be improved since the gaze movements were jittery. They also felt it would be good to lock the position of the pointers in the Mouse and Gaze conditions. Once the participants identified the target, it would be easier if they could lock the position of the pointer. This way even if their partners were moving, the pointer would still be locked in position.

Some of the participants gave the following feedback: *"A simple addition in the Mouse condition would be to have a toggleable marker to prevent distraction on larger transitions," "In the mouse system you could use a visual representation of your clicks. You could have a feature that locked the highlighted area so that you could point that place while you look somewhere else."*

Adding onto the feedback from the remote experts, the local workers were also requested to provide their comments in the post experiment questionnaire. Similar to the results from the remote experts, the results from local workers also showed that they were in favour of the Gaze condition over the other two conditions. The local workers felt that they were able to communicate well with their partners using the gaze and mouse cues. They also felt that it was time consuming when their partners tried to instruct them using voice alone. To quote one of the participants who ranked in favour of the Gaze condition, he said, *"It was easy to perform the action with the help of the gaze pointer. Time saving."* He also said, *"It took a lot of time to search the buttons that my partner was trying to explain how the button looks."*

6. Discussion

The results from the comparison of the task completion time showed that the participants were able to complete the task faster in the Gaze condition compared to the Voice or Mouse conditions. This shows that using gaze cues in addition to voice communication could be more effective for completing a remote collaboration task compared to plain voice communication. This is probably due to the efficiency of eye gaze as a cue. Local workers could be alerted to the target element much earlier than in the other cases, and could therefore start acting sooner.

Gaze following has also been shown to be an effective and important non-verbal communication cue, so integrating support for this appears to be effective as well. The gaze cue worked as an implicit input for indicating focus of interest without requiring explicit action as in the mouse pointing. When the remote expert looks around they are able to provide continuous input, compared to mouse input that has to be explicitly moved around.

The feedback from the users on the SUS survey and the overall ranking show the most favourable result towards using gaze with voice together in the collaborative process. The usability level of using voice alone was significantly lower than the other two conditions. This result indicates that both of the cues were easy enough to use and had similar effect on the task. This was also evident from the qualitative feedback from the users in the post experiment questionnaire.

From the qualitative feedback provided by the users, it is clear that they enjoyed the collaborative process. They felt that it would be a valuable addition to have mouse or gaze cues added to their regular voice communication when performing a task that involves target identification.

The results from the rankings for the three conditions showed the remote expert participants were more in favour of the Gaze condition compared to the other conditions. They especially valued the gaze cue over the other two conditions in terms of helping them feel present and connected with their remote partners. While the results from the Social Presence questionnaire also showed a trend towards higher rating with having visual cues, there was no statistically significant difference found between the Mouse and Gaze conditions.

6.1. Limitations

The study results reported the user experience mostly from the remote expert's perspective, due to loss of quantitative measurements on the local worker's side. We tried to reflect the local worker's perspective through qualitative feedback collected, yet further investigation in the future is necessary focusing on the local worker's perspective.

Future studies including more different types of participants would be also desirable. In this study, many participants knew each other beforehand, which could have affected the sense of social presence and the quality of collaboration. It would be interesting to have future studies with people who were total strangers and compare to results with people who knew each other.

Lastly, the task used in the experiment did not clearly indicate success or failure. This limits our ability to judge the effectiveness of the collaboration, as the task completion time only indicates efficiency. Further investigation using tasks in real-world settings would be beneficial to better understand the true value of sharing gaze cues in remote collaboration. It would be also desirable to test a wider range of tasks, from remote instruction to problem solving, to see which types of task benefit from remote gaze input most.

7. Conclusion and Future Work

This research investigated the effect of sharing gaze in a remote collaboration setup using an HMD/HMC system. Compared to prior work, this paper focused on sharing the gaze of the remote expert. The results show sharing gaze helped users to perform faster than voice only and mouse conditions. It also enabled remote experts to feel more connected with their partners in the voice only condition. Ranking results clearly reflect this, where participants preferred sharing gaze cues over a mouse cursor or voice only.

These results imply that providing gaze from a remote expert could significantly improve collaboration in a task involving indicating objects and locations in the real world.

There are several possible directions for future research. User studies could be done using eye trackers for both the local workers and the remote experts in a video conferencing system. We could also mimic a range of different real-life scenarios such as performing a plumbing task, simulating an army mission, reporting on construction field work, etc.

Finally, in addition to gaze cues, we would also like to further explore other types of cues (e.g., facial expressions, heart rate, or galvanic skin response) that could help improve the user experience of remote collaboration by enabling the users to better understand each other.

References

[BCD*08] BRENNAN S. E., CHEN X., DICKINSON C. A., NEIDER M. B., ZELINSKY G. J.: Coordinating cognition: The costs and benefits of shared gaze during collaborative search. *Cognition 106*, 3 (2008), 1465–1477. 2

[BHKS98] BAUER M., HEIBER T., KORTUEM G., SEGALL Z.: A collaborative wearable system with remote sensing. In *Wearable Computers, 1998. Digest of Papers. Second International Symposium on* (1998), IEEE, pp. 10–17. 1

[BKS99] BAUER M., KORTUEM G., SEGALL Z.: " where are you pointing at?" a study of remote collaboration in a wearable videoconference system. In *Wearable Computers, 1999. Digest of Papers. The Third International Symposium on* (1999), IEEE, pp. 151–158. 2

[Bro96] BROOKE J.: Sus-a quick and dirty usability scale. *Usability evaluation in industry 189*, 194 (1996), 4–7. 4

[CHN*10] CARLETTA J., HILL R. L., NICOL C., TAYLOR T., DE RUITER J. P., BARD E. G.: Eyetracking for two-person tasks with manipulation of a virtual world. *Behavior Research Methods 42*, 1 (2010), 254–265. 1

[Duc07] DUCHOWSKI A. T.: *Eye tracking methodology: Theory and practice*. Springer, 2007. 1

[FKS00] FUSSELL S. R., KRAUT R. E., SIEGEL J.: Coordination of communication: Effects of shared visual context on collaborative work. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work* (2000), ACM, pp. 21–30. 1, 2

[FSK03] FUSSELL S. R., SETLOCK L. D., KRAUT R. E.: Effects of head-mounted and scene-oriented video systems on remote collaboration on physical tasks. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (2003), ACM, pp. 513–520. 2

[GLB16] GUPTA K., LEE G. A., BILLINGHURST M.: Do you see what i see? the effect of gaze tracking on task space remote collaboration. *IEEE transactions on visualization and computer graphics 22*, 11 (2016), 2413–2422. 1, 2

[GNTH14] GAUGLITZ S., NUERNBERGER B., TURK M., HÖLLERER T.: In touch with the remote world: Remote collaboration with augmented reality drawings and virtual navigation. In *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology* (2014), ACM, pp. 197–205. 2

[HB04] HARMS C., BIOCCA F.: Internal consistency and reliability of the networked minds measure of social presence. 4

[HYS16] HIGUCH K., YONETANI R., SATO Y.: Can eye help you?: Effects of visualizing eye fixations on remote collaboration scenarios for physical tasks. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (2016), ACM, pp. 5180–5190. 2

[IKTM90] ISHII T., KUZUOKA H., TAKAHARA T., MYOI T.: Collaboration system for manufacturing system in the 21st century. In *International Conference on Manufacturing Systems and Environment* (1990). 2

[KFS03] KRAUT R. E., FUSSELL S. R., SIEGEL J.: Visual information as a conversational resource in collaborative physical tasks. *Human-computer interaction 18*, 1 (2003), 13–49. 1, 2

[KLSB14] KIM S., LEE G., SAKATA N., BILLINGHURST M.: Improving co-presence with augmented visual communication cues for sharing experience through video conference. In *Mixed and Augmented Reality (ISMAR), 2014 IEEE International Symposium on* (2014), IEEE, pp. 83–92. 1, 2

[KMS96] KRAUT R. E., MILLER M. D., SIEGEL J.: Collaboration in performance of physical tasks: Effects on outcomes and communication. In *Proceedings of the 1996 ACM conference on Computer supported cooperative work* (1996), ACM, pp. 57–66. 1

[LMDG16] LI J., MANAVALAN M., D'ANGELO S., GERGLE D.: Designing shared gaze awareness for remote collaboration. In *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion* (2016), ACM, pp. 325–328. 2

[MD95] MOORE C., DUNHAM P. J.: *Joint attention: Its origins and role in development*. Lawrence Erlbaum Associates, 1995. 2

[MHPV13] MÜLLER R., HELMERT J. R., PANNASCH S., VELICHKOVSKY B. M.: Gaze transfer in remote cooperation: Is it always helpful to see what your partner is attending to? *The Quarterly Journal of Experimental Psychology 66*, 7 (2013), 1302–1316. 2

[MKB16] MASAI K., KUNZE K., BILLINGHURST M.: Empathy glasses. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (2016), ACM, pp. 1257–1263. 1

[NCD*10] NEIDER M. B., CHEN X., DICKINSON C. A., BRENNAN S. E., ZELINSKY G. J.: Coordinating spatial referencing using shared gaze. *Psychonomic bulletin & review 17*, 5 (2010), 718–724. 1

[OOF*08] OU J., OH L. M., FUSSELL S. R., BLUM T., YANG J.: Predicting visual focus of attention from intention in remote collaborative tasks. *IEEE Transactions on Multimedia 10*, 6 (2008), 1034–1045. 2

[SB04] STEIN R., BRENNAN S. E.: Another person's eye gaze as a cue in solving programming problems. In *Proceedings of the 6th international conference on Multimodal interfaces* (2004), ACM, pp. 9–15. 2

[SKJC95] SIEGEL J., KRAUT R. E., JOHN B. E., CARLEY K. M.: An empirical study of collaborative wearable computer systems. In *Conference companion on Human factors in computing systems* (1995), ACM, pp. 312–313. 1