Using Freeze Frame and Visual Notifications in an Annotation Drawing Interface for Remote Collaboration

Seungwon Kim[1], Mark Billinghurst[1], Chilwoo Lee[2], and Gun Lee[1]
[1] The School of Information Technology and Mathematical Sciences,
The University of South Australia, Australia
[e-mail: Seungwon.Kim@unisa.edu.au, Mark.Billinghurst@unisa.edu.au, Gun.Lee@unisa.edu.au]
[2] School of Electoronics and Computer Engineering, Chonnam National University, Gwangju, South Korea
[e-mail: leecw@jnu.ac.kr]
*Corresponding author: Gun Lee

Received May 4, 2018; revised July 18, 2018; accepted August 12, 2018; published December 31, 2018

Abstract

This paper describes two user studies in remote collaboration between two users with a video conferencing system where a remote user can draw annotations on the live video of the local user's workspace. In these two studies, the local user had the control of the view when sharing the first-person view, but our interfaces provided instant control of the shared view to the remote users. The first study investigates methods for assisting drawing annotations. The auto-freeze method, a novel solution for drawing annotations, is compared to a prior solution (manual freeze method) and a baseline (non-freeze) condition. Results show that both local and remote users preferred the auto-freeze method, which is easy to use and allows users to quickly draw annotations. The manual-freeze method supported precise drawing, but was less preferred because of the need for manual input. The second study explores visual notification for better local user awareness. We propose two designs: the red-box and both-freeze notifications, and compare these to the baseline, no notification condition. Users preferred the less obtrusive red-box notification that improved awareness of when annotations were made by remote users, and had a significantly lower level of interruption compared to the both-freeze condition.

Keywords: Augmented Reality, Stabilized Annotations, Video Freeze Interaction, Notification, Remote Collaboration

A condference paper with the one third of the results in the first study was published in the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems, with the titled "Automatically freezing live video for annotation during remote collaboration".

http://doi.org/10.3837/tiis.2018.12.023 **ISSN : 1976-7277**

1. Introduction

With Augmented Reality (AR) technology, researchers can add virtual objects in the real world to present context-sensitive information [1]. One of the main purpose of it is providing information to solve a physical task, especially for supporting training [2, 3, 4, 5, 6, 7, 8] and improving remote collaboration [9]. In this paper, we study real-time remote collaboration between local and remote users, and how AR visual cues can be used to improve collaboration.

Many researchers have explored how video conferencing systems help sharing local user's surroundings with a remote collaborator [10]. However, only sharing the local surroundings was not enough for effective collaboration, because a remote user could not visually represent the information but only verbally collaborate with a local user [11]. Thus, several researchers have added virtual visual cues such as pointers [9,12], drawn annotations [9, 12, 13, 14, 15], or hand gestures [16, 17], for a remote user to present spatial information. Extending this prior research, in this paper we explore user interfaces that help users to have better communication with the annotation cue on a shared live video in a teleconferencing system.

In earlier studies, virtual annotations were anchored to the screen space of the shared video [9, 12, 18] so the drawn annotations no longer pointed to the object of interest after the viewpoint of the shared video view moved. Researchers solved this issue by using AR tracking to stabilize annotations in the real world [14, 15], so that the drawn annotations stayed in the same place regardless of the viewpoint movement. However, this still had an issue of annotations being anchored at an incorrect place if the local user changes the viewpoint of the shared view when the remote user is drawing annotations. (see **Fig. 1).**

Fig. 1. An issue in drawing an annotation: a remote user attempted to draw a triangle as marked in red (1a) and successfully drew one side of it (1b), but the next side was drawn incorrectly (1c) because the local partner was unexpectedly changing their viewpoint of the live video.

To overcome this issue, Kim et al. [15] and Gauglitz et al. [14] used the manual freeze method that a remote user paused the live video and drew annotations in a still image rather than in a live video. However, it required additional user inputs to pause and restart (freeze and unfreeze) the live video. As an alternative, we introduced a novel auto-freeze method in our previous paper [19]. In our first study, we extend our previous paper and deeply explore the auto-freeze method with the advanced technique of stabilizing annotations.

In addition to our first study that explores the remote user interface with the auto-freeze method, our second study investigates visual notifications on the local user interface. The visual notifications are for local user's awareness on the remote user drawing activities. We compare three visual notification methods on local user's interface: (1) no notification (as the baseline condition), (2) a *red-box* notification (which shows a red outline around the local user's display boundary), and (3) a *both-freeze* notification (which simultaneously freezes the local user view).

In this paper, we make the following novel contributions (to the best of our knowledge, this is the first study of):

(1) Providing an auto-freeze method for the annotation cue in remote collaboration study with the support of a visual tracking system to stabilize annotations without pausing the live video on the local user view

(2) The design of visual notifications for the local user to provide better awareness in the remote collaboration.

In the rest of this paper, we will describe the related work, methodology, user study results, then conclusion will be followed.

2. Related Work

In this section, we review prior related research in remote collaboration. We first review prior work in remote collaboration, then more specifically in remote user's drawing annotations, and awareness and notification cues.

2.1 Remote Collaboration

Collaboration is the process of people working together to achieve a common goal [20], and remote collaboration is enabled with teleconferencing systems. However, remote collaboration with the teleconferencing systems is limited by the media for sharing an environment (i.e. limited field of view in a video conferencing) and in support for communication cues (such as hand gesture). Previous studies in remote collaboration mostly focused on solving these issues and can be categorized according to the media they used, the communication cues in their systems, and the type of shared views. For example, Olson [21] and Kraut et al. [22] compared media such as videoconferencing, telephones, and text messages. Fussell et al. [9] and Kim et al. [12] investigated sharing hand gesture information with pointer and annotation communication cues. Fussell et al. [23] explored independent and dependent shared views (i.e. whether a remote user had the same view or a different view from a local user).

Compared to audio only systems, video conferencing often provides better user understanding [11]. With a shared live video, remote users can watch the local environment and local users can show their activities to the remote users. One way of doing this is where a local user wears a HMD with a camera attached to it and streams video to a remote user [23]. This is a dependent view configuration, where the local user controls the viewpoint by moving their head, and the remote and local users have the same view.

In the shared dependent view, local users can show their hand gestures and object manipulation to a remote partner, but remote users cannot. As a solution, researchers added virtual cues for the remote users such as pointers and annotations overlaid on the live video [12, 14, 15]. Our first study focuses on improving the use of annotation cues, especially by using freeze functions. Moreover, it is also crucial to ensure that local users understand the annotations as well. This is generally referred to as 'awareness' [24, 25], and our second study focuses on improving the local user's awareness of the remote user's annotations.

2.2 Annotation in Remote Collaboration

There are several early remote collaboration systems supporting drawing annotations. In VideoDraw [26], a user could draw annotations on paper and share them by capturing a live video of it and projecting on the other side. Similarly, ClearBoard [27] allowed users to share

annotations on a board style display. However, the use of these systems was limited as they required heavy and static setups.

Fussell et al. [9] introduced a system sharing a top-down third-person view from a camera on a tripod. Remote users drew annotations on the shared view and they were displayed back on the local user's desktop monitor. The system was not portable, and a local user could not simutaneously see the workspace and the shared view. Later, Kim et al. [12], Chen et al. [28], Koh et al. [29], and Rice et al. [30] developed portable remote collaboration systems with a handheld device or a HMD. With these systems the local user could simultaneously see their workspace and the shared view, but annotations were displayed on the screen space, so they lost their real world referents if the local user changes the viewpoint.

To solve the issue, several researchers stablized the annotations in the real world with Augmented Reality techniques. Kato and Billinghurst [31] developed a system with a set of AR markers that were used to determine the real world position of remotely added annotations. Gauglitz et al. [13, 14] and Kim et al. [15] used markerless tracking to stablize annotations in the real world. They both used a manual freeze function [32, 33] that paused the live video and allowed the user to draw on it. This addressed the issue of incorrectly anchoring annotations if they were drawn while the local user changed their viewpoint. However, Kim et al. [15] found that the manual freeze function could not fully solve the issue, as remote users drew more annotations in the live video than in manually frozen views.

To overcome this limitation, recently, our earlier work [19] introduced an auto-freeze method which automatically pauses the video when the user starts drawing, and this paper describes an extention of the earlier study. While Fakourfar et al. [34] followed our auto-freeze study, their auto-freeze study did not include stablized annotations in the real world and the annotations were only available on frozen 2D images so disappeared after returning back to the live video. Moreover, the local user's view was also frozen while the remote user was using the freeze function. In contrast, our system provides stablized annotations in the real world, and the local user can keep the live video with the remote user's drawings regardless of the remote user using freeze function or not. Moreover, the local participant's device was a hand held device (HHD) in Fakourfar's study, so the local participants could use only one hand, but we use a HMD (Vuzix Wrap 1200DX-VR) in our study hence the local participants can use both hands.

Recently, some researchers focused on estimating the proper depth of the annotations in the 3 dimensional world. Chang et al. [35] and Nuernberger et al. [36] found that the users preferred displaying annotations on the surface of the target objects. Later, Nuernberger et al. [37] extended their study for displaying annotations in a large area such as a city hall (i.e. annotations on a building) by using image-based reconstruction with multiple images.

2.3 Awareness and Notification

Awareness is being conscious of others' activities and it helps coordinating the next user actions [38]. However, it is challenging because the remote collaboration system can only provide a fraction of the awareness that is available in co-located collaboration [20].

Many researchers have explored ways to increase the level of awareness. One solution is to assign a specific role to a collaborator [24], which defines the collaborator's activities. For example, a remote user is assigned the role of an instructor and knows all the information needed for completing a collaboration task. In this case, the local user knows that the remote user will send instructions, so coordinates their activity to receive the instruction. A second solution is to provide additional communication cues [39], such as adding a pointer [9, 12],

annotations [13, 14, 15, 29], or hand gestures [16, 17] in the shared video. Third, providing a better interface is another solution, and it includes two steps: 1) monitoring or tracking collaborator's activities, and 2) properly notifying the tracked collaborator's activities to a user. Some researchers focused on the first step studying training systems [2, 6, 7, 8, 40], and investigated what activities a system needs to track after. Other researchers focused on the second step. Gutwin et al. [41, 42] explored audio notification when a user watched a video, and found that they were effective. Cidota et al. [43] compared an audio notification to a visual notification, and found that participants clearly preferred the visual notification over audio or no notifications. In our second study, we extend Cidota's study [43] investigating visual notification, but focus on the design of the visual notification to notify the local user of the remote user's drawing activity.

While there is no prior study exploring the design of visual notifications in remote collaboration study, there are several researchers who emphasized the requirements in designing notifications in real world task management [44]. Ho et al. [45] found that notifications through a modality different from the one used in the primary task reduced the level of disruption. However, Posner [46] and Hameed et al. [47] found that the visual notification on the peripheral vision can be effectively perceived while using a fovea vision for the primary task. Additionally, informative notification needs to be context sensitive and sufficiently salient without being interruptive [24, 48]. Based on these prior studies, we designed a peripheral visual notification that uses a red outline around the screen of the local user (see section 3.2) and we conducted a second user study with it.

3. Methodology

We conducted two user studies to explore the use of freeze interfaces and visual notifications on top of the basic annotation system. In this section, we explain the user study design including the basic drawing annotation system, experimental conditions (user interfaces), and experiment setup.

3.1 Basic Annotation System

For the basic annotation system, we adopted our previous prototype system [6, 23], and implemented experimental conditions on top of it.

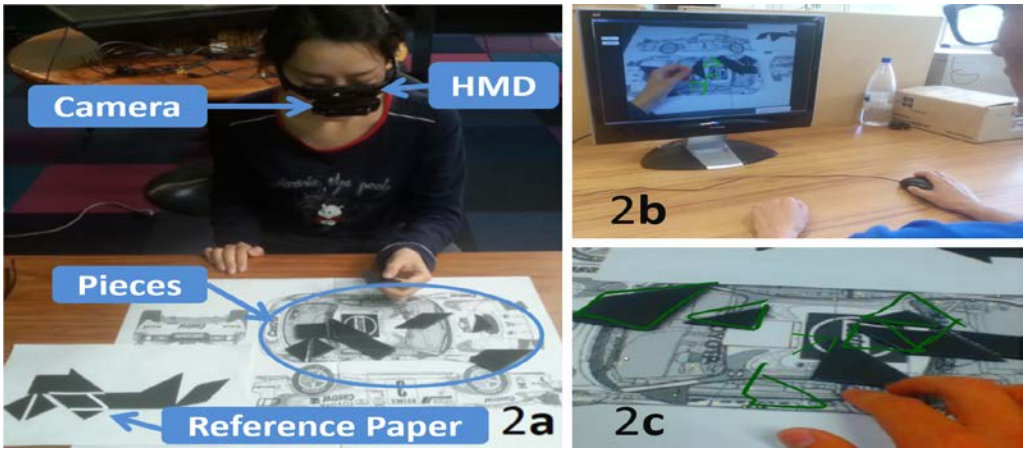


Fig. 2. System setup: a local user wears a HMD with a camera to share a live video of the workspace (2a), and a remote user uses a mouse (2b) to draw annotations in green (2c).

In the system, a local user shares his/her first-person view with a webcam attached to a HMD, and a remote user can draw annotations on the shared view using a desktop interface. On the local user's end, a live video is captured using a Logitech C920 webcam with a resolution of 640 by 480 pixels at 24 fps, and displayed on a Vuzix Wrap 1200DX-VR HMD (see **Fig. 2a**). The webcam and HMD are connected to a PC running monocular SLAM (Simultaneous Localization and Mapping) software to visually track the scene from the live video [49]. The SLAM software is highly reliable (95% success with 1,000 iterations) and provides robust tracking even when the scene is changed, for example when the user rearranges books on a desk.

The remote user at a desktop computer (see **Fig. 2b**) has the same view with the one that the local user has, and uses a mouse to draw annotations on the shared view by pressing the left button and dragging. While dragging, the system calculates the 3D positions of the mouse cursor in the real world, and saves them in a list describing the drawn shape. To calculate the 3D positions, the system creates an invisible plane, approximated by feature points from the SLAM tracking, and a ray casting method casts a ray from the mouse screen position with directional information from the SLAM projection matrix. The collision points of the ray with the invisible plane are used for forming a virtual annotation, so the virtual drawing appears at a real world location (see **Fig. 2c**). The annotations drawn by the remote user are immediately displayed on both the remote and local user screens.

Our system did not support audio streaming, however the participants in the studies could easily talk to each other as they were in the same room with a divider between them.

3.2 Experimental Conditions

On top of the basic annotation system, we implemented the experimental conditions for the two user studies. For the first study, we prepared three remote user interfaces and compared them. The three interfaces are described below:

1) *Auto-freeze*: This method combines drawing and freezing interactions. When the remote user presses the left mouse button down to start drawing, the live video is automatically paused and the remote user can draw annotations with mouse dragging interaction on the frozen image. When the remote user releases the left button to finish drawing, the view automatically returns back to the live video.
2) *Manual-freeze:* This method requires two additional inputs, independent from drawing interactions, compared to the auto-freeze method. The remote user need to freeze (pause) and unfreeze (restart) the shared live video view with a mouse double click, before and after drawing.
3) *Non-freeze:* This does not include freeze interaction, so the remote user draws annotations on a live video that changes according to the local user's head movement.

While the remote user had three interface conditions, the local user had one interface throughout the first user study. The local user's view remained live at all times and was not affected by the remote user freeze interaction. Our hypotheses of the first study were:

(H1) The auto-freeze condition helps the remote user to participate more quickly in collaboration while solving the issue of incorrectly anchored annotations.

(H2) Participants will prefer the auto-freeze condition over the other conditions.

For the second study, we implemented another three experimental conditions and compared them. The conditions were the local user interfaces as described below:

1) *Red-box:* This condition shows a virtual red outline around the shared view when the remote user is drawing (see **Fig. 3**). This design satisfied three requirements from prior work: (1) the notification should reveal information about the drawing interaction by presenting temporal information indicating when annotations are drawn [43], (2) the notification should not require focused attention nor disturb the ongoing primary task [24, 48], and (3) the notification should be easily recognized as being noticeable with its size and color [46, 47] (i.e. the size of it was much bigger than an annotation as it covered every sides of the screen).
2) *Both-freeze:* This condition pauses the local user's view together with the remote user's view while the remote user uses the freeze interface to draw annotations. This is based on participants' suggestion from the first study, and similar to Fakaurfar's auto-freeze interface [34] except in the use of annotation stabilization and the local user device type. To prevent the local users thinking that the frozen view is a system malfunction, this condition also showed the red outline around the frozen shared view, as in the red-box condition. This also made this condition directly comparable to the red-box condition, with the only difference being if the local view was frozen or not.
3) *No notification:* This is the baseline condition, and does not include any notification.

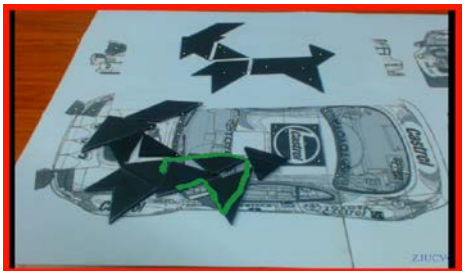


Fig. 3. An example of a local user's view in the red-box condition when the remote user is drawing.

While the local user had three interface conditions in the second study, the remote user had the auto-freeze interface from the first study with minor improvements. For example, instead of immediately returning back to a live video, the paused view was kept for two seconds after finishing drawing. This allows remote users to draw more on the same frozen view. Our hypotheses of the second study are listed below:

(H3) The local users will have better awareness with the red-box notification than with the other conditions.

(H4) Participants will prefer the red-box condition over the other conditions.

3.3 Experimental Task, Procedure, and Data Collection

For both user studies, we used the same experimental tasks while collecting the same type of data in the same procedure.

The collaboration type was a mutual collaboration [15, 50, 51] where no one has the solution at the beginning, so the participants needed to share and discuss ideas to solve the task. The experimental task was assembling a Tangram, a seven-piece puzzle arranged to form a shape of a given silhouette. To prevent bias from previous experience, we created custom Tangram puzzles. Each of them had ten puzzle pieces, and the size and shape of the pieces were different than the standard designs. We balanced the level of difficulty through a pilot

test that led us to provide a reference paper with three border lines between pieces (see **Fig. 2a**). We prepared four Tangram puzzles per user study with a similar difficulty, so each was solved in about four minutes when pilot tested with five people.

Participants solved the Tangram puzzles in pairs while communicating through speech and drawing annotations. The local participants wore a HMD, sat in front of a table, and had a reference paper and puzzle pieces on the table. The shared live video showed the workspace on the table and the remote participant watched it on a desktop computer (see **Fig. 2**).

For both user studies, we collected task completion time, data from questionnaires, interviews, activity log, and video and audio recordings of the user screens.

In each user study, the experimental procedure consisted of five sessions: a training session, three sessions under different experimental conditions, and a final interview session. In the training session, we asked participants to solve a Tangram puzzle face-to-face to let them understand the task. After training, the researcher explained the prototype system and participants performed three experimental sessions under the three conditions. The order was counter balanced using a balanced Latin-square design, and each session consisted of two minutes of practice and five minutes of experimental task. The participants filled out a questionnaire at the end of each session, including a set of rating items on a Likert-scale from 0 (Strongly disagree) to 10 (Strongly agree), and one open-ended question asking what they liked or disliked.

We prepared different questionnaires for the two user studies because the first study focused on remote user's use of annotations for better sending messages and the second study focused on local user's use of notifications for better understanding the annotations (receiving messages). In the first study, we prepared rating questions in the questionnaires from [15] and [52] (see **Table 1**). The rating questions asked participants' experience of message sending (Q1, Q2, Q3 and Q4) and receiving (Q5), and overall collaboration (Q6 and Q7). Three questions (Q2, Q3 and Q4) were prepared only for the remote participants asking their experience of drawing annotations with the given condition. After finishing all sessions, they ranked the conditions and had an individual interview to get more detailed feedback on their experience with the conditions.

Table 1. Rating questions in Study 1. Both participants answered four questions (Q1,Q5,Q6,Q7) and the remote participants answered three more questions (Q2, Q3, Q4, highlighted in grey).

Q1	I was able to express my idea properly.
Q2	It was easy to draw/annotate on the remote view.
Q3	I was able to draw annotation on the remote view as soon as I wanted to.
Q4	I had to be careful while drawing on the remote view.
Q5	I easily understood what my partner was trying to do and explaining.
Q6	I felt we collaborated well.
Q7	The interface was mentally stressful to use.

In the second study, the questionnaire included five (for the remote particpants) or seven (for the local particiapnts with two additional questions, Q3 and Q4) Likert-scale rating items (see **Table 2**), and one open question asking what they like or dislike. The rating questions asked the participants' experience of message sending (Q1) and receiving (Q2, Q3, Q4, and Q5), and overall collaboration (Q6 and Q7).

Table 2. Rating questions in Study 2. Both participants answered the five questions and local participants answered two more questions (Q3, Q4, highlighted in gray).

Q1	I was able to express my idea properly.
Q2	I easily understood what my partner was trying to do and explaining.
Q3	I knew the moment when my partner drew annotations.
Q4	I was aware of where my partner drew annotations.
Q5	I felt interrupted when my partner was drawing (or I felt I was interrupting my partner when I was drawing).
Q6	I felt we collaborated well.
Q7	The communication with my partner was mentally stressful.

4. Results and Discussion

In this section, we separately present the results of the first and second user studies, then discuss these results. In the sub-sections of each user study, we present the results under the theme of sending and receiving messages, and overall collaboration by reporting the relevant results together from the questionnaires, interviews, video recordings and log data. To analyze Likert scale rating results, we used Friedman tests (α = .05) with post hoc tests for pair-wise comparison using Wilcoxon Signed-Rank tests with Bonferroni correction (α=.0167).

4.1 Results from Study 1: Freeze Interface

To explore the freeze interfaces, we compared auto-freeze, manual freeze, and non-freeze conditions. We recruited 12 pairs (24 participants) who were friends or family, and used video conferencing more than once a month. There were 17 males and 7 females with ages ranging from 15 to 33 years old (*Mean* = 25.6; *SD* = 4.6). In this section, we use acronym N, A, and M for the non-freeze, auto-freeze, and manual-freeze conditions, respectively.

4.1.1 Sending Messages

Both the M and A conditions solved the issue of mistakenly anchored annotations. Video recording showed all remote participants except R7 (the remote participant in group 7) mistakenly anchored annotations in the N condition (*Mean* = 1.83, *SD* = 0.94), while no one did in the M or A conditions. With the issue in the N condition, participants were more careful in drawing annotations. From the Likert scale ratings (see **Table 3**), we found that they felt being more careful in drawing annotations with the N condition than with the other two conditions (Q4, A: Z=-2.675, *p*=.007, and M: Z=-2.597, *p*=.009). Needing to be more careful in the N condition affected the remote participants' behavior as they waited for the moment when the shared live video view remained still, then hurryingly finished the drawings. In the interview, seven remote participants (R1, R3, R4, R6, R10, R11, and R12) reported difficulty in drawing with the N condition, and six remote participants (R1, R6, R7, R8, R9, and R12) mentioned that they rushed to finish drawing.

Table 3. Likert scale ratings for Q1~Q4 in Study 1. (0: strongly disagree ~ 10: strongly agree)

Role	Question	Mean (*Std.Dev.*)			Friedman	Wilcoxon (between)
		N	M	A		
Local	Express idea properly (Q1)	6.67 (1.92)	7.08 (1.16)	7.58 (0.79)	$\chi^2(2)=2.294$, $p=.318$	
Remote	Express idea properly (Q1)	6 (2.04)	4.17 (1.95)	7 (1.60)	$\chi^2(2)=12.043$, ***p*=.002**	$Z=-1.606$, $p=.108$ (N and M)
						$Z=-1.481$, $p=.139$ (A and N)
						***Z*=-2.102, *p*=.014** (M and A)
	Easy to draw annotations (Q2)	6 (2.45)	4.92 (2.43)	7.5 (1.31)	$\chi^2(2)=6.488$, ***p*=.039**	$Z=-1.670$, $p=.095$ (N and M)
						$Z=-1.897$, $p=.058$ (A and N)
						$Z=-2.378$, ***p*=.017** (M and A)
	Quickly start drawing annotations (Q3)	6.83 (2.08)	4.42 (2.83)	7.92 (1.44)	$\chi^2(2)=7.946$, ***p*=.019**	$Z=-1.743$, $p=.081$ (N and M)
						$Z=-1.876$, $p=.061$ (A and N)
						***Z*=-2.524, *p*=.012** (M and A)
	Being careful while drawing annotation (Q4)	6.25 (2.98)	4.33 (2.42)	3.42 (2.77)	$\chi^2(2)=7.6$, ***p*=.022**	***Z*=-2.597, *p*=.009** (N and M)
						***Z*=-2.675, *p*=.007** (A and N)
						$Z=-0.462$, $p=.644$ (M and A)

While both A and M conditions solved the issue, the A condition had several benefits. The Likert scale ratings showed significant differences in the Friedman tests for the question about 'expressing well', 'easy drawing' and 'quickly drawing' (Q2: $\chi^2(2)=12.043$, $p=.002$; Q3: $\chi^2(2)=6.488$, $p=.039$; and Q4: $\chi^2(2)=7.946$, $p=.019$). In pair-wise comparisons, remote participants felt that they were able to express their ideas significantly better (Q2: $Z=-2.102$, $p=.014$) and significantly quicker (Q4: $Z=-2.524$, $p=.012$) with the A condition than with the M condition. The ratings about 'easy to draw annotations' showed a similar trend as the statistical result was close to significant level (Q3: $Z=-2.378$, $p=.017$).

We found similar user comments from the interviews. Regarding the A condition, R5 and R4 mentioned that "*It's quick, precise and expressive*" and "*I didn't need to switch the views (with additional interaction). It's essential for quickly drawing in the quickly changing local environment (as local users manipulated pieces)*". The additional inputs in the M condition were mentioned as the reason for being slower in annotating compared to the A condition. R5 and R6 said that it was hard to be effective with the M condition. R3 and R7 reported that they forgot to freeze or unfreeze the scene before or after drawing. From the log data, we found that in the M conditon remote participants spent 1.88 seconds on average (SD = 1.341, N = 212) after freezing until starting drawing, and 2.87 seconds on average (SD = 2.38, N = 212) after completing drawing until returning back to live video.

From the video recordings, we found three typical annotations: circle, tick mark, and piece-shape (see **Fig. 4**). The circle and tick marks were mostly used to select a piece or to indicate approximate position. The piece-shape annotation mostly indicated position and orientation of a piece. To analyze the use of the drawn shapes, we collected the number of drawings in each condition (N condition: M = 24.2, SD = 6.4, M condition: M = 17.7, SD = 5.6, A condition: M = 22.8, SD = 9.4, Friedman tests: $\chi^2(2)=5.167$, $p=.076$). Then we calculated the percentages of the drawn shapes (see the bar chart on the right of **Fig. 4**). Friedman tests (α = .05) showed a significant difference between three conditions for the use of circles ($\chi^2(2)=18.167$, $p<.001$) and puzzle piece shapes ($\chi^2(2)=16.667$, $p<.001$) but not in the use of the tick marks ($\chi^2(2)=4.5$, $p=.105$). Pair-wise comparisons showed that the remote participants drew significantly fewer piece shapes in the N condition than in the M ($Z=-2.102$, $p=.014$) and

A (Z=-2.746, p=.006) conditions, and drew significantly less circles in M condition than in A (Z=-3.059, p=.002) and N (Z=-3.059, p=.002) conditions.

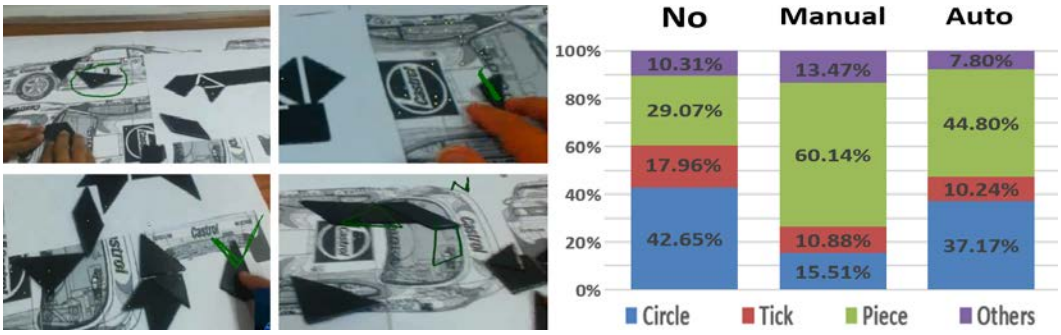


Fig. 4. Examples of shapes drawn (circle – top left, tick mark – bottom left and top right, and piece shapes – bottom right) and bar charts on the percentages of the typical shapes in each condition.

4.1.2 Verbal messages

The annotations and verbal messages were the main communication channels. When remote participants used them together, the verbal words were used for confirming or describing an action. To select a piece, the remote participants drew a circle or a tick mark together with verbal messages mostly including pronouns ('this' or 'it'). For position and orientation, verbal messages described actions to place a piece (e.g. "*It goes like this~*" by R7 and "*Put it here*" by R10) with one of the shapes described in previous section.

When remote participants used only verbal messages, it was more descriptive and included spatial information for compensating absence of annotations. Since remote participants hurryingly finished drawing in the N condition and had less piece-shape annotations, all of them, except R7, often described the orientation of the piece with only verbal messages (e.g. 'flip it' by R3, and 'turn it clockwise' by R8). With the M condition, as they needed more time to freeze and unfreeze the live video, some remote participants (R1, R3, R5, and R8) sometimes did not draw any annotations to select a piece, but they instead referred to a piece with verbal messages describing its size (small or big), shape (triangle, rectangle, or parallelogram), or position (on top, bottom, left or right) in the shared view, such as "*the top left triangle*" by R3. Additionally, since the silhouettes of the target shape resembled animals, participants used the words representing animal body parts, such as head and tail. R1, R5, R9, and L4 (the local participant in group 7) notified the starting point of assembly by mentioning a body part (e.g. "*Let's start from the head*" by R5), and used the words for positioning a piece (e.g. "*it's the tail on the right side*" by R5).

4.1.3 Receiving Messages

Table 4 shows the results of the Likert scale ratings on the question about easily understanding what the partner was doing and explaining (Q5). In a Friedman test, the local participants did feel a significant difference between the three conditions ($\chi^2(2) = 10.585$, p=.006), but the remote participants did not ($\chi^2(2)$=5.076, p=.056). Pair-wise comparisons showed local participants felt that they had a significantly better understanding of their remote partners with M (Z=-2.484, p=.013) and A (Z=-2.762, p=.006) conditions than with the N condition. However, local participants did not feel a significant difference between the M and A conditions in understanding their partners (Z=0, p=1.0).

Table 4. Likert scale ratings for Q5 in Study 1. (0: strongly disagree ~ 10: strongly agree)

Role	Question	Mean (*Std.Dev.*)			Friedman	Wilcoxon (between)
		N	M	A		
Local	Understand Partner (Q5)	6.25 (1.81)	7.25 (1.29)	7.25 (2.00)	$\chi^2(2)=10.585$, ***p*=.005**	***Z*=-2.484, *p*=.013** (N and M)
						***Z*=-2.762, *p*=.006** (A and N)
						$Z=0.000$, $p=1.0$ (M and A)
Remote	Understand Partner (Q5)	6 (2.04)	4.17 (1.95)	7 (1.60)	$\chi^2(2)=5.076$, $p=.056$	

From the interview, we found that these results might come from the mistakenly anchored annotations in the N condition affecting the experience of the local participants. Local users L1, L4, L6, and L12 mentioned that it was sometimes difficult to understand the remote partner's drawing in the N condition. On the other hand, the two freeze conditions appeared to have resolved this issue as the annotations were clear to understand.

4.1.4 Overall Collaboration

We asked questions 6 and 7 about how well they collaborated and mental stress in using the interfaces. **Table 5** shows the results.

Table 5. Likert scale ratings for Q6 and Q7 in Study 1. (0: strongly disagree ~ 10: strongly agree)

Role	Question	Mean (*Std.Dev.*)			Friedman	Wilcoxon (between)
		N	M	A		
Local	Collaborate Well (Q6)	6.25 (1.76)	7.00 (1.48)	7.25 (1.36)	$\chi^2(2)=4.471$, $p=.107$	
	Mental Stress (Q7)	5.17 (1.90)	4.33 (1.50)	4.50 (1.51)	$\chi^2(2)=6.7462$, ***p*=.004**	$Z=-2.111$, $p=.035$ (N and M)
						$Z=-1.000$, $p=.317$ (A and N)
						$Z=-1.752$, $p=.084$ (M and A)
Remote	Collaborate Well (Q6)	6 (2.45)	4.92 (2.43)	7.5 (1.31)	$\chi^2(2)=7.136$, ***p*=.028**	$Z=-1.616$, $p=.106$ (N and M)
						$Z=-1.594$, $p=.111$ (A and N)
						***Z*=-2.591, *p*=.001** (M and A)
	Mental Stress (Q7)	6.83 (2.08)	4.42 (2.83)	7.92 (1.44)	$\chi^2(2)=8.844$, ***p*=.012**	$Z=-1.347$, $p=.178$ (N and M)
						$Z=-1.813$, $p=.070$ (A and N)
						***Z*=-2.439, *p*=.015** (M and A)

The remote participants' ratings showed a significant difference among the three conditions for the questions about 'Collaborate Well' ($\chi^2(2)= 7.136$, $p=.028$) and 'Mentally Stressful ($\chi^2(2)=8.844$, $p=.012$) in Friedman tests. In pair-wise comparisons, the remote participants felt that they had better collaboration ($Z=-2.591$, $p=.01$) and less mental stress ($Z=-.2.439$, $p=.015$) with the A condition than with the M condition. In the interviews, R1, R2, and R8 mentioned that the M condition was more mentally stressful because it required additional inputs for freezing and unfreezing the shared view.

The ratings from the local participants did not show a significant difference for the question on 'Collaborate Well' (Friedman test: $\chi^2(2)=4.471$, $p=.107$) and 'Mentally Stressful' (Friedman test: $\chi^2(2) = 6.462$, $p=.04$, pair-wise comparisons: $Z=-2.111$, $p=.035$ for N and M, $Z=-1.00$, $p=.317$ for N and A, $Z=-1.75$, $p=.084$ for the M and A).

We also measured the task completion time, and participants took 275, 263, and 290 seconds on average with the N, A, and M conditions, respectively. A Shapiro-Wilk test indicated that M condition was not normally distributed (p=.154), so a Friedman test was used and found no significant difference between the conditions ($\chi^2(2)$=0.667, p=.717).

4.1.5 User Preference

After trying the three conditions, participants ranked them (see **Table 6**). Ten remote participants (83.3%) most preferred the A condition, while one each of the others preferred the M (8.3%) or N conditions (8.3%). For the least preferred condition, eight remote participants (66.6%) selected the M condition, and three of the rest picked the N condition (25%) and one (8.3%) picked the A condition. A Friedman test showed a significant difference ($\chi^2(2)$=11.167, p=.004), and, in pair-wise comparisons, the A condition was preferred significantly more over the M condition (Z=-2.769, p=.006), but no significance was found in other pairs (A and N: Z=-2.309, p=.021; M and N: Z=-1.291, p=.197).

Table 6. Participants' preference among the conditions in Study 1.

	Most Preferred			Second Most Preferred			Least Preferred		
Condition	N	A	M	N	A	M	N	A	M
Local Participants	0	**9**	3	**5**	3	4	7	0	**5**
Remote Participants	1	**10**	1	**8**	1	3	3	1	**8**

For the local participants, nine participants (75%) most preferred the A condition, while the other three (25%) picked the M condition. For the least preferred condition, seven and five of them chose the N (58.3%) and M conditions (41.6%), respectively. A Friedman test showed a significant difference between the three conditions ($\chi^2(2)$=11.167, p=.004), and the A condition was preferred significantly more compared to the N condition (Z=-2.769, p=.006) in a pair-wise comparison. There was no significant difference in other pairs (between N and M: Z=-1.155, p=.248; between M and A: Z=-2.183, p=.029).

4.2 Results from Study 2: Notification

To understand the effectiveness of notifications for local user awareness, we conducted a second user study comparing the three notification conditions with another twelve pairs who had been using video conferencing more than once a month. All pairs knew each other well as friends or family, and there were 21 males and 3 females with ages ranging from 20 to 38 years old (*Mean* =26.6; *SD*=4.6). In this section, we use acronym N, R, and B for the no notification, red-box, and both-freeze conditions, respectively.

4.2.1 Receiving Messages (User Awareness)

The results of rating questions showed that local participants had better awareness with the R and B conditions than with the N condition. **Table 7** shows the results related to the understanding partner's activities with visual notifications.

Table 7. Likert scale ratings Q2~Q5 in Study 2. (0: strongly disagree ~ 10: strongly agree)

Role	Question	Mean (*Std.Dev.*)			Friedman	Wilcoxon (between)
		N	R	B		
Local	Understand Partner (Q2)	6 (1.92)	7.5 (1.68)	7.83 (1.47)	$\chi^2(2)=10.333$ ***p*=.006**	***Z*=-2.716, *p*=.007** (N and R)
						Z=-0.303, *p*=.762 (R and B)
						***Z*=-2.434, *p*=.015** (B and N)
	Know when partner draw (Q3)	5.75 (1.42)	6.83 (1.64)	7.92 (2.11)	$\chi^2(2)=6.950$ ***p*=.031**	***Z*=-2.412, *p*=.016** (N and R)
						Z=-1.492, *p*=.136 (R and B)
						***Z*=-2.522, *p*=.012** (B and N)
	Know where partner draw (Q4)	5.67 (1.83)	6.67 (1.44)	7.92 (1.88)	$\chi^2(2)=11.167$ ***p*=.004**	*Z*=-1.491, *p*=.136 (N and R)
						***Z*=-2.714, *p*=.007** (R and B)
						***Z*=-2.449, *p*=.012** (B and N)
	Interrupted when partner draws (Q5)	2.33 (1.72)	2.83 (1.70)	6.33 (2.77)	$\chi^2(2)=13.818$ ***p*=.001**	*Z*=-1.098, *p*=.272 (N and R)
						***Z*=-2.919, *p*=.004** (R and B)
						***Z*=-2.871, *p*=.004** (B and N)
Remote	Understand Partner (Q2)	5.67 (2.06)	6.5 (2.58)	6.33 (2.19)	$\chi^2(2)=4.200$ *p*=.122	
	Interrupting partner when drawing (Q5)	4 (2.26)	4.25 (2.73)	6.67 (2.27)	$\chi^2(2)=12.350$ ***p*=.002**	*Z*=-0.299, *p*=.765 (N and R)
						***Z*=-2.409, *p*=.016** (R and B)
						***Z*=-2.814, *p*=.005** (B and N)

The local participants' ratings showed a significant difference between the three conditions in all four questions: 'understood partner's explanation' (Q2: $\chi^2(2)=10.333$, p =.006), 'knew the moment when the partner drew' (Q3: $\chi^2(2)=6.950$, p =.031), 'knew where the partner drew' (Q4: $\chi^2(2)=11.167$, p =.004), and 'being interrupted when the partner drew' (Q5: $\chi^2(2)=13.818$, p =.002). In pair-wise comparisons, we found that local participants felt significantly better understanding the partner's explanation in the R (Z=-2.716, p=.007) and B (Z=-2.434, p=.015) conditions than in the N condition. Specifically, they felt better understanding *when* the remote partners drew annotations in the R (Z=-2.412, p=.016) and B (Z=-2.522, p=.012) conditions than in the N condition. The B condition showed additional benefit of knowing *where* the remote partners drew annotations compared not only to the N condition (Z=-2.449, p=.012) but also to the R condition (Z=-2.714, p=.007), while the R condition did not show any benefit compared to the N condition (Z=-1.491, p=.136). Accordingly, in the interview, L5, L7, L10, L11, and L12 said that they knew when their partner was drawing in the R condition: *"Visual notification helped me know when he started drawing"* (L11). Similar comments were found for the B condition, but they added that they knew where their partner drew.

However, while the B condition showed better awareness on both when and where remote partners drew annotations, it had a downside of interrupting the local worker. The local participants felt a significantly higher level of interruption in the B condition than in the N (Z=-2.871, p=.004) and R (Z=-2.919, p=.004) conditions. The remote participants also felt that they were interrupting their local partners significantly more in the B condition than the N (Z=-2.409, p = .015) and R (Z=-2.814, p=.005) conditions. In the interview, all local participants complained about being interrupted as they had to stop their on-going activities (i.e. manipulate pieces and apply their ideas) and only watch the remote participant's annotations in a frozen view. Two thirds of the remote participants reported that they interrupted their partners and felt they were being rude by freezing the local participants' view.

In the R condition, interruptions were not an issue as the local participants kept their live view while knowing when the remote partner drew annotations.

4.2.2 Sending Messages

For the question about properly expressing ideas (Q1), **Table 8** shows the results across the three conditions. Friedman tests did not show any significant difference in local and remote participants' ratings (local participants': $\chi^2(2) = 0.419$, p=.811, remote participants': $\chi^2(2) = 3.231$, p =.199). This means that the participants did not feel a significant difference among the conditions in being able to properly express their ideas.

Table 8. Likert scale ratings for Q1 in Study 2. (0: strongly disagree ~ 10: strongly agree)

Role	Question	Mean (*Std.Dev.*)			Friedman	Wilcoxon (between)
		N	R	B		
Local	Express ideas (Q1)	6.5 (1.31)	6.33 (2.77)	6.25 (1.60)	$\chi^2(2)=0.419$, p=.811	
Remote	Express ideas (Q1)	4.92 (2.71)	6 (2.13)	6.33 (1.37)	$\chi^2(2)=3.231$ p=.199	

While we did not find significance in the ratings, we observed interesting results from the video recordings. The drawing activity mostly occurred at the same time as speech from the remote participants and local participants used it as a notification in the N condition where they suffered from the lack of notification. Local user's comments supported it, *"I knew when he was drawing from his speech but the red box in test was clearer in this regard"* (L12). However, speech was used to provide additional information about their annotations rather than to notify when they drew annotations, so was not efficient enough.

With the B condition, local participants reacted to the suddenly frozen view mostly with a verbal exclamation (such as *'Oh'*). This let the remote participants know that they had interrupted. We found two interesting observations in this regard. First, R7 did not draw anymore after figuring out the interruption, only using verbal speech afterwards. Second, R4 and R12 notified that they were about to freeze the view with brief words (i.e. *"wait"*).

4.2.3 Overall Collaboration

To see the effect on overall remote collaboration, we also analyzed the results about collaborating well and being mentally stressful in using the notifications (see **Table 9**). In Friedman tests, the remote participants' ratings were not significantly different across conditions in collaborating well ($\chi^2(2) = 3.619$, p=.164) and mental stress ($\chi^2(2) = 0.242$, p=.886). The local participants' ratings were not significantly different in mental stress ($\chi^2(2) = 3.152$, p=.207), but were in collaborating well ($\chi^2(2) = 6.500$, p=.039). Pair-wise comparisons showed thet local participants felt they had a significantly better collaboration in the R condition than the N condition (Z=-2.55, p=.01) but no difference between the R and B conditions (Z=-1.713, p=.09) nor between the B and N conditions (Z=-0.180, p=.86).

Table 9. Likert scale ratings for Q6 and Q7 in Study 2 (0: strongly disagree ~ 10: strongly agree)

Role	Question	Mean (*Std.Dev.*)			Friedman	Wilcoxon (between)
		N	R	B		
Local	Collaborate Well (Q6)	6.08 (2.47)	7.33 (1.92)	6.33 (1.67)	$\chi^2(2)=6.500$, ***p*=.039**	***Z*=-2.549, *p*=.011** (N and R)
						Z=-1.713, *p*=.087 (R and B)
						Z=-0.180, *p*=.857 (B and N)
	Mental Stress (Q7)	3.5 (2.11)	2.75 (1.48)	3.75 (2.26)	$\chi^2(2)=3.152$, *p*=.207	
Remote	Collaborate Well (Q6)	5.58 (2.91)	6.91 (1.31)	6.17 (1.90)	$\chi^2(2)=3.619$, *p*=.164	
	Mental Stress (Q7)	4.58 (2.50)	4.17 (2.40)	4.25 (2.45)	$\chi^2(2)=0.242$, *p*=.886	

We measured the task completion time, and participants took 272, 271, and 280 seconds on average with N, R, and B conditions, respectively. A Shapiro-Wilk test indecated that R condition was not normally distributed (*p*=.228), so a Friedman test was used and found no significant difference between the conditions ($\chi^2(2)=0.0$, *p*=1).

4.2.4 User Preference

Eight local participants (66.6%) most preferred the R condition while the rest were split equally between the B (16.6%) and N (16.6%) conditions (see **Table 10**). For the least preferred, eight of them (66.6%) selected the B condition, while the rest (33.3%) picked the N condition. A Friedman test showed a significant difference ($\chi^2(2)=8.667$, *p*=.013), and, in pair-wise comparisons, they significantly preferred the R condition over the B condition (*Z*= -2.581, *p*=.010), but no significant difference between the R and N conditions (*Z*=-2.352, *p*=.019, close to significance) nor between the B and N conditions (*Z*=-.733, *p*=.464).

Table 10. Participants' preference among the conditions in Study 2

	Most Preferred			Second Most Prefer			Least Preferred		
Condition	N	R	B	N	R	B	N	R	B
Local Participants	2	**8**	2	**6**	4	2	4	0	**8**
Remote Participants	2	**8**	2	**5**	3	4	5	1	**6**

Eight remote participants (66.6%) preferred the R condition most, while the rest were split equally between the B (16.6%) and N (16.6%) conditions. For the least preferred, half of them (50%) chose the B condition, while other five (41.6%) and one (8.3%) selected the N and R conditions, respectively. A Friedman test found a significant difference ($\chi^2(2)=6.167$, *p*=.046), but no significant difference was found in pair-wise comparisons.

4.3 Discussion

In this section, we separately discuss the results of the first and second user studies, then generally discuss our study environment and limitation.

4.3.1 Discussion on Study 1: Freeze Interface

We identified the issue of mistakenly anchored annotations in the non-freeze condition if the remote participants were drawing them when the local participants changed the shared

viewpoint. With this issue, remote participants were more careful in drawing annotations and it affected the use of verbal and annotation cues as they drew fewer piece shapes and verbally described the orientation of the pieces more often.

Freezing the shared view with manual or auto-freeze conditions solved this issue, and local particiapnts also felt that they more easily understood what the remote partners explained in the two freeze conditions. However, the manual-freeze condition required two additional interactions, and it affected the use of verbal and visual communication as they drew fewer circles to select a piece with the manual-freeze condition than the other two but verbally describing the piece selection. An interesting point was that participants reduced the use of manual-freeze interaction in piece selection rather than in piece position and orientation. This could be because verbally describing the piece selection was easier than verbally describing the position or orientation information.

The additional two interactions in the manual-freeze condition had a strong influence on these results, even though they were simple mouse left button double clicks. We compared the sequence of interactions required in the conditions to the real world annotating steps. In the real world, people draw annotations with three steps: 1) watch the environment to find where to draw, 2) draw word or symbol annotations, and 3) watch the environment again to check the annotations. In the auto-freeze condition, remote participants had a similar sequence: 1) watch a live video, 2) draw annotations on the frozen image, and 3) watch the live video again. However, in the manual freeze conditions, they had to follow a different sequence: 1) watch the live video, 2) freeze the live video, 3) draw an annotation on the frozen image, 4) unfreeze the live video, and 5) watch the live video. Freezing and unfreezing the live video are two extra steps compared to real world drawing annotations, and these unfamiliar extra steps may have required more practice and led users to feel less easy or quick to use, with more mental stress.

As a result, our hypothesis H1 (the auto-freeze condition helps the remote user to participate more quickly in collaboration while solving the issue of incorrectly anchored annotations) and H2 (participants prefer the auto-freeze condition than other conditions) were both supported.

4.3.2 Discussion on Study 2: Notification

In the second study, we explored the local user's interfaces to investigate whether the local user had better awareness or not with the visual notifications. Our hypothesis H3 that postulated the local users have better awareness with the red-box notification than with others was half-correct, as the red-box notification showed better awareness of when remote users drew than the no notification, but not against the both-freeze notification. Instead, the both-freeze notification provided better awareness of both when and where remote users drew annotations compared to the no notification but had a trade-off of interrupting the local users' on-going task.

The both-freeze condition intensified the shared activities by capturing the view of local users, forcing them to only see remote users' annotations in a frozen image view, and abandon on-going individual activities such as manipulating objects within the live video reference. In comparison, the red-box condition left the choice to local users if they would shift their attention from the individual on-going activities to the shared activities by keeping the live video. As a result, participants preferred the red-box condition than others and the hypothesis H4 was supported.

The both freeze interface was also used in Fakourfar's study [34] while exploring the auto-freeze interface. Interestingly, they did not report any interruption issues, which could be because of the two differences between their and our studies: 1) they used a tablet instead of HMD as a local user's display, and 2) the collaboration style of their experimental task was remote expert collaboration (where only the remote user has the solution and provides instructions to the local user, resulting in mostly one-way communication) while ours was mutual collaboration. With a tablet, the local users' view is not fully obscured by the frozen view, so the local users can still see the work space off the tablet screen, and can continue the individual task while leaving the tablet with the shared frozen view aside (e.g. putting the tablet down). In a remote expert collaboration, a local user follows the remote user's instructions so there is a low chance of the local user doing individual activities (such as thinking their own ideas and trying it out) rather than focusing solely on receiving and following instructions. In mutual collaboration (as in our study), there is a higher chance of the local users doing individual activities as they are not only following remote users' ideas but could also contribute their ideas to solving the task, so the both freeze notification would interrupt these individual activities.

Another interesting point is that the remote users' attitude was influenced by social etiquette even though they were connected through media. After they figured out the both-freeze notification being interruptive, they became more careful in using it by giving a verbal warning before freezing the view. It would be an interesting future study exploring the relationship between the type of media (or type of communication cue) and the influence of social etiquette on the collaborators' attitude.

In the second study, we did not measure the number of ignored annotations because it would be hard to distinguish between intentional and unintentional ignorance. If there is measuring tools for defining users' intention, then the study could be extended. Tools for measuring collaborator state such as electroencephalography (EEG) and capturing the collaborators' brain activities could be explored in a future study. For example, if a collaborator knew whether the other person was busy or not, then they could coordinate the shared activities such as sending messages.

4.3.3 Overall Discussion and Limitation

In these two studies, we explored variations of controlling the shared view (i.e. pausing the view by the freeze method and providing notifications in the local user view) between two users of a remote collaboration system that supports visual annotations in the shared video. Typically, the local user has the control of the view when sharing the first-person view, but our interfaces provided instant control of the shared view to the remote users. In the first study, the remote users were able to control their own views for better drawing interaction. Both local and remote users did not experience any significant inconvenience while the remote users gained control of their own view through freezing. We note that the remote users were uncomfortable with the manual freeze condition not because of gaining control of their views but due to the discord between the freezing and drawing interactions. In the second study, the remote users had control of the local user view and it was only acceptable if it did not interrupt the local user's individual activities.

Another interesting point about the mutual collaboration was that the result of the task completion time did not differ significantly between the conditions, not only in the first study but also in the second study. This is not aligned with the other results showing a signicficant effect between the conditions in our first and second studies. This could be because the task performance in the mutual collaboration is affected by the participants' ideas or luck as other

literatures studying about the mutual collaboration reported [15, 50, 51]. In the future, it would be interesting to further explore other type of tasks using more objective measures to better understand how the proposed interfaces affect on tasks performance.

Through these studies we found that the auto-freeze technique was useful to remote users for drawing annotations, and the red-box notification was useful for making local users aware of when their remote partners were drawing. However, there were still a few users selecting other conditions as their favorites because the quality of collaboration is not only determined by the interface but is also influenced by other factors such as the relationship between users, familiarity with a task, ideas they have for solving a task, and other reasons.

There are several factors limiting the scope of our studies. The physical objects to manipulate were small and light, so the local participants were able to quickly try out their own idea. If the task was to manipulate big or heavy objects, the local users may not have tried out their ideas as quickly, and may have become more cautious and more discussion with the remote user before manupulating the objects. Moreover, if the task objects were spread in a larger area, the task may have included navigating or walking activities in addition to manipulating objects. In terms of study design, further investigation on various types of tasks and inclusion of more objective meausures would be desireable. In addition, developing more comprehensive questionnaire to measure 'awareness' and other qualitative indicators of remote collaboration including large enough number of rating items to construct scale would be an interesting and important future study direction.

5. Conclusion and Future Work

In this paper, we described two user studies on user interfaces for better remote collaboration with visual annotation cues. In the first study, we explored solutions for the issue of incorrectly anchoring annotations when the local user changes the viewpoint. We compared two solutions (manual-freeze and auto-freeze interfaces) to a non-freeze condition, and found that the auto-freeze interface was most preferred by both the remote and local participants as it supports easy and quick drawing annotations while solving the issue of inaccurately anchored annotations. The manual-freeze method also solved the issue but required two additional inputs to pause and restart a live video, and so was least preferred.

The second study investigated how visual notifications could be used to improve local user awareness about the remote partner's new annotations. We designed two visual notifications (red-box and both-freeze notifications) and compared them to a no notification condition. The results showed that the red-box notification was the most preferred as it improved local participants' awareness of when the remote user drew annotations. The both-freeze condition improved the local participants' awareness of both when and where the remote participants drew annotations but it was too interruptive for local participants.

We will continue a future study exploring the use of the freeze method while a local user walks or navigates around during the task. Additionally, future work could include objectively measured human factors. For example, capturing gaze alignment between collaborators using gaze tracking systems (e.g. the Tobii eye tracker) would help to investigate how well they focus on the same objects. Another interesting direction would be sharing emotions between the collaborators to help the collaboration partners to react according to the shared emotions.

References

[1] Ronald T. Azuma, "A survey of augmented reality," *Presence: Teleoperators and Virtual Environments*, vol. 6, no. 4, pp. 355–385, 1997. Article (CrossRef Link)

[2] Bhaskar Bhattacharya and Eliot Winer, "A method for real-time generation of augmented reality work instructions via expert movements," *The Engineering Reality of Virtual Reality 2015, International Society for Optics and Photonics*, Vol.9392 , March, 2015. Article (CrossRef Link)

[3] S. Webel, U. Bockholt, T. Engelke, N. Gavish, M. Olbrich, and C. Preusche, "An augmented reality training platform for assembly and maintenance skills," *Robotics and Autonomous Systems*, vol. 61, no. 4, pp. 398-403, 2013. Article (CrossRef Link)

[4] Steven J. Henderson and Steven K. Feiner, "Augmented reality in the psychomotor phase of a procedural task," in *Proc. of IEEE International Symposium on Mixed and Augmented Reality* ,pp. 191-200, October 2011. Article (CrossRef Link)

[5] Rafael Radkowski, Jordan Herrema, and James Oliver, "Augmented reality-based manual assembly support with visual features for different degrees of difficulty," *International Journal of Human-Computer Interaction*, vol. 31, no. 5, pp. 337-349, 2015. Article (CrossRef Link)

[6] Nils Petersen and Didier Stricker, "Learning task structure from video examples for workflow tracking and authoring," in *Proc. of IEEE International Symposium on Mixed and Augmented Reality*, pp. 237-246, November 2012. Article (CrossRef Link)

[7] Guido Maria Re and Monica Bordegoni, "An augmented reality framework for supporting and monitoring operators during maintenance tasks," in *Proc. of International Conference on Virtual, Augmented and Mixed Reality*, pp. 443-454, June 2014, Article (CrossRef Link)

[8] Dima Damen, Teesid Leelasawassuk, and Walterio Mayol-Cuevas, "You-Do, I-Learn: Egocentric unsupervised discovery of objects and their modes of interaction towards video-based guidance," *Computer Vision and Image Understanding*, vol. 149, pp. 98-112, August 2016. Article (CrossRef Link)

[9] Susan R. Fussell, Leslie D. Setlock, Jie Yang, Jiazhi Ou, Elizabeth Mauer and Adam D.I. Kramer, "Gestures over video streams to support remote collaboration on physical tasks," *Human-Computer Interaction*, vol. 19, no. 3, pp. 273-309. November 12 2009. Article (CrossRef Link)

[10] J. R. Brubaker, G. Venolia, and J. C. Tang, "Focusing on shared experiences: moving beyond the camera in video communication" in *Proc. of Designing Interactive Systems Conference*, pp. 96-105, June 11-15, 2012. Article (CrossRefLink)

[11] S. R. Fussell, R. E. Kraut, and J. Siegel, "Coordination of communication: Effects of shared visual context on collaborative work," in *Proc. of the 2000 ACM conference on Computer supported cooperative work*, pp. 21-30, December 2000. ACM Article (CrossRef Link)

[12] Seungwon Kim, Gun A. Lee, and Nobuchika Sakata, (2013, October). "Comparing pointing and drawing for remote collaboration," in *Proc. of IEEE International Symposium on Mixed and Augmented Reality*, pp. 1-6, 2013. Article (CrossRef Link)

[13] S. Gauglitz, C. Lee, M. Turk, and T. Höllerer, "Integrating the physical environment into mobile remote collaboration," in *Proc. of the 14th international conference on Human-computer interaction with mobile devices and services*, pp. 241-250, September 2012. Article (CrossRef Link)

[14] S. Gauglitz, B. Nuernberger, M. Turk, and T. Höllerer, "World-stabilized annotations and virtual scene navigation for remote collaboration," in *Proc. of the 27th annual ACM symposium on User interface software and technology*, pp. 449-459, October 2014. Article (CrossRef Link)

[15] 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 2014 IEEE International Symposium on Mixed and Augmented Reality*, pp. 83-92, September 2014. Article (CrossRef Link)

[16] D. Kirk, T. Rodden, and D. S. Fraser, "Turn it this way: grounding collaborative action with remote gestures," in *Proc. of the SIGCHI conference on Human Factors in Computing Systems*, pp. 1039-1048, April 2007. Article (CrossRef Link)

[17] R. S. Sodhi, B. R. Jones, D. Forsyth, B. P. Bailey, and G. Maciocci, "BeThere: 3D mobile collaboration with spatial input," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 179-188, April 2013. Article (CrossRef Link)
[18] J. Rekimoto and K. Nagao, "The world through the computer: Computer augmented interaction with real world environments," in *Proc. of Proceedings of the 8th annual ACM symposium on User interface and software technology*, pp. 29-36, December 1995, Article (CrossRef Link)
[19] S. Kim, G. A. Lee, S. Ha, N. Sakata, and M. Bllinghurst, "Automatically freezing live video for annotation during remote collaboration," in *Proc. of Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, pp. 1669-1674, April 2015. Article (CrossRef Link)
[20] Joerg Hauber, "Understanding Remote Collaboration in Video Collaborative Virtual Environments," *PhD Thesis, University of Canterbury*, 2008. Article (CrossRef Link)
[21] Gary M. Olson and Judith S. Olson, " Distance matters," *Human-computer interaction*, vol. 15, no. 2, pp. 139-178, December 09 2009. Article (CrossRef Link)
[22] R.E. Kraut, M. D. Miller and J. Siegel, "Collaboration in performance of physical tasks: Effects on outcomes and communication," in *Proc. of Proceedings of the ACM conference on Computer supported cooperative work*, pp. 57-66, November 1999. ACM. Article (CrossRef Link)
[23] 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 Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 513-520, April 2003. Article (CrossRef Link)
[24] P. Dourish, and V. Bellotti, "Awareness and coordination in shared workspaces," in *Proc. of Proceedings of the ACM conference on Computer-supported cooperative work*, pp. 107-114, December 1992. Article (CrossRef Link)
[25] C. Gutwin and S. Greenberg, "Workspace awareness for groupware," in *Proc. of Proceedings of the Conference on Human Factors in Computing Systems*, pp.208–209, 1996. Article (CrossRef Link)
[26] John C. Tang and Scott L. Minneman, "VideoDraw: a video interface for collaborative drawing," *ACM Transactions on Information Systems (TOIS)*, vol. 9, no.2, pp.170-184, April 1991. Article (CrossRef Link)
[27] Hiroshi Ishii, Minoru Kobayashi and Kazuho Arita, "Iterative design of seamless collaboration media," *Communications of the ACM*, vol.37, no.8, pp.83–97, August 1994. Article (CrossRef Link)
[28] S. Chen, M. Chen, A. Kunz, A. E. Yantaç, M. Bergmark, A. Sundin, and M. Fjeld, "SEMarbeta: mobile sketch-gesture-video remote support for car drivers," in *Proc. of Proceedings of the 4th Augmented Human International Conference on - AH*, pp.69–76, 2013 Article (CrossRef Link)
[29] W. L. Koh, J. Kaliappan, M. Rice, K. T. Ma, H. H. Tay, and W. P. Tan "Preliminary investigation of augmented intelligence for remote assistance using a wearable display," in *Proc. of TENCON 2017-2017 IEEE Region 10 Conference*, pp. 2093-2098. November 5-8 2017. Article (CrossRef Link)
[30] M. Rice, S. C. Chia, H. H. Tay, M. Wan, L. Li, J. Ng, and J. H. Lim, "Exploring the use of visual annotations in a remote assistance platform," in *Proc. of Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pp. 1295-1300, May 2016. Article (CrossRef Link)
[31] H. Kato and M. Billinghurst, "Marker tracking and HMD calibration for a video-based augmentedreality conferencing system," in *Proc. of Proceedings 2nd IEEE and ACM International Workshop on Augmented Reality*, October 1999. Article (CrossRef Link)
[32] S. Güven, S. Feiner, and O. Oda, "Mobile augmented reality interaction techniques for authoring situated media on-site," in *Proc. of IEEE International Symposium on Mixed and Augmented Reality*, pp.235–236, October 2006. Article (CrossRef Link)

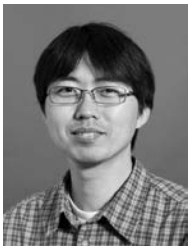
[33] G. A., Lee, U. Yang, Y. Kim, D. Jo, J.H. Kim and J. S. Choi, "Freeze-Set-Go interaction method for handheld mobile augmented reality environments," in *Proc. of Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology*, pp.143–146, 2009. Article (CrossRef Link)
[34] O. Fakourfar, K. Ta, R. Tang, S. Bateman, and A. Tang, "Stabilized annotations for mobile remote assistance," in *Proc. of Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 1548-1560, May 2016. Article (CrossRef Link)
[35] Y.S Chang, B. Nuernberger, B. Luan and T. Höllerer, "Evaluating gesture-based augmented reality annotation," in *Proc. of 2017 IEEE Symposium on 3D User Interfaces*, pp. 182-185, March 2017. Article (CrossRef Link)
[36] B. Nuernberger, K. C. Lien, T. Höllerer, and M. Turk, "Interpreting 2d gesture annotations in 3d augmented reality," in *Proc. of 2016 IEEE Symposium on 3D User Interfaces*, pp. 149-158, March 2016. Article (CrossRef Link)
[37] B. Nuernberger, K. C. Lien, G. Grinta, C. Sweeney, M. Turk, and T. Höllerer, "Multi-view gesture annotations in image-based 3D reconstructed scenes," in *Proc. of Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*, pp. 129-138, November 2016. Article (CrossRef Link)
[38] Saul Greenberg and Carl Gutwin, "Implications of we-awareness to the design of distributed groupware tools," *Computer Supported Cooperative Work (CSCW)*, vol. 25, issue. 4-5, pp. 279-293, October 2016. Article (CrossRef Link)
[39] Stephan Lukosch, Heide Lukosch, Dragos Datcu, and Marina Cidota, "Providing information on the spot: Using augmented reality for situational awareness in the security domain," *Computer Supported Cooperative Work (CSCW)*, vol. 24, issue 6, pp. 613-664, December 2015. Article (CrossRef Link)
[40] Jesus Gallardo, Crescencio Bravo, and Ana Isabel Molina, "A framework for the descriptive specification of awareness support in multimodal user interfaces for collaborative activities," *Journal on Multimodal User Interfaces*, vol. 12, no. 2 pp. 145-159, November 28 2017. Article (CrossRef Link)
[41] C. Gutwin, O. Schneider, R. Xiao, and S. Brewster, "Chalk sounds: the effects of dynamic synthesized audio on workspace awareness in distributed groupware," in *Proc. of Proceedings of the ACM 2011 conference on Computer supported cooperative work*, pp. 85-94, March 2011. Article (CrossRef Link)
[42] C. Gutwin, S. Bateman, G. Arora, and A. Coveney, "Looking Away and Catching Up: Dealing with Brief Attentional Disconnection in Synchronous Groupware," in *Proc. of In Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pp. 2221-2235, February 2017. Article (CrossRef Link)
[43] Marina Cidota, Stephan Lukosch, Dragos Datcu, and Heide Lukosch, "Comparing the effect of audio and visual notifications on workspace awareness using head-mounted displays for remote collaboration in augmented reality," *Augmented Human Research*, vol.1, no. 1, 10 October 2016. Article (CrossRef Link)
[44] Simon Y. Li, Farah Magrabi, and Enrico Coiera, "A systematic review of the psychological literature on interruption and its patient safety implications," *Journal of the American Medical Informatics Association*, vol. 19, no. 1, pp.6-12, September 2011. Article (CrossRef Link)
[45] Chih-Yuan Ho, Mark I. Nikolic, and Nadine B. Sarter, "Supporting timesharing and interruption management through multimodal information presentation," in *Proc. of Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 45, no. 4, pp. 341-345, October 2001. Article (CrossRef Link)
[46] Michael Posner, "Orienting of attention," *Quarterly journal of experimental psychology*, vol. 32, no.1, pp.3-25, May 2007. Article (CrossRef Link)
[47] Shameem Hameed, Thomas Ferris, Swapnaa Jayaraman, and Nadine Sarter, "Using informative peripheral visual and tactile cues to support task and interruption management," *Human factors: The Journal of the Human Factors and Ergonomics Society*, vol. 51, no. 2, pp. 126-135, June 2009. Article (CrossRef Link)

[48] David D. Woods, "The alarm problem and directed attention in dynamic fault management," *Ergonomics*, vol. 38, no. 11, pp. 2371-2393, March 2007. Article (CrossRef Link)
[49] W. Tan, H. Liu, Z. Dong, G. Zhang, H. Bao, "Robust monocular SLAM in dynamic environments," in *Proc. of the 2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 209-218, May 2013. Article (CrossRef Link)
[50] Seungwon Kim, Mark Billinghurst, and Gun Lee, "The Effect of Collaboration Styles and View Independence on Video-Mediated Remote Collaboration," *Computer Supported Cooperative Work (CSCW)*, vol. 27, no. 3-6, pp 569-607, June 2018. Article (CrossRef Link)
[51] G. A. Lee, S. Kim, Y. Lee, A. Dey, T. Piumsomboon, M. Norman, and M. Billinghurst, "Improving Collaboration in Augmented Video Conference using Mutually Shared Gaze," in *Proc. of ICAT-EGVE 2017 - International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments*, pp.197-204, 2017. Article (CrossRef Link)
[52] Kunal Gupta, Gun A. Lee, and Mark Billinghurst, "Do you see what i see? the effect of gaze tracking on task space remote collaboration," *IEEE transactions on visualization and computer graphics*, vol. 22, no. 11, pp.2413-2422, July 2016. Article (CrossRef Link).

Dr. Seungwon Kim is a postdoctoral fellow at Empathic Computing Laboratory, University of South Australia. He received his PhD degree in HITLab NZ, New Zealand, in 2016 with the supervision of Prof. Mark Billinghurst. During his PhD, he developed a drawing annotation interface that stabilizes the annotations in the real world. He received his Master in 2010 and his Bachelor in 2008 at University of Tasmania. His research interests include remote collaboration using augmented virtual communication cues and sharing experience between distance users.

Dr. Mark Billinghurst is a Professor of Human Computer Interaction at the University of South Australia, researching Empathic Computing. Prior to this he was Director of the HIT Lab NZ at the University of Canterbury (2002-2015), developing innovative computer interfaces. He received a PhD degree in Electrical Engineering from the University of Washington (2002), and has produced over 350 publications in areas such as Augmented Reality, multimodal interaction and mobile interfaces. In 2013 his work in Augmented Reality received the IEEE VR Technical Achievement Award, and he was elected as a Fellow of the Royal Society of New Zealand.

Dr. Chil-Woo LEE received the B.S. and M.S. degrees in electronic engineering from Chung-Ang University in 1986 and 1988 respectively. He received a Ph.D. in electronic engineering from University of Tokyo, Japan in 1992. Since 1996, he is a professor at Dept. of Computer Engineering, Chonnam National University. He was a senior researcher at Laboratories of Image Information Science and Technology (LIST) for four years, form 1992 to 1996, and during the years he was also a visiting researcher at Osaka University. His research area is image recognition and image synthesis, but his research interest is not limited to them and includes Computer Vision, Computer Graphics, and visual human interface system.



Dr. Gun Lee is a Senior Research Fellow at the Empathic Computing Laboratory, University of South Australia, investigating interaction and visualization methods for sharing virtual experiences in Augmented Reality (AR) and Virtual Reality (VR) environments. Recently, using AR and wearable interfaces to improve remote collaborative experience has been one of his main research themes. He received his PhD degree in Computer Science and Engineering at POSTECH investigating immersive authoring methods for creating VR and AR content using 3D user interfaces.