Do You See What I See? The Effect of Gaze Tracking on Task Space Remote Collaboration

Kunal Gupta, Gun A. Lee, and Mark Billinghurst

Abstract—We present results from research exploring the effect of sharing virtual gaze and pointing cues in a wearable interface for remote collaboration. A local worker wears a Head-mounted Camera, Eye-tracking camera and a Head-Mounted Display and shares video and virtual gaze information with a remote helper. The remote helper can provide feedback using a virtual pointer on the live video view. The prototype system was evaluated with a formal user study. Comparing four conditions, (1) NONE (no cue), (2) POINTER, (3) EYE-TRACKER and (4) BOTH (both pointer and eye-tracker cues), we observed that the task completion performance was best in the BOTH condition with a significant difference of POINTER and EYETRACKER individually. The use of eye-tracking and a pointer also significantly improved the co-presence felt between the users. We discuss the implications of this research and the limitations of the developed system that could be improved in further work.

Index Terms—Computer-supported collaborative work, Computer conferencing, teleconferencing, and videoconferencing

1 Introduction

In recent years, wearable computers and cameras have become more common, and researchers are exploring how they can be used to support remote collaboration. With desktop computers the focus for teleconferencing systems is often on "talking head" video conferencing, where one person can see the other's face [14]. However, wearable systems also allow the sharing of first person video and a view of the remote user's task space [31]. Researchers have studied how these systems can be used to allow remote helpers to help local workers better perform real world tasks. For example, using virtual graphics to show users how to interact with real objects [18], view remote hands in the local worker's view [23] and place Augmented Reality cues on real objects [3].

In these systems the remote helper typically sees the local worker's workspace through the camera view of a head worn camera. However, while the camera captures the view of the local worker's environment, it doesn't show exactly what they are looking at. With a wide field of view camera, the local worker may be looking at one part of their workspace, while the remote helper is focusing on a different part. In this research we explore how providing more accurate focus of attention information could affect remote collaboration. We combine an eye-tracker with a head mounted display (HMD), and a head worn camera (HWC) to create a system that provides richer communication cues between the collaborators.

The main difference in this system from other head worn collaborative interfaces is that the use of an eye-tracker allows the remote helper to know exactly where the local worker is looking. Previous research has shown that people look towards objects before manipulating them [16] and that gaze is a strong indicator of areas of interest in a scene [28]. Thus our system could provide richer non-verbal communication, and so it is important to explore what the effect of this could be on the remote collaboration.

There have been many studies that explore the use of head worn cameras for remote collaboration (e.g., [7], [2]). However there have only been a few examples of earlier research that has used a head worn eye-tracker (e.g., [8], [9]) and none that have combined a HMD and HWC with an eye-tracker. Thus, our research is novel and makes the following contributions:

(1) it is one of the first examples of a wearable system for remote collaboration that combines eye tracking, a HMD and HWC,

(2) it is one of the first user studies exploring the effect of sharing gaze cues on wearable remote collaboration,

(3) it provides a good example of how to evaluate remote task space collaboration with eye-tracking using a variety of qualitative and quantitative experimental measures.

In the remainder of the paper we first provide an overview of the related work, then describe the prototype we developed, followed by a user evaluation of the prototype, a discussion of the results, and finally a conclusion and directions for future research.

- *Kunal Gupta is with HIT Lab NZ. E-mail: kgupta2789@gmail.com.*
- *Gun A. Lee is with HIT Lab NZ. E-mail: gun.lee@hitlabnz.org.*
- *Mark Billinghurst is with University of South Australia. E-mail: mark.billinghurst@unisa.edu.au.*

2 Related Work

Our research builds on earlier work on task space collaboration, head worn collaborative systems, and eye-tracking. In this section we review research in each of these areas and highlight the research gap that our work addresses.

2.1 Task Space Teleconferencing

Teleconferencing is a subject that has been well studied for several decades, but most of the focus has been on trying to reproduce face-to-face collaboration. Applications provide a "Talking Head" experience that allows a user to see a remote person's face, but not what they are seeing and doing [14]. In our research we are interested in developing collaborative systems that help people perform tasks in the real world and so a shared view of the local workers task space could be particularly important.

Previous research on tools for enhancing "Task Space" collaboration has included using a fixed camera placed over the shoulder of a person performing a task with the video streamed to a remote helper [9], several cameras that provide multiple views of the workspace [12], and small cameras that can be placed in the workspace and controlled by the remote collaborator [22].

More recently, work has been conducted on using handheld tablets with cameras for supporting task space collaboration. For example, Kim et al. [19] developed a system that allowed a local user with a tablet to communicate with a remote helper on a desktop computer, and explored if pointing or drawing on live video or still images were better for remote collaboration. Tablets provide intuitive touch based interaction, but one of the limitations of the system was that the virtual annotations on the video were not fixed to real objects, and so if the user moved the tablet they no longer aligned with the point of interest. This was addressed by the work of

Gauglitz et al. [11] who used SLAM based visual tracking to align Augmented Reality (AR) annotations to the real world. In their tablet based remote collaboration system, when the remote helper makes annotations they remain fixed in place in the local worker's view.

Using fixed cameras can limit the worker to perform a task in a particular workspace. Handheld and fixed cameras also have the limitation that it can be difficult for the remote helper to know exactly what the local worker is focusing on. To overcome these limitations the camera can be placed on the worker's head, enabling the remote helper to see a first person point of view of the work environment and for the worker to move around freely. In the next section we review research in head worn collaborative systems.

2.2 Head Worn Collaborative Systems

One of the first examples of using a Head Worn Camera for remote collaboration was the SharedView system [15], [21]. In this case the view from the camera mounted on the head of a local factory worker was shared with a remote helper and viewed on a monitor. The remote helper could use hand gestures to provide assistance, which were filmed and then shown back in a HMD that the local worker was also using. This allowed the local worker to see the hand gestures of the remote helper helping them complete the task. In a user study they found that this system enabled the local worker to complete a task faster than with a fixed camera system.

Since then many people have explored the use of head worn systems for remote task space collaboration. For example, Fussell et al. [7] used a HWC to view objects of interest in remote collaboration on physical tasks, where a local worker was sharing their view with a remote helper. However, there was no way for the remote helper to point into the local worker's view and so they were limited to giving verbal feedback. To address this, there is other research supporting the use of telepointers in HWC based remote collaboration. For example, Bauer et al. [2] demonstrated that pointing to real world objects using a telepointer along with verbal and visual communication significantly enhances the remote collaboration experience. Kim et al. [20] showed how SLAM tracking can be used to allow a remote helper to place AR annotations which appear fixed in place in the local worker's HMD view of their task space. They also compared using a HMD to tablet for remote collaboration, finding that users liked the hands-free nature of the HMD and having the camera closer to their eyes.

The development of lightweight head worn computers such as Google Glass have made remote collaboration easier than ever before, in a wider range of settings. For example, Armstrong [1] showed how Glass could be used to provide remote assistance during surgery, and HELPlighting [13] have developed a commercial product named VIPAAR for remote consultation based on Glass. In both of these cases Google Glass is used to view video of the local user's workspace enhanced by virtual cues from remote helpers.

Head worn cameras can be used to share a local worker's view with a remote helper, but what the remote helper can see is fixed to the local worker's viewpoint. So if the local worker turns his or her head the remote view also follows, reducing the ability of the remote user to view points of interest in the local workers workspace. One way to address this limitation is to use shoulder or body worn cameras that allow the remote helper to have an independent view into the local worker's workspace [30]. A second limitation is that the remote expert doesn't know exactly where the user is looking in the shared view. This could be addressed by the use of gaze tracking, but there has been little research on this for head worn systems. In the next section we review gaze tracking in more detail.

2.3 Gaze Tracing in Teleconferencing

In face-to-face conversation eye gaze provides information about where the person is directing his or her focus of attention and so it should also be an important cue in remote collaboration. There has been some research on the use of eye-tracking in desktop and virtual reality teleconferencing systems. For example, Carletta et al. [6] demonstrated a collaborative desktop game in which two users could control objects using their eye gaze. In EyeCVE [33] the user's gaze was tracked and used to control their virtual avatar's eye movements, facilitating awareness of other people and remote mutual eye-contact in an immersive virtual environment.

Using a monitor mounted eye-tracker, Brennan et al. [4] compared performance in a visual search task between remote people who could either just talk to each other, or use speech and a gaze cue. They found that sharing gaze information significantly improved performance because it allowed one person to know the other's focus of attention. More recent research has replicated this result in remote puzzle-solving tasks [26], [32] and simulated bomb diffusion [24]. Overall, for desktop systems it seems that gaze provides an excellent cue for inferring user attention and improving remote collaboration.

In contrast, there has been little research on using head worn eye-tracking systems and remote collaboration. One of the few systems that explores this is the work of Fussell et al. [8], [9] who developed a system with a HWC with an attached eye tracker for the local worker. This sent workspace video along with the local worker's eye gaze details to the monitor of a remote helper. However, the local worker was not wearing a head mounted display, and so the remote helper was not able to provide virtual cues as feedback to help them. They found that the performance of using a HWC with eye gaze information was not significantly different than using an audio only mode, and performance was significantly worse than with a fixed scene camera showing the entire workspace. In a later experiment, Ou et al. [28] found that the focus of attention can be predicted from monitoring eye gaze in a remote collaboration task, and eye-tracking provides a benefit.

The most directly relevant research is our own earlier work that combines a HMD, HWC and wearable eye-tracker in a similar system to that described in this paper [25]. However in this case, the system was only evaluated in a very preliminary pilot study, with a very different task, was not using remote expert instruction, and did not have a divided attention activity.

2.4 Summary

As we have shown, there have been previous examples of head worn systems developed that support remote collaboration. However there is little research available that has investigated eye-tracking in head worn systems for remote collaboration. Table 1 summarizes some of the relevant research papers we have mentioned and the features of these systems. In this table, *HWC* = head worn camera, *HMD* = head mounted display, *ET* = eye-tracker, *Rem* = Remote conferencing, *FtF* = Face to Face conferencing. As can be seen from the table, apart from our own work [25], we could find no research that combines a head mounted camera with a head worn display and an eye-tracker for remote collaboration (HWC, HMD, ET and Rem.), and especially research that has conducted formal user studies with such a system. So this work fills an important gap in the research literature.

In the next section we describe the prototype interface we developed, followed by a user study conducted to explore the effect of showing gaze information on remote collaboration.

Table 1. Previous wearable collaborative systems and the configuration we are targeting with our research (shaded grey).

Papers	*HWC*	*HMD*	*ET*	*Rem*	*FtF*
[8],[9],[28],[29],[36]	X		X	X	
[6],[10],[33]			X		X
[4],[24],[26], [27],[32],[34]			X	X	
[1],[2],[7],[13], [15],[20],[21]	X	X		X	
[9],[12],[11], [19],[22],[30]				X	
[25]	X	X	X	X	

3 Prototype Design

The prototype system that we developed combines the following elements: (1) a head mounted eye-tracker, (2) head mounted camera, (3) head mounted display, and (4) remote viewing software that allows a remote expert to see and annotate on the local worker's view. Figure 1 shows the system components.

The HMD chosen was a Brother AirScouter [5], a high quality, optical see-through monocular display with an 800 by 600 pixel resolution and 22.4 degree field of view. It has an adjustable mounting mechanism so that the display can be aligned with the user's sight line. For the eye tracker we used the Microsoft Lifecam HD 5000 with a custom 3D printed enclosure pointing at the user's eye. Finally, a Logitech Webcam C920 camera was mounted on the system facing outwards as the HWC. Figure 2 shows the final assembled prototype.

The software part of the prototype consisted of two components. One was the Eye Tracking Application that tracks the eye's pupil and overlays a gaze marker on the real world video captured from HWC. The second was the Annotation Application that allows the remote helper to move a mouse pointer over the remote video view of the local worker and show the result in the display of the HMD.

For the Eye Tracking Application we modified the open source eye tracking software developed by Pupil Labs [17]. This software uses computer vision techniques to track the eye pupil (see Figure 3) in the video stream from the Microsoft LifeCam HD 6000 webcam and map it over the video from the HWC capturing the real world. The software can track the eye at a frame rate of 30 fps.

For the remote helper, we wanted to show the local worker's view with their eye gaze indicated on top of it. The Pupil Lab's default eye tracking viewer application provides this functionality except that it adds some additional user interface elements for camera calibration and recording, and graphs showing CPU processing, etc. So we removed these additional components leaving only the video and eye gaze marker parts. The local worker's gaze point is shown as a red circle over the live video. Figure 4 shows the remote helper's view on their desktop computer.

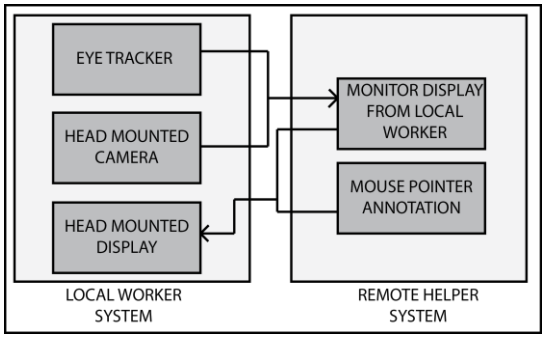


Fig 1. System components.



Fig. 2. Prototype hardware.

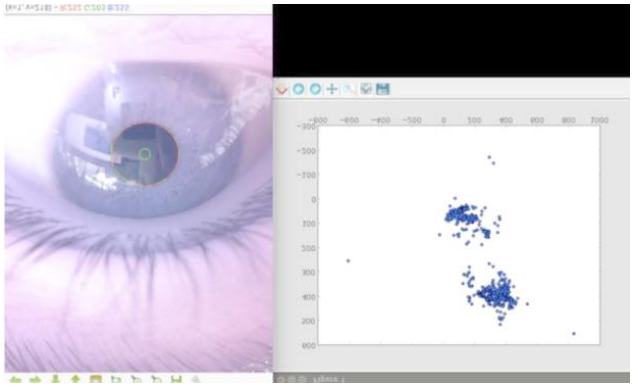


Fig. 3. Pupil Labs eye tracking software.

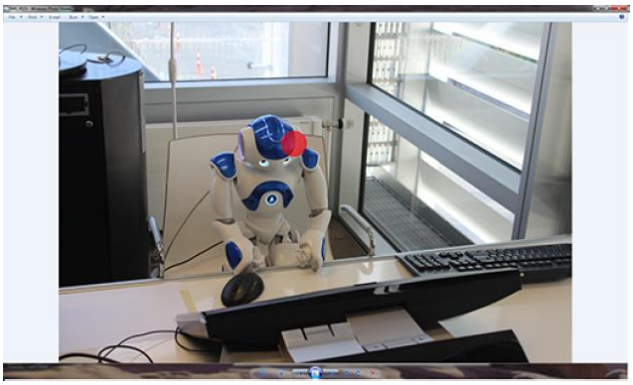


Fig. 4. User Interface on the remote helper's end showing the video and eye-tracking red dot shared from the local end user.

To support collaboration we created an Annotation Application to enable the remote helper to provide virtual communication cues back to the local worker. In the prototype system the head worn camera from the local worker is plugged directly into the remote helper's computer and the live video view overlaid with their eye-tracking information. We then developed software that added a green pointer circular annotation on the screen that was activated when the remote helper clicked with a mouse on their monitor. The pointer was programmed to stay on the display for 3 seconds after the remote helper released the mouse button, a period found from pilot studies to be neither intruding in the next instruction nor ignored by the local worker. However, the pointer was not fixed to the clicked target in the real world, unlike [11], but was shown in screen space as in [20].

Finally, we developed a screen cast application in the .NET language that could share the remote helper's computer display to the local worker's head mounted display. The Air Scouter HMD was plugged in as an extended display to the remote helpers computer, so we assigned the casted screen to the Air Scouter's display. We used jpeg compression to transfer screen images at a frame rate of 24 frames per second to cast the remote helper's display to the HMD.

Using this software, the remote helper can watch the video stream from the local worker's HMC along with the eye gaze marker and can click on the display whenever they wanted to provide visual feedback to the worker to help them perform their task. The remote helper screen with the visual feedback is then shown back into the local worker's HMD, forming a video see-through configuration. So the local worker sees exactly what the remote helper is seeing on their display, except for an additional visual cue showing the remote helper's pointer. The HMD view wasn't overlaid on the real environment as in a traditional Augmented Reality interface. The users can also talk to each other, so they can share visual and audio cues between each other to support the remote collaboration.

4 Experimental Evaluation

To evaluate the prototype, we designed an experiment comparing various interfaces conditions for the remote helper and local worker with respect to the time taken to complete a particular task and their subjective feedback. The main objective of this evaluation was to compare the co-presence levels and performance in systems for remote collaboration that provide different cues. Previous research has shown that gaze and pointing cues can provide a significant difference in co-presence in remote collaboration interfaces (e.g. [20][29]). Similarly, sharing gaze and pointing cues can significantly improve task performance in remote collaborative tasks (e.g. [4][18]). Based on this, the hypotheses of the experiment were:

• *H1: Co-Presence* - There will be a significant difference in co-presence measures between conditions that provide additional attention cues and those that don't.

• *H2: Performance* - There will be a significant difference in performance time between conditions that provide additional attention cues and those that don't.

4.1 Experiment Design

The experimental setup was designed to support a remote collaboration task in a controlled environment. We arranged a room with a local worker on one side and a remote helper on the other separated by a divider so they couldn't see each other. The prototype HMD was setup on the local worker's side, connected to a computer, with the monitor from the computer placed on the remote helper's side, so they could both see the shared task space video. The participants could clearly hear each other and spoke to each other constantly during the experiment to complete the task. This audio condition was the same for all tasks, so there was no audio confound. All of the interface conditions used the same setup (see Figure 5).

The experimental task involved the local worker constructing three-dimensional structures using LEGO Duplo pieces with assistance from the remote helper. We created four different structures of 17 pieces. Simple tasks using Lego bricks or similar have previously been used in many papers related to remote task space collaboration (e.g. [7][9][10][21]). In pilot testing 5 users built each model following a manual, taking about the same amount of time (within 10 seconds), showing a similar level of difficulty. We counterbalanced all the models across all conditions to further reduce model effects, and similar times were found for each model in the same condition. Figure 6 shows one of the models after completion.

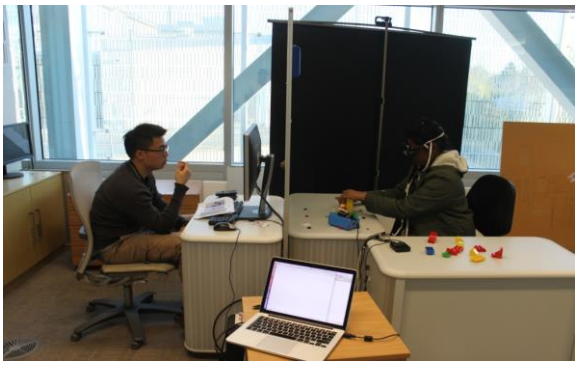


Fig. 5. Participants performing the experimental task.

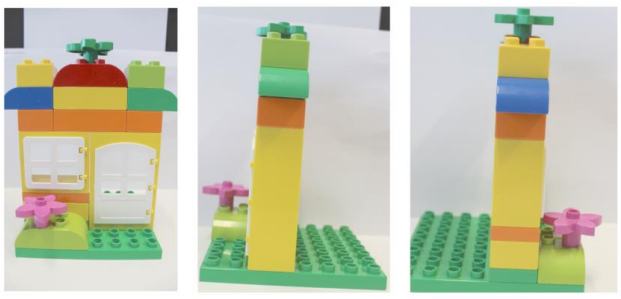


Fig. 6. Example of a typical Lego brick model.

To encourage collaboration we added two constraints. First, for the local worker we used two tables arranged in an 'L' shape, where one table held the loose LEGO blocks and the model was constructed on the other table (see Figure 5). The local worker was told to take only one piece at a time from the block table to main workspace and that they had to use that block before taking the next one. These rules encouraged local workers to rotate their head from left to right and change their focus of attention between areas.

We also introduced an additional subtask to divide the local worker's attention; a countdown timer that they needed to pay attention to while constructing the LEGO structure. The timer was created using a 4-digit 7-segment LED display connected to an Arduino board. It counted down from 40 seconds to 0 and continued into negative time unless a button was pressed that reset the timer to a random number. The timer was placed on the building table so that it was visible within the shared camera view. We asked both participants to keep track of it and to press the button before it reached 0. To make sure the participants could not avoid paying attention to the timer, the reset button was active only when less than 4 seconds were left on the clock. The number of times the participants pressed the button outside of this range was recorded.

We evaluated the different interfaces and interaction techniques for the prototype system using a within-subject study of 2 by 2 (i.e. four) conditions. We had two independent variables, (1) POINTER - having a virtual pointer available for the remote helper to point on the local worker's HMD view, and (2) EYETRACKER – sharing eye gaze information from the local worker to the remote helper. Table 2 summarizes the interfaces used in the four conditions. The four conditions were:

No Cue (*NONE*): In this condition, only audio and live video were shared as a part of the remote collaboration.

Pointer Cue (*P*): A virtual pointer was used by the remote helper to aid the local worker. When the remote helper clicks on the shared video on his display, a green virtual pointer appears at the clicked position, and is also shown in the local worker's HMD. The pointer is provided as an additional cue to the basic audio and video cues in the NONE condition.

Eye Tracker Cue (*E*): An eye gaze marker (red circle) is displayed on the remote helper's monitor to show the local worker's focus of attention. The eye gaze was provided as an additional cue to the basic audio cues in the *NONE* condition.

Both Pointer and Eye-Tracker Cues (*BOTH*): Both eye-tracker and pointer cues were provided in

Table 2. Experimental conditions.

	EYETRACKER: No	EYETRACKER: Yes
POINTER: No	*NONE*	*E*
POINTER: Yes	*P*	*BOTH*

Table 3. Likert scale rating questions.

Q#	Statement
Q1	I felt **connected** with my partner.
Q2	I felt I was **present** with my partner.
Q3	My partner was able to sense my presence.
Q4	My partner (*or for Remote Helper*: I) **could tell** when I (*or for Remote Helper*: my partner) **needed assistance**.
Q5	I **enjoyed** the experience.
Q6	I was able to **focus** on the task activity.
Q7	I am confident that we **completed** the task correctly.
Q8	My partner and I **worked together** well.
Q9	I was able to **express** myself **clearly**.
Q10	I was able to **understand partner**'s message.
Q11	**Information** from partner was **helpful**.

addition to the cues provided in *NONE* condition.

The order of conditions tried by the participants was arranged in a Balanced Latin Square design in order to counterbalance the carryover effects between conditions.

Participants were recruited from the university. On arrival, they were randomly assigned as a local worker or a remote helper, and we gave them a pre-task questionnaire asking for demographic information, including their previous experience with video conferencing, and using LEGO blocks. They were told about the main objective of this research, and the experimenter demonstrated the interfaces by explaining the cues provided with each interface.

Before the experimental sessions, we had each pair complete a practice face-to-face collaborative task in which the remote helper was provided with an instruction manual and asked to assist the local worker to complete the LEGO structure. The practice model was different from any of the models used in the actual test. Through this, participants learnt how to construct LEGO models together.



Fig. 7. Local Worker assembling Lego.

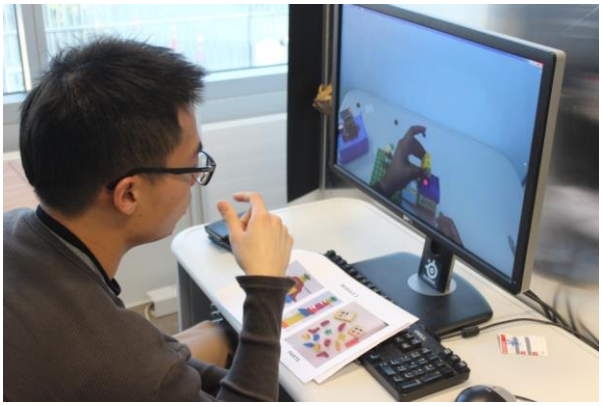


Fig. 8. Remote Helper guiding the Local Worker using the eye-tracker.

After the practice task, participants were separated to sit at their desks and perform the tasks using the provided interface. For each condition, we had the remote helper create the structure first following instructions provided. Then the same LEGO blocks were provided to the local worker to perform the experimental task following the instructions given by the remote helper. After each condition, both subjects were asked to complete a questionnaire with Likert scale rating items (see Table 3) on a scale of 1 (totally disagree) to 7 (totally agree) and the time to complete the task was recorded. The questionnaire was based on a prior published survey (from [20]), slightly modified to reflect our experiment. After all the conditions, participants were asked to rank the interface conditions on various aspects of their experiences and interviewed.

We invited 30 participants (15 pairs) to participate in the experiment, but data from only 13 pairs was used. One pair stopped the experiment early, and one pair was removed because one of the participants gave the same response to all the questions.

The participants were from 21 to 33 years old, and 20 (77%) were male and 6 (23%) were female. All of them had good English understanding and speaking. On average, they used video-conferencing once a year. However none of them had previous experience of constructing LEGO over video-conferencing. Figures 7 and 8 show the participants performing the experimental task.

4.2 Results

This section reports on the results from the experiment. First, we report on the task performance time, and then the questionnaire data gathered from the participants. Finally, we summarize the qualitative feedback collected through asking open questions of the participants, and report on the results of conversational analysis.

The key results from the experiments include the following:

- Both the pointer and eye tracking visual cues helped participants to perform the task significantly faster.
- The pointer cue significantly improved both local workers' and remote helpers' perceived quality of communication, collaboration, and co-presence.
- The eye tracking cue not only significantly improved the communication and collaboration quality, but also other aspects such as sense of being focused for the local workers, and enjoyment for the remote helpers.
- The *BOTH* condition, which has both of the visual cues, was ranked as the best in most of the aspects of user experience, while the baseline *NONE* condition was ranked as the worst.
- Analysing the user conversation revealed that visual cues made the conversation more efficient, changed the choice of wording in deictic expressions, and helped participants' feel more connected.

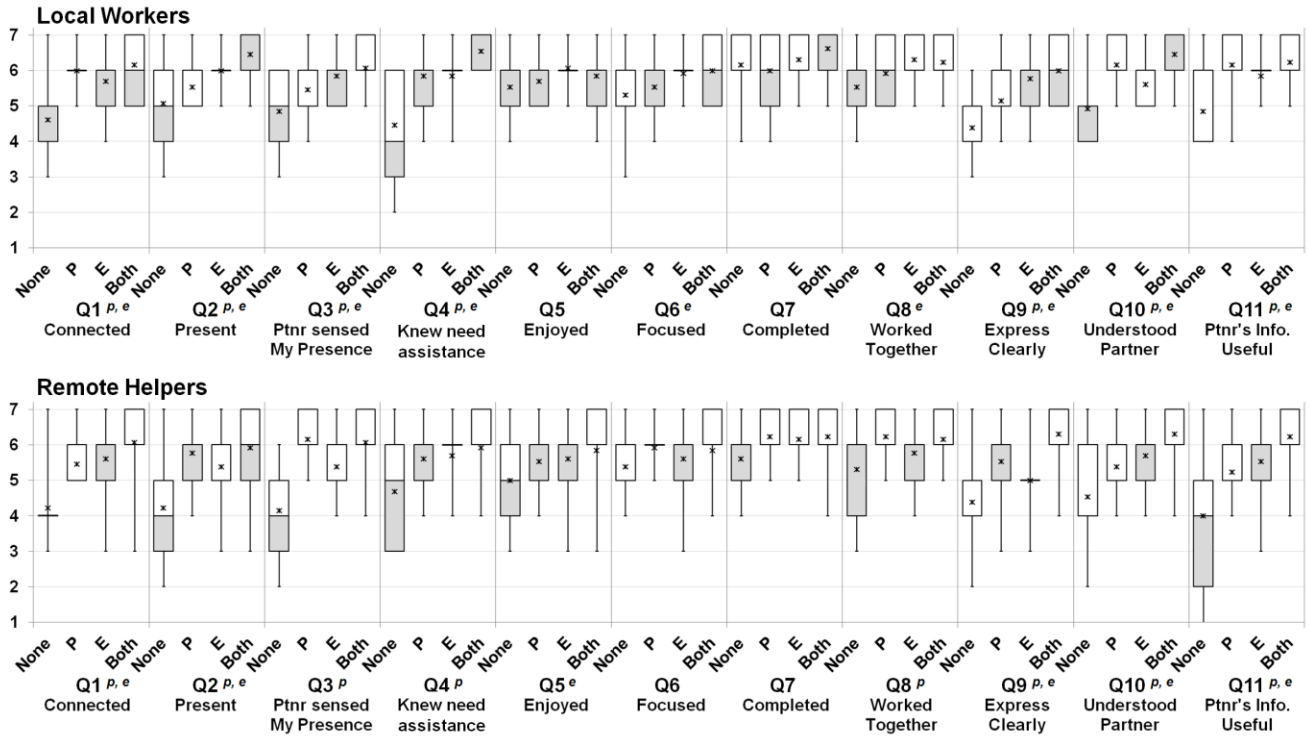


Fig. 11. Results of Likert scale rating (1: strongly disagree ~ 7: strongly agree, *: mean, significant main effects are marked as superscript to the questions: *p* for POINTING and *e* for EYETRACKING).

4.2.1 Task completion time

There was a significant different in performance time between the various interfaces. Figure 9 shows the average performance time across the different conditions. A repeated measure two-way ANOVA (a = 0.05) revealed that there was a significant main effect of both POINTER ($F(1,12)=4.908$, $p=.047$) and EYETRACKER ($F(1,12)=5.811$, $p=.033$) on the average time taken to complete a task using these interfaces. No significant interaction was found between POINTER and EYETRACKER ($F(1,12)=0.566$, $p=.466$).

Descriptive statistics showed that on average the participants took about 15% less time to complete the task when having a POINTER (with: M=217.5 second, *Standard Error*=16.5; without: M=252.0, SE=15.3), and about 10% less time when having an EYETRACKER (with: M=223.1, SE=13.7; without: M=246.4, SE=15.7). Figure 9 shows that participants took less time to complete the task in *E* (M=245.7, SE=17.2) and *P* (M=234.5, SE=20.7) conditions compared to the baseline *NONE* condition (M=258.3, SE=19.6). Overall participants completed the task fastest while using both the POINTER and EYETRACKER together (M=200.5, SE=14.1), saving about 22% of time compared to the *NONE* condition.

4.2.2 Questionnaire: Likert scale rating

After finishing the task in each condition, participants were asked to answer a questionnaire that included 11 questions (see Table 3). The Cronbach's alpha indicated very good internal consistency among Likert items (a=.937) and excluding each item didn't affect the reliability much (a ranging from .928 to .937). Figure 10 shows the results with taking all Likert items aggregated into a single scaled index of overall collaborative experience (from 0 to 100). The questions were answered by both remote helpers and local workers, and we report the results for each group separately.

For factorial analysis of the Likert scale rating results, we used the Aligned Rank Transform (ART)

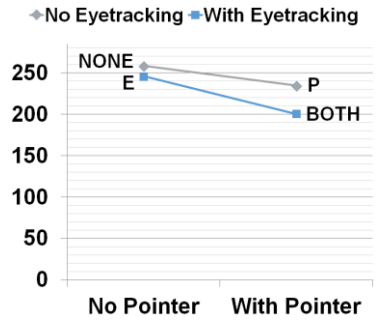


Fig. 9. Average Task Performance Time (seconds).

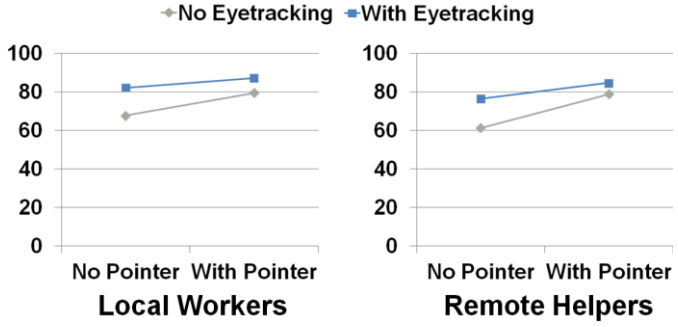


Fig. 10. Likert scale rating on collaborative experience

for non-parametric factorial analysis using ANOVA procedures (α=.05) [35]. In cases where we wanted to investigate pair-wise comparison between conditions, we used Wilcoxon Signed Rank tests.

The factorial analysis with ART ANOVA showed significant main effects of both POINTER ($F(1,12)=7.414$, $p=.019$) and EYETRACKER ($F(1,12)=26.822$, $p<.001$) factors on collaborative experience for the local workers, and also for remote helpers (POINTER: $F(1,12)=11.914$, $p=.005$, EYETRACKER: $F(1,12)=15.929$, $p=.002$). No significant interaction between the two factors was found for the local workers ($F(1,12)=2.023$, $p=.180$), while there was significant interaction for the remote helpers ($F(1,12)=5.157$, $p=.042$). The *BOTH* condition was rated highest and the *NONE* condition was rated lowest by both local workers and remote helpers.

We also analysed the rating results of each individual item. Figure 11 and Table 4 summarize the results of the individual rating questions. In most of the cases the baseline condition (*NONE*) was rated lowest, while rating for the *BOTH* condition was mostly the highest. The rating for the POINTER and EYETRACKER conditions were mostly in between the *NONE* and *BOTH* conditions.

Factorial analysis using ART ANOVA showed that there was a significant main effect of the POINTER factor in Q1~Q4 and Q9~Q11 for the local worker participants, but not for Q5~Q8. For the remote helper participants, a significant main effect of the POINTER factor was found in Q1~Q4 and Q8~Q11, but not for Q5~Q7. This shows having a pointer significantly affected both the local workers' and the remote helpers' subjective quality of communication and co-presence (Q1~Q4 and Q8~11) but not on enjoyment (Q5), focus (Q6), or self-confidence (Q7).

The EYETRACKER factor had a significant main effect for the local workers on Q1~Q4, Q6, and Q8~Q11, but no significant effect on Q5 or Q7. For the remote helpers, the EYETRACKER had a significant main effect on Q1, Q2, Q5, Q9~Q11, but not for Q3, Q4, and Q6~Q8. This reveals that, compared to the POINTER factor, the EYETRACKER had a significant effect not only on communication quality and co-presence (Q1~Q4 and Q8~Q11), but also on other aspects such as sense of being focused (Q6) for the local workers and enjoyment (Q5) for the remote helpers.

Although no significant main effect of EYETRACKER was found for Q3, Q4 and Q8 for the remote helpers, there were significant interactions between the independent variables, so we further investigated the interaction by comparing the *NONE* and *E* conditions in these questions. We found a significant difference for both Q3 (Z=-2.654, p=.008) and Q8 (Z=-2.121, p=.034) indicating the EYETRACKER factor did affect Q3 and Q8 but the effect was saturated when combining it with the POINTER which had very strong effect on these questions. For Q4, no significant effect of the EYETRACKER factor was found when comparing the *NONE* condition to the *E* condition (Z=-1.913, p=.056).

Most of the significant interaction effects between the POINTER and the EYETRACKER can be explained with a ceiling effect as most of the ratings for the *BOTH* condition were saturated at the highest value (7) on rating scale. There were two cases (Q3 and Q8 for remote helpers) where the descriptive variable suggested the *BOTH* condition was rated lower than the *P* condition, however a pair-wise comparison showed that the difference was not significant (Q3: Z=-0.276, p=.783; Q8: Z=-0.447, p=.655).

4.2.3 Questionnaire: Ranking

After finishing all four conditions, participants were given a questionnaire to rank the conditions based on their experiences with respect to different criteria described (1 = the best, 4 = the worst). Table 5 shows the list of the criteria for ranking.

Table 5. Ranking criteria.

No.	Ranking criteria
	Which condition was best ...
C1	at helping you to **enjoy** 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 your partner at the same workspace?
C5	for you (or the partner) to **know** that the partner (or you) **needed assistance**?
C6	at helping you **understand** the partner's message?

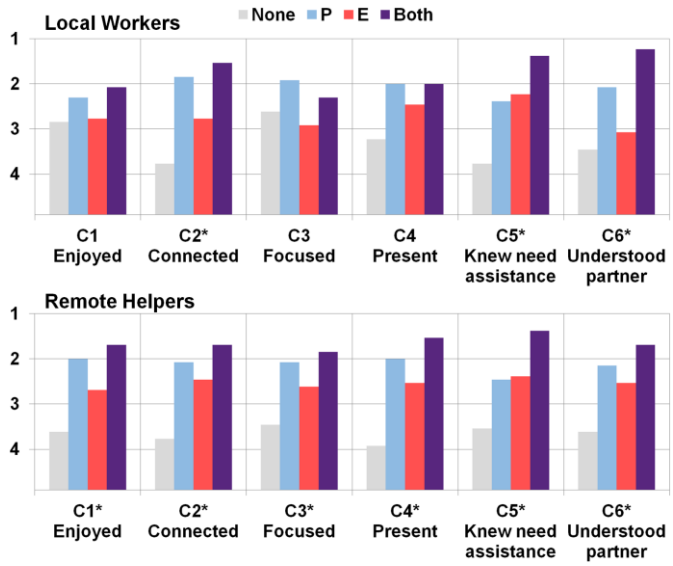


Fig. 12. Average ranking results (*: significant difference).

Table 4. Results of ART Repeated measure two-way ANOVA on rating questions (significant result in bold).

Q#	User L/R	POINTER $F_{1,12}$	POINTER p	EYETRACKER $F_{1,12}$	EYETRACKER p	Interaction $F_{1,12}$	Interaction p
Q1	L	9.763	**.009**	8.220	**.014**	8.291	**.014**
	R	15.096	**.002**	17.153	**.001**	6.052	**.030**
Q2	L	5.086	**.044**	14.153	**.003**	0.205	.659
	R	9.412	**.010**	7.926	**.016**	5.781	**.033**
Q3	L	6.858	**.022**	9.179	**.010**	1.185	.298
	R	22.511	**<.001**	4.648	.052	8.359	**.014**
Q4	L	8.372	**.013**	19.761	**.001**	5.501	**.037**
	R	9.828	**.009**	3.434	.089	8.715	**.012**
Q5	L	0.011	.917	4.195	.063	0.436	.521
	R	2.753	.123	5.589	**.036**	0.230	.640
Q6	L	0.321	.581	5.334	**.040**	0.260	.619
	R	4.110	.065	1.423	.256	0.589	.458
Q7	L	0.161	.695	4.248	.062	3.636	.081
	R	2.585	.134	2.500	.140	2.678	.128
Q8	L	0.386	.546	7.303	**.019**	1.581	.233
	R	5.172	**.042**	0.777	.395	9.060	**.011**
Q9	L	6.381	**.027**	17.388	**.001**	3.275	.095
	R	13.119	**.004**	14.944	**.002**	0.059	.813
Q10	L	14.690	**.002**	14.739	**.002**	1.389	.261
	R	9.273	**.010**	18.381	**.001**	1.523	.241
Q11	L	11.766	**.005**	9.946	**.008**	4.206	.063
	R	7.531	**.018**	18.036	**.001**	1.937	.189

Figure 12 shows the average results of ranking questionnaire. In most cases the *BOTH* condition was ranked in the first place, while the baseline condition (*NONE*) was ranked in the worst place. The other two conditions, *P* and *E*, were mostly ranked in-between with no clear distinction against each other.

To see if the conditions were ranked significantly different, we used Friedman tests (α=.05). The results showed that the remote helpers ranked the conditions significantly different in all criteria (C1: $\chi^2(3)=17.031$, $p=.001$; C2: $\chi^2(3)=19.062$, $p<.001$; C3: $\chi^2(3)=12.046$, $p=.007$; C4: $\chi^2(3)=24.969$, $p<.001$; C5: $\chi^2(3)=18.256$, $p<.001$; C6: $\chi^2(3)=15.738$, $p=.001$). With the local workers, only C2 ($\chi^2(3)=22.907$, $p<.001$), C5 ($\chi^2(3)=22.256$, $p<.001$), and C6 ($\chi^2(3)=24.535$, $p<.001$) showed a significant difference in ranking, while the rest of the criteria showed no significant difference (C1: $\chi^2(3)=3.185$, $p=.364$; C3: $\chi^2(3)=4.302$, $p=.231$; C4: $\chi^2(3)=7.031$, $p=.071$). This indicates that while POINTER and EYETRACKER affected the remote helpers on ranking in all aspects, they affected local workers in terms of the quality of fluent communication (C2: feeling connected, C5: feeling their partner knew their needs, C6: understanding their partner's message), but not much on enjoyment (C1), focus (C3), or sense of presence (C4).

In cases where significant difference was found in ranking, we performed post-hoc analysis using Wilcoxon signed-rank tests with the Bonferroni correction (α=.0083) to investigate if certain conditions were ranked significantly different compared to the others. In all the cases except for C5 with remote helpers ($Z=-2.254$, $p=.024$), the condition *P* was ranked significantly higher than the baseline condition *NONE* (local workers on C2: $Z=-3.270$, $p=.001$; C5: $Z=-3.140$, $p=.002$; C6: $Z=-2.738$, $p=.006$; remote helpers on C1: $Z=-3.022$, $p=.003$; C2: $Z=-3.236$, $p=.001$; C3: $Z=-2.946$, $p=.003$; C4: $Z=-3.219$, $p=.001$, C6: $Z=-2.961$, $p=.003$).

The *E* condition was also ranked significantly higher than the *NONE* condition by the local workers on C5 ($Z=-2.676$, $p=.007$) and by the remote helpers on C2 ($Z=-2.951$, $p=.003$), C4 ($Z=-3.286$, $p=.001$), and C5 ($Z=-2.697$, $p=.007$).

The *BOTH* condition was ranked significantly higher than the *NONE* condition by the local workers on C2 ($Z=-2.217$, $p=.002$), C5 ($Z=-3.196$, $p=.001$), and C6 ($Z=-3.223$, $p=.001$), and by the remote helpers on C2 ($Z=-2.951$, $p=.003$), C4 ($Z=-3.203$, $p=.001$), and C5 ($Z=-3.007$, $p=.003$). The *BOTH* condition was ranked significantly higher than the *E* condition by the local workers only on C6 ($Z=-3.223$, $p=.001$).

4.2.4 Questionnaire: Preference and qualitative feedback

When asked in which condition they could better understand where their partner was focusing, almost 85% (11/13) of the local workers preferred the condition in which pointer feature was available (i.e. *P* and *BOTH* conditions) as it was helping them to understand their partner's focus and clear instructions. One participant said: *"With Pointer, I can relate to what he is talking about, because I could understand him more."* In the case of the remote helpers, around 70% (9/13) chose the *BOTH* condition since they were able to see the place where their partner was looking and also point on the video. One participant said: *"The eye tracker helps me to look in the same view of my partner, and I know what he is doing and will do next"*. However, 15% (2/13) said that they didn't notice much difference with eye tracking as the focus of attention was not much of their interest.

When asked which condition helped them more in performing the task efficiently and quickly, in terms of the information provided, about 77% (10/13) of the local workers said that the *BOTH* condition was the best. One participant said, *"The eye tracker was giving my partner more information about where I looked at, while the pointer was for giving me the instruction from my partner, where I should look at and which piece I should take"*. Most of the remote helpers (85%, 11/13) also picked the *BOTH* condition as the best.

In terms of the timer tasks, 9 of the local workers (69%) said that the remote helper used the pointer cue to help them know when to press the button. One of the local workers said that, *"if I keep looking at the display, with the help of his mouse click on the block, I can finish the task quickly"*. On the other hand, 8 of the remote helpers (73%) suggested that eye tracking information was very useful to keep track of the timer since they knew if the local worker was paying attention to the timer or not, and so if they needed to remind them to check the timer regularly.

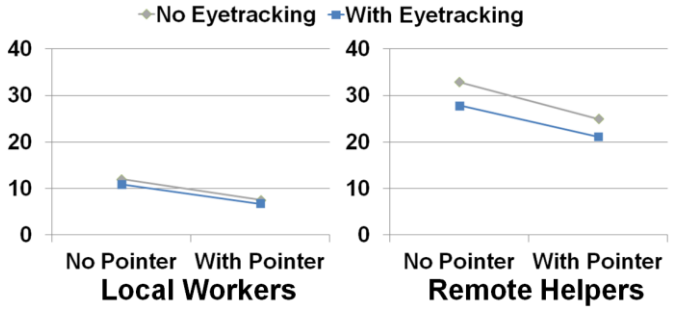


Fig. 13. Results of conversation analysis (number of phrases).

4.2.5 Behavioral differences

To understand user behaviour, we recorded the user conversation for later analysis. From this recording we made a written transcript of the conversation and counted the phrases of what the participants said during completing the experimental tasks. Using an approach commonly followed in other work (e.g. [7]), a phrase was defined as a distinct spoken utterance. Once written transcripts of the conversation were made it was simple to select out the types of speech used, for example deictic phrases from non-deictic and the speech behaviours used while pointing. Unfortunately the recording of one pair of participants was lost, so we analyzed the data from only 12 pairs of participants. Figure 13 shows the average number of phrases spoken by each participant.

Overall, remote helpers spoke more than local workers, and use of a pointer reduced the amount of communication needed. A two-way repeated measure ANOVA showed a significant effect of POINTER on the number of phrases said by both local worker ($F(1,11)=6.532$, $p=.027$) and remote helper ($F(1,11)=8.479$, $p=.014$). No statistically significant effect of EYETRACKER (local worker: $F(1,11)=0.538$, $p=.479$; remote helper: $F(1,11)=3.057$, $p=.108$) nor interaction between the two factors (local worker: $F(1,11)=0.012$, $p=.914$; remote helper: $F(1,11)=0.175$, $p=.683$) was found.

The remote helper's role was to give instruction to the local worker, so the remote helper mostly led the conversation and talked significantly more, saying on average three times more phrases than the local workers ($F(1,94)=103.004$, $p<.001$). In comparison, the local workers mostly said short phrases to either confirm that they understood the remote helper's instruction (e.g. *"Okay", "Oh, yeah",* or *"Yes"*) or to ask the remote helper to confirm if what he or she had done was correct (e.g. *"Here?", "This one?",* or *"Like this?"*).

We also noticed the remote helpers using different words for identifying objects or specifying directions. When needing to identify an object, in the conditions without the pointing cue, participants described the object in terms of colour, shape, or size (e.g. *"green piece", "square one", "yellow big one"*). In the conditions with the pointing cue, most of the time participants simply used the phrase *"this one"* while pointing at the object. A similar pattern was observed when specifying directions, such as participants saying *"put it here"* or *"next to this"* while using the pointer. In the conditions where the pointer cue was not available, participants used words that described direction more frequently (e.g. *"in front of your hand", "move right", "more left"*).

Another interesting observation was that when using the pointer cue the remote helpers tended to use the word *"this"* to refer to the objects on the local worker's side, compared to when without the pointer they referred to the objects on the shared video using the word *"that"* more.

5 Discussion

From the experiment results, we found that the time taken by the participants to complete the task was significantly improved by using both the pointer and eye tracking cues. On average, the *BOTH* condition which uses both of the cues took least amount of time, while the *NONE* condition was the slowest. The visual cues helped both the remote helpers and local workers to perform the task more quickly. The remote helper was able to use the pointer cue to give direct guidance to the local worker. One of the local workers said, *"The information provided by my partner in this condition helped in easily performing the task",* indicating the pointer cue enabled them to work significantly better than without it. Similarly, the eye tracker cue showed the local worker's focus of attention to the remote helper, making it easier for the remote helpers to instruct the local workers quickly by recognizing where they were looking.

The questionnaire results were in line with the task completion time results, showing a significant effect of both pointer and eye tracking cues on the answers to the questions related to communication quality, including Q1 (feeling connected), Q9 (being able to express oneself clearly), Q10 (understanding partner's message), and Q11 (partner's information being helpful) in rating questions, and C2 (feeling connected), C5 (knowing partner's need), and C6 (understanding partner's message) in ranking questions. These results show that with better communication through using the visual cues, the participants were able to perform more efficiently.

The user interface for the local worker in the *E* condition was the same as in the *NONE* condition (i.e. the instructions from the remote helper to local worker were only in verbal form since they had no pointer). However, participants still took less time to complete the task than in the *NONE* condition, and they rated the *E* condition higher in most questions asking about communication quality. These results show the benefit of making the remote helper aware of the local worker's focus of attention.

In the experiment, we provided separate points in the real world to focus on (the timer, block table) that was meant to divert the local worker's attention, reflecting real world use cases. One remote helper explained the benefit of knowing the status of focus of attention of local worker as, *"I know that she was checking the timer regularly, so I focused only on instructing her about the next move."* Based on our observations, in the *NONE* condition participants had difficulty coordinating with pressing the timer button on time and simultaneously working on the main task (constructing LEGO).

The results from the questionnaire show that the eye-tracking cue not only influenced the quality of the communication (Q1, Q9, Q10, Q11, C2, and C6) but also other aspects of the user experience. The remote helpers rated that they enjoyed the task significantly more when the eye-tracking cue was present (Q5 and C1). On the other hand, local workers reported they were able to focus more on the task (Q6) when the eye-tracking cue was shared. These results are in line with the results from the local workers saying that they felt their partner could tell when they needed assistance better when the eye tracking cue was present (Q4 and C5), hence the remote helpers disturbing the local worker less, and the local workers were able to focus more on the task. Overall the local workers felt their teams worked together better with the eye tracking cue being present (Q8).

While it was the eye-tracking cue that mainly affected the local workers' feeling of working together (Q8), the remote helpers were mainly affected by the pointer cue. This indicates that the local and remote helpers have a different focus on deciding the quality of collaboration, depending on their role. The local workers mainly received instruction, and so seemed to value more the empathy aspects of collaboration, such as the remote partner understanding their situation and providing proper feedback. On the other hand, the remote helpers seem to be more focused on giving and expressing their instructions in a clear and efficient way.

Rating results on the co-presence question (Q2) showed that both the local workers' and remote helpers' sense of being present together with their partner was improved significantly by using both of the visual cues. The pointer cue significantly improved both the local workers' and remote helpers' subjective rating of their partner's sensing their presence (Q3).

6 Implications and Limitations

This research is one of the first studies of the impact of using eye-tracking and pointer cues in remote task space collaboration using head worn systems. As such there are many implications that can be drawn from the results.

One of the most interesting is that eye-tracking can be used to change the nature of remote collaboration with head worn systems. In previous research the remote helper was able to view the local workers view and could see when they were directly interacting with objects in the real world. In this case the local worker often had to explicitly reach for an object in order for the remote helper to know that they were about to interact with it. However, adding gaze-tracking means that the remote helper can be aware of the local users implicit intentions before they physically start interacting with the world around them, looking at objects before selecting them. This means that the remote helper can now be aware of both implicit and explicit communication cues, and respond accordingly. For example, telling the local worker that they are paying attention to the wrong object before they reach out to pick it up.

A second implication is that providing gaze cues alone can significantly improve the remote collaboration even without supporting pointing from the remote helper. Current eye-tracking hardware is small enough to be placed inside head mounted displays and so there could be a great opportunity there for developers wanting to create intuitive systems for remote collaboration. From Figure 12 we can see that using eye-tracking was seen to be just as beneficial as using remote pointing by itself.

Finally, the study implies that providing communication cues like gaze and pointing are not only important for improving task performance, but also play a very important role in creating a sense of co-presence and creating deeper understanding between users. Most of the users preferred the two conditions with the pointer present because it provided a way for the remote helper to convey non-verbal cues, which were important for creating a connection between the collaborators.

Although the results from the user study are very interesting there are a number of limitations. The prototype head mounted system is clearly too bulky and currently is tethered to a desktop PC making it far from wearable. However new HMDs with integrated scene cameras and eye-tracking hardware could overcome this.

The task was limited compared to what might be a realistic remote assistance task, such as machine maintenance. However it did have some of the key elements such as object identification and assembly. This may limit the applicability of the results, but more realistic scenarios could be tested in the future.

The experimental measures were not as detailed as they could have been. We analysed performance time and user subjective feedback, but a deeper understanding of the impact of gaze tracking could have been found from a more detailed behavioural analysis. In the future we could perform a more thorough conversational analysis (at the word level), and also looked at gaze and gesture interaction patterns.

None of these limitations present obstacles that cannot be overcome, but rather highlight directions for future research.

7 Conclusion

The main objective of our evaluation was to determine the effect of sharing eye tracking information from a local worker to a remote helper in terms of co-presence and performance in a remote assistance task. To explore this we developed an eye tracking tool that could be mounted on an existing HMD and prototyped a remote collaboration system.

We stated two main hypotheses that we wanted to investigate in the user study, about co-presence (H1) and performance (H2). In

terms of connectedness, presence, and needing assistance, we found that there was a significant main effect in POINTER and EYETRACKER factors most of the time with a significant interaction of POINTER and EYETRACKER on each other. This supports hypothesis H1, that providing these cues did indeed improve the feeling of co-presence between remote users.

From assessing the task completion time, we found both POINTER and EYETRACKER factors had a significant impact on the task performance for the given task. However, there was no interaction between the two factors. Overall, the participants' task performance was the best in the *BOTH* condition compared to the other conditions. Based on these findings we conclude that the study results support our second hypothesis H2, that providing additional attention cues did indeed improve performance time.

This research is one of the first using gaze tracking for remote collaboration in head mounted systems, so there are many areas of potential future research. In particular we can conduct more research on future user studies, bidirectional communication, and adding more communication cues, among others.

The current user study is with a simple task that doesn't require much collaborative problem solving. In most cases the local worker just gave minimal feedback showing that they had understood the remote helper's instructions. In the future it would be interesting to explore user behaviour in more parallel tasks where the users don't have the problem solution and have to work together more closely. This should encourage more two-way communication between the remote and local participants.

Another area that could be explored is in providing symmetric communication cues between both participants. In our prototype we had gaze information being sent from the local worker to the remote helper, but the local worker didn't have any idea of where the remote user was looking. In the future we could provide both users with a similar interface so that we could track the gaze of the remote helper and also send this back, making a fully symmetric system. It would be interesting to see what impact this would have on communication.

Finally, gaze information is just one of many physiological cues that could be send between participants. It would be interesting to explore the impact of sending other cues such as heart rate, galvanic skin response or EEG. These could be used to infer levels of stress, cognitive load or other physiological states about the local worker, which could be important in applications such as emergency response, and increase understanding between users.

REFERENCES

[1] D. G. Armstrong, T. M. Rankin, N. A. Giovinco, J. L. Mills, and Y. Matsuoka, "A Heads-Up Display for Diabetic Limb Salvage Surgery A View Through the Google Looking Glass," *Journal of diabetes science and technology*, vol. 8, no. 5, pp.951-956, 2014.

[2] M. Bauer, G. Kortuem, and Z. Segall, ""Where Are You Pointing At?" A Study of Remote Collaboration in a Wearable Videoconference System," in *Proc. of Intl. Symp. Wearable Computers*, 1999, p.151.

[3] M. Billinghurst and H. Kato, "Collaborative augmented reality," *Commun. of the ACM*, vol. 45, no. 7, pp. 64–70, 2002.

[4] S. E. Brennan, X. Chen, C. A. Dickinson, M. B. Neider, and G. J. Zelinsky, "Coordinating cognition: The costs and benefits of shared gaze during collaborative search," *Cognition*, vol. 106, no. 3, pp. 1465-1477, 2008.

[5] Brother. Brother Air Scouter. www.brother.co.uk.

[6] J. Carletta, R. L Hill, C. Nicol, T. Taylor, J. P. De Ruiter, and E. Gurman Bard, "Eyetracking for two-person tasks with manipulation of a virtual world," *Behavior Research methods*, vol. 42, no 1, pp. 254–265, 2010.

[7] S. R. Fussell, R. E. Kraut, and J. Siegel, "Coordination of communication: Effects of shared visual context on collaborative work," in *Proc. of the ACM Conf. on Computer Supported Cooperative Work*, 2000, pp. 21–30.

[8] S. R. Fussell and L. D. Setlock, "Using Eye-Tracking Techniques to Study Collaboration on Physical Tasks: Implications for Medical Research," Carnegie Mellon University, unpublished.

[9] S. R. Fussell, L. D. Setlock, and R. E. Kraut, "Effects of head-mounted and scene-oriented video systems on remote collaboration on physical tasks," in *Proc. of the SIGCHI conference on Human factors in computing systems*, 2003, pp. 513–520.

[10] S. R. Fussell, L. D. Setlock, and E. M. Parker, "Where do helpers look?: Gaze targets during collaborative physical tasks," in *CHI'03 Extended Abstracts on Human Factors in Computing Systems*, 2003, pp. 768-769.

[11] S. Gauglitz, B. Nuernberger, M. Turk, and T. Höllerer, "In touch with the remote world: Remote collaboration with augmented reality drawings and virtual navigation," in *Proc. of the 20th ACM Symp. on Virtual Reality Software and Technology*, 2014, pp. 197-205.

[12] W. Gaver, A. Sellen, C. Heath, and P. Luff, "One is not enough: Multiple views in a media space," in *Proc. of Interchi '93*, 1993, pp. 335-341.

[13] HELPlighting, VIPAAR. http://www.helplightning.com/

[14] H. Ishii, M. Kobayashi, and K. Arita, "Interactive design of seamless collaboration media," *Commun. of the ACM*, vol. 37, no. 8, 83-98,1994.

[15] T. Ishii, H. Kuzuoka, T. Takahara, and T. Myoi, "Collaboration system for manufacturing system in the 21st century," in *Intl. Conf. on Manufacturing Systems and Environment*, 1990.

[16] R. S. Johansson, G. Westling, A. Bäckström, and J. R. Flanagan, "Eye–hand coordination in object manipulation," *Journal of Neuroscience*, vol. 21, no. 17, pp. 6917–6932, 2001.

[17] M. Kassner, W. Patera, and A. Bulling, *Pupil: An open source platform for pervasive eye tracking and mobile gaze-based interaction*, in *Proc. of the 2014 ACM Intl. Joint Conf. on Pervasive and Ubiquitous Computing: Adjunct Publication*, 2014, pp. 1151-1160.

[18] H. Kato and M. Billinghurst, "Marker tracking and hmd calibration for a video-based augmented reality conferencing system," in *Proc. 2nd IEEE and ACM International Workshop on Augmented Reality*, 1999, pp. 85–94.

[19] S. Kim, G. A. Lee, N. Sakata, E. Vartiainen, and M. Billinghurst, "Comparing pointing and drawing for remote collaboration," in *Proc. IEEE Intl. Symp. on Mixed and Augmented Reality – Work-In-Progress*, 2013, pp. 1-6.

[20] S. Kim, G. Lee, N. Sakata, and M. Billinghurst, "Improving co-presence with augmented visual communication cues for sharing experience through video conference," in *Proc. of IEEE Intl. Symp. on Mixed and Augmented Reality*, 2014 pp. 83-92.

[21] T. Kurata, N. Sakata, M. Kourogi, H. Kuzuoka, and M. Billinghurst, "Remote collaboration using a shoulder-worn active camera/laser," in Proc. Intl. Symp. Wearable Computers, 2004, pp. 62–69.

[22] H. Kuzuoka, K. Toshio, and T. Masatomo, "GestureCam: A video communication system for sympathetic remote collaboration," in *Proc. of the 1994 ACM Conf. on Computer Supported Cooperative Work*, 1994, pp. 35-43.

[23] H. Kuzuoka, "Spatial workspace collaboration: a SharedView video support system for remote collaboration capability," in *Proc. of the SIGCHI conference on Human factors in computing systems*, 1999, pp. 533–540.

[24] J. Li, M. Manavalan, S. D'Angelo, and D. Gergle, "Designing Shared Gaze Awareness for Remote Collaboration," in *Proc. of the 19th ACM Conf. on Computer Supported Cooperative Work and Social Computing Companion*, 2016, pp. 325-328.

[25] K. Masai, K. Kunze, and M. Billinghurst, "Empathy Glasses," in *Proc. of the 2016 CHI Conf. Extended Abstracts on Human Factors in Computing Systems*, 2016, pp. 1257-1263.

[26] R. Müller, J. R. Helmert, S. Pannasch, B. M. Velichkovsky, "Gaze transfer in remote cooperation: Is it always helpful to see what your partner is attending to?" *The Quarterly Journal of Experimental Psychology*, vol. 66, no. 7, pp .1302-1316, 2013

[27] M. B. Neider, X. Chen, C. A. Dickinson, S. E. Brennan, and G. J. Zelinsky, "Coordinating spatial referencing using shared gaze," *Sychonomic bulletin & review*, vol. 17, no. 5, pp. 718–724, 2010.

[28] J. Ou, L. M. Oh, S. R. Fussell, T. Blum, and J. Yang, "Predicting visual focus of attention from intention in remote collaborative tasks," *IEEE Trans. on Multimedia*, vol. 10, no. 6, pp. 1034–1045, 2008.

[29] J. Ou, L. M. Oh, J. Yang, and S. R. Fussell, "Effects of task properties, partner actions, and message content on eye gaze patterns in a collaborative task," in *Proc. of the SIGCHI conference on Human factors in computing systems*, 2005, pp. 231–240.

[30] N. Sakata, T. Kurata, T. Kato, M. Kourogi, and H. Kuzuoka, "WACL: Supporting telecommunications using wearable active camera with laser pointer," in *Proc. Intl. Symp. Wearable Computers*, 2003, pp. 53-56.

[31] J. Siegel, R. E. Kraut, B. E. John, and K. M. Carley, "An empirical study of collaborative wearable computer systems," in *Proc. Conf. companion on Human factors in computing systems*, 1995, pp. 312–313.

[32] R. Stein and S. E. Brennan, "Another person's eye gaze as a cue in solving programming problems," in *Proc. of the 6th international conference on Multimodal interfaces* . 2004, pp. 9–15.

[33] W. Steptoe, R. Wolff, A. Murgia, E. Guimaraes, J. Rae, P. Sharkey, D. Roberts, and A. Steed, "Eye-tracking for avatar eye-gaze and interactional analysis in immersive collaborative virtual environments," in *Proc. of the 2008 ACM conference on Computer supported cooperative work*, 2008, pp. 197–200.

[34] C. A. Stitzlein, J. Li, and A. Krumm-Heller, "Gaze analysis in a remote collaborative setting," in *Proc. of the 18th Australia conference on Computer-Human Interaction: Design: Activities, Artefacts and Environments*, 2006, pp. 417–420.

[35] J. O. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins, "The aligned rank transform for nonparametric factorial analyses using only anova procedures," in *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*, 2011, pp. 143-146.

[36] J. Wong, L. M. Oh, J. Ou, C. P. Rosé, J. Yang, and S. R. Fussell, "Sharing a single expert among multiple partners," in *Proc. of the SIGCHI Conference on Human Factors in Computing Systems*, 2007, pp. 261–270.