

**The Interface for  
Remote Parallel Experience Collaboration  
Using Augmented Visual Communication Cues**

A thesis submitted in partial fulfilment  
of the requirements for the Degree of  
Doctor of Philosophy

at the  
University of Canterbury

by  
Seungwon Kim

**Supervising Committee**

Associate Professor Christoph Bartneck	Senior Supervisor
Dr. Gun Lee	Co-Supervisor
Professor Mark Billinghurst	Associate Supervisor

Human Interface Technology Laboratory New Zealand

**University of Canterbury**

**2016**



## Abstract

Today people use video conferencing systems to share their task space for many collaborative purposes. There is a body of research on adding visual communication cues in the shared task space, but this mostly focuses on remote expert collaboration in which communication is mostly one way communication from a remote expert to a local operator. This thesis explores the parallel experience collaboration where both users discuss their ideas via a video conferencing system that includes verbal and visual communication cues such as pointers or drawing annotations.

The first study (Chapter 5), explores the use of two different visual communication cues (pointer and drawing annotation) and how these affect the parallel experience collaboration. The user study found that while both cues increased the level of connectedness, the participants preferred the pointer interface over the annotation interface.

The second and third studies (Chapter 6) investigate solutions for the issues in the annotation interface. One such issue is that if the local user (who is sharing his or her task space) changes the viewpoint of a shared live video while the remote user is drawing an annotation, the annotation is drawn at the wrong place. In the second study, the author investigates freeze functions as a solution; comparing auto freeze and manual freeze conditions to the non-freeze condition. The user study revealed that the auto freeze function solved the issue without additional inputs neither losing the “liveness” of the shared live video. In the third study, the author investigated solutions for local users’ task management. The local users sometimes missed new annotations from the remote users while focused on their own task. As potential solutions, the author proposed the *red box* and the *both freeze* notifications and conducted a user study to compare these with the non-notification condition. The results found that the participants significantly preferred the red box notification, because it solved the problem without causing significant interruption.

Since the independence of the remote user's view from the local user's view helps them have better remote expert collaboration, the fourth study (Chapter 7) explores the use of the independent view in parallel experience collaboration. As with remote expert collaboration, parallel experience collaboration participants preferred the independent view because it provided a stable view rather than moving view according to the local user's head movement.

In conclusion, through a series of user experiments the author found that the interface for drawing annotation in parallel experience collaboration would be better to 1) support quick and easy use of a drawing annotation interface, 2) include appropriate notification that helps a local user to know when a remote user is making an annotation and 3) support an independent view 4) with the function that a remote user can start drawing while toggling from the dependent view to independent views.

During this Ph.D., several contributions have been achieved. The author introduces one of the earliest prototypes for anchoring drawings in the real world without any previously known tracking data or visual markers (described in Chapter 4); describes the use of visual communication cues: pointers and drawing annotation, for a parallel experience collaboration, and found four user states (*collaborating together*, *playing in parallel*, *passive*, and *do-it-alone* states) (in Chapter 5); suggests an auto freeze interface for easy and quick draw annotation and a notification interface without causing a significant level of interruption for local user's task management (in Chapter 6); has found the benefit of independent view in parallel experience collaboration (in Chapter 7); and additionally, from the four user studies, introduces a communication model for parallel experience collaboration including verbal and visual communication cues.

Deputy Vice-Chancellor's Office  
Postgraduate Office



## Co-Authorship Form

This form is to accompany the submission of any thesis that contains research reported in co-authored work that has been published, accepted for publication, or submitted for publication. A copy of this form should be included for each co-authored work that is included in the thesis. Completed forms should be included at the front (after the thesis abstract) of each copy of the thesis submitted for examination and library deposit.

**Please indicate the chapter/section/pages of this thesis that are extracted from co-authored work and provide details of the publication or submission from the extract comes:**

### **Chapter 3 / Pages 31- 48**

- 1 . **Kim, S.**, Lee, G. A., Sakata, N., & Billinghamurst, M. (2013, October). Comparing pointing and drawing for remote collaboration. In *Mixed and Augmented Reality (ISMAR)*, 2013 IEEE International Symposium on (pp. 1-6). IEEE., (Work-In-Progress)
- 2 . **Kim, S.**, Lee, G. A., Sakata, N., Dunser, A., Vartiainen, E., & Billinghamurst, M. (2013, October). Study of augmented gesture communication cues and view sharing in remote collaboration. In *Mixed and Augmented Reality (ISMAR)*, 2013 IEEE International Symposium on (pp. 261-262). IEEE (Poster)

### **Chapter 5 /Pages 60 -85**

- 3 . **Kim, S.**, Lee, G., Sakata, N., & Billinghamurst, M. (2014, September). Improving co-presence with augmented visual communication cues for sharing experience through video conference. In *Mixed and Augmented Reality (ISMAR)*, 2014 IEEE International Symposium on (pp. 83-92). IEEE.(Long)

**Chapter 6 / Section 6.1 to 6.3 / Pages 87 -101**

- 4 . **Kim, S.**, Lee, G. A., Ha, S., Sakata, N., & Billinghamurst, M. (2015, April).  
Automatically Freezing Live Video for Annotation during Remote Collaboration.  
In Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human  
Factors in Computing Systems (pp. 1669-1674). ACM (Conference Extended  
Abstracts)

**Chapter 6 / Pages 86-120**

- 5 . **Kim, S.**, Lee, G., & Billinghamurst, M., “Using Freeze Frame and Visual Notifications  
for Drawing Annotations in Augmented Reality Remote Collaboration”, Submitted  
at ‘International Journal of Human-Computer Studies’

Please detail the nature and extent (%) of contribution by the candidate:

Publications are ordered by numbers in the previous section.

Publication (1): 80% - Developed the system, conducted the study, and lead the writing of the paper.

Publication (2): 80% - Developed the system, conducted the study, and lead the writing of the paper.

Publication (3): 80% - Developed the system, conducted the study, and lead the writing of the paper.

Publication (4): 80% - Developed the system, conducted the study, and lead the writing of the paper

Publication (5): 80% - Developed the system, conducted the study, and lead the writing of the paper.

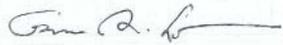
**Certification by Co-authors:**

If there is more than one co-author then a single co-author can sign on behalf of all

The undersigned certifies that:

- The above statement correctly reflects the nature and extent of the PhD candidate’s contribution to this co-authored work
- In cases where the candidate was the lead author of the co-authored work he or she wrote the text

Name: *Gun Lee*

Signature: 

Date: *29 April 2016*

## **Publications from this dissertation**

Material from this dissertation has been previously published or submitted in the peer-reviewed papers, posters, and extended abstracts listed below. The chapters of this thesis that relate to each publication are noted.

- 1 . Kim, S., Lee, G., & Billinghurst, M., “Using Freeze Frame and Visual Notifications for Drawing Annotations in Augmented Reality Remote Collaboration”, Submitted at ‘International Journal of Human-Computer Studies’, Described in the Chapter 6 (the second and third studies)
- 2 . Kim, S., Lee, G. A., Ha, S., Sakata, N., & Billinghurst, M. (2015, April). Automatically Freezing Live Video for Annotation during Remote Collaboration. In Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (pp. 1669-1674). ACM, Described in the Chapter 6 (the second study)
- 3 . Kim, S., Lee, G., Sakata, N., & Billinghurst, M. (2014, September). Improving co-presence with augmented visual communication cues for sharing experience through video conference. In Mixed and Augmented Reality (ISMAR), 2014 IEEE International Symposium on (pp. 83-92). IEEE., Described in the Chapter 5 (the first study)
- 4 . Kim, S., Lee, G. A., Sakata, N., & Billinghurst, M. (2013, October). Comparing pointing and drawing for remote collaboration. In Mixed and Augmented Reality (ISMAR), 2013 IEEE International Symposium on (pp. 1-6). IEEE., Described in the Chapter 3 (the first pilot study)
- 5 . Kim, S., Lee, G. A., Sakata, N., Dunser, A., Vartiainen, E., & Billinghurst, M. (2013, October). Study of augmented gesture communication cues and view sharing in remote collaboration. In Mixed and Augmented Reality (ISMAR), 2013 IEEE International Symposium on (pp. 261-262). IEEE., Described in the Chapter 3 (the first and second pilot studies)

During my doctoral studies, the following papers were also published, but are not a part of this thesis.

- 6 . Lee, G. A., Dunser, A., Kim, S., & Billinghamurst, M. (2012, November). CityViewAR: A mobile outdoor AR application for city visualization. In Mixed and Augmented Reality (ISMAR-AMH), 2012 IEEE International Symposium on(pp. 57-64). IEEE.
- 7 . Kim, S., Junuzovic, S., & Inkpen, K. (2014, November). The Nomad and the Couch Potato: Enriching Mobile Shared Experiences with Contextual Information. In Proceedings of the 18th International Conference on Supporting Group Work (pp. 167-177). ACM.

## **Acknowledgements**

I wish to express sincere appreciation to the following people for their advising, supervising, guiding and support. Throughout this Ph.D., I learnt not only about developing a prototype, designing a user study, and writing papers, but also being smarter for reading and understanding the world.

I would like to thank Mark Billingham. He gave me the opportunity of an internship at HITLab NZ and this Ph.D. study. I have become smarter as my study progressed, and I am sure that it will advance me not only economically but in every aspect of my life. Dr Billingham helped me plan and design studies, and write papers. I was lucky to have him as my supervisor.

I thank Gun Lee, not only for his help with my study but also for my life in Christchurch. He was the one I relied on; I talked with him almost every day about the study and also personal issues. He contributed to every single step of my study and suggested the best alternative for every issue or discussion.

I thank Christoph Bartneck. He was my senior supervisor from June 2015 to April 2016. Discussions with Dr Bartneck expanded my horizons in academic research and he helped me to think outside of my usual perspective.

Lastly, I would like to dedicate this work to my family who always believed in me. This would not have been possible without their unwavering support.

# Table of Contents

<b>Abstract .....</b>	<b>I</b>
<b>Publications from this dissertation.....</b>	<b>V</b>
<b>Acknowledgements .....</b>	<b>VII</b>
<b>Table of Contents.....</b>	<b>VIII</b>
<b>List of Figures .....</b>	<b>XIII</b>
<b>List of Tables.....</b>	<b>XVIII</b>
<b>1 Introduction .....</b>	<b>1</b>
1.1 Motivation.....	1
1.2 Research Area.....	3
1.3 Thesis Structure .....	6
<b>2 Related Work.....</b>	<b>9</b>
2.1 Augmented Visual Communication Cues.....	9
2.1.1 Early Works .....	9
2.1.2 Laser Pointer.....	10
2.1.3 Using Graphical Pointer and Annotation.....	11
2.1.4 Hand Gesture .....	15
2.1.5 Summary.....	17
2.2 Sharing Task Space View .....	19
2.2.1 Early Works .....	19
2.2.2 Head Mounted Camera View & Scene Camera View.....	20
2.2.3 Independent View .....	20
2.2.4 Summary.....	22
2.3 Communication mechanisms.....	22

2.3.1	Early Works .....	2 3
2.3.2	Interpersonal Communication.....	2 3
2.3.3	Collaboration .....	2 5
2.3.4	Summary .....	2 6
2.4	Chapter summary .....	2 6
<b>3</b>	<b>Pilot Studies of Remote Expert.....</b>	<b>2 8</b>
3.1	Introduction to the Pilot Studies .....	2 8
3.2	Prototype.....	2 9
3.3	Pilot Study Design .....	3 1
3.4	First Pilot Study Results .....	3 2
3.5	First Pilot Study (Discussion).....	3 4
3.6	Second Pilot Study Design & Result .....	3 5
3.7	Second Pilot Study (Discussion).....	3 8
3.8	General Discussion .....	3 8
3.9	Chapter Summary .....	4 0
<b>4</b>	<b>The Research Approach.....</b>	<b>4 2</b>
4.1	Relationship between studies.....	4 2
4.2	System Architecture.....	4 3
4.2.1	Local End Point .....	4 4
4.2.2	Remote End Point .....	4 6
4.3	Experiment Style .....	4 7
4.3.1	Experimental Task and Environment .....	4 7
4.3.2	Experimental Procedure.....	4 8
4.3.3	Data Collection .....	4 9

4.4	Chapter Summary .....	5 1
<b>5</b>	<b>Study about the Use of Visual Communication Cues.....</b>	<b>5 2</b>
5.1	User Study Design.....	5 2
5.2	Results .....	5 5
5.2.1	User Enjoyment & Focus on the Task.....	5 6
5.2.2	Connectedness & Being Together .....	5 8
5.2.3	Express Idea (sending/encoding messages).....	6 0
5.2.4	Understand Partner (receiving/decoding messages) .....	6 3
5.2.5	User Preference.....	6 5
5.2.6	Live Video vs. Snapshot Image .....	6 6
5.2.7	HMD vs. HHD.....	6 6
5.3	Observation of User Behavior .....	6 7
5.4	Discussion.....	7 1
5.5	Conclusion and Summary .....	7 3
<b>6</b>	<b>Studies about Freeze Frame Interactions and Visual Notifications.....</b>	<b>7 5</b>
6.1	User Study on Freeze Function.....	7 6
6.1.1	User Interface.....	7 6
6.1.2	User Study Design .....	7 7
6.2	Results of Freeze Study .....	8 0
6.2.1	Focusing on the assembling task .....	8 1
6.2.2	Sending/Encoding Messages (drawing annotation cue) .....	8 3
6.2.3	Verbal messages .....	8 6
6.2.4	Receiving/Decoding Messages .....	8 7
6.2.5	Overall Collaboration .....	9 0

6.3	Summary of Freeze Function and an Issue on the Local End.....	9	2
6.4	User Study on Visual Notification.....	9	4
6.4.1	Previous Studies in Notification .....	9	4
6.4.2	User Interface.....	9	6
6.4.3	User Study Design .....	9	7
6.5	Results of the Notification Study.....	9	9
6.5.1	Focusing on the task .....	1	0 0
6.5.2	Receiving/Decoding Messages .....	1	0 2
6.5.3	Verbal Communication & User Behavior.....	1	0 5
6.5.4	Sending/Encoding Messages .....	1	0 7
6.5.5	Overall Collaboration & User Preference.....	1	0 8
6.6	Summary of Notification Study.....	1	1 3
6.7	Discussion.....	1	1 3
6.8	Conclusion.....	1	1 5
<b>7</b>	<b>Exploring View Independence.....</b>	<b>1</b>	<b>1 7</b>
7.1	Prototype and User Interface .....	1	1 7
7.2	User Study Design.....	1	2 2
7.3	Results of Independent Views .....	1	2 4
7.3.1	Focusing and understanding the task.....	1	2 4
7.3.2	Receiving/Decoding Messages .....	1	2 9
7.3.3	Sending/Encoding Messages .....	1	3 1
7.3.4	Connectedness & Being Together .....	1	3 4
7.3.5	Overall Collaboration & User Preference.....	1	3 6
7.4	Discussion.....	1	4 0

7.5	Conclusion.....	1 4 1
<b>8</b>	<b>Discussion.....</b>	<b>1 4 3</b>
8.1	Drawing Annotation Interface .....	1 4 3
8.2	Remote Expert & Parallel Experience.....	1 4 3
8.3	Connectedness .....	1 4 4
8.4	Parallel Experience Collaboration Model.....	1 4 6
8.5	Limitation of the studies .....	1 4 8
<b>9</b>	<b>Conclusion.....</b>	<b>1 5 0</b>
9.1	Summary.....	1 5 0
9.2	Future Research .....	1 5 2
	<b>References.....</b>	<b>1 5 3</b>
	<b>Appendices .....</b>	<b>1 5 9</b>
	Appendix A : Questionnaire for Demographic Information.....	1 5 9
	Appendix B : Material for Study about the Use of Visual Communication cues .....	1 6 1
	Appendix C : Material for the Study about Freeze Frame.....	1 7 4
	Appendix D : Material for the Study about Notificatoion .....	1 8 3
	Appendix E : Material for the Study about Independent view .....	1 8 7

## List of Figures

Figure 1.1: Examples of traditional ‘talking head’ video conferencing (left) and video conferencing beyond the face-to-face talk.. .....	3
Figure 1.2: A new trend of using video conferencing: sharing task space .....	3
Figure 1.3: The local users’ views of the systems.....	4
Figure 2.1: ClearBoard supports remote gaze awareness while sharing annotation.....	1 0
Figure 2.2: NaviCam (Rekimoto & Nagao 1995).....	1 0
Figure 2.3: System architecture of GestureCam (Kuzuoka et al. 1994) .....	1 1
Figure 2.4: Wearable Active Camera/Laser (WACL) system (Sakata et al. 2003) .....	1 1
Figure 2.5: ‘DOVE’ system (Fussell et al. 2004).....	1 2
Figure 2.6: AR conferencing for remote collaboration. (Kato & Billinghurst 1999) .....	1 3
Figure 2.7: The device for a local user (a) and views for remote and local users (b and c) (Gauglitz et al. 2012) .....	1 3
Figure 2.8: Screenshot of the remote user's interface (Gauglitz et al. 2014b)	1 4
Figure 2.9: Depth interpretations of 2D drawings. (Gauglitz et al. 2014a) ...	1 5
Figure 2.10: Schematic diagram of ‘VideoDraw’ (left) (Tang & Minneman 1991) and VideoArms in action showing two groups working together (right) (Tang 2004) .....	1 6
Figure 2.11: System architecture (Left) and worker's view (Right) (Kirk et al. 2007) .....	1 6
Figure 2.12: Remote Tele-Expert System (Alem et al. 2011).....	1 7
Figure 2.13: (Sodhi et al. 2013) .....	1 7
Figure 2.14: JackIn. (Kasahara & Rekimoto 2014) .....	2 1
Figure 2.15: The context compass in Social Panoramas (Reichherzer et al. 2014)	

.....	2 2
Figure 2.16: Shannon and Weaver's transmission model of communication (Shannon & Weaver 1949).....	2 3
Figure 2.17: The modified Shannon-Weaver model for interpersonal communication, simplified from Devito (1998, page 12).....	2 4
Figure 2.18: Hauber's communication model for remote collaboration.....	2 5
Figure 3.1: Four conditions in pilot studies .....	3 0
Figure 3.2: Experiment set up (left) and a given instruction picture to a remote user for the first (middle) and second pilot studies (right).....	3 1
Figure 3.3: Average performance time in the first pilot study .....	3 3
Figure 3.4: Average of mouse cursor movement by remote participants.....	3 4
Figure 3.5: A task for the second pilot study.....	3 6
Figure 3.6: The sequence of collaboration in remote assistance with a live video .....	3 7
Figure 3.7: Stacked annotations with the 3D task.....	3 8
Figure 3.8: The communication model for remote expert .....	4 0
Figure 4.1: Overview of the proposed system structure .....	4 4
Figure 4.2: The local and remote users' environments. ....	4 5
Figure 4.3: An experimental task with a reference paper and puzzle pieces shown in a shared live video .....	4 8
Figure 4.4: Hauber's communication model for remote collaboration.....	5 0
Figure 5.1: Prototype system used by a remote user (Left) and a local user using a HMD (top right) or a HHD (bottom right).....	5 3
Figure 5.2: Results of Likert scale ratings from participants for the question about how much they felt they enjoyed or focused on assembling task. ....	5 6
Figure 5.3: Results of Likert scale ratings from participants for the question about how much they felt they were connected and together with. ....	5 9
Figure 5.4: Results of Likert scale ratings from participants for the question	

about how much they expressed ideas well. ....	6 1
Figure 5.5: Results of Likert scale ratings from participants for the question about how much they understood their partner well.....	6 4
Figure 5.6: Results of participants' preference .....	6 5
Figure 6.1: The process of drawing an annotation in the real world (top left), with the auto freeze condition (top right), and the manual freeze condition (bottom).....	7 7
Figure 6.2: The prototype systems used by a remote user (top left) and a local user wearing a HMD (right).....	7 8
Figure 6.3: Results of Likert scale ratings from local (left) and remote (right) participants for the question about how much they felt they focused on assembling task .....	8 2
Figure 6.4: Results of Likert scale ratings from the local participants (left) for the Q2 and from the remote participants (right) for the Q2, Q3, Q4, Q8, and Q9. ....	8 3
Figure. 6.5: Example of wrong drawings by the remote participant in group 3 with the non-freeze condition.....	8 5
Figure 6.6: Example of drawings.....	8 6
Figure. 6.7: Results of Likert scale ratings from local (left) and remote (right) participants for the question about understanding what their partners were trying to do and explaining. ....	8 8
Figure 6.8: Results of Likert scale ratings from local (left) and remote (right) participants for the question about how easily they understood what their partners were trying to do and explaining.....	9 0
Figure 6.9: User preference among the three conditions. ....	9 1
Figure. 6.10: Local user view when the remote user drew in the red box condition. The remote user is pointing out a part that is not matched with the reference.....	9 7
Figure 6.11: Results of Likert scale ratings from local (left) and remote (right)	

participants for the question 1, 5, and 9. ....	1 0 1
Figure 6.12: Results of Likert scale ratings from local (left) and remote (right) participants for the question 2, 4, 7, and 10 .....	1 0 2
Figure 6.13 Results of Likert scale ratings from local (left) and remote (right) participants for the question about properly expressing ideas .....	1 0 7
Figure 6.14: Results of Likert scale ratings from local (left) and remote (right) participants for the question about collaborating well and interface mentally stressful. ....	1 0 9
Figure. 6.15: Participants' preference among the three conditions .....	1 1 0
Figure 6.16: The average of the percentages among piece-manipulation types (local user alone, collaborative, remote user alone) in each condition (No: no notification, Red: red box, Both: both freeze). ....	1 1 1
Figure 7.1: The distance between HMD and the point where the users see ( $M$ = Mean, $SD$ = Standard Deviation, $N$ = number of measuring distance).....	1 1 8
Figure 7.2: Experimental environment at the local end for the independent view .....	1 1 9
Figure 7.3: Experimental environment at remote end for remote expert collaboration.....	1 2 0
Figure 7.4: The view of the local user. The brown tetragon represents the area that the remote partner could see. ....	1 2 1
Figure 7.5: Results of Likert-scale ratings of Q2, Q4, and Q12 from remote participants .....	1 2 5
Figure 7.6: The amount of navigation by a keyboard and a mouse .....	1 2 7
Figure 7.7: Results of Likert-scale ratings of Q2 and Q12 from local participants .....	1 2 9
Figure 7.8: Results of Likert-scale ratings of Q6, Q7, Q8, and Q13 from the remote participants .....	1 3 0
Figure 7.9: Results of Likert-scale ratings of Q6, Q7, Q8, and Q13 from the local participants .....	1 3 1

Figure 7.10: Results of Likert-scale ratings of Q5 from the remote and local participants .....	1 3 2
Figure 7.11: The number of drawings in four conditions.....	1 3 3
Figure 7.12: Number of strokes in four conditions .....	1 3 3
Figure 7.13: Results of Likert-scale ratings of Q9 and Q10 from the remote and local participants .....	1 3 5
Figure 7.14: Results of Likert-scale ratings of Q1, Q3, Q11 and Q14 from remote participants .....	1 3 6
Figure 7.15: Results of Likert-scale ratings of Q1, Q11, and Q14 from local participants .....	1 3 8
Figure 7.16: User performance with the remote assistance task.....	1 3 9
Figure 7.17: Participants' preference for each type of tasks .....	1 3 9
Figure 8.1: The communication model for parallel experience.. .....	1 4 7

## List of Tables

Table 2.1: Classifying the previous studies according to a type of remote collaboration they focused (A) and visual tracking system they used (B).....	1	8
Table 3.1: Average number of errors with the interfaces .....	3	3
Table 3.3: Average task completion time (in seconds) and the number of times the ‘Clear’ function was used while using annotation. ....	3	7
Table 5.1: Questions in the questionnaire after each session .....	5	4
Table 6.1: Questions asked after each session. Both participants answered the six questions (no color) and remote participants answered four more questions (highlighted in gray).....	7	9
Table 6.2 The number of wrong drawings that remote participants made.....	8	5
Table 6.3. Questions in the questionnaire after each session (Q7, highlighted in gray, was only for local participants). ....	9	8
Table 6.4: Results of the number of strokes, and the percentage of strokes in two second.....	1	0 8
Table 7.1: Interaction mapping for navigation in an independent view.....	1	2 0
Table 7.2: Questions in the questionnaire after each session (the questions 3 and 4, highlighted in grey, were only for the remote user). ....	1	2 3



# **1 Introduction**

Collaboration is important for problem solving or playing together, but sometimes the people collaborating are far apart. To address this problem, Alexander Graham Bell (1876) invented the telephone, enabling verbal communication between people in different places. Talking over a telephone makes it easy to share information, but it is limited to voice only communication.

Following the success of the telephone, the first video conferencing system was developed in the late 1930s and the first commercial system, AT&T's PicturePhone, became available in 1964 (Schnaars & Wymbys 2004). With early video conferencing, people were able to see each other as well as talk together. However the limited network bandwidth made the experience very poor until the late 1990s. Today, video conferencing has become widespread with powerful computing processors and video compression techniques.

The videoconferencing system helps users to understand the feelings of the remote person and gives an increased feeling of being together compared to audio-only communication over a phone (Isaacs & Tang 1994). However, the use of videoconferencing is limited to bust shot sharing (Inkpen et al. 2013) and verbal communication (Kato & Billinghamurst 1999).

This dissertation presents novel interface designs that could provide a better experience for remote collaboration while addressing the basic questions in the previous paragraph. The remainder of this chapter describes why it was studied (motivation), the research area, and an overview of the PhD work (thesis structure).

## **1.1 Motivation**

The study of this PhD was firstly designed with the potential of applying Augmented Reality (AR) technology on the videoconference system. Azuma (1997) reported the benefits of AR as enhancing a user's perception of and interaction with the real world. The augmented virtual objects display

information that helps the user to understand the real world. Billinghurst and Kato (1999, 2002) applied the AR technology into a video conference system with the benefits. Their study showed that AR is a promising tool for collaboration due to its ability to enhance the reality and provide spatial conferencing cues.

The AR technology has been advanced to anchor visual communication cues, such as a pointer and annotation, in the live video of the video conferencing system. Comport et al. (2006) introduced markerless visual tracking technologies so it is possible to anchor augmented virtual objects in the real world without any preparation (or learning). Recently, Tan et al. (2013) developed a robust SLAM system to enable stable camera tracking, so stable anchoring of virtual objects is possible even in heavily changed scenes of the real world.

There is a new trend emerging in video conferencing. Most current videoconferencing systems focus on sharing bust shots, so-called 'talking heads' (Inkpen et al. 2013) (see figure 1.1). However O' Hara et al. (2006) and Brubaker et al. (2012) discovered that people are increasingly starting to go beyond talking heads, such as showing what the remote partner is seeing or what is around them. Gaver (1993) demonstrated that for some collaborative tasks participants prefer seeing a video of the activity space rather than the people they were talking with. In this case the 'Task Space' video is more useful than the 'Talking Heads' video. Some recent examples of sharing task space that have attracted media attention are the iPad Bridesmaid<sup>1</sup> and a deployed soldier who watched the birth of his son on Skype<sup>2</sup>.

---

<sup>1</sup> [http://news.cnet.com/8301-17938\\_105-20097249-1/ipad-bridesmaid-attends-wedding-via-facetime/](http://news.cnet.com/8301-17938_105-20097249-1/ipad-bridesmaid-attends-wedding-via-facetime/)

<sup>2</sup> [http://www.youtube.com/watch?v=Z\\_w8RzYVIM4](http://www.youtube.com/watch?v=Z_w8RzYVIM4).



Figure 1.1: Examples of traditional ‘talking head’ video conferencing (left) and video conferencing beyond the face-to-face talk. A friend attends a wedding ceremony (middle), and a deployed soldier watches the birth of his son (right).

This new trend is possible as video conferencing devices become more portable and people can remain within the activities while sharing the task space. Some researchers (Kim et al. 2014; Kurata et al. 2004; Inkpen et al. 2013) investigated providing mobility while sharing the task space view (see Figure 1.2). They developed portable and wearable systems so that users could share the task space while remaining involved in the activity or event.



Figure 1.2: A new trend of using video conferencing: sharing task space (Inkpen et al. 2013; Kim et al. 2014)

In short, applying advanced AR technologies in the new way of using a video conferencing system may enable people to have a better remote collaboration experience.

## 1.2 Research Area

The ways of using a video conferencing system can be categorized into ‘Talking Head’ and ‘Task Space Sharing’ as mentioned in the previous section.

This research focuses on ‘Task Space Sharing’ where the shared video shows the local user’s surroundings.

The experience of sharing task space has been widely studied. Some of the studies did not add any virtual object on the shared task space view (Inkpen et al. 2013; Yarosh et al. 2010; Kim et al. 2014; Inkpen et al. 2012) while others did (Poelman et al. 2012; Fussell et al. 2004; Kirk et al. 2007). The added virtual objects were mainly for helping communication or collaboration. This author focuses on adding virtual objects in the shared task space view.

One of the main purposes of adding virtual objects into the shared task space is representing non-verbal hand gesture information (such as pointing, positioning, or orientating an object). This could be important when a remote user needs to share spatial information with a local user and the added virtual objects are visual communication cues in the collaboration. Various approaches have been investigated for sharing spatial information, such as using a laser pointer on a remote controlled robotic arm (Kuzuoka et al. 1994; Sakata et al. 2003) for pointing at an object or area, or superimposing the remote user’s hand into the video of the task space shared by the local user (Kirk et al. 2007; Huang & Alem 2011; Sodhi et al. 2013) (see Figure 1.3).



Figure 1.3: The local users’ views of the systems Kuzuoka et al. 1994 (left) and Sodhi et al. 2013 (right).

This author also focuses on augmenting visual communication cues for

sharing spatial information in a task space video. Based on direction of communication (between local and remote users' surroundings) and shared information, the remote collaboration experience can be categorized into mainly three types: (1) remote teaching, (2) remote expert, and (3) parallel experience. In remote teaching, the shared task space is mostly the remote teacher's space. When remote teachers share their real world space, they mostly employ the required material (or objects), and directly manipulate (or do something to) the material rather than using virtually augmented visual communication cues.

In the remote expert and parallel experience, both the remote and local users share the real world space of the local user (task space) and the local user does not know the solution for the task. The main difference between the remote expert and parallel experience is whether the remote user has the solution for the task. In remote expert, the remote user knows the solution so can provide instructions to the local user. In parallel experience neither user knows the solution, so they need to share their ideas to find a solution together.

Generally, there are two main purposes for remote communication (or collaboration): (1) transferring information (Fussell et al. 2004) and (2) sharing experience and feeling closer (Inkpen et al. 2013). Most previous studies using augmented visual communication cues (Fussell et al. 2004; Kirk et al. 2007; Alem et al. 2011; Gauglitz et al. 2012; Gauglitz et al. 2014b; Sodhi et al. 2013) focused on transferring information for improving task performance in remote expert systems.

Since these studies focused on effective information transferring from a remote expert to a local novice, the communication is mostly one way. However, Griffin (2005) noted that communication between people is not only one-way, but also frequently involves bilateral interpersonal communication.

For user discussion and sharing ideas in the parallel experience collaboration, the required communication is bilateral. The parallel experience

with the use of visual communication cues has not been studied as much as the other two types. Therefore, this dissertation explores parallel experience collaboration.

A parallel experience should include both the sharing experience and feeling closer. In good face-to-face communication, a person considers their partner's situation to decide when they will start to talk, and how they will control the speed and tone of their words. Understanding one's partner is a key factor of successful collaboration. Likewise, since the 'Task Space' is what both users watch, having a similar understanding of the task is also required for parallel experience. In this thesis, the term 'connectedness' is defined as how well users understand each other and have a similar level of task understanding.

In short, this dissertation explores how the interface helps users to have better parallel experience collaboration while using augmented visual communication cues in a shared task space view.

### **1.3 Thesis Structure**

The rest of this dissertation includes chapters on related work, pilot studies, the research approach, and four user evaluation studies, followed by discussion and conclusion chapters. The related work is presented in the second chapter and mainly explores previous research in 1) augmented visual communication cues, 2) view sharing systems, and 3) communication mechanisms. In the pilot studies (Chapter 3) the author explores the use of visual communication cues in a remote expert system. The main purpose of these pilot studies was to gain a deeper understanding of the previous studies, and to explore how visual communication cues such as pointing and drawing annotations can aid in transferring spatial information.

In Chapter 4, the author describes the research approach including experiment design, software architecture, and the relationship between the experiments. The first user study explored how effectively the visual

communication cues are used, and how they have an effect on parallel experience collaboration (Chapter 5).

The second user study (Chapter 6) explored the design of the shared view on the remote user end by comparing manual freeze and automatic freeze conditions to a non-freeze condition, and exploring which freeze function resulted in better parallel experience collaboration. The third study (Chapter 6) addressed the design of the shared view on the local user end for the issue of managing the remote user's annotation while the local user conducted his/her own work, by comparing three conditions with different visual notification methods: (1) no notification, (2) red box notification, and (3) both freeze notification.

The fourth study (Chapter 7) was about navigable independent views for remote users. Kuzuoka et al. (1994) noted that an independent field of view is one of the main factors affecting remote collaboration, so this study explores the use of independent views by comparing independent and dependent views in two different types of remote collaboration: remote expert and parallel experience.

During this PhD, several contributions have been achieved. The author introduces one of the earliest prototypes for anchoring drawings in the real world without any previously known tracking data or visual markers (described in Chapter 4). From the first study (Chapter 5), the author found that quick and easy inputs are required in the use of visual communication cues in the parallel experience collaboration, and four user states (*collaborating together, playing in parallel, passive, do-it-alone* states) during remote parallel experience collaboration. From the second study (Chapter 6), the author found that auto freeze function is required to support quick and easy inputs for the use of drawing annotation interface in the parallel experience collaboration. In the third studies (Chapter 6), the author introduced a better user interface for the local user, and found that a visual notification that minimizes interruption and alerts when the remote user draws

an annotation helps a local user to manage the collaboration. From the fourth study (Chapter 7), the author found that the benefit of independent view in both remote expert and parallel experience collaborations but the benefit is more prominent in the parallel experience collaboration than in the remote expert collaboration.

Additionally, from the four user studies, the author introduces a communication model for the parallel experience collaboration with the use of the visual communication cue.

The overall goal of this thesis is to suggest better interface for helping users to have better parallel experience collaboration in the use of the visual communication cue.

## **2 Related Work**

This chapter describes relevant prior research in remote collaboration. First, previous studies in remote collaboration using visual communication cues (such as pointers, annotation, or hand gestures) are described. Second, previous research of shared view interfaces is discussed. Thirdly, to see the communication in remote collaboration, previous research in communication mechanisms for collaboration are explored.

### **2.1 Augmented Visual Communication Cues**

Most of the previous remote collaboration studies using augmented visual communication cues such as a pointer, annotation, and hand gesture, focused on transferring spatial information from a remote expert to a local user. To the best of our knowledge, there is no previous parallel experience collaboration study using the technology anchoring the visual communication cues in real world.

#### **2.1.1 Early Works**

One of the earliest works for sharing visual communication cues for remote task space collaboration is “VideoDraw” (Tang & Minneman 1991) in which users could share drawings on a paper by projecting a live image of the paper onto each other’s monitor. Ishii et al. (1994, 1992) introduced “ClearBoard” (see Figure 2.1) in which users could see their partner’s facial expressions and achieve eye contact while sharing drawing annotations. However, these systems required the same static video conferencing set up at both ends and did not support anchoring the user’s virtual annotations in the real world.



Figure 2.1: ClearBoard supports remote gaze awareness while sharing annotation (Ishii et al. 1992; Ishii et al. 1994)

One of the early works displaying and overlaying virtual annotations on the real world was demonstrated in 1995 (Rekimoto & Nagao 1995). They developed a device called NaviCam that was a handheld screen with a small video camera attached to detect real-world visual markers. NaviCam allowed the user to view the real world together with information generated by the computer or another user (see Figure 2.2). However, the user's view needed to include a color barcode to trigger the context sensitive information, and the user's drawing was not anchored in the real world but attached on the user's screen.



Figure 2.2: NaviCam generates information about a recently arrived book (left). The use of NaviCam for remote assistance (right) (Rekimoto & Nagao 1995)

### 2.1.2 Laser Pointer

Other early studies used a laser pointer attached on a remote controlled robotic arm as an abstract physical representation of a hand. Kuzuoka et al. (1994) developed the 'GestureCam' in which a remote user could point at any object in the shared area (see Figure 2.3). A laser pointer and a camera were attached on the GestureCam, and the remote user could control the point of view and the laser pointer with another actuator that was synchronized to the orientation of the GestureCam. In a user study, they found that a pointer could

show what a remote expert was interested in to the local user.

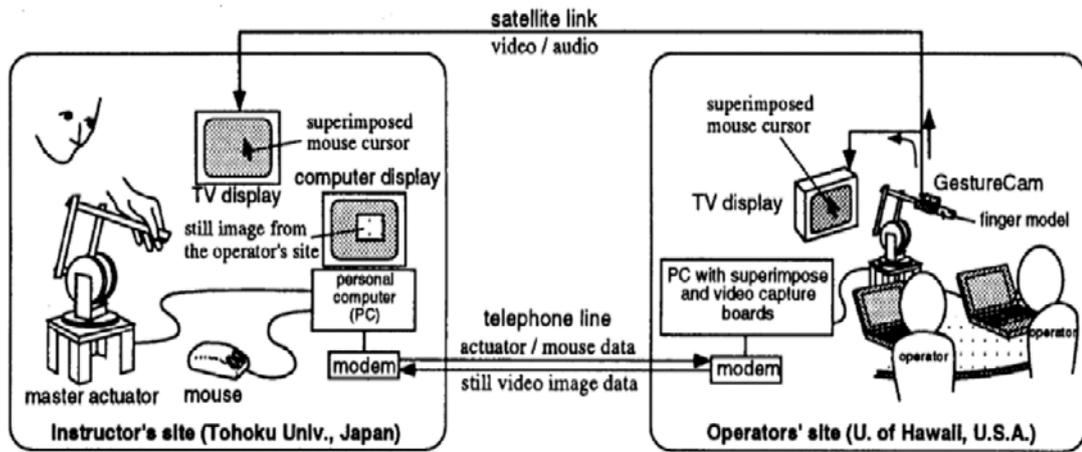


Figure 2.3: System architecture of GestureCam (Kuzuoka et al. 1994)

Sakata et al. (2003) and Kurata et al. (2004) continued the study of a laser pointing cue. They developed the Wearable Active Camera/Laser (WACL) system that involved the worker wearing a steerable camera/laser head. A remote user could control the robotic head on which a camera and a laser pointer were attached; changing view by rotating the robotic-head and pointing by turning the laser pointer on and off (see Figure 2.4).

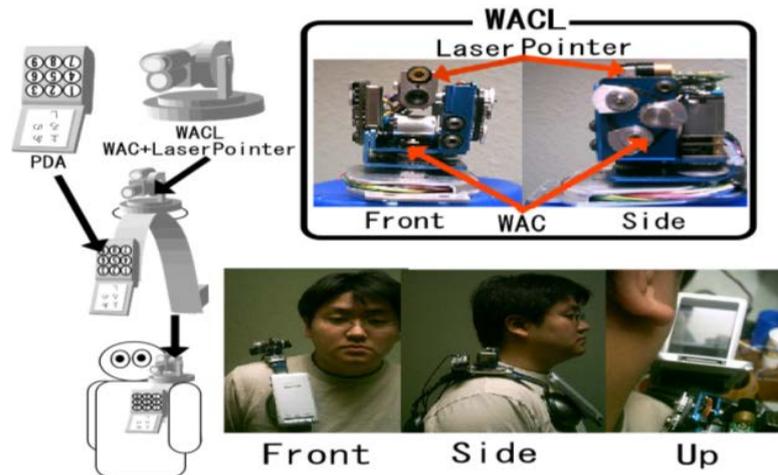


Figure 2.4: Wearable Active Camera/Laser (WACL) system (Sakata et al. 2003)

### 2.1.3 Using Graphical Pointer and Annotation

Fussell et al. (2004) compared pointer and drawing annotation interfaces with a remote collaboration system, named DOVE (see figure 2.5). Their first

study demonstrated that a graphical pointer interface did not improve performance compared to no visual communication cue condition while an annotation interface did. In their second study, the annotation interface where annotations were automatically erased after several seconds improved the task performance compared to a normal annotation interface. However, their system did not include visual tracking, so the pointer and annotation were not anchored in the real world but attached to the screen of the users' view. This meant that if the viewpoint of the shared view was changed, the virtual annotations would no longer be aligned to the physical objects they were drawn on.

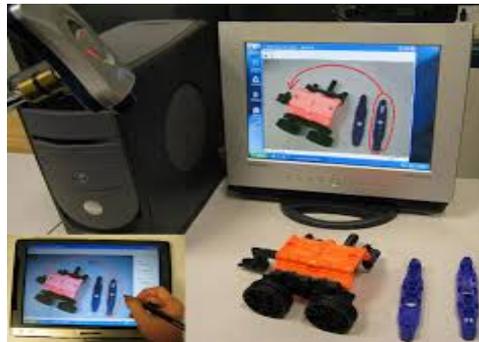


Figure 2.5: 'DOVE' system, a drawing tool on the Helper's tablet PC (left front insert) and on the Worker's monitor (right) (Fussell et al. 2004)

To solve the issue of misalignment, Kato and Billingham (1999) applied Augmented Reality (AR) techniques to allow a remote user to anchor virtual annotations on a set of tracked AR markers (see Figure 2.6). Moreover, remote collaborators could be shown as a virtual video in the user's real workspace for a live video conferencing experience.



Figure 2.6: AR conferencing for remote collaboration. A remote collaborator can draw annotations on the marker based image panel (left) and remote collaborators can be arranged in the view by positioning the real markers they are attached to (right) (Kato & Billinghurst 1999)

Gauglitz et al. (2012) introduced a mobile AR remote collaboration system (see Figure 2.7). They used markerless AR tracking to allow users to anchor a world-stabilized fixed shape annotation (either 'X' or 'O') in a shared view of the real world. They compared their system to video-only and static marker conditions (screen-stabilized) and found that using the world-stabilized annotations significantly improved the task performance compared to the video only interface and is significantly preferred by users. However, their prototype was limited in generality because the annotation tool was designed to provide pointing information with fixed shape annotations rather than free form annotation drawing, and it was compatible only with planar scenes.

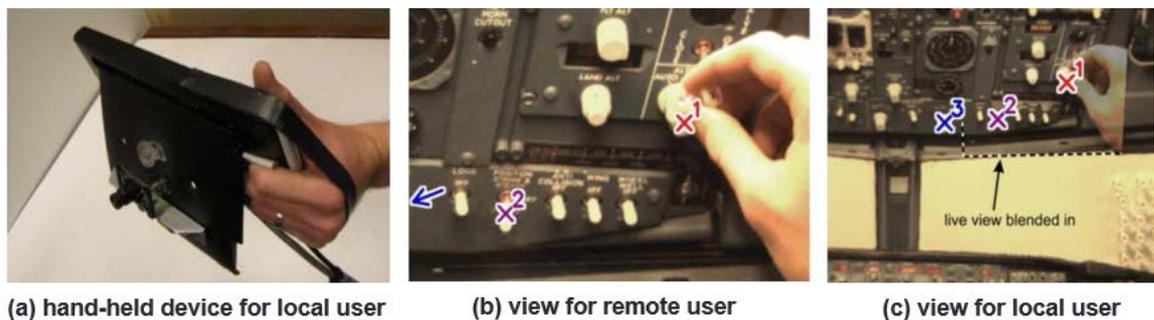


Figure 2.7: The device for a local user (a) and views for remote and local users (b and c) (Gauglitz et al. 2012)

Gauglitz et al. (2014b) presented another system for live mobile remote collaboration (see Figure 2.8). A remote user could communicate via AR

spatial annotations that were immediately visible to the local user. Their system used real-time visual tracking and modeling, thus it did not require any preparation or instrumentation of the environment. In a user study, participants significantly preferred the condition with world-stabilized annotations over the annotation on screen condition, and the world-stabilized annotation produced better performance than the video only condition. However, their system still provided pointing information rather than sketching or drawing.



Figure 2.8: Screenshot of the remote user's interface when he/she has an independent view from the one that the local user has (Gauglitz et al. 2014b)

Recently, Gauglitz et al. (2014a) presented a study exploring how to display drawings in 3D space. They compared four different virtual annotation conditions: (1) Spray paint, (2) Minimum depth paint on a plane orthogonal to viewing direction, (3) Median depth paint on a plane orthogonal to viewing direction, and (4) Dominant surface plane (see Figure 2.9). In the user study, the dominant plane condition was the most preferred.

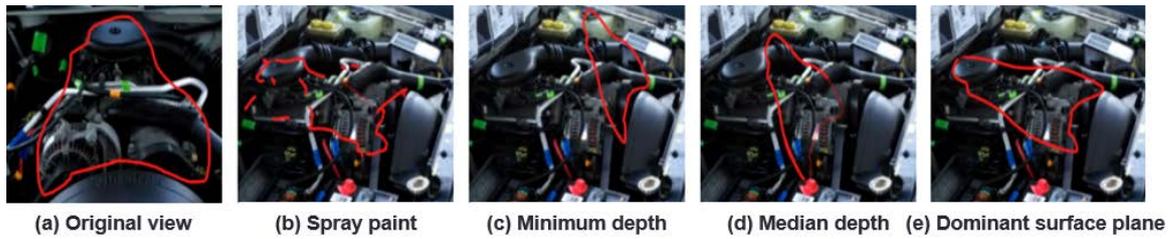


Figure 2.9: Depth interpretations of 2D drawings. (a) shows the viewpoint from which the drawing was created, and (b-e) show different depth interpretations from a second view point (Gauglitz et al. 2014a)

Gauglitz et al. (2012, 2014a, 2014b) mainly focused on how to place annotations in the real world rather than studying the role of visual communication cues. Their first two studies (2012, 2014b) mainly focused on how to anchor the fixed form annotation ('O' or 'X') while providing an independent view for a remote user. In his third study (2014a), their system started to support drawing annotation (none fixed form annotation) and they concluded that participants preferred the dominant surface plane to display drawing annotations on. Similar to these studies the proposed system in this thesis also supports anchoring visual communication cues in the real world and displaying drawing annotation on the dominant surface plane; however, its focus is the use of visual communication cues; a pointer and drawing annotations (not fixed form annotation), in parallel experience collaboration.

#### 2.1.4 Hand Gesture

Other researchers studied hand gesture as a visual communication cue. Early hand gesture systems used projection based 2D artifacts. 'VideoDraw' (see left of Figure 2.10) provides a 'virtual sketchbook' that allows participants to see each other's drawings and projects video of their accompanying hand gestures directly onto one another's drawing space (Tang & Minneman 1991). Tang et al. (2004) introduced 'VideoArms' (see right of Figure 2.10), which captures hand gesture and digitally recreates it as 'virtual embodiments' on the other end screen. However, these systems shared a virtual environment or a designated area.

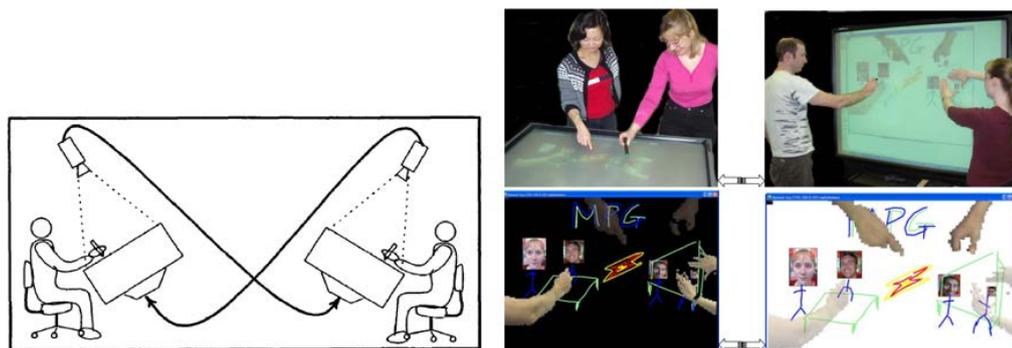


Figure 2.10: Schematic diagram of 'VideoDraw' (left) (Tang & Minneman 1991) and VideoArms in action showing two groups working together (right) (Tang 2004)

Kirk et al. (2007) designed a remote gesturing system that captures a top-down view of the shared workspace and projects the top-down view on the other side (see Figure 2.11). In the experiment, they compared a handgesture condition to a voice only (with a live video) condition. The result suggested that the gesture communication cue improved performance. Almeida et al. (2012) also presented a system that shared hand gesture on the shared workspace, conducted pilot studies with the system, and had similar results to Kirk's. However, the users of these systems had to stay in a fixed place because the systems did not support mobility.

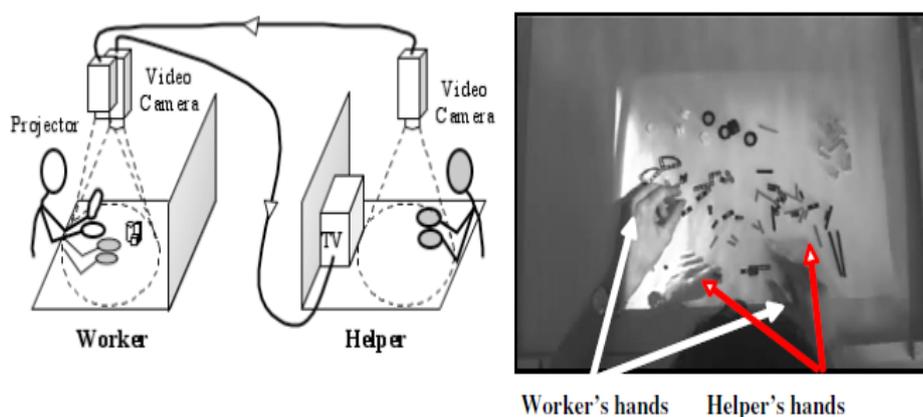


Figure 2.11: System architecture (Left) and worker's view (Right) (Kirk et al. 2007)

Alem et al. (2011) developed a mobile augmented reality system for remote assistance (see Figure 2.12). The system takes a live video displayed on a touch-enabled display. The hand gesture of a remote user above the display is also taken as a live video and displayed on a local worker's Head Mounted

Display (HMD).



Figure 2.12: Remote Tele-Expert System (Alem et al. 2011)

Sodhi et al. (2013) presented a mobile remote collaboration system with two active depth cameras mounted onto a monopod (see Figure 2.13). With a camera facing the front of a local user, the system captures and shares a 3D local physical environment with a remote user. With another camera, the hand gesture they make is captured and mapped into the other users' environment. In their preliminary user test, users found the task easy and effortless to perform with their prototype.

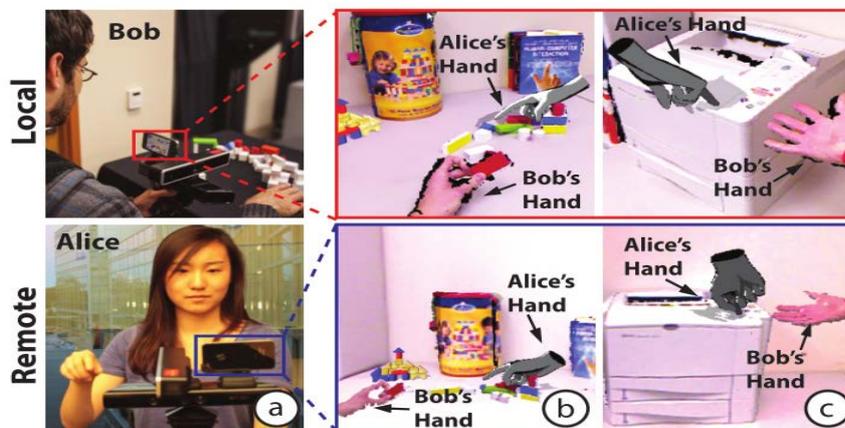


Figure 2.13: Bob shares the local work space with Alice and manipulates objects according to Alice's instructions. Alice provides an instruction with her hand while watching the shared environment displayed on a mobile device (Sodhi et al. 2013)

### 2.1.5 Summary

Augmented visual communication cues have been studied for decades and have been used for transmitting spatial information from a remote expert to a local user. Three visual communication cues (a pointer, annotation, and hand

gesture) have been the main cues studied for information transmission, and the interface with a richer communication cue was always preferred.

Table 2.1 categorizes the previous studies according to the types of remote collaboration they focused on and whether or not the previous systems support anchoring visual communication cues in the real world with robust model free tracking.

Table 2.1: Classifying the previous studies according to a type of remote collaboration they focused (A) and visual tracking system they used (B).

A	B	Robust model and marker free tracking	Model or marker based tracking	Non-tracking system
<b>Remote Expert</b> (Both users have different amounts of information, Asymmetric)	Gauglitz et al., 2012 Gauglitz et al., 2014a Gauglitz et al., 2014b Sodhi et al., 2013		Almeida et al, 2012 Kato & Billinghurst, 1999	'NaviCam', Rekimoto et al. 1995 'GestureCam', Kusuoka et al. 1994 'DOVE', Fussell et al., 2004 Kirk et al., 2007 Alem et al, 2011
<b>Parallel Experience</b> (Both users do not have any information at first, Symmetric)	This is the area the author focused			'VideoDraw', Tang and Minneman, 1991 'Clearboard', Ishii et al., 1993, 1994

This table shows that there have been few studies of visual communication cues in a parallel user experience task. The visual communication cues would be used not only when a remote user has all the information (remote expert), but also when both users do not have any information (parallel experience). Moreover, as Carey (1989) observed, if a local user merely unpacks and accepts the message from a remote user, then the remote user is not only

sharing ideas but also controlling the local user. This neglects the local user's ability of interpreting local context and having ideas to solve the task, and prevents meaningful discussion with the remote user.

## **2.2 Sharing Task Space View**

Visually sharing context is an important factor in remote collaboration, and the shared view is where users play or work together. How much information the users take from the shared view has an effect on the remote collaboration, so there are several studies about the sharing task space view while using visual communication cues.

Some researchers (Yarosh et al. 2010; Inkpen et al. 2013) studied the effect of view sharing methods in parallel experience collaboration that do not employ augmented visual communication cues. In this section, the author describes studies of view sharing methods with the support of visual communication cues.

### **2.2.1 Early Works**

Kuzuoka et al. (1994) developed 'GestureCam' that enabled a remote user to have an independent field of view from the local user's view, and could navigate and have an independent view by physically controlling the local end camera. They stated that dependency of the shared view could be a factor in reducing effectiveness for transmitting spatial information. Sakata et al. (2003) and Kurata et al. (2004) followed Kuzuoka's suggestion and developed the controllable independent view system.

Additionally, Kuzuoka et al. (1994) conducted two pilot studies with GestureCam to see the difference between narrow and wide fields of views. While it was hard for the remote user to find objects with the narrow field of view, they could see many more objects with the wide field of view and had better situation awareness.

### **2.2.2 Head Mounted Camera View & Scene Camera View**

Some remote collaboration systems (Alem et al. 2011; Rekimoto & Nagao 1995) use head mounted cameras that send a view of the local user's environment to a remote collaborator. This may help both users to have an identical view and focus on the same area of the work space, but the remote user can only see what the local user is looking at.

Fussell et al. (2003) compared five different ways of sharing task space: collocated side-by-side, audio-only, head-mounted camera, scene camera, and scene plus head cameras conditions. From a user experiment, they found that participants had the best work quality with the side-by-side condition, and they had better work quality with the scene camera which showed top-down view covering all the task space than with the other conditions. Kirk et al. (2007) also adopted the top-down scene in their remote collaboration system. However, the scene-based task space view had a low level of portability because the camera for sharing view needed a support fixture.

### **2.2.3 Independent View**

To provide independency on remote user view from local user view, Alem et al. (2011) added four initially empty additional windows beside a main live video window, where at any moment live video could be copied as a still image. With this system a remote user could select and save still images showing an important moment of the live video into the four additional windows, and use these later by choosing and displaying them in the main large window.

Similarly Gauglitz et al. (2012, 2014b, 2014a) used a freeze function (a still image view). Remote users could have an independent still image view with the freeze function, enabling them to draw an annotation without the worry of the local user suddenly changing viewpoint. Additionally, within the still image view (or frozen view), the remote user could navigate the scene independently from the local user's current view. To construct the navigable

frozen independent view, the system continually stored and stitched new key frames with their associated camera poses from the live video stream.

A similar view sharing system which supports a navigable independent view was introduced by Kasahara and Rekimoto in 2014 (See Figure 2.14). A remote user could control his or her view by hand-pointing interaction in the still image mode. The system also stores and stitches several still images to construct a navigable independent view. With a hand gesture detection device, the system acquires a pointed image among the stitched still images and the pointed image moves to the center of the whole view.

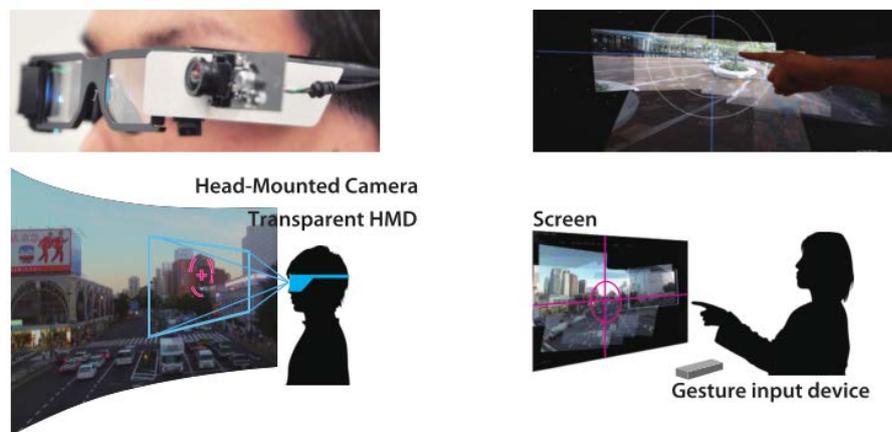


Figure 2.14: JackIn, A remote user can see and control the view point of the integrated wider scene around a local user and point remotely into the scene. (Kasahara & Rekimoto 2014)

Sodhi et al. (2013) implemented a system supports independent viewpoint (see Figure 2.13). The system supported physical navigation so remote users could navigate the local area as if in the local area; for example, a user turning right had the accordance view of the turning movement. This would provide intuitive and natural navigation in the local environment to a remote user, but the field of view is comparatively small and it is difficult to have navigational techniques beyond physical movement such as quickly covering large distances or adopting a bird's eye view.

While Gauglitz et al. (2012, 2014b, 2014a), Kasahara (2014), and Sodhi et al. (2013) implemented real time shared view construction, Reichherzer et al.

(2014) developed a system for sharing prestructured panorama images between a high level of portability by using a Google Glass head mounted display and a tablet computer (see Figure 2.15). The panorama view was controlled by touch screen interaction for a remote tablet user, or by body and head movement (e.g. turning head to focus on a section) for a Google Glass local user. To show where the other user was watching, the system included a context compass, which increased the shared engagement.

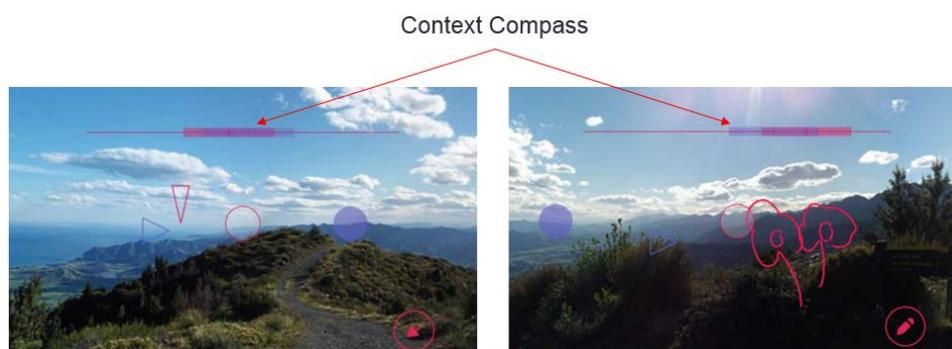


Figure 2.15: The context compass in Social Panoramas (Reichherzer et al. 2014)

Seo et al. (2014) introduced a real-time panorama video streaming system to provide immersive and realistic content. In a user study, they compared a panoramic view to a narrow dependent view. Participants preferred the panoramic view because it allowed the remote participants to choose wherever they wanted to watch.

#### 2.2.4 Summary

The previous studies described a variety of methods to provide view independency to a remote user, and remote users have preferred to have an independent view from the local user view in remote expert collaboration. The independent view usually come with navigable and large field of view to provide freedom of the view changes.

### 2.3 Communication mechanisms

Collaboration directly relates to human communication, so this section describes previous remote communication models.

### 2.3.1 Early Works

One of the earliest and most influential communication models was introduced by Shannon and Weaver (1949), who mainly considered information transmission from a source to a destination as the basic mechanism of communication (see Figure 2.16).

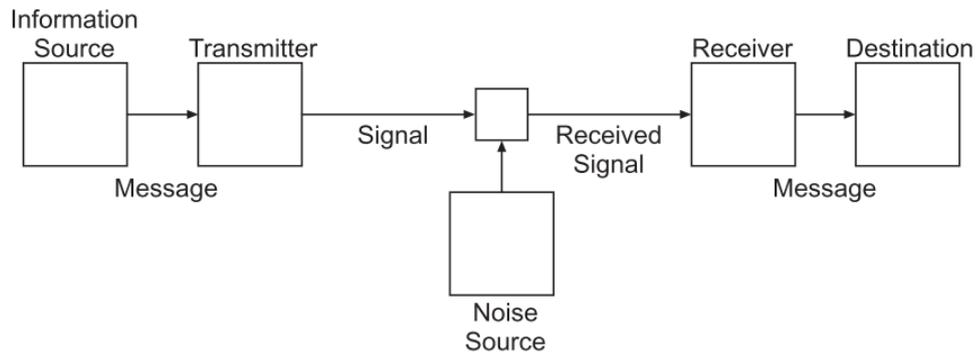


Figure 2.16: Shannon and Weaver's transmission model of communication (Shannon & Weaver 1949)

In this model, the information source and destination are typically persons with a reason for engaging in communication. The Transmitter and Receiver include the interface of the proposed media that the users use. The Transmitter accepts a message and transforms it into a signal that can be sent to a remote receiver and converted back into the message. The interfering noise causes loss of information and the message at the destination is always a distorted version of the message sent.

Shannon and Weaver (1949) designed this model from an engineering perspective and only considered the noise from a physical connection such as telephony circuit. However, Roszak (1986) pointed out that Shannon's model is incapable of distinguishing messages with valuable meaning from pure nonsense because it is not concerned about the meaning of messages; only transmitting messages.

### 2.3.2 Interpersonal Communication

Several researchers have described human communication using a model of

interpersonal communication. For example, Gouran et al. (1994) characterized interpersonal communication as: (1) there are a few participants involved, and the participants are in close physical proximity to each other, (2) there are many sensory channels used, and (3) the feedback is immediate. Griffin (2005, page 52) defined interpersonal communication as the process of creating and sharing meaning between people. Hauber et al. (2008) stated that interpersonal communication is dynamic and bilateral, so cannot be conceived as a linear and literal transmission of information from a sender to a receiver.

Devito (1998) developed an adapted version of the Shannon-Weaver model, incorporating several modifications to take into account interpersonal communication (see Figure 2.17). A sender encodes an idea into a message and sends it to a receiver, who decodes the message and tries to reconstruct an image of the idea. Semantic noise occurs while encoding an idea into a message and when decoding the message. Physical noise occurs while the message is in the channel. Encoding is the process of representing an idea in an appointed form of communication tool (such as language) and decoding is the process of interpreting the message, so encoding and decoding are referred to as complex cognitive processes instead of technical processes. Communication is successful if the image created at the receiver side corresponds to the initial idea of the sender.

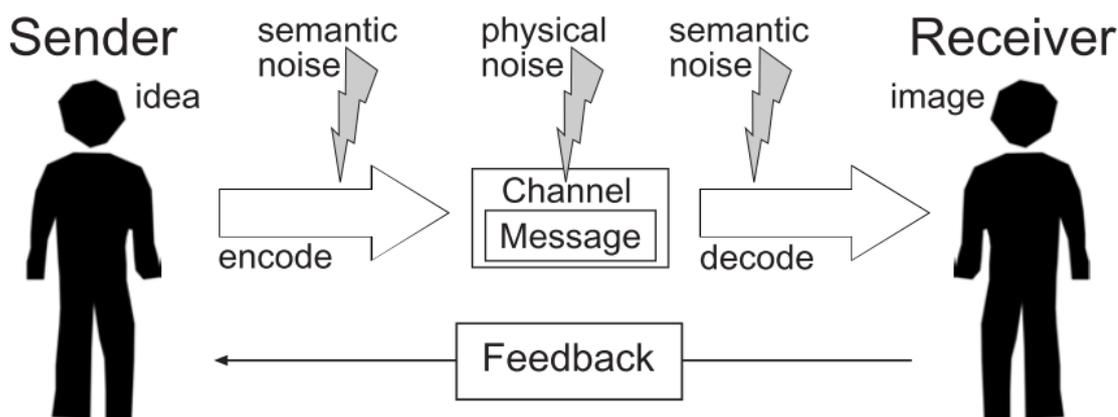


Figure 2.17: The modified Shannon-Weaver model for interpersonal communication, simplified from Devito (1998, page 12).

### 2.3.3 Collaboration

Collaboration generally refers to the process of interdependent activities to achieve a common goal (Hauber, 2008). It includes the processes of creating shared meaning through interpersonal communication, which serves as the basic joint action that leads to solving common tasks. A number of interpersonal communications for creating many shared meanings constructs collaboration.

According to Hauber (2008), collaboration is driven by the exchange of messages (see Figure 2.18). Collaborators can be both senders and receivers, so they encode/send messages and receive/decode messages. During collaboration, every collaborator develops his or her own mental situation model which represents how much he/she understands the current state of the shared task and the other user. Hauber emphasized that the largest part of collaborative effort for collaborators is to expand each other's situation models and maximize the overlap. In short, collaboration would be more efficient if collaborators had a higher level of understanding both of the shared task and their collaboration partners.

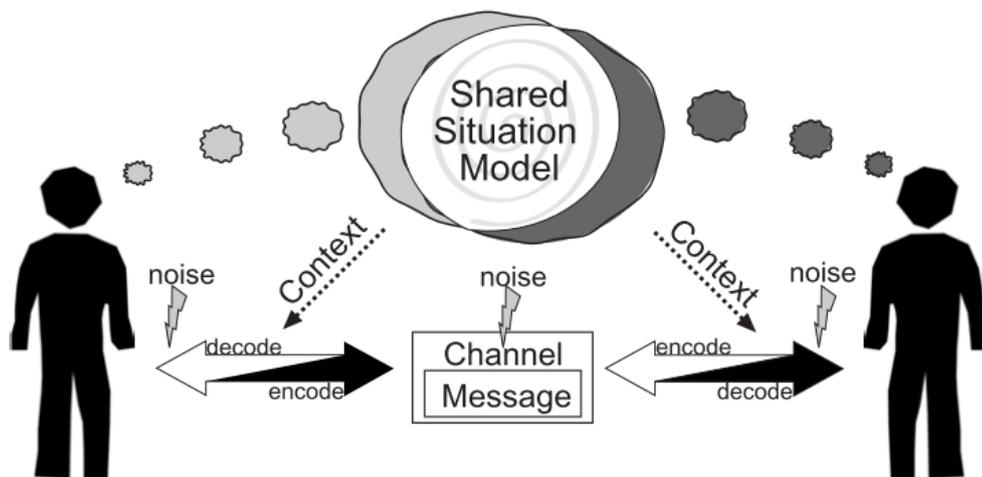


Figure 2.18: Hauber's communication model for remote collaboration

Clark and Brennan (1991) also report that collaborators try to have least collaborative effort to achieve shared understanding. For example, when

transferring information through verbal communication they try to use the least number of words; or when they indicate a place or an object, they just point with their hand rather than verbally describing the place or object with dozens of words.

However, the least collaborative effort rule is not directly applied to verbal communication. In verbal communication, the words used could be different according to the place, the time, people spoken to, and so on. This means that the least collaborative effort rule is not the only consideration for communication or collaboration. Additionally, people do not consciously try to find the least collaborative way because that in itself causes additional cognitive effort; rather they unconsciously present messages in easily applicable way. The easy way of presenting messages does not always lead to the least collaborative effort. Therefore, the author presupposes that richer visual communication cues (which can transfer more information and have less collaborative effort) are not always the choice of people for remote collaboration.

#### **2.3.4 Summary**

In this section, the author described previous communication models and their mechanisms of collaboration. Previous researchers emphasized the aspect of human-to-human communication (interpersonal communication) rather than merely transferring messages as in the engineering perspective. In interpersonal communication there are semantic noises in message encoding and decoding steps and the communication is bilateral. The communication in collaboration is interpersonal communication, and Hauber (2008) additionally emphasizes the situation model that represented the level of understanding of the shared task and collaboration partners.

#### **2.4 Chapter summary**

This chapter reviewed previous research in remote collaboration. Section 2.1 described previous studies with different visual communication cues for

remote collaboration. There were mainly three types of visual communication cues studied, pointers (Sakata et al. 2003), drawing annotations (Fussell et al. 2004; Gauglitz et al. 2012), and hand gestures (Kirk et al. 2007; Sodhi et al. 2013). These showed that richer communication cues which can transfer more information can lead to better performance for remote expert collaboration (for example, an annotation interface shows better performance than a pointer interface).

Section 2.2 reviewed previous literature about sharing the local environment. Most previous studies emphasized the independency of the remote user's view from the local user view. Alem et al. (2011) used still images (independent view) for keeping the moment of a live video view. Gauglitz et al. (2012, 2014a) implemented a navigatable independent view with the freeze function that allows a remote user to pause a live video for adding annotation in the shared view. The author also studied the still images showing interesting moment of view in the first study (Chapter 5), freeze functions in the second study (Chapter 6), and large navigatable independent view in the fourth study (Chapter 7).

In section 2.3, the author reviewed an early communication model (Shannon & Weaver, 1949) and the modified models of it applied to interpersonal communication (Devito 1998, Hauber et al. 2008). Shannon's model was designed from an engineering perspective so the concept of interpersonal communication was added to include the aspect of a human-to-human communication by Devito. Hauber modified Devito's communication model for remote collaboration, emphasizing the importance of understanding the current states of the shared task and collaboration partner (situation models). This author studied parallel experience collaboration that requires bilateral communication like interpersonal communication, and interfaces that may help collaborators have better understanding of the shared task and collaboration partners.

### **3 Pilot Studies of Remote Expert**

Before studying parallel experience collaboration, the author initially studied the remote expert collaboration with visual communication cues, a pointer and annotation. The main purpose of this study was to gain a deeper understanding of the use of visual communication cues in the remote expert collaboration. With this study, the author designed a communication model of a remote expert.

#### **3.1 Introduction to the Pilot Studies**

The author explored how pointer and drawing annotation communication cues can be used under different view sharing techniques in a remote expert system, and conducted two pilot studies with two tasks. The first study compared four conditions: Pointers on Still Image (PS), Pointers on Live Video (PV), Annotation on Still Image (AS), and Annotation on Live Video (AV). There were three key results:

- 1.** The annotation interface is more effective to transfer object position and orientation information than the pointer interface.
- 2.** Live video becomes more important when quick feedback is needed.
- 3.** Users provide more inputs with the pointer cues than with the annotation cues.

In a second follow-on study, the author compared the conditions PV and AV with a more complicated task, with two key results:

- 1.** The pointing interface requires good verbal communication to be effective for orientation information
- 2.** The drawing annotation interface should to be erased before the previous annotations cause a messy view.

In the remainder of this chapter the author describes the prototype interfaces and the two pilot studies in more detail.

## **3.2 Prototype**

The author developed a pair of software applications to support remote collaboration: (1) An Android tablet application and (2) a desktop Personal Computer (PC) application (see Figure 3.1). The Android application is for a local user and it streams live video or transfers images to a PC application. It allows the local user to position a pointer or draw annotations on the shared live video or images. The personal computer (PC) application is for a remote user and it receives and displays the live video or transferred images. It allows the remote user to point or annotate on them with mouse inputs.

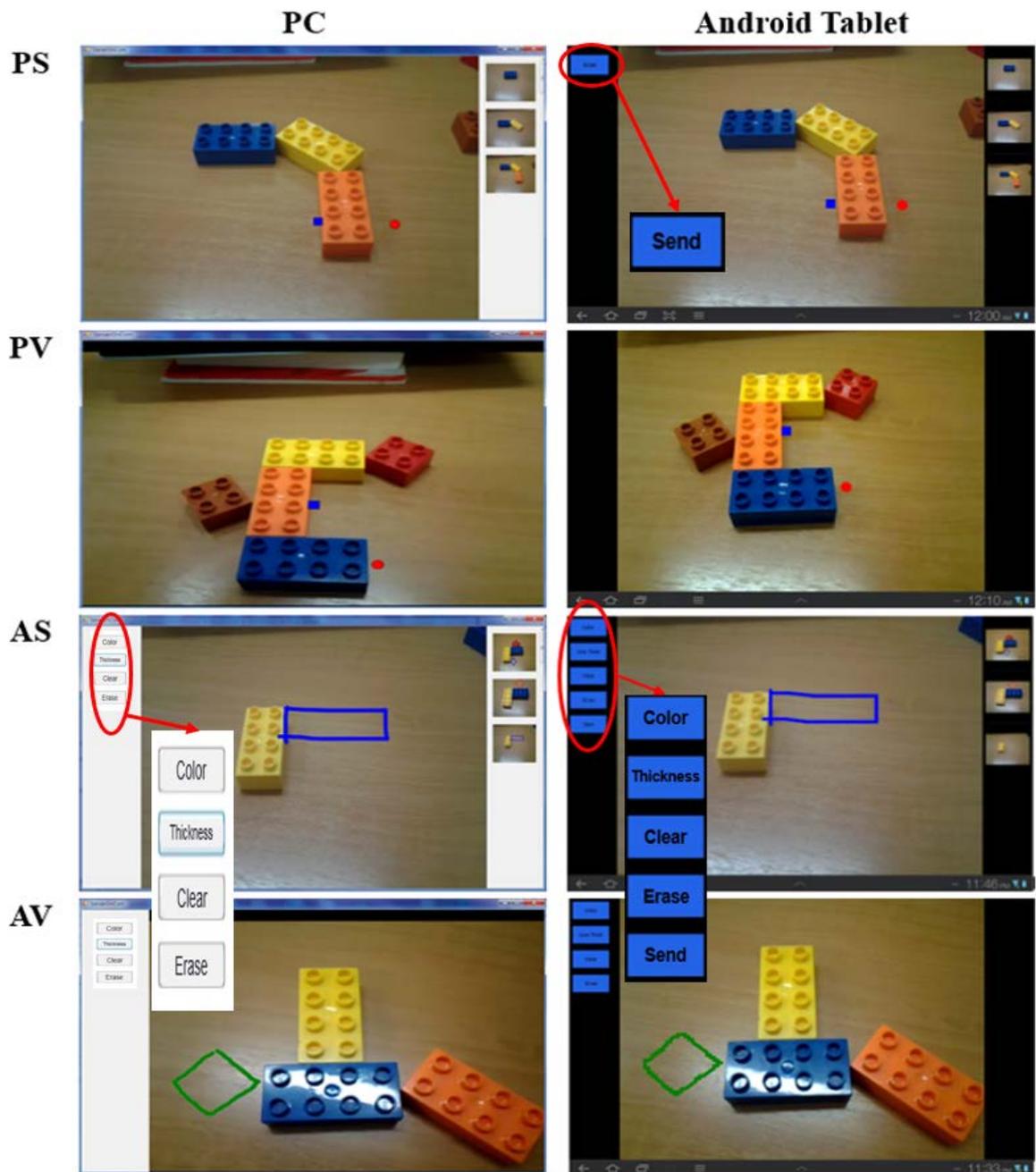


Figure 3.1: Four conditions in pilot studies, Pointers on Still Image (PS), Pointers on Video (PV), Annotation on Still Image (AS), Annotation on Video (AV)

In still image conditions PS and AS, a local user (a tablet user) takes a picture and the image is immediately shared with a remote user (a PC user). The shared still images are listed on the right side of the both screens, and any user can choose one of them to share. In live video conditions PV and AV, the

tablet application streams a live video captured with an embedded camera to the PC.

For using visual communication cues, the author adopted the touch screen interaction for a local tablet user and the mouse interaction for the remote PC users. When using pointers in conditions PS and PV, each user controls a coloured pointer (red for a local user and blue for a remote user) on the shared still image or live video. The controlled pointers are immediately synchronized between the views of the local and remote systems. With the annotation interface in conditions AS and AV, both users can draw annotations on top of the still image or live video, and the annotation is also immediately synchronized between the local and remote systems.

Additionally, there are ‘Clear’ and ‘Erase’ buttons for the annotation interface. Clicking the ‘Clear’ button erases all annotations and clicking the ‘Erase’ button erases the most recently drawn annotation. Each user can also choose the colour and thickness of annotation with the button ‘Colour’ and ‘Thickness’.

### 3.3 Pilot Study Design

With the prototype systems, the author conducted two pilot studies. The two studies had the same user study design except for the level of task difficulty (see middle and right of Figure 3.2) and the conditions compared in them.



Figure 3.2: Experiment set up (left) and a given instruction picture to a remote user for the first (middle) and second pilot studies (right).

In two pilot studies, the participants were able to talk to each other while sitting back to back in the same room so they couldn’t see each other (see left

of Figure 3.2). The remote user (expert) had an instruction paper showing how to construct a block model and the goal was for the remote user to tell the local user how to complete a model with 10 Lego pieces (5 small squares and 5 large rectangles).

Both pilot studies had the same procedure and data collection. Each study began with a pair of participants answering a questionnaire asking demographic information about them, and being informed about the purpose of the study. Before starting each experimental session, participants had two minutes training session with a condition. The author recorded the performance in terms of task completion time and the number of mistakes made (determined as when a local user placed a piece in a wrong position or orientation and the local user changed their focus to another area or piece), collected user preference, and took video recordings of the laptop and Android tablet screens.

### **3.4 First Pilot Study Results**

The four conditions (described in 3.2) were compared with four pairs of participants. Each pair used all four conditions with four object manipulation tasks, and the order of conditions was counterbalanced using a Latin square design. All participants were HIT Lab NZ (Human Interface Technology Laboratory New Zealand) students who had previous experience of using Augmented Reality applications.

From this pilot study, participants were able to complete the tasks faster with an annotation cue than with a pointer cue (see Figure 3.3). Live video helped participants complete the tasks quicker than still images. A two-way repeated measure ANOVA for the task completion times showed significant main effects between annotation and pointer interfaces ( $F(1,3)=54.7074$ ,  $p=.0051$ ) and between a live video and still images ( $F(1,3)=26.0584$ ,  $p=.0145$ ). There was no significant interaction between the two factors ( $p = .42$ ) The interesting result was that there was little difference in

performance between the live video and the still images when using annotations, but a big difference between them when using pointer cues. Additionally, the communication was mostly one way from the remote user to the local user, transmitting information through verbal and visual communication cues, and the local user only sometimes giving feedback with a few words such as ‘Okay’ or ‘Yes’. None of the local participants used any of the visual cues (drawing annotation or pointer manipulation), choosing instead to directly show their manipulation.

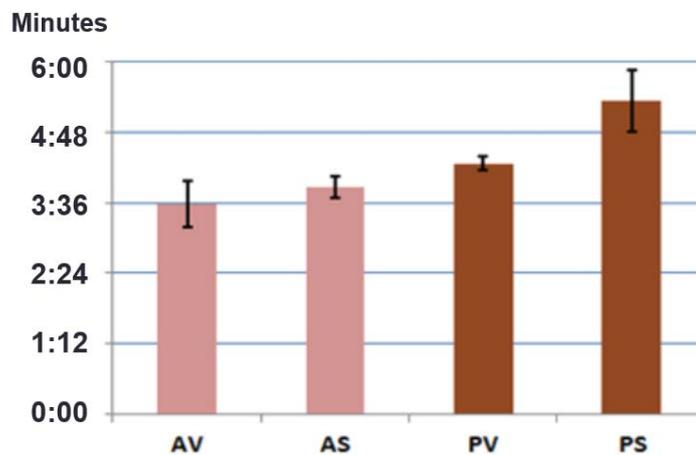


Figure 3.3: Average performance time in the first pilot study

Table 3.1: Average number of errors with the interfaces

	AV	AS	PV	PS
<b>Mistakes</b>	2	1.75	7.75	4.25

More mistakes were made when using the pointer cue than annotation cue (see Table 3.1,  $F(1,3)=14.83116$ ,  $P=0.031$ ). Figure 3.4 shows the average number of mouse movements for drawing annotation or positioning a pointer by remote participants. The results show they had fewer inputs with annotation than with pointers ( $F(1,3)=33.3959$ ,  $p=0.010$ ).

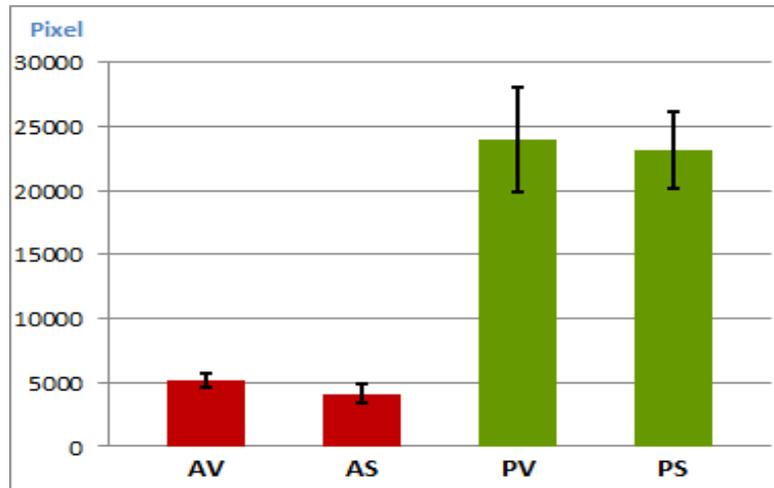


Figure 3.4: Average of mouse cursor movement by remote participants

When asked to rank the interfaces, participants preferred the annotation interface over the pointer interfaces, and video over still image (Friedman test  $\chi^2(3) = 14.850, p = 0.002$ ).

An interesting point was that the remote users always transferred only one block manipulation information with the pointer interface, but they sometimes transferred more than one block manipulation information at a time with the annotation interface.

### 3.5 First Pilot Study (Discussion)

The results show that participants had significantly better performance with the annotation than with the pointer, and they preferred the annotation cue to the pointer cue. This agreed with the result of Fussell's study (2004), and appeared to be because the drawing annotation is a richer communication cue than the pointer. For one block spatial information, the remote participants drew the 2D shape of a block at the desired position and orientation with one time annotation drawing, but they kept tracing the shape of a block repeatedly with a pointer. This would be the reason why participants had significantly more inputs with the pointer interface than the annotation interface.

This difference could be because of the permanency of drawn annotations

and volatility of the pointer. The annotation was still displayed after the drawing was complete, so the remote participant did not need to draw the shape of blocks repeatedly. However, the pointer only showed a single point of interest at a time, so the pointer could not include 2D shape information at a time. The local user needed to figure out the shape of a block based on the movement of the pointer and remember the position and orientation of its shape. Moreover, the permanency of the drawing annotation would allow the remote user to transfer more than one block information with the annotation interface.

When local participants manipulated blocks the remote users immediately checked if the manipulation was correct or not with the live video. However, with the still image interface, the local user needed to take a picture for sharing the updated state of the work space with the remote user. In other words, the still image interface required more user inputs and took more time to share the workspace. This would be one of the main reasons why participants preferred the live video and had better performance with it.

### **3.6 Second Pilot Study Design & Result**

In the first pilot study, participants preferred the annotation interface over the pointer interface. The main reason for this was that the annotation is a richer communication cue than the pointer. With the pointer interface the remote users gave more inputs (more mouse dragging interaction) and the local user was required to figure out the spatial information (position and orientation) from the movement of a pointer.

The second pilot study addresses the question of what will happen if the task requires more inputs (more mouse dragging interaction) with a more difficult, complex task. Figure 3.5 shows one of the tasks for the second pilot study that includes placing blocks on top of each other and rotating them about any axis.

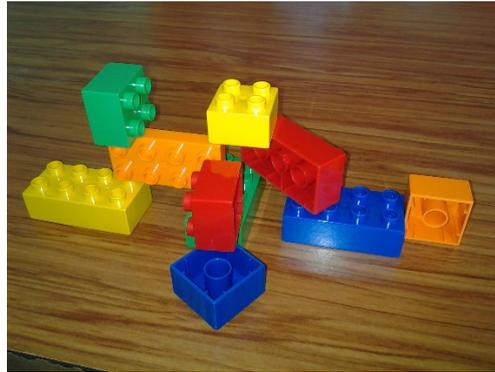


Figure 3.5: A task for the second pilot study

Since the still image interface was found not as effective for sharing the task space in the first pilot study, the author only compared two conditions, ‘Annotation on Video’ (AV) and ‘Pointers on Video’ (PV), in the second pilot study. Another four pairs of participants were recruited and each pair used both conditions with two object manipulation tasks. The order of conditions was counterbalanced. All participants were HIT Lab NZ students who had previous experience with augmented reality applications. The experimental procedure was the same as in the first pilot study as described in Section 3.3.

The author did not find much difference in user preference and task completion time between the two interfaces. Among the remote participants, two preferred the annotation cue and two the pointer cue. Among the local participants, three preferred the annotation cue and one the pointer cue. Using the Wilcoxon Signed-Rank Test ( $\alpha = .05$ ), no significant difference in preference between the conditions ( $Z = -.707$ ,  $p = .480$ ) was found.

In terms of task completion time, no significant difference was found between two conditions (See Table 3.3,  $Z = 0$ ,  $p = 1.0$ ). Corresponding to Fussell’s study (2004), which showed that a cleared shared view resulted in faster task completion time with automatically erased annotation condition, this result showed that participants generally had faster completion time when they had a clearer shared view, with more uses of the ‘Clear’ function.

Table 3.3: Average task completion time (in seconds) and the number of times the 'Clear' function was used while using annotation.

	<b>AV</b>	<b>PV</b>	<b>Clear</b>
<b>Group 1</b>	274	370	20
<b>Group 2</b>	298	358	18
<b>Group 3</b>	373	242	1
<b>Group 4</b>	295	275	8

Verbal communication was observed as a key factor for conveying the orientation information in the pointer condition. This is due to the fact that it could be hard to represent the orientation information with the pointer interface. Local participants in groups 1 and 2 sometimes did not understand the verbal descriptions well, so they made more mistakes and needed more time to finish the task while using the pointer interface.

The three main steps were mostly conducted sequentially (see Figure 3.6). First, the remote user read the instruction paper. Second, both users communicated with verbal and visual cues. Third, the remote user tried to understand the local environment while checking the local user's object manipulation. When they made a mistake in communication, they restarted from the second process. In collaboration, the processes were recursively conducted in sequence. The feedback from the local user to the remote user was not counted as a main step because it did not happen in every communication sequence.

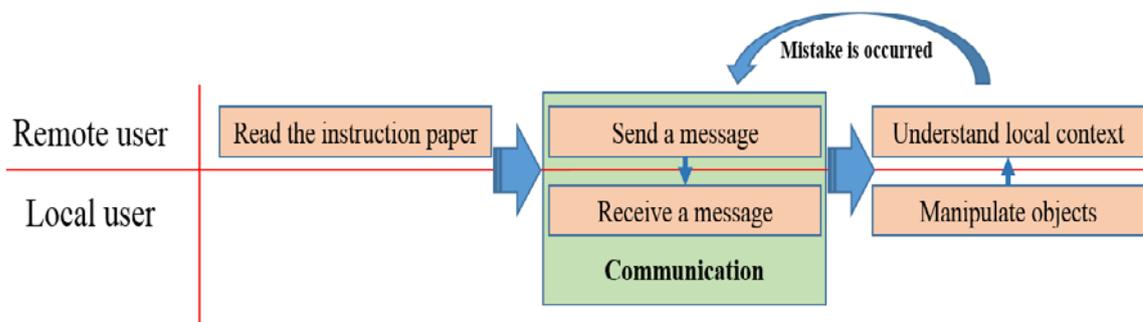


Figure 3.6: The sequence of collaboration in remote assistance with a live video

### 3.7 Second Pilot Study (Discussion)

The remote participants in groups 1 and 2 usually cleared their drawn annotations after checking if their partner had manipulated the block correctly. The other two groups kept the previous annotation and overlapped their drawings (see Figure 3.7). With the overlapped annotation, the local worker found it difficult to figure out the proper position and orientation of the block from new drawings. This would have led to being less effective in the use of annotation in groups 3 and 4.



Figure 3.7: Stacked annotations with the 3D task

### 3.8 General Discussion

The author recruited four groups of participants for each pilot study. This was a small number of participants to derive elaborate results; however the purpose of these pilot studies was to gain deeper understanding about remote collaboration, and the results of these pilot studies correspond with those of Fussell's study (2004).

The annotation interface was better for transferring spatial information than pointer interface. The position and orientation of real world objects can be represented and understood with several lines and the annotation interface is capable of drawing and displaying those lines. In contrast, the pointer interface cannot display several lines at a time; just a single point. Thus the annotation interface would produce less noise in encoding and decoding the

spatial information than the pointer interface. This may lead to fewer mistakes in collaboration with the annotation interface than with the pointer interface.

However, inappropriate use of annotation in the second pilot study led to more mistakes. Since the annotations overlapped, local users were required to distinguish new annotations from old ones. They got confused, misunderstood annotation messages, and made mistakes.

In our pilot studies, none of the local users suggested any object manipulation but passively followed the instruction from the remote users. Even though the local users sometimes gave verbal feedback such as ‘okay’ and ‘yes’, or questioning about the instruction when they did not understand well, the communication remained mostly one way. In remote expert collaboration, a remote user has a solution for the common goal but the local user does not. This asymmetric condition inspires the remote user to actively instruct the local user and the local user mostly to follow the instruction.

In designing a communication model for the remote expert collaboration based on Hauber’s communication model (described in Figure 2.18), one user is a remote expert and the other a local operator (see Figure 3.8). Unlike interpersonal communication, the communication is mostly one way from the remote user to the local user. There are verbal and visual communication channels, and encoding and decoding processes are conducted based on their local context understanding. Both communications are complementary for sharing meaning, and give effect to each other. Every encoding and decoding includes ‘noise’, and navigating the shared work space (i.e. selecting and sharing a still image in the first study) requires the remote user’s inputs.

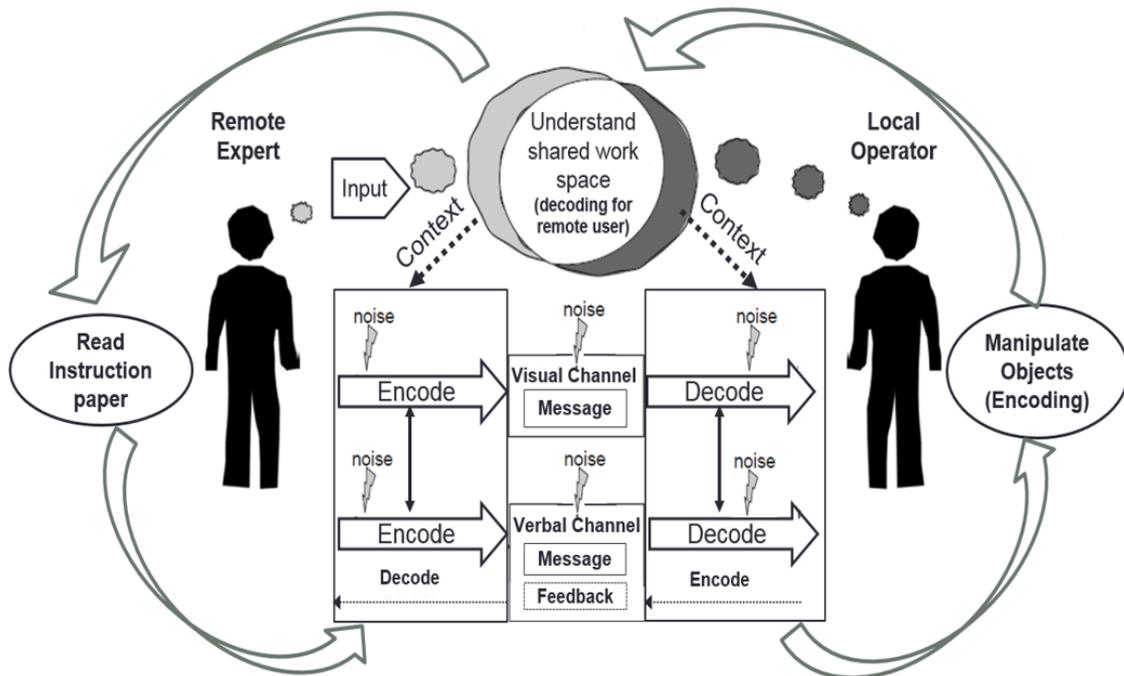


Figure 3.8: The communication model for remote expert

While considering the sequences described in figure 3.6, the author added ‘Read instruction paper’ on the remote end, and ‘Manipulate object’ on the local end. Since the steps were in order, the author also added arrows showing the order. This round was recursively conducted during collaboration until users achieve their goal.

Understanding the current state of a workspace can be the alternative of the ‘*situation model*’ (in Hauber’s communication model) in remote expert, because the situation is remote expert collaboration and occurs in the shared workspace.

### 3.9 Chapter Summary

In this chapter, the author conducted two pilot studies to gain a deeper understanding of the use of visual communication cues and different view sharing methods in remote expert collaboration. The first pilot study compared four conditions: PS, PV, AS, and AV. The second pilot study was followed with two conditions, PV and AV with a more complex task. The

findings from the two pilot studies are listed below:

- 1) There was better performance with the annotation interface than with the pointer interface in the aspect of transmitting spatial information.
- 2) A live video was better than still images for providing fast feedback from the remote user to the local user.
- 3) Even though the annotation interface was more suitable for transferring spatial information, overlapping several annotations caused misunderstandings in the local users and reduced the performance and connectedness in remote expert collaboration.
- 4) In remote expert collaboration, the communication was mostly one way from a remote user to a local user except for the local user's simple feedback.

Through the pilot studies, the author also revised Hauber's communication model by adding two more steps 'Read instruction paper' and 'Manipulate objects'.

## **4 The Research Approach**

Through the pilot studies in chapter three, the author gained a deeper understanding about the use of visual communication cues for the remote expert collaboration, which is a specific type of collaboration where a remote user has a solution for the common goal and the result cannot be generalized. In this Ph.D., the author explores the remote collaboration where both users do not have a solution for the common task.

This research started with the doubt that a richer communication cue is better for parallel experience collaboration. The first main user study (Chapter 5) was designed to determine whether the richer communication cue; annotation, is better than the pointer cue in parallel experience collaboration. The second and third studies (Chapter 6) were about the user interfaces on remote and local ends for better parallel experience collaboration while using drawing annotations. In the fourth study (Chapter 7), the effect of an independent view in parallel experience was investigated.

For these studies, the author developed software prototypes and designed a set of experiments. This chapter will address the relationship between the studies, a base system architecture of prototypes, and the experimental design. Details of each interface design used in the studies will be addressed in the corresponding chapters of each study.

### **4.1 Relationship between studies**

In the first study the author explored the use of pointer and annotation interfaces, and compared the interfaces with a voice only condition (as a baseline) in parallel experience collaboration. Participants preferred the pointer interface to the annotation interface because it was quick and easy to use and had good compatibility with a live video.

The second study was about the freeze function in the use annotation interface. If a local user changes the viewpoint of a shared live video while a remote user is drawing an annotation, the annotation is drawn at the wrong

place. To overcome this issue, previous researchers (Gauglitz et al. 2012 and 2014b; Kasahara & Rekimoto 2014) suggested manually freezing a live video with additional user inputs, and the author followed this practice in the first study. However the participants in this study still drew many annotations in a live video.. The author suggested using an auto freeze method to solve the issue and compared it with a manual freeze and a no freeze condition.

The third study was for solving an issue whereby the local users sometimes unintentionally missed new drawings in the first and second studies. The author investigated interfaces to help the local users have better notification management to handle the issue. The author compared three conditions with different visual notification methods: (1) no notification (as the baseline condition), (2) red box notification, and (3) both freeze.

In the first three studies, the remote user always shared the same view with a local partner (dependent view). However, independent and navigable views for remote users were considered one of the key factors for improving performance in remote expert collaboration. The author designed the fourth study to explore whether or not having an independent navigable view is a positive factor in parallel task experience. In this study, the author compared four conditions: Remote expert collaboration with Dependent view (RD), Remote expert collaboration with Independent view (RI), Parallel experience collaboration with Dependent view (PD), and Parallel experience collaboration with Independent view (PI).

## **4.2 System Architecture**

To study the visual communication cues and shared views in the remote collaboration system, the author implemented several user interfaces. Each pair of interfaces (for local and remote users) was designed according to the purpose of the study by extending a base system. In this section, the base system architecture is described (see Figure 4.1) while the interfaces for each study are explained in corresponding chapters. The base system architecture

provides two end point communication tools; one for the local user and the other for the remote user. The local users share a (task space) live video with a remote user, and the remote user can draw or display visual communication cues on the shared live video. All user experiments were conducted in a setting where verbal communication could be achieved without a voice chat application (the local user in a room-sized booth and the remote user outside the booth so they could not see each other while having verbal communication).

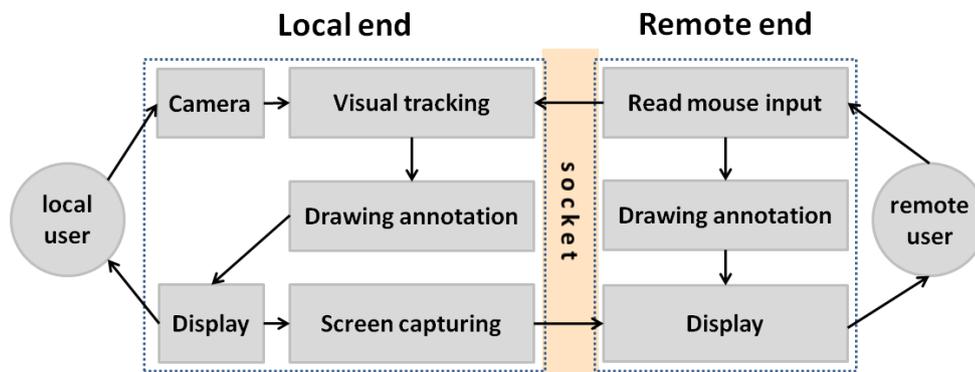


Figure 4.1: Overview of the proposed system structure

A key technology behind the fundamental system is robust monocular SLAM (Simultaneous Localization And Mapping) software, developed by Tan et al. (2013). While the local user watches the live video from the camera, the SLAM software tracks the current scene. The live video is shared with the remote user, who can position a pointer or draw annotations with mouse inputs. These inputs are sent to the local end system and the pointer or annotation is anchored in the real world with the SLAM tracking data.

#### 4.2.1 Local End Point

The local end interface was designed to meet the following requirements: 1) It should capture and share the local context with the remote user, 2) the remote user's input should be displayed immediately, and 3) the local user should be able to easily express ideas to the remote user while interacting with or manipulating real world objects.

The base system used a Logitech Webcam C920 (Logitech QuickCam in the first study) to capture a video of the local scene with a resolution of 640 by 480 at 24 frames per second. To display the captured scene, the system used a Vuzix Wrap 1200DX-VR head mounted display (HMD) (Figure 4.2; right). The camera was mounted on the front of the HMD where it could capture at the first person's view. The device is light enough to wear and provides hands free movements, so a user can make hand gestures or manipulate physical objects. The camera and HMD were connected to a PC (3.4 GHz Dual Core CPU, 16 GB RAM, and NVIDIA GeForce GTX 670 GPU) on which a robust monocular SLAM software was running to visually track the scene from the live video.

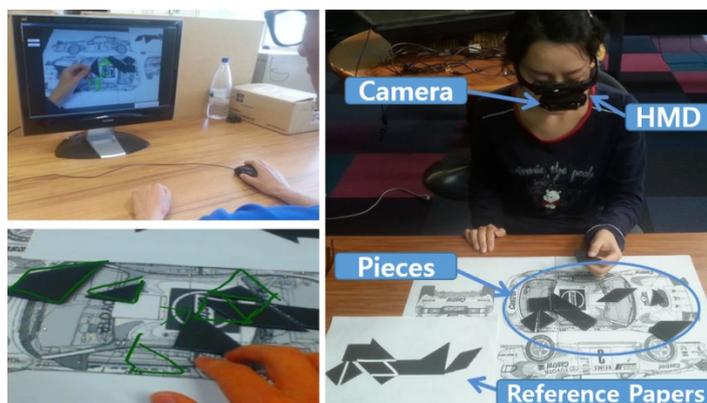


Figure 4.2: The local and remote users' environments. A local user wears an HMD with a camera to share the view (right), while a remote user draws annotations on the shared view with a mouse (top left). An example of drawing (bottom left)

The SLAM tracking allowed the base system to work in an arbitrary physical scene without any prior knowledge, and kept robustly tracking even with more than half of the scene changed (e.g. when the user moved objects in the scene). At the beginning of the video conference call, the system constructed a map of visual features while the local user showed the space (e.g. on a desk) for a couple of seconds. Once the tracking algorithm recognized the space, the SLAM system kept generating feature points and providing tracking data about the current camera feed.

### 4.2.2 Remote End Point

For the live video, both end systems were connected through a wired Gigabit Ethernet LAN, and the video stream showed little delay; good enough for live communication for the study.

For the remote end interface, the author used a desktop computing environment (see left of Figure 4.2). While watching the local user's object manipulation on a shared live video, the remote user used a mouse interface to position a pointer or draw annotations on the shared view. The mechanisms for these actions is as follows: To operate the mouse interface, a pointer is displayed while a remote user presses the mouse right button down and can be repositioned with the mouse dragging interaction; an annotation is drawn when the mouse left button is pressed down and the mouse dragged over the screen. When the remote user positions a pointer or draws annotations, the system calculates the relative three-dimensional (3D) position of the mouse cursor in the real world.

To calculate the relative 3D position of a pointer or drawing in the real world space, the author used a ray casting method. A ray is calculated with the view projection matrix and the position of the mouse cursor in screen space. Among the feature points generated from the SLAM library, the system selects the three feature points that are close to the ray. These selected points are used to make an invisible plane, and the system checks the collision between the ray and the invisible plane. This 3D collision point is used for placing a pointer or forming a virtual annotation, hence the drawing appears attached to a location or object in the real world. Moreover, since a drawn line consists of several points, the base system saves several 3D collision points to form a line in a list for the annotation interface. The virtual pointer or annotations drawn by the remote user are immediately displayed on both the remote and local ends.

Based on the findings in previous pilot studies, the author added 'Clear' and 'Undo' functions to allow the remote user to erase annotations. All annotations

are erased when the user clicks on the 'Clear' button, while the 'Undo' button erases the most recently drawn annotation.

### **4.3 Experiment Style**

Even if the purposes of the studies varied according to the study topic, the overall goal of the user studies was to design an interface that helps better parallel experience collaboration while using visual communication cues. To evaluate the interfaces, the author designed and conducted user experiments and the common structure and design of these are described in this section.

#### **4.3.1 Experimental Task and Environment**

For designing the experimental task the author considered the features of the prototype system. As the prototype system shares a live video, the task should not require too much camera movement, otherwise the remote users might get dizzy. Moreover, focusing on a parallel experience, the experimental task should require active collaboration between the participants and encourage the use of visual communication cues to share spatial information.

Considering these factors, the author chose a Tangram<sup>3</sup> assembly task (see right of Figure 4.2). A Tangram is a seven-piece puzzle that can be arranged to form different shapes. Since it requires spatial object manipulation, it would encourage the use of a visual communication cue for sharing spatial information. The task also involves an appropriate amount of view movement, where the local user has to search for puzzle pieces and place them together to form a target shape, without too many drastic motions which could discourage using visual annotations and cause dizziness on the remote user's end.

To prevent bias from the participants' previous experience of solving Tangrams, the author created a custom Tangram puzzle with ten pieces of different sizes or shapes from the original design. If a task is too difficult

---

<sup>3</sup> <http://en.wikipedia.org/wiki/Tangram>

participants would not have any ideas to share; if too easy they would not need to communicate as they could easily solve it without sharing their ideas. The level of difficulty was balanced through a pilot test which led us to provide a reference paper with three white lines implying borders between pieces (see Figure 4.2 4.3). Five Tangram puzzle silhouettes with a similar level of difficulty (each solved in about four minutes when pilot tested with five people) were thus created for the experiments.

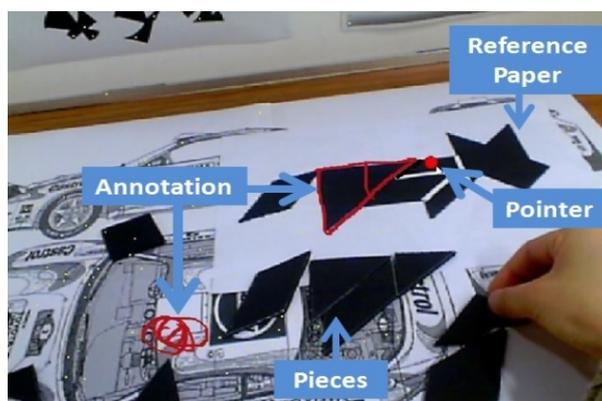


Figure 4.3: An experimental task with a reference paper and puzzle pieces shown in a shared live video

In the experiment, the participants were to solve the Tangram puzzles in pairs while using the prototype system for communication. Both local and remote users were asked to share and discuss their ideas, while the local user directly manipulated the Tangram pieces and the remote user used visual communication cues. The local user sat in front of a table in a small room and the table was prepared as the playing space, with all the materials for the task (i.e. puzzle pieces, a reference paper, and HMD as described in section 4.2 and 4.3). The remote user sat in front of a desk with a personal computer system running the prototype system to communicate with the local user.

### 4.3.2 Experimental Procedure

The experiment began with a pair of participants answering a questionnaire asking for demographic information, and being informed about the purpose of the study. The experiment then continued with a practice session of face-to-

face collaboration with a sample Tangram puzzle. This was to let the participants understand the task, and also to observe their face-to-face collaboration. After trying the sample puzzle, the author explained how the prototype system worked.

The experimental sessions followed after the face-to-face collaboration. During the experimental sessions, pairs were separated and used the prototype system to perform the Tangram assembly tasks under different experimental conditions. The order of the conditions was counter balanced using a Latin square design. Each session consisted of two minutes of training followed by five minutes of solving the Tangram. During the training period, the participants used the prototype system to get familiar with the user interface and communicating with their partners. During the problem solving period, participants performed the experimental task under the given condition. After each session, they filled out a questionnaire asking about their experience with the given condition. Following the questionnaire, a brief interview was held for each participant separately to provide more details about their experience in that session.

After finishing all of the experimental conditions, the experiment wrapped up with final interviews.

All experiments took about 70 minutes for each pair of participants (50 minutes for the experiment and 20 minutes for the interviews) except for the first study, which took 90 minutes (70 minutes for the experiment and 20 minutes for the interviews).

### **4.3.3 Data Collection**

As the task was a remote parallel experience where no-one had the goal solution at the beginning of the task, participants needed to work together, sharing and discussing ideas. This means that task performance could be affected by the ideas that the participants had. Thus the author refrained from measuring task performance and instead used questionnaires and interviews to

collect participants' subjective feedback about their experience of collaboration (every pair of participants completed a model within five minutes with a given condition).

A questionnaire on participants' demographic information was filled out when the experiment began. After each experimental session with a given condition, each participant answered questions with Likert-scale questions (see Appendix) ranging from 0 (Strongly disagree) to 10 (Strongly agree) about their experience of a given interface. For the questions in the questionnaires, the author considered the simple model for remote collaboration by Hauber (2008, see Figure 4). According to this model there are four main factors (both users who encode and decode messages, channels, and the situation model) in remote collaboration. Therefore, this author prepared questions asking about main factors except channels that were verbal and annotation cues for all conditions in each study.

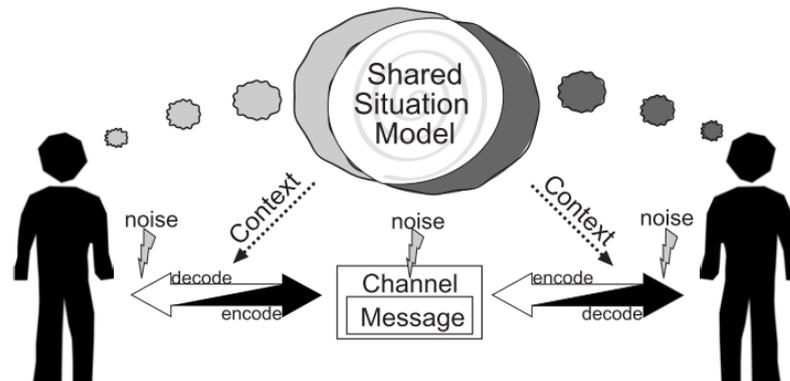


Figure 4.4: Hauber's communication model for remote collaboration

Following the questionnaire, each participant was interviewed to give more details on what he or she liked or disliked about the given interface. Then the participants ranked the conditions, followed by intensive interviews; initially for each of the participants separately and then both partners together.

To complement the collected subjective feedback, the author also logged the mouse interaction of remote users. These logs included the amount of mouse movement while using visual communication cues on the shared views,

and use of the Clear and Undo functions. For further investigation, the author also recorded the screens of the both users.

#### **4.4 Chapter Summary**

In this chapter, the author described how the user studies were prepared. The general structure of this Ph.D. was set with three goals listed below:

- 1) Study the use of visual communication cues in parallel experience (in the first study described in Chapter 5)
- 2) Suggest better interfaces to solve the weak points of visual communication cues for better parallel experience collaboration (in the second and third studies described in Chapter 6)
- 3) Study the use of independent views in both parallel experience and remote expert collaborations (in fourth study described in Chapter 7)

The author described the base remote collaboration system used to design interfaces as conditions in user studies, and conducted four user experiments for the listed study goals. All the experiments followed the experiment design style given in this Chapter, and the experimental tasks and environments, experimental procedure and data were described.

## **5 Study about the Use of Visual Communication Cues**

The pilot studies described in Chapter 3 showed us that the annotation interface transferred more information than a pointer interface and the annotation interface was more effective than a pointer interface in the remote expert collaboration. However, the use of both visual communication cues in the parallel experience collaboration had not been studied yet and overusing of annotation could have caused an issue of local participant misunderstanding in the second pilot study. In this first main user study, the author studied the use of the annotation and pointer interfaces in parallel experience.

The experiment compared three video-conferencing conditions with different combinations of communication cues: (1) only live video (including voice communication, as the baseline condition), (2) voice + pointer, and (3) voice + annotation.

This study is novel in a number of ways: (1) comparing pointer and annotation visual communication cues integrated with a mobile AR system using robust tracking (2) investigating augmented visual communication cues contributing to user connectedness and the sense of being together, and (3) focusing on parallel experience with complex object manipulation.

### **5.1 User Study Design**

This user study was built upon the base prototype system and experiment style described in chapter 4. The main independent variable was the type of interface used, and there were five sessions: face-to-face session, three sessions with three experimental conditions, and a session for free exploration with the prototype system in which both visual communication cues were available.

In addition to comparing visual communication cues, the author also compared using a HMD with using a handheld display (HHD) as a display device for a local user. The HHD was a Microsoft Surface Pro which was

light enough to hold so that the local user could make free movements, just as with the Vuzix Wrap 1200DX-VR HMD. The type of visual communication cues was a within subject independent variable, while the type of display was a between subject independent variable. The author did not use the internal camera of the HMD in order to make sure that both conditions had the same system performance in the user study; instead choosing to use the same camera for both displays. The camera and the HMD were directly connected to the PC, and the HMD received the video stream from the PC through a dedicated Wifi (IEEE 802.11n) connection to display it on the screen.

The still image interface was added as another option for sharing the local view in addition to a live video. Drawing annotations may appear in a wrong place if the local user changed the view point while the remote user was drawing. This was not because of the problem in the tracking system (note that the prototype system was using a robust tracking technique [35]), but more of a systematic problem where the movement of the view resulted in displacing the drawing point in the real world. To overcome this problem, the author adopted the previous solution; still image interface (manual freeze) that takes snapshots of the shared video on the remote end so that the remote users can draw and point on the snapshot images instead of the live video.

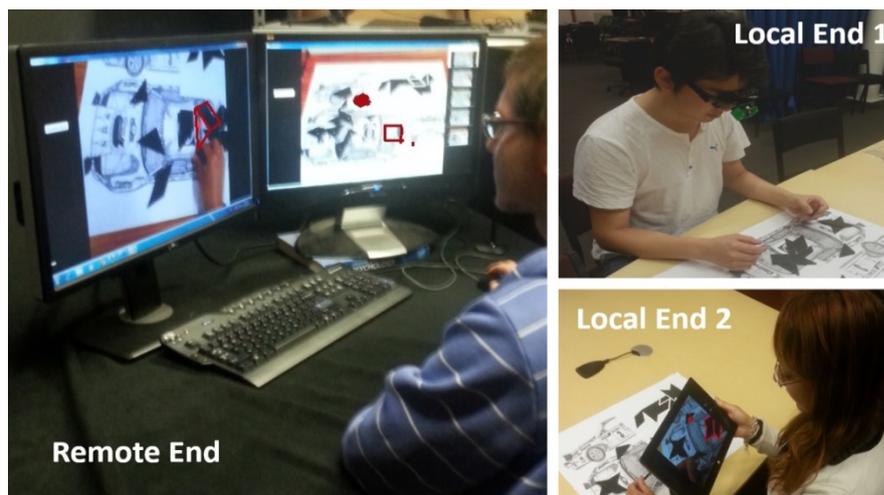


Figure 5.1: Prototype system used by a remote user (Left) and a local user using a HMD (top right) or a HHD (bottom right).

The two monitors on the desk (see left, Figure 5.1) show the live video stream (left) and the current snapshot image (right). The user can draw or point on both the live video and the snapshot image (note that the local end user always sees the live video). The pointing and drawing on the still images synchronize immediately in a local user view as they do on the live video. To take a snapshot image of the shared live video, the remote user has to click on a snapshot button or double click on the live video whenever they want. The snapshot images are then added on a list at the right end of the screen, and the user can revisit them as needed by clicking on them.

After each of the experimental sessions for the different conditions, each pair of participants answered a questionnaire with six questions (see Table 5.1) on a Likert-scale from 0 (Strongly disagree) to 10 (Strongly agree).

Table 5.1: Questions in the questionnaire after each session

Q1	I enjoyed assembling a model.
Q2	I felt connected with my partner.
Q3	I was able to focus on the activity (building models).
Q4	I felt I was there with my partner (or my partner was here with me).
Q5	I was able to express my idea properly.
Q6	I easily understood what my partner was expressing.

Since there are four essential factors (both users who encode and decode messages, channels, and the situation model) to achieve the remote collaboration (Hauber, 2008), the author prepared the questions asking about main factors except channels that were verbal and annotation cues for all three conditions. Since the situation model is the solving an object manipulation task, the author had questions (Q1 – question 1 and Q3) asking how much they enjoyed and could focus on the task. Both users could be a message sender or receiver and encode or decode messages, so Q5 was prepared for sending/encoding messages and Q6 for receiving/decoding messages. Since the situation model also includes understanding a collaboration partner, Q2

and Q4 were prepared.

Since the experimental conditions are not mutually exclusive, another section of free exploration was prepared for five minutes. In the session, participants were allowed to use both pointer and annotation interfaces freely as they performed the experimental task.

## 5.2 Results

The author recruited twenty four participants (in pairs) who had already been using a video conferencing system regularly (more than once a month). All recruited pairs knew each other well as friends or family. There were 17 males and 7 females, ranging in age from 15 to 33 years old ( $M = 24.8$ ;  $SD = 5.2$ ). Among the twelve participants who played the role of a local user, half of them (in groups 1, 2, 6, 7, 11, and 12) used the HMD and the others used the HHD.

The main findings from this user study are listed below:

- 1) Adding a visual communication cue increased the level of feeling connectedness and being together.
- 2) Both users preferred the pointer cue rather than the annotation cue because it was quick to communication and had good compatibility with the live video.
- 3) There were four user states in parallel experience collaboration: *collaborating together*, *playing in parallel*, *passive*, *do-it-alone* states while there were two user states in face-to-face collaboration: *collaborating together*, and *playing in parallel* states.

The author analyzed the participants' answers of the questionnaires that included the six Likert-scale questions about how the given interface helped remote parallel experience collaboration, and analyzed the post-experimental questionnaire asking participants to choose the most and least preferred conditions.

To compare the Likert-scale ratings between the three conditions, the author ran the Friedman test ( $\alpha = .05$ ). For those results showing a significant difference between the three conditions, post hoc tests for pair wise comparison were conducted using the Wilcoxon Signed-Rank Test with Bonferroni correction applied ( $\alpha = .0167$ ).

Comments in the interviews and log information of the remote participants' activity ( $N = 12$ ) were also analyzed.

### 5.2.1 User Enjoyment & Focus on the Task

For the user enjoyment and focus on the task, figure 5.2 shows the results comparing the three conditions using the Likert scale ratings. First, the author analyzed the results of the local ( $N=12$ ) and remote participants ( $N=12$ ) separately, then the combined results of all participants ( $N=24$ ). Participants could focus on the task well when they had good communication while keeping the liveness of the shared video.

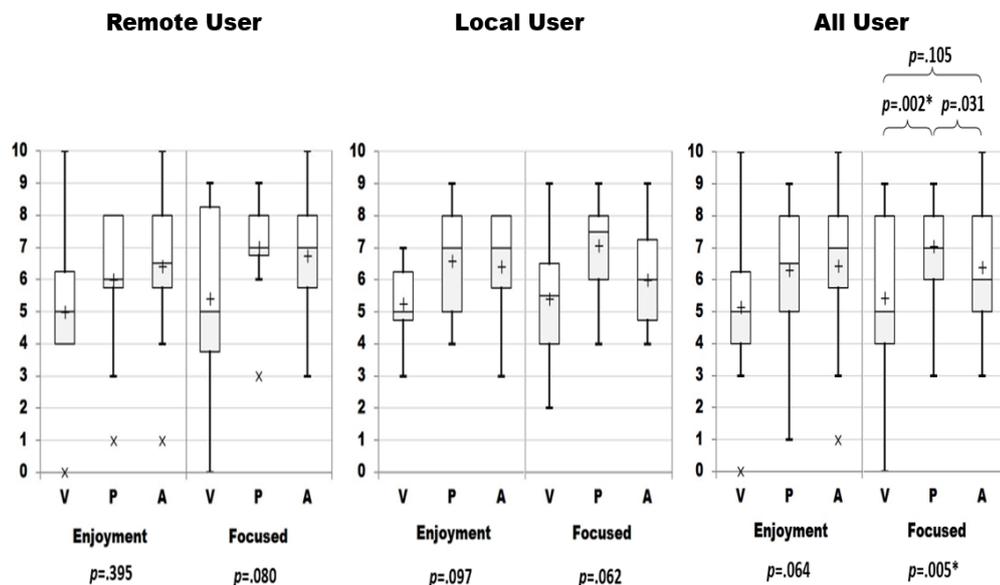


Figure 5.2: Results of Likert scale ratings from participants for the question about how much they felt they enjoyed or focused on assembling task (0: strongly disagree ~ 10: strongly agree; \*: statistically significant difference, +: mean, x: outlier, V: video only, P: pointer, A: annotation).

There was no significant difference between the three conditions according to the Friedman test for the ratings of user enjoyments (remote:  $\chi^2(2)=1.857$ ,  $p=.395$ ; local:  $\chi^2(2)=4.667$ ,  $p=.097$ ) and of user focus on the task (remote:  $\chi^2(2)=5.056$ ,  $p=.080$ ; local:  $\chi^2(2)=5.571$ ,  $p=.062$ ). However, comparing the ratings from all users ( $N=24$ ), the author found the significant difference in Friedman test outcomes for the user focus ( $\chi^2(2)=10.564$ ,  $p<.005$ ). Pair wise comparison with post hoc tests showed a significant difference between the baseline (video with voice communication,  $Mean = 5.41$ ,  $SD = 2.51$ ) and pointer conditions ( $Mean = 7.04$ ,  $SD = 1.51$ ) ( $Z=-3.025$ ,  $p=.002$ ), but it did not in other pair wise comparisons. This suggests that participants had a better level of focus on the task with the pointer interface than with the baseline condition.

The participants reported the good compatibility between a pointer and a live video as a benefit of the pointer. R2 and R8 respectively commented that *"When (the view of the) live video is moving, I can follow video with pointer but not with annotation"* and *"Using a pointer means not moving your eyes and mouse pointer to the (snapshot on the) second screen to draw annotations. Essentially, I could focus on both watching the live video and using the pointer at the same time."*

Several participants pointed out the inappropriate combination between the annotation and the live video. R2, 7, 9 and 10 reported that it was difficult to draw on live video. L1 and L2 commented that they were not able to move the view while the partner was drawing. On the other hand, annotation on snapshot images had problems of not showing scene updates and having a fixed viewpoint. R5 and R11 reported that they got confused with where they drew the annotation when returning to the live view from the snapshot. R11 said *"Sometimes things have changed and what I've drawn (on a still image) was not matched (with the scene on the live video because the local partner manipulated the puzzles before R11 returned back to the live video)"*. The problem also affected the experience of the local users. When the local users

were looking at an area distant from the snapshot image, they were not able to find the annotation immediately. L8 reported "*When he drew on a different area, I had to find it.*"

The author compared how much time they spent between the live video view and the snapshot image view (Wilcoxon Signed-Rank Test with  $\alpha = .05$ ) while using a pointer interface or an annotation interface. The author simply used each view's window having focus (for input) as the criterion to decide which view the user was focusing on. Although the measurement would have been more accurate by using eye tracking, the measure the author took was reasonably accurate as the user kept managing the input focus to use the interface. The results showed that participants spent significantly more time focusing on the live video view than on the snapshot image view with both pointer ( $Z=-3.059, p=.002$ ) and annotation conditions ( $Z=-2.510, p=.012$ ).

However, in comparing the percentage of how much time focusing on the live video view between pointer and annotation conditions, there was significant difference (Wilcoxon Signed-Rank Test :  $Z=-2.090, p=.037$ ). On average they focused on the live video 92.15 percent with the pointer condition and 73.35 percent with the annotation condition.

### **5.2.2 Connectedness & Being Together**

Since the situation model was remote collaboration for an object manipulation task, understanding the current state of a collaboration partner and having a similar level of understanding of the current workspace (connectedness) were important factors. For the user connectedness and being together, figure 5.3 shows the results comparing the three conditions using the Likert scale ratings. For the ratings from remote participants, the Friedman test showed a significant difference between three conditions in user connectedness ( $\chi^2(2)=6.837, p=.033$ ) but no significant difference in pair wise comparisons.

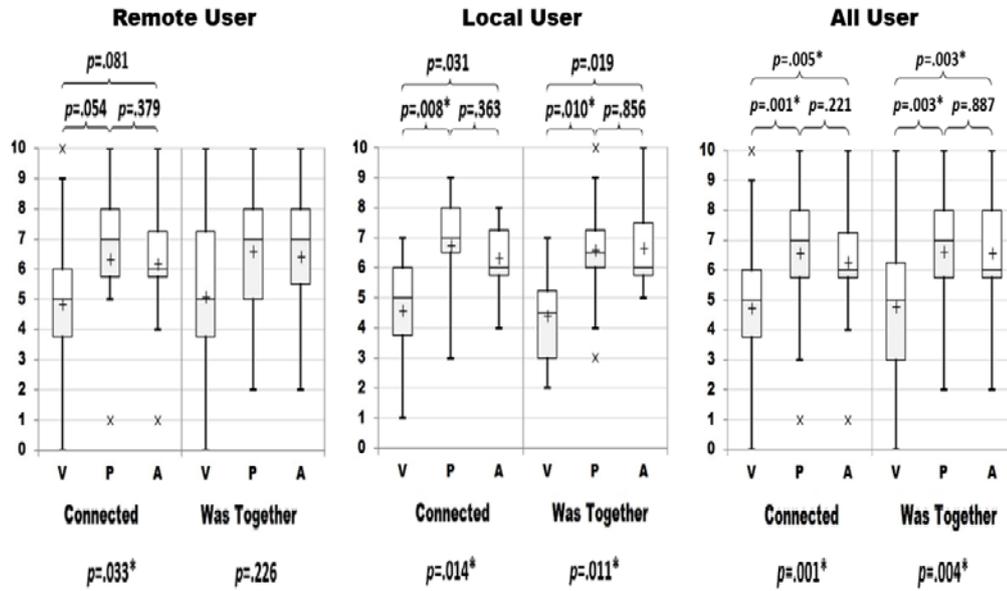


Figure 5.3: Results of Likert scale ratings from participants for the question about how much they felt they were connected and together with (0: strongly disagree ~ 10: strongly agree; \*: statistically significant difference, +: mean, x: outlier, V: video only, P: pointer, A: annotation).

For the ratings from local participants, the Friedman test showed a significant difference between three conditions in user connectedness ( $\chi^2(2)=8.512$ ,  $p=.014$ ) and being together ( $\chi^2(2)=8.977$ ,  $p=.011$ ). The condition using pointer cue was found to be significantly better than the baseline condition in terms of feeling connected (Q2:  $Z=-2.671$ ,  $p=.008$ ,  $M = 6.75$  for a pointer and 4.58 for baseline) and being together (Q4:  $Z=-2.572$ ,  $p=.010$ ,  $M = 6.58$  and 4.42).

To compare the ratings from all users ( $N=24$ ), the Friedman test showed a significant difference for user connectedness ( $\chi^2(2)=15.140$ ,  $p<.001$ ) and being together ( $\chi^2(2)=10.932$ ,  $p=.005$ ). Pair wise comparisons showed a significant difference between the baseline and pointer conditions in connectedness ( $Z=-3.202$ ,  $p=.001$ ) and being together ( $Z=-2.991$ ,  $p=.003$ ), and between the baseline and annotation conditions in connectedness ( $Z=-2.784$ ,  $p=.005$ ) and being together ( $Z=-2.946$ ,  $p=.003$ ). This suggests that participants felt more connectedness and being together when using the pointer and annotation interfaces than when using the baseline condition.

To the question asking when they felt connected or being together with their partners, participants mostly picked when they communicated well. Four participants (R4, R8, L5 and L7) said they felt connected and had a stronger sense of being together when their communication was bidirectional: "*At the moment we decided the shape (orientation) of a piece. We shared ideas and felt being together*" (R8) and "*I felt connected when we were doing it together but not at those moments I just followed what he said*" (L5). Understanding each other well was another factor that contributed to the feeling of connectedness (R2, 8 and L7): "*There was a moment when she made a mistake. I only told her she made a mistake but she fixed it exactly in the same way I thought of*" (R2). In contrast, participants felt disconnected when they misunderstood each other. R6 mentioned that "*I felt disconnected when he did not understand me.*"

### **5.2.3 Express Idea (sending/encoding messages)**

For the user expressing ideas, figure 5.4 shows the results comparing the three conditions using the Likert scale ratings. For the ratings from remote participants, the Friedman test showed a significant difference between three conditions in the level of expressing idea well ( $\chi^2(2)=10.773, p=.005$ ) and there were significant differences between pointer and baseline conditions ( $Z=-2.657, p=.008$ ) and between annotation and base line conditions ( $Z=-2.524, p=.012$ ). However the author did not find a significant difference in Friedman test for the ratings from local participants ( $\chi^2(2)=1.590, p=.452$ ).

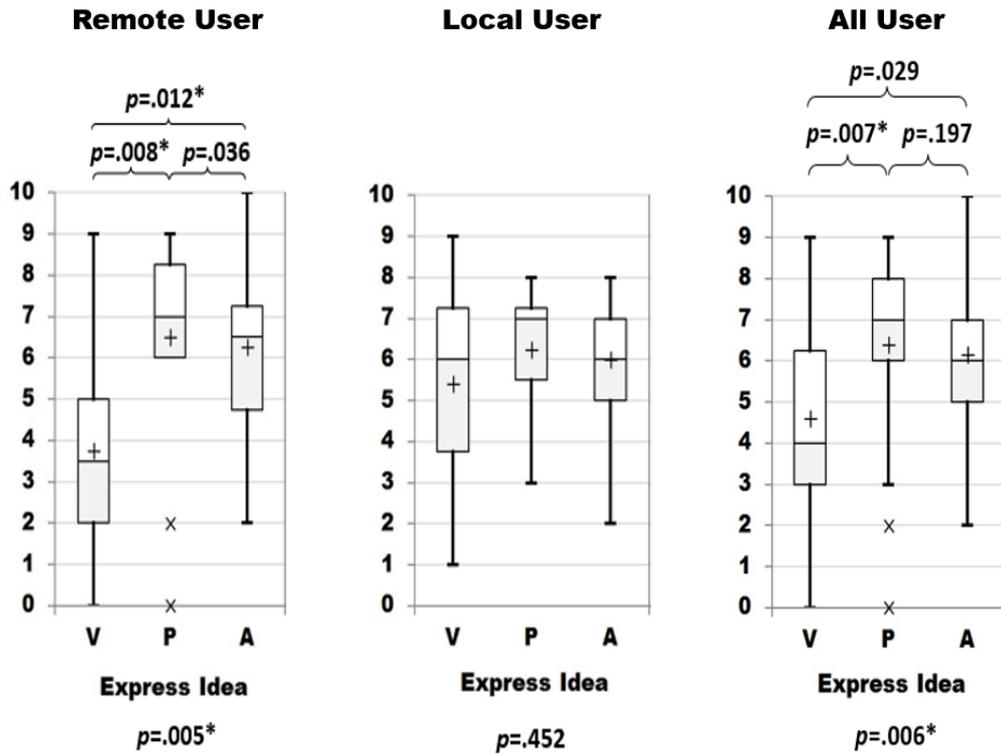


Figure 5.4: Results of Likert scale ratings from participants for the question about how much they expressed ideas well (0: strongly disagree ~ 10: strongly agree; \*: statistically significant difference, +: mean, x: outlier, V: video only, P: pointer, A: annotation).

In comparing the ratings from all users ( $N=24$ ), the author found a significant difference in the Friedman test for user connectedness ( $\chi^2(2)=10.289$ ,  $p=.006$ ). Pair wise comparison showed a significant difference between the baseline and pointer conditions ( $Z=-2.706$ ,  $p=.007$ ) but not between other pairs of conditions.

Only live video condition had an issue in expressing ideas. Voice communication was not enough to describe spatial concepts. All remote participants except R3, and two thirds of the local participants (L1, 5, 6, 8, 9, 10, 11 and 12) reported on the difficulty of communication: "*Describing the location of shapes and their orientation was pretty difficult to do verbally*" (R9) and "*Without help (of visual cue), it's hard to communicate which object you were talking about*" (L9).

Participants emphasized the harmony of visual and verbal communication

cues when using pointer. R11 commented "*I felt it strikes a good balance between being able to point in space and still encouraging communication.*" R1 was more analytical about the benefit: "*Since I can have more verbal communication, I participated more and become active ...*" This benefit was also revealed on the local end: "*There might have more mistakes with a pointer but had more communication than annotation so it feels like I was always with my partner*" (L10).

Three participants commented on the immediacy of the pointer. R10 (the remote user in group 10) mentioned "*I pointed directly without any hesitation,*" and R4 said "*It was possible to give quick feedback.*" This benefit was also identified on the local end with L4 (the local user in group 4) saying, "*Pointer was simpler, direct, precise and easy to understand.*"

The main benefits of annotation mentioned were: easy to explain the position and orientation information (R1, 2, 3, 4 and 6). Some remote participants pointed out the benefit of its permanence. R1 mentioned that "*I was able to draw an outline and it remained there.*" This was also noticed at the local end: "*It stayed on the scene*" (L3) and "*I felt he was here because his annotation was there*" (L9). The richness of visual information with annotation was also emphasized by R2, 4 and 6: "*Annotation was more descriptive than pointer and highlighting the shape was very clear*" (R4).

However, the annotation interface tended to slow down the conversation. Three participants mentioned that annotation required more time than pointer (R4, 7 and 10). R10 said "*I tried to draw carefully and accurately otherwise he got confused.*" Some of the local users also noticed that working with annotation was slower. L9 commented "*He slowed me down more. Actually, I knew what he wanted to say after halfway through his drawing.*" Six participants (R7, 11, L2, 5, 7 and 9) mentioned that they disliked having less verbal communication with drawing.

The author compared the amount of mouse movement when using different visual communication cues. While the log data showed that the majority of

the remote users (9 out of 12; 75%) made more mouse movements when using the pointer, no statistically significant difference was found ( $Z=-0.706$ ,  $p=.48$ ) between using the pointer ( $M = 4074.7$  pixels,  $SD = 4852.1$ ) and annotation ( $M = 3362.3$  pixels,  $SD = 3873.4$ ). Based on the observation that the mouse movement is highly affected by the characteristic of a pointer and annotation, and the user's habit and style of using the interface, the author compared the percentage of mouse movement when using pointer or annotation to the total amount of mouse movement, but no significant difference was found ( $Z=-1.020$ ,  $p=.308$ ).

#### **5.2.4 Understand Partner (receiving/decoding messages)**

For the understanding partner, figure 5.5 shows the results comparing the three conditions using the Likert scale ratings. For the ratings from remote participants, the Friedman test showed a significant difference between three conditions in the level of understanding partner well ( $\chi^2(2)=9.220$ ,  $p=.010$ ) and there were significant differences between pointer and baseline conditions ( $Z=-2.791$ ,  $p=.005$ ) but not between annotation and base line conditions ( $Z=-2.524$ ,  $p=.028$ ). For the ratings from local participants, the Friedman test showed a significant difference between three conditions ( $\chi^2(2)=6.045$ ,  $p=.049$ ) but no significant difference in pair wise comparison.

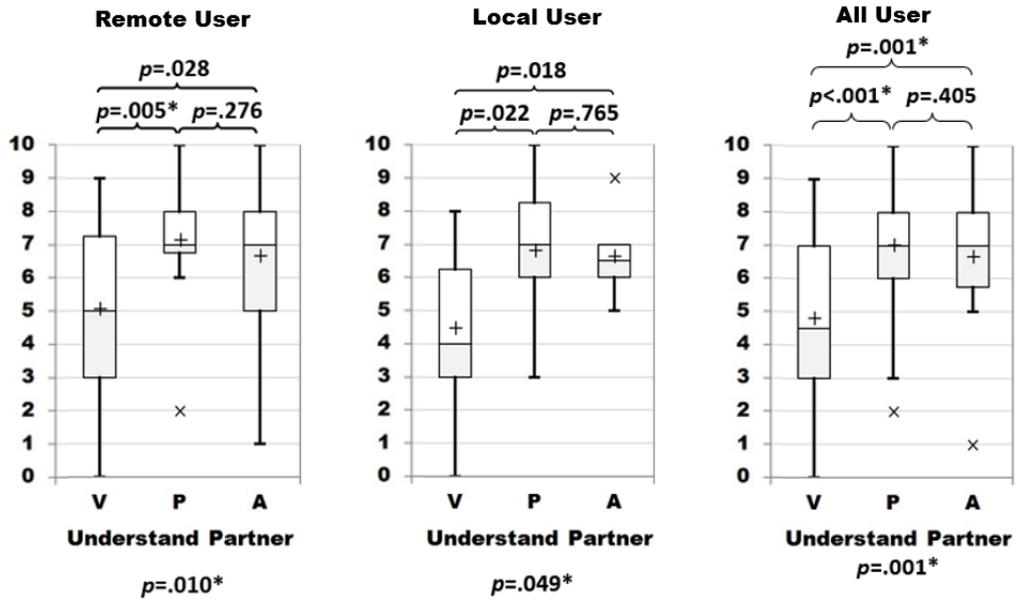


Figure 5.5: Results of Likert scale ratings from participants for the question about how much they understood their partner well (0: strongly disagree ~ 10: strongly agree; \*: statistically significant difference, +: mean, x: outlier, V: video only, P: pointer, A: annotation).

In comparing the ratings from all users ( $N=24$ ), the author found the significant difference in the Friedman test for user connectedness ( $\chi^2(2)=13.929, p<.001$ ). Pair wise comparison showed a significant difference between the baseline and pointer conditions ( $Z=-3.622, p<.001$ ) and between the baseline and annotation interface ( $Z=-3.188, p=.001$ ).

From the video recording, the author counted the number of misunderstandings the local participants had during the experiment. The local participants misunderstood remote participants' messages on average 6.25 times with the video only condition (baseline), 3.33 times with the pointer interface, and 4.08 times with the annotation interface. The Friedman test showed significant difference with the number of misunderstandings ( $\chi^2(2)=12.333, p=.002$ ), and pair wise comparison with Wilcoxon Signed-Rank ( $\alpha = .05$ ) showed a significant difference between baseline and pointer interfaces ( $Z=-2.679, p=.007$ ) and between baseline and annotation interfaces ( $Z=-2.405, p=.016$ ).

### 5.2.5 User Preference

After trying all of the conditions, participants were asked to pick their most preferred condition (results shown in Figure 5.6). Two thirds of the participants (16 = 8 remote + 8 local users) chose the condition with the pointer cue, while most of the remainder (7 = 3 remote + 4 local users) chose the one with the annotation cue, except the one remote user who picked the baseline condition. For the least preferred condition, two thirds of the participants (16 = 9 remote + 7 local users) chose the baseline condition which only had voice communication with shared live video, while most of the remainder (7 = 3 remote + 4 local users) chose the condition with annotation, and one local user picked the pointer as least preferred.

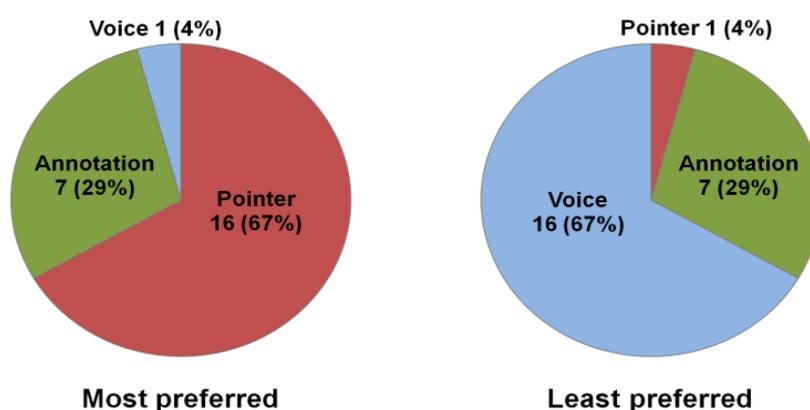


Figure 5.6: Results of participants' preference

The results from the Friedman Test ( $\chi^2(2)=18.750, p<.001$ ) and post hoc tests with the Wilcoxon Signed-Rank Test showed participants ranked (based on their preference) the condition with pointer significantly higher than the baseline condition ( $Z=-3.985, p<.001$ ) but no significant difference was found between the baseline and annotation conditions ( $Z=-2.245, p=.025$ ) or between the pointer and annotation conditions ( $Z=-2.245, p=.025$ ).

In the interview, the author found why they preferred the pointers: immediacy, harmony with verbal communication, and good compatibility with live video. While permanence and richness of visual information were identified as benefits of annotation, inappropriate combination with a live

video, the long time required to draw and reduced opportunity for verbal conversation were reported as disadvantages of annotation.

### **5.2.6 Live Video vs. Snapshot Image**

Since the prototype system provided two shared views at the remote end, a live video and still images, the author asked participants to give a comment for the views. For the live video, showing the current state of playing space was mentioned as a benefit of live video by all remote participants. However, drawing annotation on the live video while the local users changed the shared viewpoint, the annotation was drawn at wrong place and this disadvantage was reported by several participants (R2, R7, R9, R10, L1, and L2).

One of the main benefits of having snapshot images was that drawing annotations was easier than in live video, which was mentioned by half of the remote users (R2, 4, 7, 8, 9 and 10). On the contrary, not getting live updates was one of the main disadvantages of the snapshot images (R2, R6, R7 and R9): *"It's not real time and not showing the current state of the work space"* (R7). This disadvantage affected the use of visual communication cues when the local user was looking at a different area from the snapshot image as described earlier in section 5.2.3.2. R8 highlighted this problem: *"My message was not seem to be displayed because he was watching others."*

### **5.2.7 HMD vs. HHD**

Even though the author did not find statistically significant difference between HMD and HHD conditions from the questionnaire results, the author found interesting comments from the local users about the benefits and drawbacks of each device in the interview. Since each participant only used one of them during the experiment, the author showed the other device and let them try it before the interview.

All of the local users except L3 and L12 mentioned having both hands free as a benefit of using HMD. L9 pointed out a benefit in terms of more free hand movement, *"Hand movement would be more natural (not being blocked*

*as with the tablet).*" Another benefit of using a HMD was having the camera closer to the view of their eyes (L4, 5 and 8): "*Camera position is much closer to my eyes than the tablet, so I might (have) focused more on the task*" (L4) and "*Can have wider view (because the camera is farther away from the work space)*" (L5). L10 reported sharing the exactly same view with the partner as a benefit.

On the other hand, wearing a HMD made some of the participants uncomfortable (L1, 2, 7 and 10). L7 (who used a HMD during the experiment) and L10 (who tried during the interview) said that wearing it with eye glasses was uncomfortable. L1 and L2 reported on unnatural head movement: "*I needed to be careful to move my head*" (L2). Besides problems in physical movement, an interesting comment was reported by L8 on mental stress: "*Visual communication cue on HMD could give some pressure on me because it turned up in front of me (my eyes).*"

As a benefit of using HHD, L1 (tried during interview) and L10 (used during the experiment) mentioned having less tired eyes. For the disadvantage, four participants (L3, 5, 9 and 10) mentioned "*it is heavy to hold,*" and three participants (L1, 3 and 4) pointed out having a worse shared view: "*It would be hard to sit (with HHD) to provide a good (shared) view*" (L1), "*It might reduce the field of view*" (L3), and "*Holding a tablet and manipulating the pieces made the view shaking*" (L4). L6 said it affected natural hand movement: "*It blocks my hand movement.*" During the experiment, some of the participants who used the HHD watched the shared space directly rather than through the screen, and this behavior caused problem: "*I didn't really get what my friend said*" (L6). This problem also affected the remote users (R8 and 10): "*I felt the live video did not follow what he saw, but followed how he held the tablet*" (R10).

### **5.3 Observation of User Behavior**

Based on experimental observations, the author compared user behavior

patterns between face-to-face collaboration and remote collaboration.

When collaborating face-to-face, the collaboration was smooth and effective. Both local and remote participants were actively participating in solving the puzzle and discussing with each other. The most interesting observation was that their behavior smoothly switched between two states: *playing in parallel* and *collaborating together*. When *playing in parallel*, each of them concentrated on what they were doing and ignored what their partners did. When *collaborating together* they focused on the same puzzle piece and discussed on possible ways of solving the puzzle and tried them out together.

Similar behavior patterns were found while sharing the experience remotely. First, the author observed the cases of participants *collaborating together* to solve the puzzle. In such cases, participants kept communicating with each other both verbally and using visual cues to express how they could place a piece to match the silhouette. These cases mostly occurred when they were using visual communication cues. Group 5 showed a good example of this while they were working with a parallelogram shaped puzzle piece using a pointer: "(L) Do you know which one is this? (R) I think it's this one, cause it's only uh ~ parallel. (L) But this is too long. (R) Oh~ what about this one? (L) ya ya ya". Another good example of this while using a drawing annotation was found in group 8: "(L) I think this guy goes like this, what you reckon? (R) Uh, have a look at this triangle (L) Oh~, right, right! (R) that can be like~ this (drawing a triangle)". Most of the groups showed this collaborative pattern, except for group 3 and 6 that never worked collaboratively, and for group 11 and 12 when using the drawing annotation. In the interview, L5 shared the experience of collaborating together saying "*Pointer was easiest and I felt I was doing it together*" and R7 mentioned "*We did it together and (felt) connected*" after finishing the session with annotation.

The state of *playing in parallel* was less common in remote collaboration compared to the *collaborating together* state. It was found in cases where the remote participant was using the drawing annotation in a well-organized way,

for instance with R1 and R10 drawing borders or numbers as mentioned in section 5.2.3.2 (see Figure 5.8). This involved the remote user having a good view of the work space (e.g. a snapshot of the overall space) so that he or she could *play in parallel* by making a drawing annotation on it while the local user was focusing on his or her own interest.

The author also observed new user behavior patterns in the remote collaboration case that were not observed during the face-to-face experience. Unlike face-to-face collaboration where both participants actively participated in assembling the puzzle, from time to time the remote users just passively watched what the local users were doing while the local users were busy to solve the puzzle by themselves. The author observed this happening mostly in the voice only conditions. In such cases, the remote users explained their idea but were often not understood by the local users. Then, the remote user turned into an observer and just encouraged or commented to the local user verbally, saying, for example, "yes", "right", "looks a bit weird" or "*the triangle is too small*".

Such behaviour pattern of remote users switching into passive observers was found in all groups except in group 6. R2 mentioned in the interview, "*I cannot keep talking because when I talked about a triangle, he thought I was talking about another triangle so (he) changed the view to make that triangle (to come) at the centre of the view*". Moreover, R1, R11 and R12 fell into this *passive state* with annotation as well. In the interview, R1 and R11 mentioned "*I don't know how to draw and talk together*" and "*It was frustrating to draw while my partner is trying this and that,*" respectively. The local users assembled the puzzle alone in such cases. After the baseline condition, L7 said "*I solved the puzzle almost by myself*" and L12 said "*I can (was able to) contribute more than the other two (conditions)*".

However, in some cases remote users became extremely active and the local users just followed passively what the remote participants told them. In these cases, the remote user talked continuously and gave instructions, and the

local user manipulated the puzzle following the remote participant's instructions. Such cases mostly occurred when the remote user was able to express their ideas not only verbally but also using visual communication cues. Participants in group 2 showed a typical example when using annotation: "(R) *This goes here. No, No, No. Look at my triangle. Ye, ye, ye. And this one is like this. Wait, wait. One and two (while drawing two triangles).*" Another example was found in group 9 while they also used annotation: "(R) *So, small triangle goes here and I (am) guessing that a rectangle goes here, is it? Okay!! Do you have a small triangle? A small triangle should go here.*" Overall, group 2, 5, 6, and 9 showed this behavior while using the drawing annotation and group 4 and 6 experienced this while using the pointer. L5 and L9 said that "*Annotation was like I got an order from him*" and "*Annotation seems like giving me an order. I need to wait for his command,*" respectively.

There were two extreme cases that the author would like to note. One was the remote user in group 3 acting passively in all three experimental conditions, just watching what the local user does. On the contrary, the remote user in group 6 was very active in all three conditions, and the partner just followed his instructions.

In summary, in the face-to-face collaboration participants smoothly transitioned between the two states: *playing in parallel* and *collaborating together*. When participants were sharing their experience remotely, in addition to the two states, their behaviour showed another two more states: *passive* and *do it alone*. The *passive state* is when a user is mostly watching what the partner does and briefly commenting on the partner's work for encouraging the partner or pointing out a mistake. The *do it alone state* is the opposite of the *passive state*, when a participant plays mostly alone, taking only a few words from the partner. This is different from the *playing in parallel* state when each user focuses on his or her own task. In contrary, in the *do it alone state* both users still have shared focus on the same task. The *passive state* and the *do it alone state* were usually observed mutually

between the remote and local users. Interestingly for remote users, both the *passive* and *do it alone states* were observed more often with annotation than with pointer.

## 5.4 Discussion

The user study results showed that both of the augmented visual communication cues (pointer and annotation) significantly increased the feeling of connectedness and being together compared to the plain video conferencing condition. However, the pointer condition was preferred by the participants and had higher ratings over the annotation condition. Permanence and richness of visual information were still mentioned as the benefits of annotation, but the benefits of pointer interface: (1) immediacy, (2) good harmony with verbal communication, and (3) good compatibility with live video; were considered more functional in a parallel experience remote collaboration.

The author identified different behaviour patterns of the users in remote collaboration during the experiment compared to face-to-face collaboration. Two additional states (*passive* and *do it alone*) were introduced in the remote collaboration, while also finding behaviour patterns similar to the face-to-face collaboration (*playing in parallel* and *collaborating together*). The difference may come from the fact that only the local user was able to manipulate the pieces in remote collaboration, while both users were able to manipulate the pieces in face-to-face collaboration. The character of users and the relationship between them would also affect their behaviour patterns.

While the results of this study may seem to contradict those of previous studies (Fussell et al. 2004, the pilot studies described in Chapter 3) that showed drawing annotation gaining more favour over using a pointer, the author would like to point out the key difference between this study and previous works. Compared to previous studies that investigated remote expert collaboration where a remote expert gave instructions to a local user, this

study focused on sharing experience (parallel experience) where active collaboration with bilateral communication between the two users was necessary. This difference made the author's study involve more *collaborating together* moments, rather than the combination of *do it alone* and *passive states* that are mostly the case where a remote expert gives instruction and a local worker has to follow.

In that sense shared pointers had a key benefit of a remote user being more active and immediate way to communicate compared to the annotation that was found to be slower by many participants. Since the pointer had the benefit of immediacy, remote users could jump into the process of piece manipulation quickly or suggest an idea before the local user moved onto another manipulation process. This might convey less information but it would tell local users that remote users were participating in the process more frequently, and prevent local users from moving into the *do it alone* state. Moreover, since a pointer can convey a limited amount of information, remote users would not easily move into the *do it alone* state and local users could have room to participate more actively with more verbal conversation.

On the other hand, using the drawing annotation was slow for two reasons. First, annotation required more time to encode and decode messages because of containing more information than a pointer interface. Second, the annotation interface required time to draw correctly on a live video or still images. With a live video, a remote user had to wait for the moment when the shared live video was not moving to anchor the annotation correctly. With the still images interface, a remote user spends time to take a still image with additional inputs. Thus, remote users have fewer chances to jump into the current process of object manipulation before the local user moves onto another. This may be the reason why more remote users sometimes fell into the *passive* state with annotation than with a pointer. Moreover, since drawing annotation is slower, the local user might spend more time to understand the remote users' intention rather than having time to think about solving the

puzzle, which would lead the local users to be more passive in the collaboration.

While an option of taking a snapshot of the shared view helped prevent incorrectly placed annotations, it had two drawbacks: (1) mismatched views for the two users, and (2) requiring additional inputs for a still image. The author investigated possible solutions such as making the switch between the live video and snapshot image more instant, in the second study described in the next chapter.

## **5.5 Conclusion and Summary**

In this chapter the author described the first main user study on investigating the use of augmented visual communication cues; a pointer and drawing annotation, for sharing parallel experience through video conferencing. A user experiment was conducted to compare the three conditions of sharing experience with different combinations of communication cues in video conferencing. The results showed both using pointer and annotation could significantly improve the shared experience compared to live video only condition in terms of feeling connected, being together, and understanding the partner. Participants most preferred the pointer cue among the three conditions. Further discussion included benefits and problems of different visual communication cues, and behaviour patterns that were introduced in remote collaboration compared to face-to-face collaboration.

The annotation interface was less preferred than the pointer interface in a parallel experience collaboration. Even though using drawing annotations has the benefits of permanency and richness, users preferred the pointer interface with the benefits of immediacy, harmony with verbal communication, and compatibility with live video. Annotations required more time to input, and drawing annotation had the issue with the live video in which the annotation would be anchored in the wrong place if the remote user drew an annotation while the local user was changing the viewpoint.

In the second study, the author designed an interface that solves the issue of annotation anchoring at wrong place. The third study would involve designing an interface for better notification to tell the local user when the remote user is drawing an annotation. This would prevent the local user from falling into a do it alone state and help users stay in the collaborating together state. The next chapter will report on these two user studies.

## **6 Studies about Freeze Frame Interactions and Visual Notifications**

In this chapter the author reports on the second and third user studies. The second study investigates the usefulness of different freezing techniques on the shared live video view of a remote user. The third study investigates the effect of different visual notifications on the local end.

If a local user changes the viewpoint of a shared live video while the remote user is drawing an annotation, the annotation can be drawn at a wrong point. To solve this issue, a still image interface (manual freeze method) was introduced in the first study (Chapter 5), but users did not prefer using still images over a live video view. In the second study, the author introduces another approach, the auto freeze method, to solve the issue and compare this with the manual freeze condition and the live video condition (baseline).

Compared to the second study focusing on freeze techniques on the remote end, the third study investigates the effect of different visual notifications on the local end. In the second study the author noticed that local participants sometimes missed the remote user's drawings while handling multiple tasks such as receiving drawings from the remote user while manipulating objects themselves. To address this issue, the author compared three conditions with different visual notification methods: (1) no notification (as a baseline condition), (2) red box notification, and (3) both freeze.

In short, the author explores: (1) the benefits and drawbacks of auto-freeze compared to manual-freeze and non-freeze conditions (the second study), and (2) how to design visual notifications to effectively allocate the user's attentional resources in a parallel experience to enhance collaboration (the third study).

To the best of the author's knowledge, both of these studies are the first user studies comparing such conditions, and can be regarded as novel contributions of this thesis.

## **6.1 User Study on Freeze Function**

The author conducted a user study comparing manual and auto freeze approaches to a baseline non-freeze condition. In this study, the local user's view remained live at all times and was not affected by the remote user freezing his/her view.

### **6.1.1 User Interface**

For the baseline, non-freeze condition, the author adopted the interface from the first study where a remote user draws an annotation on a live video. To draw a line, the remote user conducts three steps: pressing the left mouse button, dragging, and releasing the button.

As well as using these mouse interactions for drawing, manual freeze requires two additional interactions to freeze and unfreeze the shared view. In order to allow users to interact with only a mouse, the author used double clicking to toggle between freeze and unfreeze states. In this way, a sequence of interactions (1. freeze, 2. draw annotation, and 3. unfreeze) could be executed with a single device and not require the mouse cursor to be in a specific position to click GUI buttons for freezing or unfreezing the view, hence reducing the amount of mouse dragging.

With manual freeze, as the interaction for freezing is independent from drawing interactions, remote users can freeze the view first, then position the mouse cursor at the starting point of drawing on the frozen view, without worrying about positioning it in a live video whose viewpoint can move. However, they need to perform more interaction steps for the manual freeze technique.

The automatic freeze method reduces two user interaction steps by combining drawing and freezing interactions. When the remote user presses the left mouse button down to start drawing, the live video is immediately frozen. When the remote user releases the left button to finish drawing, the view automatically returns back to live video. Thus the remote user has fewer

interactions and avoids the need to remember how to freeze and unfreeze the view. However, since the interactions are combined, users have to position the cursor at the start point of drawing before freezing, which could be challenging to do while the viewpoint of the live video is moving.

In the real world, people watch the environment, focus on annotating words or symbols, and watch the environment again (See Figure 6.1). In the auto freeze condition, remote participants watch the live video, draw annotations, and watch the live video if they mainly focus on our interfaces during the experiment. The sequence of drawing annotations is very similar to how they annotate in the real world. However with manual freeze, remote participants watch the live video, freeze the live video, draw the annotation, unfreeze the live video, and watch the live video. In this sequence, freezing and unfreezing the live video are extra steps that are not used in real world annotation drawing.

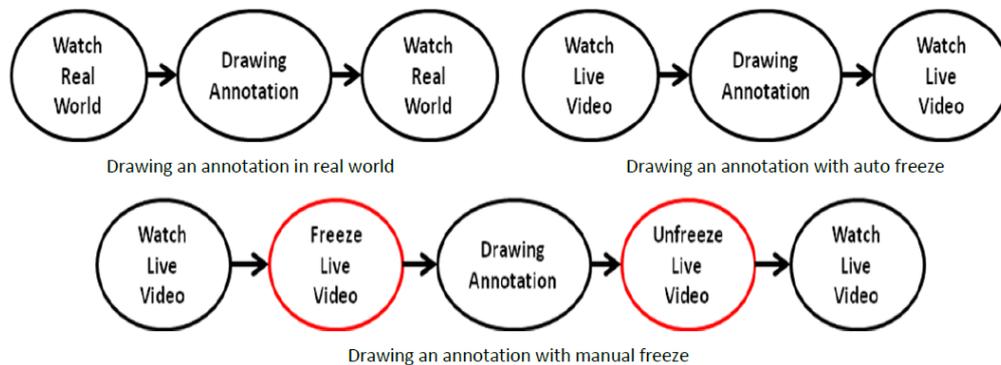


Figure 6.1: The process of drawing an annotation in the real world (top left), with the auto freeze condition (top right), and the manual freeze condition (bottom)

### 6.1.2 User Study Design

For the user study the author adopted the same experimental tasks and environment as the first study, and the experimental procedure and method of data collection were identical to the first study except the proposed prototypes (experimental conditions).

The author chose the customized Tangram that was used in the first study

(see Figure 6.2) as the experiment task. In the experiment, the participants had to solve the Tangram puzzles in pairs while located in separate rooms and using the prototype system for communication. The experimental environment was the same as in the first study, but only the Head-Mounted Display (HMD) was used as a display for the local user and only one monitor was used on the remote end. The local user sat on a chair in front of a table while wearing the HMD, and the remote user sat in front of a desk with a desktop computer. The local users had a reference paper and puzzle pieces on the table and shared their ideas by sharing their view while manipulating the puzzle pieces. The remote users used mouse interaction to share their ideas with drawing annotations as described in section 6.1.1.

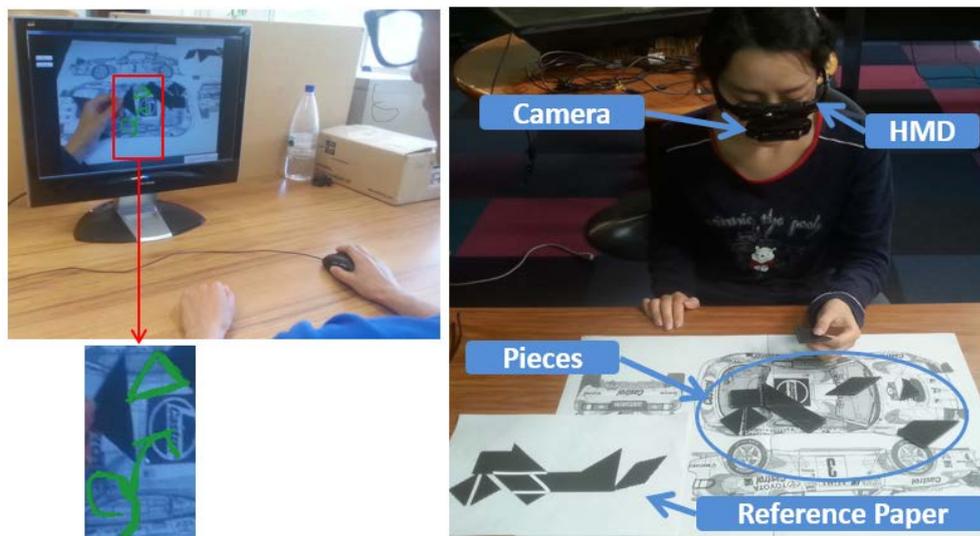


Figure 6.2: The prototype systems used by a remote user (top left) and a local user wearing a HMD (right).

The experimental procedure was identical to the one in the first study. The experiment began with participants answering a questionnaire about demographic information, and the purpose of the study was explained. The experimental procedure consisted of five sessions: face-to-face session, three sessions under three different experimental conditions, and a final interview. First, the participants were asked to solve a sample Tangram puzzle face-to-face, then the author showed and explained how to use the prototypes.

Participants were then separated into different rooms and completed three sessions under the three different experimental conditions: (1) non-freeze, (2) manual freeze, and (3) auto freeze. The order of the conditions was counterbalanced using a Latin square design. Each session consisted of two minutes training and five minutes of solving the Tangram. After finishing each experimental session, the participants separately filled out a questionnaire and had a brief interview. After the three sessions with experimental conditions, the experiment was wrapped up with final interviews. Overall the experiment took about 70 minutes for each pair of participants (55 minutes for the experiment and 15 minutes for the interviews).

For each of the three experimental conditions, a local participant and a remote participant answered six to ten questions respectively (see Table 6.1), using a Likert-scale from 0 (Strongly disagree) to 10 (Strongly agree).

Table 6.1: Questions asked after each session. Both participants answered the six questions (no color) and remote participants answered four more questions (highlighted in gray).

Q1	I was able to focus on the activity (building a model).
Q2	I was able to express my idea properly.
Q3	It was easy to draw/annotate on the remote view
Q4	I was able to draw annotation on the remote view as soon as I wanted to.
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
Q8	I felt time pressure while drawing on the remote view
Q9	I had to be careful while drawing on the remote view
Q10	The interface was physically stressful to use

For the questions in the questionnaires, the author also considered the simple remote collaboration model by Hauber (2008). Since the situation model is the solving an object manipulation task, the author prepared a question (Q1 – question 1) asking how much they focused on the task. Since

the purpose of our study was comparing three different remote user's interfaces for encoding messages, the author prepared four questions (Q3, Q4, Q8, and Q9) for only the remote user and one question (Q2) for both users about their experience of encoding message. For the experience of receiving (decoding) messages question 5 was prepared. To see how the different encoding experiences have effect on overall remote collaboration, the author also prepared questions 6, 7, and 10.

Each participant was interviewed to collect more details about what he or she liked and disliked with the given interface, then this user study wrapped up with final interviews. To complement the subjective feedback, the author also collected an activity log of the remote participants which included the amount of mouse movement while drawing annotations, the number of drawing strokes, and the number of times using the clear and undo functions. The author also recorded the screens of both participants and their verbal communication.

## **6.2 Results of Freeze Study**

The author recruited twenty-four participants (in pairs) who were friends or family and who had experience of using videoconferencing regularly (more than once a month). There were 17 males and 7 females with ages ranging from 15 to 33 years old ( $M = 25.6$ ;  $SD = 4.6$ ).

The main findings from this user study are below:

- 1) The auto freeze interface solved the issue that annotation can be drawn in the wrong place if a local user changes the viewpoint of a shared live video while the remote user is drawing an annotation.
- 2) Since the auto freeze interface did not require any additional input, the remote participants could do their annotation drawings more quickly.
- 3) Since the remote participant mostly focused on drawing annotations in the frozen view and quickly returned to the live video while using the auto freeze interface, the collaboration was not significantly affected

from losing the liveness.

In this section, the author analyzes and describes the experimental data according to the main factors of remote collaboration (shared situation task, encoding and decoding messages). To analyze the Likert scale ratings of the questionnaires between the three conditions, the author ran the Friedman Test ( $\alpha = .05$ ), and for those showing a significant difference between the three conditions the author ran post hoc tests for pair wise comparison using the Wilcoxon Signed-Rank Test with Bonferroni correction applied ( $\alpha = .0167$ ). For deeper understanding the results of questionnaire analysis, the author lists the related answers from interviews and analyzes the video recordings and log data as supplements.

### **6.2.1 Focusing on the assembling task**

For the question 1 about focusing on the assembling task, figure 6.3 shows the results comparing the three conditions using the Likert scale ratings from the remote and local participants. Friedman tests showed that the ratings were not significantly different between the three conditions (Remote participants' rating:  $\chi^2(2) = 4.682, p = .096$ ; Local participants' ratings:  $\chi^2(2) = 0.774, p = .679$ ).

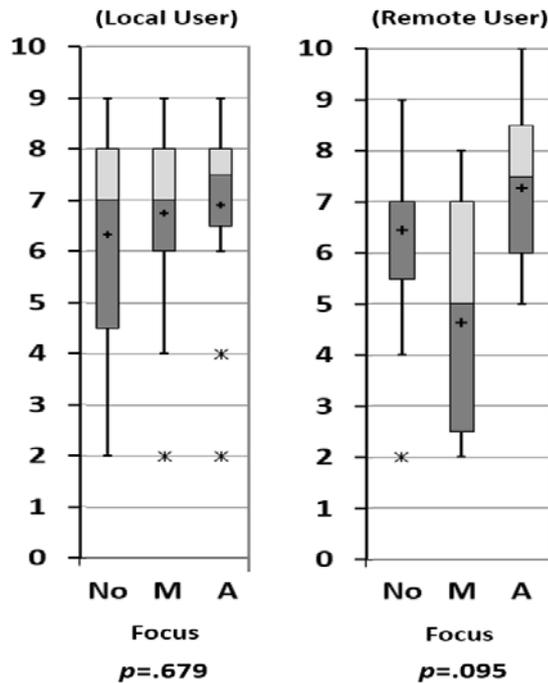


Figure 6.3: Results of Likert scale ratings from local (left) and remote (right) participants for the question about how much they felt they focused on assembling task (0: strongly disagree ~ 10: strongly agree; \*: statistically significant difference, +: mean, x: outlier, No: non-freeze, M: manual freeze, A: auto freeze).

However, remote participants reported their experience about the level of the focusing on the task in the interview. With manual freeze condition, R8 (the remote user in group 8) and R12 commented they could not focus on solving the puzzle because of needing to double click (manually freeze). Other remote participants pointed out the problem of losing the focus on the shared task space while manually unfreezing the view. R11 said "*I was confused because she already started to manipulate pieces (before he returned back to the live video) and I didn't know what she was doing*". This could be because the local participants manipulated the pieces while the remote participant had the frozen view and the remote participants sometimes lost the critical moments when using the manual freeze. Moreover, R5 reported "*Big difference between manual freeze and non-freeze in the aspect of live viewing*".

In contrast, the remote participants hardly lost their focus with the non-freeze and auto freeze conditions. R9 pointed out that a benefit of the non-

freeze condition is that it keeps showing live video without losing track of their partner's view. R6, 7, and 9 commented that they hardly felt the difference between the non-freeze and auto freeze conditions in terms of maintaining the focus of the remote environment. This could be because of the participants mostly focusing on drawing while the view was frozen and the view instantly and automatically returned back to the live video after drawing annotation.

### 6.2.2 Sending/Encoding Messages (drawing annotation cue)

The author prepared four questions (Q3, Q4, Q8, and Q9) only for the remote participants and one question (Q2) for both participants about their experience of encoding message. Figure 6.4 shows the results comparing the three conditions using the Likert scale ratings.

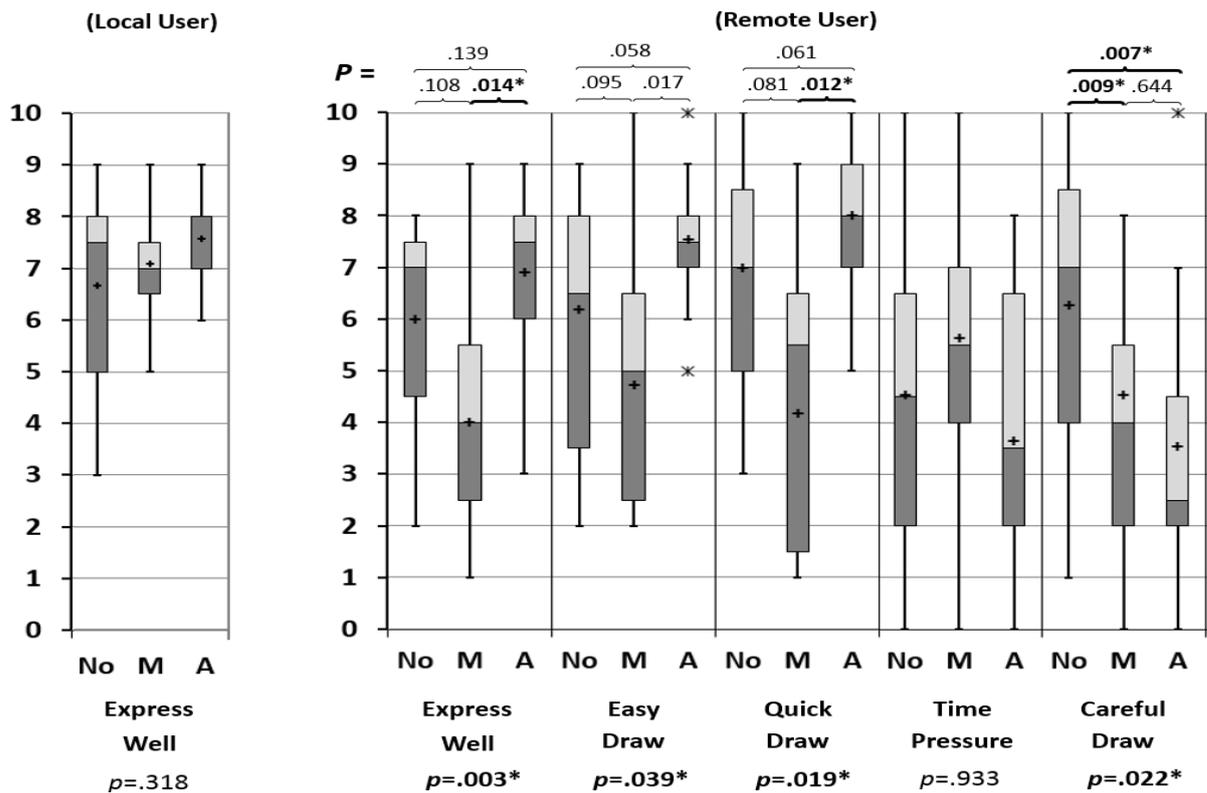


Figure 6.4: Results of Likert scale ratings from the local participants (left) for the Q2 and from the remote participants (right) for the Q2, Q3, Q4, Q8, and Q9. (0: strongly disagree ~ 10: strongly agree; \*: statistically significant difference, +: mean, x: outlier, No: non-freeze, M: manual freeze, A: auto freeze).

In Friedman test, the local participants' ratings about 'Express Well' (Q2) did not show a significant difference between the three conditions (Q2:  $\chi^2(2)=2.294$ ,  $p=.318$ ). This would mean that local participants did not feel significant difference in expressing their idea according to the condition the remote partner used. However, all remote participants' ratings except the one about 'felt time pressure when drawing annotation' (Q8:  $\chi^2(2)=0.14$ ,  $p=.933$ ) showed significantly different between three conditions in question 2, 3, 4, and 9 (Q2:  $\chi^2(2)=12.043$ ,  $p=.002$ ; Q3:  $\chi^2(2) = 6.488$ ,  $p=.039$ ; Q4:  $\chi^2(2)=7.946$ ,  $p=.019$ ; Q9:  $\chi^2(2)=7.6$ ,  $p=.022$ ).

Pair wise comparisons with post hoc tests showed a significant difference between the manual freeze and auto freeze conditions in question 2 about 'express their idea' ( $Z=-2.102$ ,  $p=.014$ ) and in question 4 about 'quickly drawing' ( $Z=-2.524$ ,  $p=.012$ ) (Kim et al., 2015). Remote participants felt that they significantly quickly drew annotation and expressed their idea significantly better with the auto freeze condition than with the manual freeze condition. From the interviews, the author also found similar comments. After using auto freeze, R5 and R4 mentioned that "*It's quick, precise and expressive*" and "*I didn't need to switch the views. It's essential for quickly drawing in the quickly changed local environment (as local users manipulated pieces)*" respectively. For the manual freeze, the additional inputs were required for freezing or unfreezing the shared view and R3, 5, 6, and 7 reported that they forgot to freeze or unfreeze the scene before or after drawing. Moreover, R5 said "*It's not natural. It was hard to be effective and made me slow down*" and a similar comment was reported by R6. The additional inputs and forgetting to freeze or unfreeze the scene would need more time (1.88 seconds before drawing annotation and 2.87 seconds after completing drawing annotation as described in the previous section 4.3.1) with the manual freeze condition than auto freeze condition.

Even though, remote participants' ratings were not significantly different between the manual and auto freeze conditions ( $Z=-2.378$ ,  $p=.017$ ) for the

question 3 about ‘easy drawing’, it was nearly close to significant difference and the author found several comments about ‘easy drawing’. R1, R3, and R8 mentioned that drawing with auto freeze was easier than other conditions.

Remote participants felt that they were more careful when drawing annotation with the non-freeze condition than with the other two conditions (Auto freeze:  $Z=-2.675$ ,  $p=.007$ , and Manual freeze:  $Z=-2.597$ ,  $p=.009$ ). In the interview, many participants commented on the importance of freezing the live video for drawing. Seven remote participants (R1, 3, 4, 6, 10, 11, and 12) reported difficulty with drawing in the non-freeze condition because their local partners changed their view point when they drew annotation. This difficulty may cause unintended drawings (see Figure 6.5) and eleven remote participants (except R7) made at least one unintended drawings with non-freeze condition ( $M = 1.83$ ,  $SD = 0.94$ , see table 6.2).



Figure. 6.5: Example of wrong drawings by the remote participant in group 3 with the non-freeze condition. The remote user attempted to draw a triangle like the red one in the left picture, but he drew as in the right picture (the green line).

Table 6.2 The number of wrong drawings that remote participants made.

	Unintended drawings		Unintended drawings		Unintended drawings
<b>R1</b>	3	<b>R5</b>	1	<b>R9</b>	2
<b>R2</b>	1	<b>R6</b>	2	<b>R10</b>	3
<b>R3</b>	3	<b>R7</b>	0	<b>R11</b>	2
<b>R4</b>	2	<b>R8</b>	1	<b>R12</b>	2

To overcome this problem, remote participants came up with their own solutions with non-freeze condition. All remote participants waited until their local partner did not make any movement (so there was not any view point movement), and quickly finished the drawing annotation. In the interview, R8, 9, and 12 reported they had to draw quickly before the viewpoint of shared video changed. R1, 6 and 7 mentioned they used the drawing interface as a pointer (only for pointing information) instead of drawing the shapes of pieces. The form of quick drawing annotation using like a pointer was a circle or tick ('v' shape or a small line) shape instead of the shapes of pieces (see left and middle pictures of Figure 6.6).

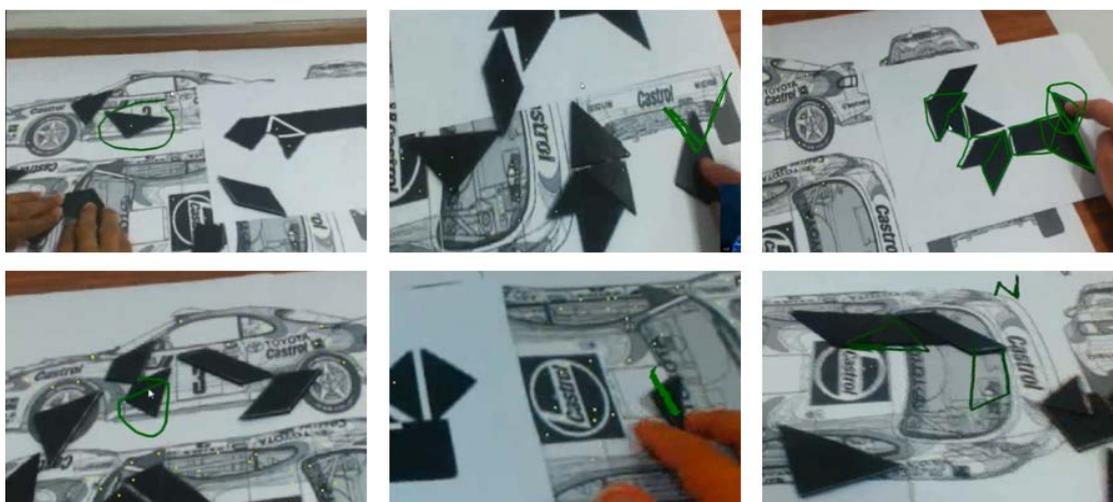


Figure 6.6: Example of drawings. Remote participants drew circle or tick shapes for pointing information (left and middle pictures) and drew shapes of pieces for pointing and orientation information (right pictures)

The circle and tick shapes were used for selecting a piece or referring to a position. The piece shape was used to representing a position and orientation of pieces. Then, how did participants describe orientation of a piece when using non-freeze condition that remote participants used significantly less piece shape drawings than others?

### 6.2.3 Verbal messages

The drawing annotation and verbal message were complementary. With non-

freeze condition, all remote participants except R7 and R11 sometimes drew a small line or circle at the point where a selected piece needs to be placed and described the orientation of the piece with verbal messages (e.g. ‘flip it’ by R3, and ‘turn it clockwise’ by R8) instead of drawing piece shapes for positioning a piece when positioning a piece. For positioning a piece with manual and auto freeze conditions, they mostly drew a piece shape at the proper position while verbal messages described piece-putting action (e.g. “*It goes like thi~~s*” by R7 and “*Put it he~re*” by R10).

The local participants showed how they manipulated pieces through a live video. When they understood the remote partner’s messages or when they had an idea for manipulating a piece, they mostly did not have a verbal message but just showing their actual manipulation. They picked up a piece and tried several times with different orientation mostly next to the completed part. Their verbal communication started mostly when they did not understand the remote partner’s messages well (e.g. ‘like this?’ by L2, ‘this triangle?’ by L4), when they disagreed with the remote partner’s idea (e.g. ‘*It’s not fit on this*’ by L5), and when the remote partner mentioned disagreement with their manipulation (e.g. ‘*You’re right. Need a small triangle*’ by L7).

Since our target models were the silhouettes of animals, both local and remote participants used the words meaning the animal or a part of body, such as ‘head’, ‘leg’ and ‘tail’. Group 1, 4, 7, and 12 discussed or mentioned an animal which the silhouette represents to before or after the assembling task. R1, R5, R9, and L4 notified the starting point of assembling to their partner by mentioning a part of model (e.g. ‘Let’s start from the head’ by R5). Before or after positioning a piece, R5 and R 7 mentioned a part of body which the selected piece was for or which the positioned piece was for.

#### **6.2.4 Receiving/Decoding Messages**

For the question about easily understanding what the partner was doing and explaining, figure 6.7 shows the results comparing the three conditions using the Likert scale ratings. Friedman tests showed that remote participants did

not feel significant difference between the three conditions in the question ( $\chi^2(2) = 5.076, p=.056$ ), while local participants did feel significant difference for the question ( $\chi^2(2) = 10.585, p=.006$ ). Pair wise comparison using the Wilcoxon Signed-Rank Test showed that local participants felt that they understood their partners significantly better when a remote participant used manual freeze ( $Z=-2.484, p=.013$ ) or auto freeze ( $Z=-2.762, p=.006$ ) conditions than using non-freeze condition (Kim et al., 2015).

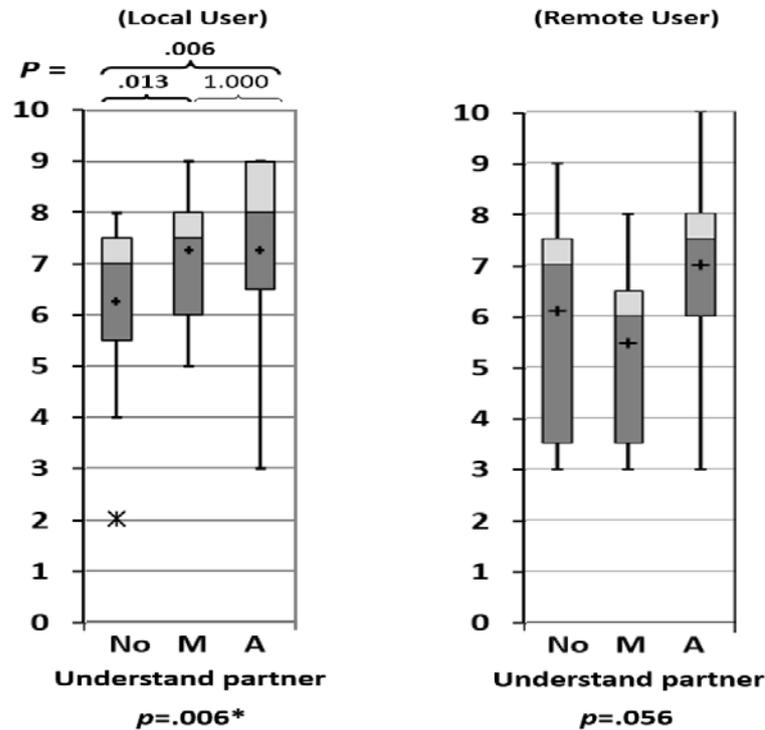


Figure. 6.7: Results of Likert scale ratings from local (left) and remote (right) participants for the question about understanding what their partners were trying to do and explaining (0: strongly disagree ~ 10: strongly agree; \*: statistically significant difference, +: mean, x: outlier, No: non-freeze, M: manual freeze, A: auto freeze).

From the interview, the author found that the issue of drawing annotation on a live video (without freeze function) appeared to affect the experience of the local participants. L1 (local user in group 1), 4, 6, and 12 mentioned it was sometimes difficult to understand the remote partner's drawing with the non-freeze condition. As described in the previous section 4.3.3, verbal messages were alternative for it by remote participants, but these seem like less clear

than the correct piece-shape drawing for positioning pieces. The freezing techniques may solve the issue of the annotation drawing while the viewpoint of the shared live video is changed. Being 'clear to understand' (L 9 and 10 with manual freeze, L1, 6, 8 and 12 with auto freeze) were reported as the benefits of the manual and auto freeze conditions.

### 6.2.5 Overall Collaboration

Generally, all collaboration started from the one end of the target model rather than from the middle of model, and participants mostly tried to continue the assembling beside to the previously completed part. If they could not find a proper piece for the next, they kept assembling from the other end of the model.

To see how the different remote participants' encoding experiences had effect on overall remote collaboration, the author also prepared question 6, 7, and 10. Figure 6.8 shows the results comparing the three conditions using the Likert scale ratings.

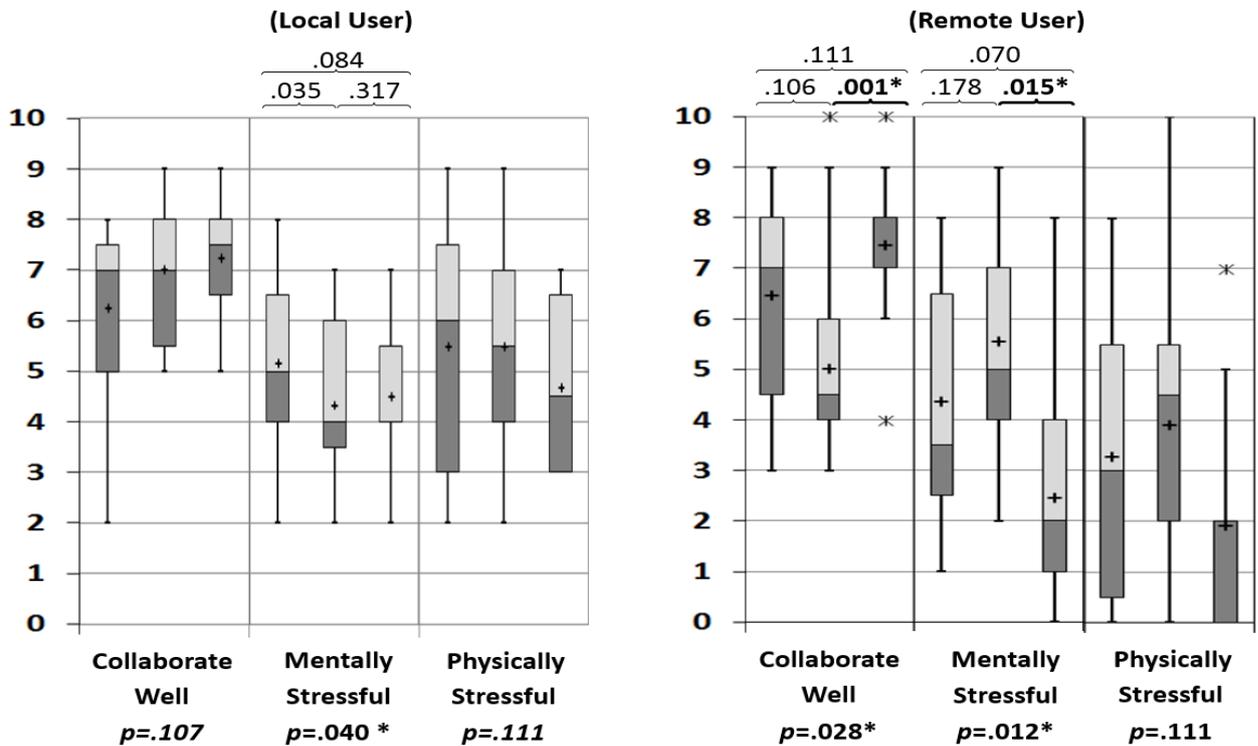


Figure 6.8: Results of Likert scale ratings from local (left) and remote (right) participants for the question about how easily they understood what their partners were trying to do and explaining (0: strongly disagree ~ 10: strongly agree; \*: statistically significant difference, +: mean, x: outlier, No: non-freeze, M: manual freeze, A: auto freeze).

With the Friedman test, the local participants' ratings about 'Collaborate Well' ( $\chi^2(2)=4.471$ ,  $p=.107$ ) and 'Physically Stressful' ( $\chi^2(2)=1.086$ ,  $p=.581$ )

did not show a significant difference between the three conditions, while the ratings about ‘Mentally Stressful’ did ( $\chi^2(2) = 6.462, p=.04$ ). In pair wise comparison, none of pair were significantly different for ‘Mentally Stressful’. These would mean that local participants did not feel significant difference between the conditions in those aspect.

With the Friedman test, the remote participants’ ratings about ‘Collaborate Well’ ( $\chi^2(2)=7.136, p=.028$ ) and ‘Mentally Stressful’ ( $\chi^2(2)=8.844, p=.012$ ) showed significant difference among the conditions, while did not for the ratings about ‘Physically Stressful’ ( $\chi^2(2)=4.389, p=.111$ ). Pair wise comparisons showed a significant difference between the manual freeze and auto freeze conditions in question about ‘Collaborate Well’ ( $Z=-2.591, p=.01$ ) and ‘Mentally Stressful’ ( $Z=-.2.439, p=.015$ ), but did not show a significant difference in the other pair wise comparisons. The remote participants felt that they had better collaboration with auto freeze condition than with manual freeze condition while having higher level of mental stress with manual freeze than auto freeze condition. From the interview, R1, 2, and 8 mentioned that manual freeze was mentally stressful because it required the additional inputs for freezing and unfreezing the shared view.

After trying all the conditions, participants chose their most and least preferred conditions (Kim et al., 2015)(see Figure 6.9). The author analyzed the remote ( $N=12$ ) and local ( $N=12$ ) participants' preferences separately.

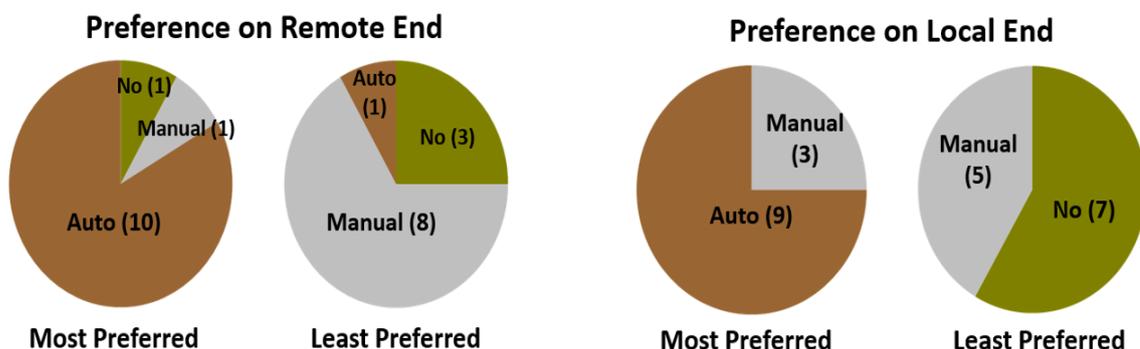


Figure 6.9: User preference among the three conditions (most preferred = 1, least preferred = 3, No: non-freeze, M: manual freeze, A: auto freeze).

Ten remote participants (83.3%) picked auto freeze as the most preferred condition, while each of the rest most preferred the manual freeze (8.3%) or non-freeze condition (8.3%). For the least preferred, eight remote participants (66.6%) selected manual freeze, while three of the rest picked non-freeze (25%) and one (8.3%) picked the auto freeze condition. Comparing the ranking based on participant's preference (most preferred = 1, least preferred = 3), a Friedman test showed that the three conditions were significantly different ( $\chi^2(2)=11.167$ ,  $p=.004$ ), and post hoc tests showed auto freeze was significantly preferred over manual freeze ( $Z=-2.769$ ,  $p=.006$ ), but no significant difference was found between auto and non-freeze ( $Z=-2.309$ ,  $p=.021$ ) and manual and non-freeze conditions ( $Z=-1.291$ ,  $p=.197$ ).

On the local end, nine participants (75%) chose auto freeze as the most preferred condition, while three participants (25%) picked the manual freeze. For the least preferred condition, seven and five local participants chose non-freeze (58.3%) and manual freeze (41.6%) respectively. A Friedman test showed there was a significance difference between the three conditions in terms of the user's preference ( $\chi^2(2)=11.167$ ,  $p=.004$ ), and post hoc tests showed auto freeze was significantly preferred compared to non-freeze ( $Z=-2.769$ ,  $p=.006$ ), while there was no significant difference between the non-freeze and manual freeze conditions ( $Z=-1.155$ ,  $p=.248$ ) and between the manual freeze and auto freeze conditions ( $Z=-2.183$ ,  $p=.029$ ).

To sum up, remote participants preferred auto freeze condition compared to manual freeze condition, and local participants preferred auto freeze condition compared to non-freeze condition.

### **6.3 Summary of Freeze Function and an Issue on the Local End**

Throughout the user study, wrong drawing occurred in the non-freeze condition whenever a remote participant drew an annotation while the local participant was changing the viewpoint of the shared view. Freezing the

shared view solved the issue, but the manual freeze condition required two additional interactions; freeze and unfreeze the view, which required more time. The auto freeze technique provides the benefit of freezing the shared view for accurate drawing without needing additional user input yet still promoting the live viewing experience of shared video. Auto freeze was the most preferred condition for both the remote and local participants, since auto freeze provided better communication between the local and remote participants with the drawing interaction.

A local participant (L7) commented about how she managed turn taking with her partner. She would try out her own idea if she was holding a piece that she needed, and follow her partner's drawing if she did not have a piece in her hand. This suggests that she sometimes ignored her partner's drawing on purpose when she had a piece in her hand. This intentional annotation ignorance would be the local participants' self-solution, as a notification management (Interruption handling strategies, Li et al., 2012), to avoid losing track of the ongoing task.

However, according to observations, she sometimes missed her partner's drawing even when not holding a piece due to not recognizing new drawings. The author observed this happen not only in group 7, but also in most of the groups. This could be because of the local participants mostly focusing on the center of the shared view shown on the HMD, while the remote participants could easily have their focus on other parts of the shared view shown on the desktop monitor. In other words, the drawings on the periphery of the shared view were natural to the remote user, but much less so to the local users.

With this unintentionally missed annotations, local users were consequently unable to apply their notification management on all the drawings from their remote partners. Moreover, a local user is the only one who can manipulate pieces in the task. If a local user (intentionally or unintentionally) ignores the remote user's participation, it is not collaboration but is the task doing alone. To promote collaboration, the interface should help local users not to fall into

conducting the task alone. In the next section, the author investigates interfaces that help promote collaboration with better notification management. As a solution to this problem, R2 and R7 suggested freezing the local user's view as well when the remote user freezes the shared view.

## **6.4 User Study on Visual Notification**

While focusing on variations of freeze techniques on the remote end in the second study, the author investigated the effect of different visual notifications on the local end in this third study. In the second study the author noticed local participants sometimes missing the remote user's drawings. In the third study, the author explores into whether a more obvious visual notification solves this issue and helps users notice the remote user's drawings. For this purpose, the author conducted a user study comparing three conditions with different visual notification methods: (1) no notification (as the baseline condition), (2) red box notification, and (3) both freeze condition.

### **6.4.1 Previous Studies in Notification**

The role of notifications in remote collaboration has not been extensively studied. In this section, the author reviews previous notification studies in remote collaboration and psychology.

Recently, Gauglitz et al. (2014a, 2014b, 2012) introduced visual notifications with an arrow shaped virtual object. The arrow shows where a remote user is drawing an annotation and helps a local user to effectively follow the remote user's instructions. However, this form of notification was designed for remote expert collaboration where the local user is only required to follow remote user's drawing. In parallel experience collaboration, the local user has to decide between following the remote user's drawing and manipulating pieces with his/her own ideas, so Gauglitz's visual notification design may not be suitable for parallel experience collaboration where the two processes must be managed by the local users.

Managing the two processes (following the remote users' drawings and

manipulating pieces with their own ideas) requires the local users to allocate their limited attentional resources between the two processes. This allocation process is defined as interruption management and includes detection, interpretation, and integration of interruptions with ongoing task performance (Latorella 1999).

Traditionally, interruptions are presented visually, but the development of new tactile and auditory technologies makes it possible to use non-visual channels (Lu et al. 2013). Li et al. (2012) found that notifications presented in a different modality from the primary task reduce disruption in task performance. However, Wickens et al. (2011) reported that presenting the same message simultaneously in more than one sensory channel results in competition for attention resources, hence can be slower than using a single modality. Posner (1980) reported that peripheral visual cues can be perceived in parallel with foveal cues and Hameed et al. (2009) reported that peripheral visual notification was effective despite using a visual modality with an ongoing visual task.

While the studies above focused on interruption modality, Edwards (Edwards & Gronlund 1998) and Gillie (Gillie & Broadbent 1989) studied interruption similarity. If an interruption is similar to the ongoing task, recovery back to the ongoing task is more difficult (Edwards & Gronlund 1998) and leads to low task efficiency (Gillie & Broadbent 1989). In addition to the interruption similarity, when a notification occurs is considered as a factor influencing interruption management (Adamczyk & Bailey 2004). Demands on cognitive resources are more intense during task execution than in between tasks, so notifications during execution are more likely to be disruptive than those occurring in between tasks. If this is so, informative interruption cuing needs to be context sensitive and sufficiently salient without being disruptive (Sarter, 2002; Woods, 1995).

According to this research, drawing an annotation itself may not be the best method for notification because it requires foveal vision attention that could

be in use for the ongoing task, may have a high similarity with the ongoing task (object manipulation), and could include too much information to be handled while focusing on the ongoing task. As an alternative, the author designed a peripheral notification that uses a red outline around the screen, and conducted a user study to compare it with other alternatives. To the best of the author's knowledge, this is the first formal user study on visual notification methods in an AR remote collaboration interface.

#### **6.4.2 User Interface**

The author adopted the same system configurations as in the first and second studies except for the slightly modified user interfaces as needed. For the remote end interface, the author used the auto freeze interface from the second study that had been the condition most preferred by the users. One improvement made was that instead of immediately returning to the live video after finishing drawing, the system kept the frozen view for two more seconds. This modification was added based on the findings from the first study that sometimes users wanted to draw more on the same frozen view rather than returning immediately to the live view.

While the remote participants used the auto freeze interface throughout the experiment, the user interface for the local end had three different visual notification methods. The first was the baseline condition where no notification was provided. This was the same interface as used in the second user study.

The second condition, named 'red box', showed a virtual red outline around the shared view when the remote user's view was frozen for drawing annotations (see Figure 6.10). The red box condition was based on Hameed's study (Hameed et al. 2009) which reported the effectiveness of a peripheral visual notification when the visual modality of foveal vision is in use for an ongoing task (manipulating puzzle pieces in this study). While designing this notification method, the author considered three requirements: (1) the notification should reveal information about the annotation drawn, (2) the

notification should not require focused attention or disturb the ongoing task (Woods, 1995), and (3) the notification should be easily recognized by the user (peripheral visual notification). The red box notification presents the timing information of when the drawing is happening, appears in the periphery of the shared view, and is easily noticeable because of its colour and size (much bigger than a drawing because it covered every side of the screen).

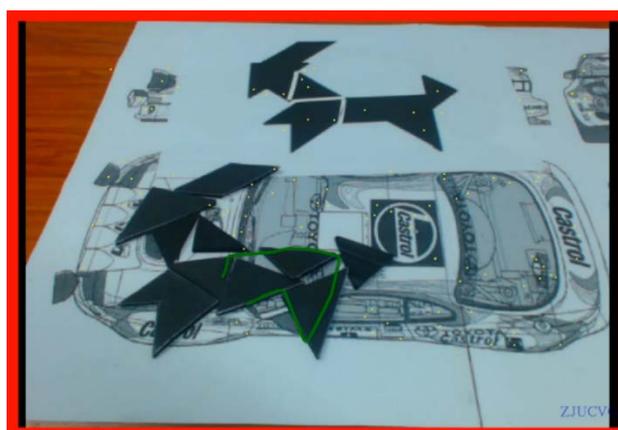


Figure. 6.10: Local user view when the remote user drew in the red box condition. The remote user is pointing out a part that is not matched with the reference.

The notification method used in the third condition, named ‘both freeze’, was freezing the local user’s view together with the remote user’s view. This both freeze condition was designed based on participants' suggestions from the second study. To prevent the local users from thinking the frozen view was a system malfunction, the both freeze condition also showed the red outline around the frozen shared view, as in the red box condition. This also made the condition directly comparable to the red box condition, with the only difference being whether the view was frozen or not.

### 6.4.3 User Study Design

The author used the same experimental task (solving Tangram puzzles) and procedure as in the first and second studies. The participants solved Tangram puzzles in pairs while located in separate rooms and using the prototype system for communication. The experiment started with a questionnaire asking demographic information and then the purpose of the study was

explained. The participants were asked to solve a sample Tangram puzzle face-to-face, then the author showed and explained how to use the prototypes. Participants performed three sessions under different experimental conditions: (1) no notification, (2) red box notification, and (3) both freeze. The order of the conditions was counterbalanced using a Latin square design, and each session consisted of two minutes training and five minutes of solving the Tangram. After each session, the participants answered a questionnaire and had a brief interview. When all sessions had been completed, the experiment concluded with the final interview. Overall the experiment took about 70 minutes for each pair of participants (55 minutes for the experiment and 15 minutes for the interviews).

Similar to the first and second studies, the author collected subjective feedback (questionnaires and interviews), activity logs of remote participants, and video recordings of each screen. The questionnaire included nine Likert-scale rating questions for the remote participants and ten questions for the local participants (see Table 6.3).

Table 6.3. Questions in the questionnaire after each session (Q7, highlighted in gray, was only for local participants).

Q1	I was able to focus on the activity (building a model).
Q2	I felt interrupted when my partner was drawing (or I felt I was interrupting my partner when I was drawing)
Q3	I was able to express my idea properly
Q4	I easily understood what my partner was trying to do and explaining.
Q5	I was able to focus on my partner's explanation. (I felt my partner was focusing on my explain)
Q6	I felt we collaborated well
Q7	I knew the moment when my partner wanted to explain his idea
Q8	The communication with my partner was mentally stressful
Q9	We focused on the same piece or area while collaborating
Q10	I was aware of where my partner was drawing (or My partner was aware of where I was drawing)

For the questions, the author also considered the main factors of remote collaboration (both users who encode and decode messages, channels, and the situation model) except channels like the one in the first study. Since the situation model is the solving of an object manipulation task in our case, the author had three questions about their focusing (Q1, Q5, and Q9) on the task. Since the purpose of our study is comparing three different local end interfaces in decoding messages, the author prepared three questions (Q2, Q4, and Q10) for both users and one question (Q7) only for local users about their experience of decoding messages. For the experience of sending (encoding) messages question 3 was prepared. To see how the different decoding experiences have effect on overall remote collaboration, the author also prepared questions 6 and 8.

## **6.5 Results of the Notification Study**

The author recruited another twenty four participants (in pairs) who had been using video conferencing regularly (more than once a month). Participants in each pair knew each other well as friends or family. There were 21 males and 3 females with ages ranging from 20 to 38 years old ( $M = 26.6$ ;  $SD = 4.6$ ).

The main findings from this user study are listed below:

- 1) The red box condition helped local participants allocate their attention more effectively between doing their own task and paying attention to their partner.
- 2) The local participants felt a low level of interruption with the red box condition but a high level of interruption with the both freeze condition.

As in the previous study, the author analyzed the data according to three main factors of remote collaboration (shared situation task, encoding and decoding messages). To analyze the Likert scale ratings of the questionnaires between the three conditions, the author ran the Friedman Test ( $\alpha = .05$ ), and for those showing a significant difference between the three conditions ran post hoc

tests for pair wise comparison using the Wilcoxon Signed-Rank Test with Bonferroni correction applied ( $\alpha = .0167$ ). For deeper understanding of the results, the author listed the related answers from interviews and analyzed the video recordings and log data as supplements.

### **6.5.1 Focusing on the task**

For the questions about focusing (Q1, Q5, and Q9), figure 6.11 shows the results comparing the three conditions using the Likert scale ratings. With Friedman test, the local participants' ratings about focusing on the assembling task (Q1) did not show significant difference between the three conditions ( $\chi^2(2) = 2.250, p = .325$ ), but the remote participant's ratings for Q1 showed significant difference between the three conditions ( $\chi^2(2) = 8.000, p = .018$ ). Pair wise comparisons with the remote participants' ratings showed a significant difference between the both freeze and no notification condition ( $Z = -2.102, p = .014$ ) but other comparison did not show significant difference. These would mean that remote participants felt they focused significantly more with both freeze condition than no notification condition. In the interview, R10 mentioned that his partner focused on his drawings more with the both freeze condition, and R6 and R8 reported that their partner waited for them to finish drawing with the both freeze condition. Local participants also commented on the benefits of the both freeze condition as 'more focused on drawings' (L 1, 3, and 11).

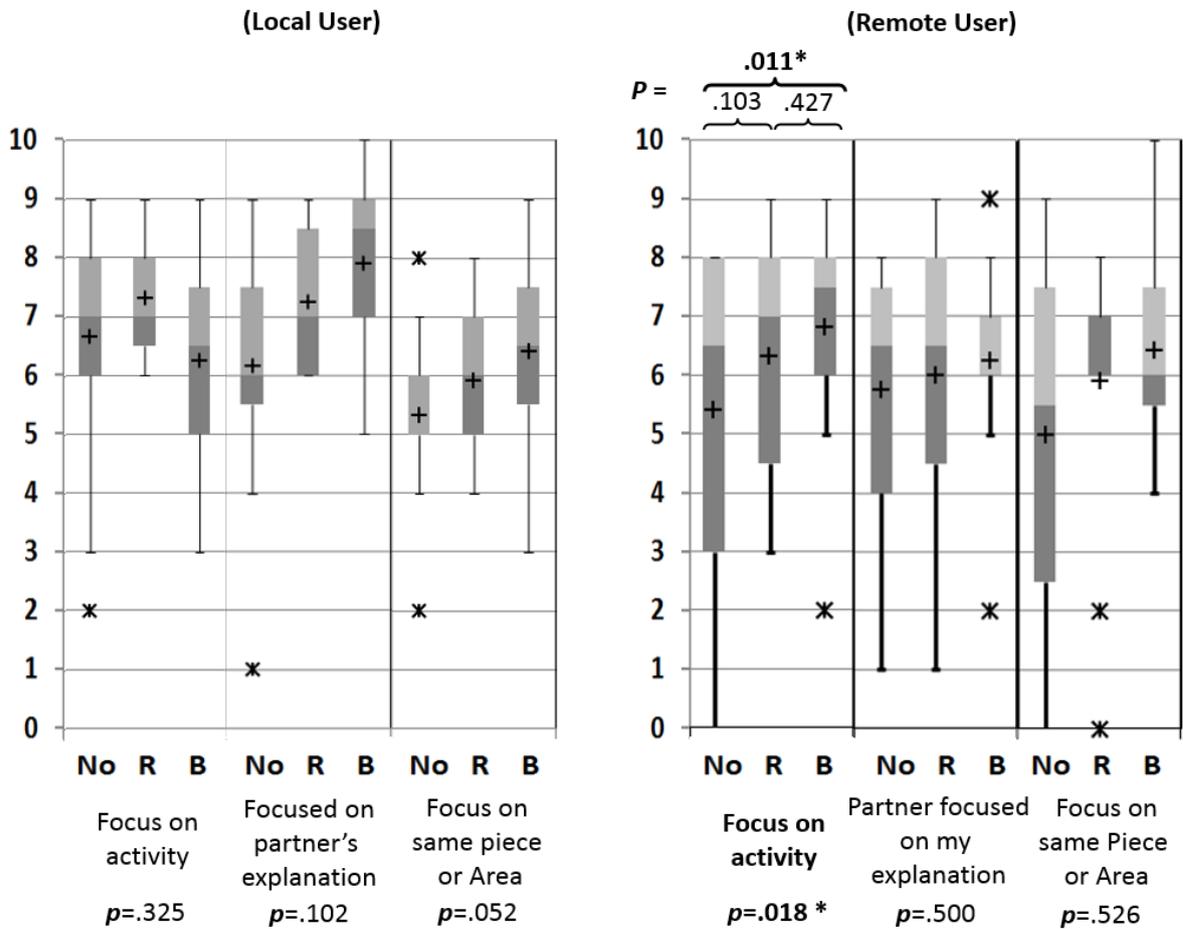


Figure 6.11: Results of Likert scale ratings from local (left) and remote (right) participants for the question 1, 5, and 9 (0: strongly disagree ~ 10: strongly agree; \*: statistically significant difference, +: mean, x: outlier, No: no notification, R: red box, B: both freeze).

The Friedman tests with ratings about focusing on the partner's explanation did not show significant difference (remote participants:  $\chi^2(2) = 1.385$ ,  $p = .500$ , local participants:  $\chi^2(2) = 4.562$ ,  $p = .102$ ). For the question about focusing on the same piece or area, both participants' ratings did not show significant different in Friedman tests as well (remote participants:  $\chi^2(2)=1.286$ ,  $p = .526$ , local participants:  $\chi^2(2)=5.895$ ,  $p = .052$ ).

Even though there was not significant difference, the author found one interesting point. For focusing on the activity the local participants gave lower points to the both freeze condition than others while the both freeze condition

took averagely highest points compared to other conditions in focusing on partner’s explanation and focusing on the same piece or area.

### 6.5.2 Receiving/Decoding Messages

The author prepared three questions (Q2, Q4, and Q10) for both participants and one question (Q7) only for the local participants about the local participants’ decoding experience. Figure 6.12 shows the results comparing the three conditions using the Likert scale ratings.

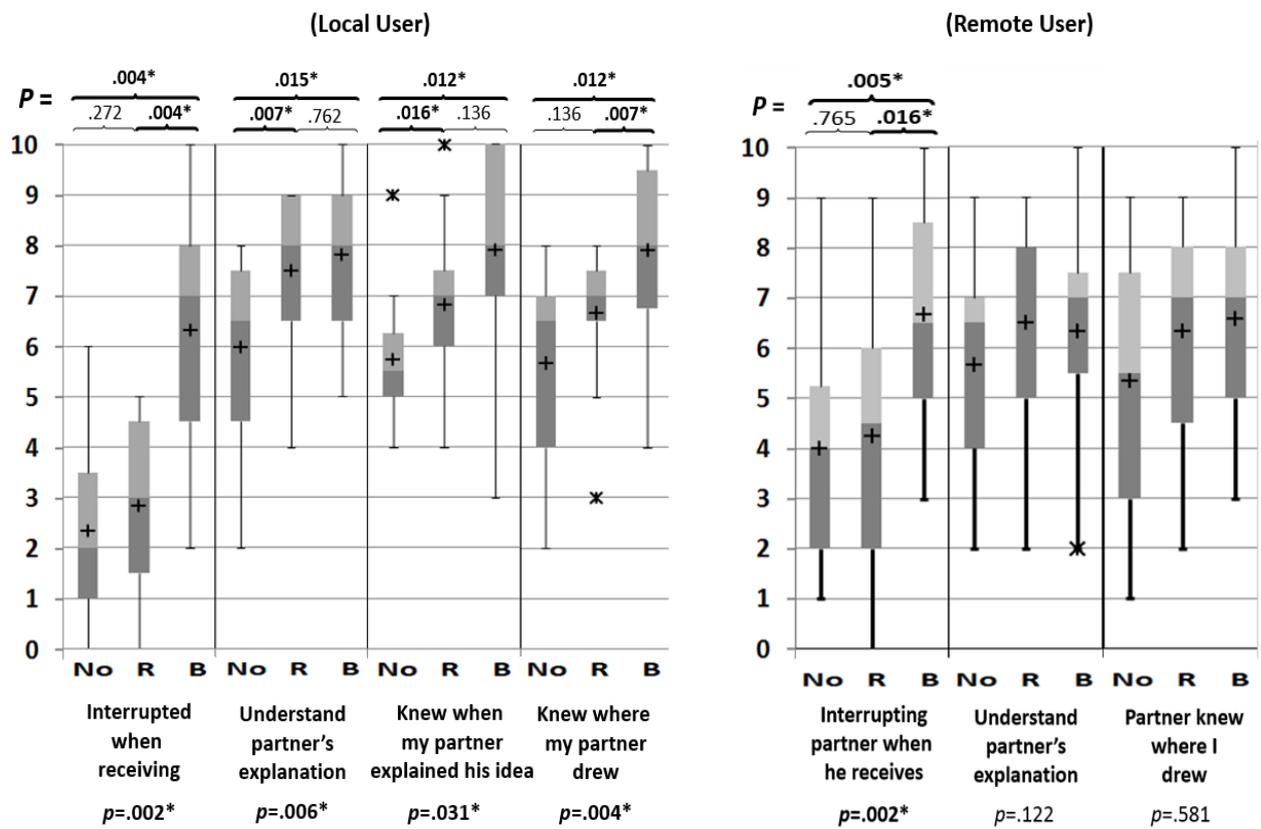


Figure 6.12: Results of Likert scale ratings from local (left) and remote (right) participants for the question 2, 4, 7, and 10 (0: strongly disagree ~ 10: strongly agree; \*: statistically significant difference, +: mean, x: outlier, No: no notification, R: red box, B: both freeze).

In Friedman test, local participants’ ratings showed significant difference between the three conditions in question about ‘got interrupted when the partner drew’ (Q2:  $\chi^2(2)=13.818, p =.002$ ), ‘understood partner’s explanation’

(Q4:  $\chi^2(2)=10.333, p =.006$ ), 'knew the moment when the partner drew' (Q7:  $\chi^2(2)=6.950, p =.031$ ), and 'knew where the partner drew' (Q10:  $\chi^2(2)=11.167, p =.004$ ).

Pair wise comparison with post hoc tests showed significant differences between the no notification and red box conditions (Q4:  $Z=-2.716, p=.007$ ; Q7:  $Z=-2.412, p=.016$ ) and between the no notification and both freeze conditions (Q4:  $Z=-2.434, p=.015$ ; Q7:  $Z=-2.522, p=.012$ ) in question 4 and 7. However, there was no significant difference between the red box and both freeze conditions. In question 2 and 10, significant differences were found between the both freeze and no notification conditions (Q2:  $Z=-2.871, p=.004$ ; Q10:  $Z=-2.499, p=.012$ ) and between the both freeze and red box conditions (Q2:  $Z=-2.919, p=.004$ ; Q10:  $Z=-2.714, p=.007$ ) in question 2 and 10. However, there was no significant difference between the no notification and red box conditions in the same questions.

These results would mean that the local participants felt they understood their partner's explanation better with the red box condition than with the no notification condition while knowing when the remote partner drew annotation. With the both freeze condition, the local participants felt the same benefits of the red box condition (better understanding their partners' explanation and better knowing when the remote partners drew annotation), but felt one more benefit, knowing where the annotation was drawn compared to no freeze condition.

In the interview, the author found the corresponding comments to the results from the questions. L5, 7, 10, 11, and 12 mentioned that they knew when their partner was drawing with the red box condition: "*Visual notification of drawing helped me know when he started drawing*" (L11). Similar comments were found for the both freeze condition by L3, 6, 7, 10, and 12, but they added that they easily knew what their partner was doing or where their partner drew, in addition to knowing when. This benefit was also found from the comments of remote participants. Regarding the red box

condition, R3 mentioned his perceived impression, "*Highlighting the working space let my partner know when I drew*". With the both freeze condition, R12 reported his perceived second hand impression, "*Freeze view possibly got my partner to look closely at my drawings*"

However, local participants felt that they got significantly more interrupted with the both freeze condition than with the other two conditions (no notification:  $Z=-2.871$ ,  $p=.004$ , and red box:  $Z=-2.919$ ,  $p=.004$ ). The remote participants also felt that they were significantly more interrupting local partners with the both freeze condition than the other two conditions (red box:  $Z=-2.814$ ,  $p=.005$ , and no notification:  $Z=-2.409$ ,  $p = .015$ ).

With the both freeze condition, the remote participants tended to lead the collaboration by capturing the visual modality of the local participants and forcing them to focus on their drawing. This could be an effective way of transferring spatial information through visual communication cues from a remote user to a local user. However, two thirds of remote participants (R1, 2, 3, 5, 6, 7, 10, and 12) reported that they felt they interrupted their partner and were being rude when the view froze once that started drawing. Moreover, all of the local participants complained about being interrupted and how they could not do anything but only watched the remote participant's explanation. The author observed that the local participants did not even make small movements while they were viewing the frozen view. In the case of the red box condition, interruptions were not an issue because the view did not freeze. The local participants were able to keep doing what they did while perceiving whether the remote partner was currently drawing or not.

To sum up, local participants felt better understanding partner's explanations while knowing when drawing annotation with red box and both freeze conditions compared to the no notification condition. However, both freeze condition had an issue of interrupting local participants. Next, what the local participants' self-solution was for the lack of notification with no notification condition, what the reaction of the participants was when being

interrupting or interrupted, and what the effect of these results was on the local participants' ignorance are described.

### **6.5.3 Verbal Communication & User Behavior**

From our observation the drawing usually came with verbal communication from remote participants simultaneously, and local participants sometimes used the verbal cue as the notification in the no visual notification condition. The author also found local participants' comments relating to this, "*I knew when he was drawing from his speech but the red box in test was clearer in this regard*" (L12). However, drawing did not always start with verbal communication from the remote participant, since they typically used speech to provide more information about their drawings and not to notify the local participant that they were starting to draw.

In addition to the verbal communication for the task, local participants verbally reacted to the suddenly and unexpectedly frozen view mostly with a monosyllabic voice (such as 'Oh'). This monosyllabic voice would let remote participants know that the both freezing the view significantly interrupted the local partner. One interesting observation in the experiment while participants using both freeze condition was that R7 did not use the drawing annotation anymore after freezing the shared view (maybe after he figured out that both freeze condition severely interrupted his partner), but verbally communicated with their partner. According to our observation, other remote participants also tried to use both freeze method less after they figured out the interruption to the local partners. Since the author could not find when the remote participants figured out the interruption of both freezing, the author counted the number of both freezing in the first half and the second half of the collaboration task. Even though the author did not find the significant difference in the use of the drawing annotation between the first half and the second half of the collaboration with the both freeze condition (Wilcoxon Signed-Rank test,  $Z=-1.679$ ,  $p=.093$ ), 58.33 percent of both freezing were conducted in the first half while 41.66 percent of both freezing were

conducted in the second half.

When they started the experimental task, all local participants arranged three areas on the desk. The two areas were for the unused pieces and a reference paper, and the other one area was an empty area for assembling a model. They arranged the empty area closest to them and the pieces and reference paper were placed above or next to the empty area. From the video recordings, the author counted how many time the remote participants explained in different area rather than the area where local partners focused on and how many time the local participants ignored the remote partner's explanation. Remote participants suggested ideas in a different area averagely 5.17 times, 4.82 times, and 5.08 times with the no notification, red box, and both freeze conditions respectively. In Friedman test, there was not significant difference between three conditions for the suggesting in a different area.

From the recorded videos, the author found that 21.96 percent, 11.36 percent, and 5.3 percent of remote participants' suggestions were ignored by the local participants with no notification, red box, and both freeze conditions respectively (the ignorance with the both freeze condition was only occurred when the remote participants verbally suggested ideas while the ignorance was occurred regardless of the types of the used communication cues with no notification and red box conditions). In Friedman test with the percent, there was significant difference between the three conditions ( $\chi^2(2)=13.762$ ,  $p=.001$ ). In pairwise comparison with Wilcoxon Signed-Rank test, the author could not find the significant difference for the local participant's ignorance between no notification and red box conditions ( $Z=-2.365$ ,  $p=.018$ : but it was very close to the level of significant,  $p=.016$ ) and between red box and both freeze conditions ( $Z=-2.309$ ,  $p=.021$ ), while the author found the significant difference between no notification and both freeze conditions ( $Z=-2.840$ ,  $p=.005$ ). This would mean that fewer remote participant's suggestions were ignored the both freeze conditions than the no notification condition. Moreover, (even though the percent did not show the significant difference)

less percent of remote participants' suggestions was ignored with the red box condition compared to the no notification condition.

#### 6.5.4 Sending/Encoding Messages

For the question about properly expressing idea (Q3), figure 6.13 shows the results comparing the three conditions using the Likert scale ratings. Friedman test with the local and remote ratings for the question did not show significant difference (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 significant difference in properly expressing idea.

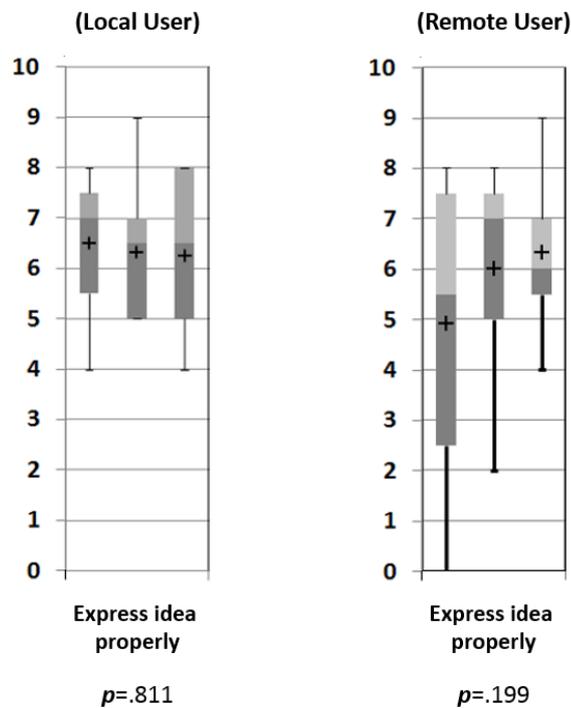


Figure 6.13 Results of Likert scale ratings from local (left) and remote (right) participants for the question about properly expressing ideas (0: strongly disagree ~ 10: strongly agree; \*: statistically significant difference, +: mean, x: outlier, No: no notification, R: red box, B: both freeze).

Through logging data on the remote end, the author collected the number of strokes the remote participants made, and the number of strokes during the additional two seconds (see Table 6.4).

Table 6.4: Results of the number of strokes, and the percentage of strokes in two second

	Mean			Standard Deviation		
	No	Red	Both	No	Red	Both
<b>Strokes</b>	11.1	10.9	14	7.1	6.6	4.2
<b>Percentage of strokes in two seconds</b>	21.6%	29.29%	38.1%	12%	14%	10%

With a Friedman test the author did not find significant difference between the three conditions in terms of the number of strokes ( $\chi^2(2)=2.085, p=.353$ ). However, the author did find a significant difference in the percentage of strokes made during the additional two seconds ( $\chi^2(2)=10.255, p=.006$ ). In the pair wise comparison between the both freeze and no notification conditions ( $Z=-2.654, p=.005$ ) had a significant difference, and no significant difference was found between the both freeze and red box conditions ( $Z=-1.961, p=.050$ ) and between no notification and red box conditions ( $Z=-2.045, p=.041$ ). This means that remote participants drew more in the both freeze condition during the additional two seconds than in the no notification and red box conditions and, they stayed in the encoding or sender state longer with the both freeze condition than with the other conditions while the local partner stayed in the decoding or receiver state.

In the previous section 5.3.2, the author mentioned that both freeze condition would be an effective way of transferring spatial information through a visual communication cue from a remote user to a local user by capturing the visual modality of the local participants. This may have an effect on drawing more annotations during the additional two seconds with the both freeze condition than in the no notification condition.

### 6.5.5 Overall Collaboration & User Preference

To see how the different local participants' decoding experiences have effect on overall remote collaboration, the author also prepared question 6 and 8. Figure 6.14 shows the results comparing the three conditions using the Likert scale ratings.

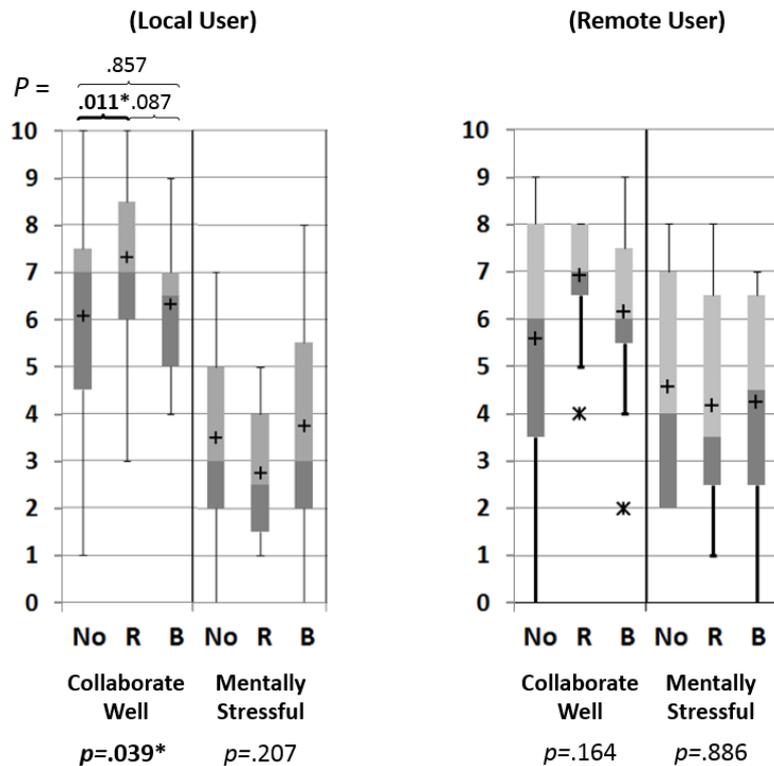


Figure 6.14: Results of Likert scale ratings from local (left) and remote (right) participants for the question about collaborating well and interface mentally stressful (0: strongly disagree ~ 10: strongly agree; \*: statistically significant difference, +: mean, x: outlier, No: no notification, R: red box, B: both freeze).

With the local participants' ratings, the Friedman test showed significant difference between the three conditions for the question about collaborating well (Q6:  $\chi^2(2)=6.500$ ,  $p =.039$ ), while it did not for the question about interface mentally stressful (Q8:  $\chi^2(2)=3.152$ ,  $p =.207$ ). In pair wise comparison for Q6, the red box condition showed a significant difference compared to the no notification condition ( $Z=-2.549$ ,  $p=.011$ ) in question 6, but no significant difference was found between the no notification and both freeze conditions, nor between the red box and both freeze conditions. This means that the local participants felt they collaborated significantly better with the red box condition compared to with no notification condition.

With the remote participants' ratings, the Friedman test did not showed a significant difference between the three conditions for the question about collaborating well ( $\chi^2(2)=3.619$ ,  $p =.164$ ) and interface mentally stressful

( $\chi^2(2)=1.286, p=.526$ ).

The author also analyzed the user preference (see Figure 6.15). From the twelve local participants, eight of them (66.6%) picked red box as the most preferred condition while the rest were split equally between the both freeze (16.6%) and no notification (16.6%) conditions. For the least preferred, eight local participants (66.6%) selected the both freeze condition, while the rest of them (33.3%) picked the no notification. A Friedman test (most preferred = 1, least preferred = 3) showed a significant difference between the three conditions ( $\chi^2(2)=8.667, p=.013$ ). A pair wise comparison showed that they significantly preferred the red box over both freeze ( $Z=-2.581, p=.010$ ), but no significant difference was found between the red box and no notification conditions ( $Z=-2.352, p=.019$ ) and between the both freeze and no notification conditions ( $Z=-.733, p=.464$ ).

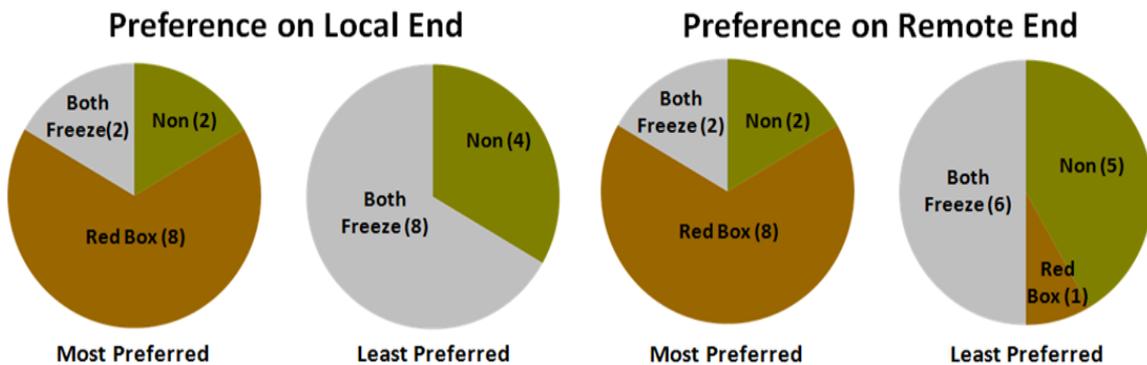


Figure. 6.15: Participants' preference among the three conditions

Eight of the remote participants (66.6%,  $N=12$ ) chose red box as the most preferred condition, while the rest were split equally between the both freeze (16.6%) and no notification (16.6%) conditions (see right of Figure 6.14). For the least preferred, half of remote participants (50%) chose the both freeze condition, while five (41.6%) and one (8.3%) of the other half selected no notification and red box, respectively. A Friedman test found a significant difference between the three conditions ( $\chi^2(2)=6.167, p<.046$ ), but no significant difference were found in the pair wise comparison.

From the video recordings, the author counted how many piece-manipulation participants were collaborative or one of them took the charge of the manipulation alone. The author categorized the piece-manipulation into three types: (1) collaborative, (2) local user alone, and (3) remote user alone. If there was not any suggestion for a piece-manipulation but only acceptance from a partner, the author considered it as one of manipulation alone types according to who led it. (Accepting the other user's idea could be considered as collaboration, but the author did not count it as a collaborative type for better understanding the use of the experimental conditions. Moreover, we did not count the abandoned piece-manipulation). Figure 6.16 shows how much percentage participants positioned pieces among the three positioning types in each condition.

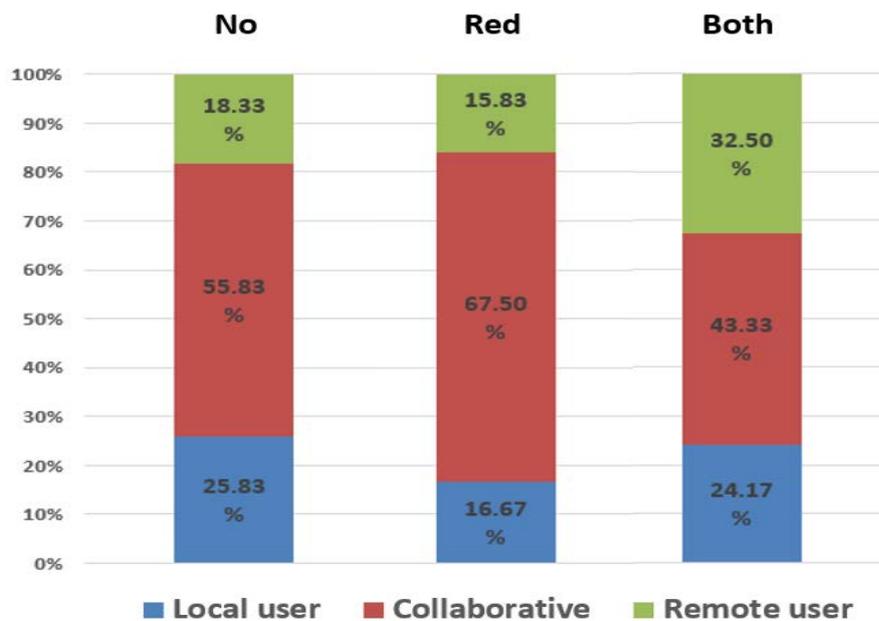


Figure 6.16: The average of the percentages among piece-manipulation types (local user alone, collaborative, remote user alone) in each condition (No: no notification, Red: red box, Both: both freeze).

With the percentages, the author ran the Friedman Test ( $\alpha = .05$ ) to compare the three conditions, and for those showing a significant difference between the three conditions the author ran post hoc tests for pair wise comparison using the Wilcoxon Signed-Rank Test with Bonferroni correction applied ( $\alpha$

= .0167). Friedman test did not show significant difference in local participant positioning alone between three conditions ( $\chi^2(2)=4.350, p =.128$ ), but it did in both participants collaboratively positioning ( $\chi^2(2)=13.762, p =.001$ ) and remote participant positioning alone ( $\chi^2(2)=8.333, p =.016$ ).

In pair wise comparison, participants positioned pieces significantly more collaboratively with the red box condition than with both freeze condition ( $Z=-2.931, p=.003$ ) while there were not significant difference in being collaborative between no notification and red box conditions ( $Z=-2.095, p=.036$ ) and between no notification and both freeze conditions ( $Z=-2.030, p=.042$ ). Similar with these results, participant gave the comments about suitability of collaborating together with red box condition. Five local participants (L1, 2, 3, 6, and 7) mentioned that they were able to collaborate while effectively allocating their focus between their partner's drawings and what they did. L1 commented that he was able to choose between paying attention to his partner's drawing and doing his own work. Similarly, R7 and 10 reported that their partner focused on drawings and also kept track of what they were doing. R11 was more analytic, "With condition of freeze (the both freeze condition) or none (non-freeze condition), I felt one of us took charge of the puzzle. In the condition with red box, it felt like a more balanced collaborative work".

On the contrary, remote participants positioned pieces alone significantly more with the both freeze condition than with the red box condition ( $Z=-2.694, p=.007$ ) while other comparisons did not show significant difference. This would be because remote participants captured the focus of the local user by freezing the shared view.

To sum up, the local participants felt they collaborated significantly better with the red box condition and (even though it is our definition of being collaborative or not) the participants were in collaborative state more with the red box condition than with the both freeze condition while they were in remote user alone state more with the both freeze condition than the red box

condition.

## **6.6 Summary of Notification Study**

From the questionnaire, local participants felt that red box and both freeze conditions had the benefits of understanding the remote participants' explanation and knowing when the remote participants drew annotation. However, the remote participants interrupted the local participants doing ongoing task by capturing the visual modality of the local participants. The both freeze condition significantly reduced the local participants' ignorance for the remote participants' suggestion but it did not help participants to be in collaborative state but in remote user alone state. On contrary, the author found that the red box notification helped local users allocate their attention more efficiently between doing their own task and listening to their partners rather than ignorance, and participants became more collaborative. This was achieved by making the notification appear in the peripheral view, and not freezing the local participant's view, hence not interrupting the local participant's ongoing task.

While the both freeze condition was highly interruptive, making it the least preferred condition, the author found that it still could be useful in certain situation where the remote user needs to grab the full attention of the local user, e.g. in emergency.

## **6.7 Discussion**

During the experiment the author observed that the drawing interaction was very intuitive. All of the remote participants found it easy to use the mouse because they were already familiar with it. None of the participants mentioned the incongruity between the virtual drawings and the real world scene. The robust tracking method used enabled virtual annotations to appear as if they were attached to the surface of real objects. However, the colour and shading of the virtual drawings made them stand out from the real world background.

Through two user studies the author found the benefits of the auto freeze

technique as a useful tool for when sharing spatial information and the benefits of a red box notification for effective allocation of the local user's attentional resource in a complex shared object manipulation task. Even though these two conditions were preferred by most of the participants, there were still a few participants selecting other conditions as their favourites. This could be because the quality of collaboration is not only determined by the interface in use but is also influenced by other factors such as the participants' personality, relationship between the participants, familiarity to the task, ideas they have for solving a task, etc. Since collaboration with object manipulation required a high level of perception, anything that effects user's perception could affect the study results.

The author also note that if the task was a remote assistance/instruction rather than a parallel experience, the effects of having auto freeze and red box could have been different. In remote assistance tasks, remote users would consistently need to present instructions with visual communication cues and the local users would only need to follow them, compared to parallel experience where both need to interpret the current state of an issue, have an idea to solve it, and share their ideas.

Previously Shannon et al. (1949) and Devito (1998) defined models for verbal communication that involved information transmission and verbally forming or interpreting message on remote or local end. In addition to these subtasks in verbal communication, our studies involved four additional subtasks for the users: ① watching and understanding the current state of workspace annotated with visual cues (for both users), ② thinking of ideas for solving the Tangram (for both users), ③ interacting with a given interface to express ideas by annotating virtual objects on the workspace (for remote users), and ④ manipulating physical objects (for local users). With these additional subtasks the communication became more complex, and the two studies investigated how to reduce the effort or cognitive load associated with these subtasks. In the first study, we introduced an auto freeze technique for

drawing annotations on a shared video easier, more efficient and less stressful, and as a result, remote participants may have less cognitive load for expressing their intention with annotating virtual objects (subtask ③) on the shared view.

In the second study, we compared the both freeze and red box conditions to the no notification condition. With the both freeze condition, local participants were forced to only focus on ① watching and understanding the current state of the workspace annotated with visual cues. This could reduce the required cognitive load by aborting the other subtasks, but it also forced local participants to abandon their on-going task. For example, if the shared view was frozen with the both freeze condition when the local participants were in progress of ② thinking ideas or ④ manipulating pieces, they should abandon the progress of ② and ④, but be forced to do subtask ①. This helped local participants to understand their partner, when and where the annotation was drawn, but make them uncomfortable because of abandonment. On the other hand, the red box condition helped local participants to keep doing on-going task while ① having better understanding of the workspace (when annotations were drawn). In other words, the red box interface helped the local participant to handle the part of subtask ① while focusing on subtasks ② or ④, and it led to have better subtasks management.

Interestingly, subtask ① and ③ were added for using visual cues for communication in remote collaboration. Our results showed that participants preferred the conditions that reduced the effort or cognitive load for the subtasks ① and ③ that were introduced for using visual communication in remote collaboration.

## **6.8 Conclusion**

In the first study (in chapter 5), the annotation interface was slower than the

pointer interface and had an issue that annotation can be drawn in the wrong place if a local user changes the viewpoint of a shared live video while the remote user is drawing an annotation. To overcome this drawback, the author investigated the shared view with a freezing function in the second study. The second study compared three conditions, non-freeze, manual freeze and auto freeze and found that auto freeze was most preferred by both remote and local participants. The experiment showed that auto freeze provided the benefits of quick drawing and keeping the live viewing of the shared space without any additional inputs for freezing.

The third study investigated visual notification methods on the local end to help the local users effectively manage their attention resource between the ongoing task and the remote user's drawings. In the study the author compared no notification, red box, and both freeze conditions, and the results showed that the red box was the most preferred method. The local participants were able to understand more about what the remote participants were doing with the red box notification, which led them to feel more collaborative. While the both freeze condition was the least preferred, it was considered to be useful for certain cases where full attention to the remote user's message was critical.

In the next chapter, the author investigates view independency on the remote end with view navigation techniques. The dependency of the shared view on the remote end could force remote users to look at what the local users were seeing. Remote users could not watch the area that they were interested in without help from the local user. To overcome this problem, the author investigated whether having an independent view on the remote end can improve the shared experience in parallel experience collaboration.

## **7 Exploring View Independence**

Several researchers mentioned the benefit of having an independent view for better remote expert collaboration (Kuzuoka et al. 1994; Sakata et al. 2003; Kasahara & Rekimoto 2014; Gauglitz et al. 2014b). The remote users in their experiments were able to navigate an independent view and give instruction to the local user. In a remote expert collaboration, the remote users know what to do and where to look in the shared workspace while the local users do not have the same knowledge, so the remote user having an independent view and deciding where to look could be critical for effective remote expert tasks.

In this chapter, the author explores how a remote user can use the independent view and its effect in both remote expert and parallel experience tasks. The author conducted a user study comparing four conditions made up of two different views (dependent and independent) with two different tasks (remote expert and parallel experience).

### **7.1 Prototype and User Interface**

In the study, the author compared four conditions: 1) Remote expert with Dependent view (RD), 2) Remote expert with Independent view (RI), 3) Parallel experience with Dependent view (PD), and 4) Parallel experience with Independent view (PI).

For the dependent view, the author adopted the auto freeze user interface that was the preferred condition from the second user study. To design the interface for the independent view, the author first considered how people watched a workspace while making a model with pieces. They sometimes watched closer or further (zooming in and out the view), and sometimes watched here and there (panning the view). For designing a zooming function in the independent view, the author measured how far people see the workspace (a desk in this studies). Assuming people mostly look straight ahead, the author measured the distance from the eye (in which position was calculated by the SLAM system) to the point where the center of the view

showed. To calculate the distance to the center point, the author adopted the method of positioning a pointer in 3D space used in the first study (chapter 5).

The author recruited 12 participants for measuring the distance from their eyes to the workspace while they made a model of Tangram alone with the given instruction paper. The distance was calculated every half second. Since the SLAM system does not use the general length units that people use such as centimeter or inch, the author first measured the corresponding value of the distance with 30 centimeters in the SLAM system, then used a proportional equation. Figure 7.1 shows the results of how far the participants looked at the workspace.

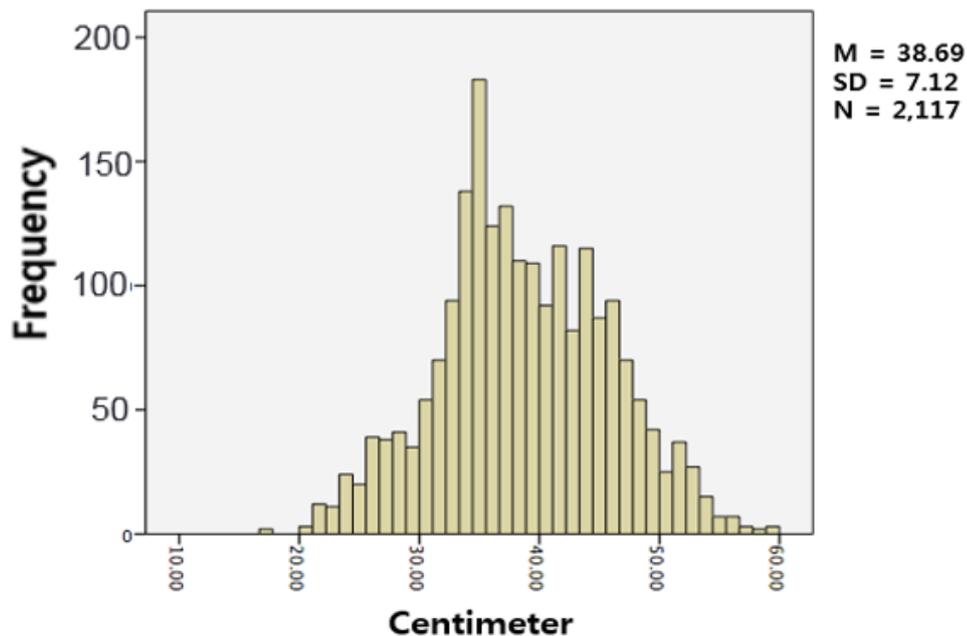


Figure 7.1: The distance between HMD and the point where the users see ( $M$  = Mean,  $SD$  = Standard Deviation,  $N$  = number of measuring distance)

Through this small test, the author collected 2117 samples and found that the average of the distance between the user and the workspace was around thirty eight centimeters. The range of the distance was from eighteen to sixty centimeters.

In the experimental prototype there were two webcams (Logitech C920s)

for users as shown in Figure 7.2. The camera for the local user was attached to a HMD as in the previous studies; and similar to the setup used by Fussell et al. (2003), the second camera for the remote user was positioned beside the local user to show the workspace as if the remote user was sitting beside the local user. The camera for the remote user was seventy centimeters away to cover the whole workspace (a desk, 93cm by 50cm). The prototype system ran two SLAM simultaneously for both views and shared the information of the SLAM tracking and drawing annotation. The drawing annotation was simultaneously drawn at both local and remote views regardless of dependency or independency of the remote user view. To anchor the annotation, the author used the same technique used for the previous studies.

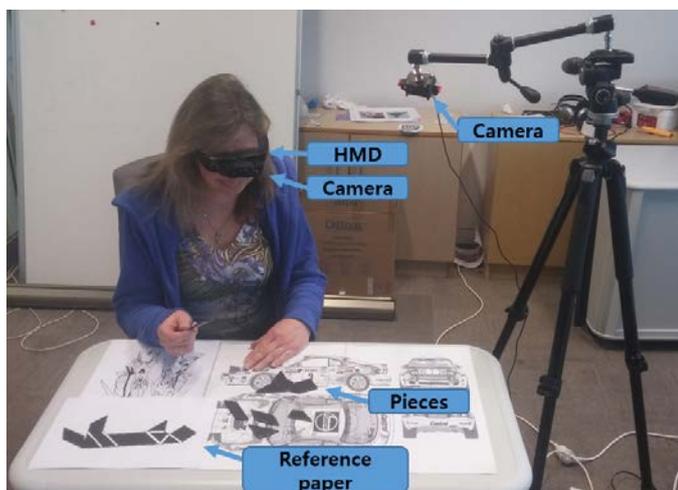


Figure 7.2: Experimental environment at the local end for the independent view

The live video from the camera on a tripod, which was seventy centimeters away from the centre of the workspace, was the original video feed for the remote user's independent view. The system calculated how much of the original live video feed would be displayed with the current zoom level, and which part of the original live feed would be displayed with the current panning information.

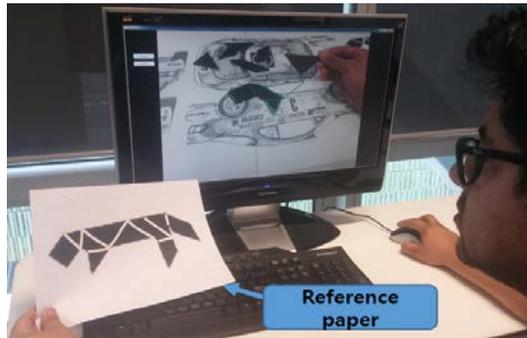


Figure 7.3: Experimental environment at remote end for remote expert collaboration

Just as in the previous studies, the author provided a desktop environment for the remote user system (see Figure 7.3). The remote user could control the level of zoom in the live camera view with the ‘+’ and ‘-’ keys or with mouse scrolling. For panning, the ‘a’, ‘s’, ‘d’, and ‘w’ keys or mouse dragging were used. Table 7.1 shows the interaction mapping.

Table 7.1: Interaction mapping for navigation in an independent view

	<b>Panning</b>				<b>Zoom</b>	
	Left	Down	Right	Up	In	Out
<b>Keyboard</b>	‘a’	‘s’	‘d’	‘w’	‘+’	‘-’
<b>Mouse</b>	Dragging interaction. (The scene always follows the mouse cursor when mouse left button is pressed)				Scroll down	Scroll up

The measurement described in Figure 7.1 was used to decide the zoom levels. There were sixty zoom levels, and the fully zoomed out view was equivalent to watching the workspace from sixty centimeters away while the fully zoomed in view was equivalent to watching from twenty centimeters away. A stroke of mouse scrolling had an effect of two or three levels of zoom

in/out interaction. While key pressing ('+' for zoom in and '-' for zoom out) for the zooming interaction had an effect of one step of zoom in/out, keep pressing the keys had an effect of continuously increasing or decreasing the zoom level. Remote users initially had a view equivalent to watching from thirty eight centimeters away when they started using the prototype system, as the average distance participants had in the measurement test was 38.69 centimeters.

For panning, the system kept the position of the user's view in the original video feed. The pivot point was changed according to the remote user's keyboard or mouse input and the relevant points of each corner of the remote user's view were calculated based on the pivot point and the level of zoom.

In addition, both end users were able to see a brown borderline of a tetragon on the workspace which represented the area where his/her partner was watching (see Figure 7.4). The corner points of the tetragon were corresponding 3D points in the real world for the corner points of the user's view in screen space. With the 2D corners points in the screen space, the corresponding 3D points were calculated by the same method of positioning annotation and pointers in the 3D space as described in Chapter 4.



Figure 7.4: The view of the local user. The brown tetragon represents the area that the remote partner could see.

## 7.2 User Study Design

The user study design adopted the same experimental tasks, Tangram, but the types of the tasks were either remote expert or parallel experience depending on the experimental condition. The experimental environment was the same as the previous studies except for the remote user had a reference paper when solving remote expert task as shown in Figure 7.3.

The participants solved Tangram puzzles in pairs while located separately and using the prototype system for communication. The experiment started with a questionnaire asking demographic information and the purpose of the study was described. Then the participants were asked to solve a sample Tangram puzzle face-to-face, followed by a demonstration and explanation of how to use the prototype system. Participants were then separated and completed four sessions under different experimental conditions: Remote expert with Dependent view (RD), Remote expert with Independent view (RI), Parallel experience with Dependent view (PD), and Parallel experience with Independent view (PI). The order of the conditions was counterbalanced using a Latin square design. Each session consisted of two minutes training and five minutes of solving the Tangram. After each experimental session, the participants gave feedback through answering a questionnaire and having a brief interview. Finally, the experiment concluded with the final interview. Overall the experiment took about 80 minutes for each pair of participants (60 minutes for the experiment and 20 minutes for the interviews).

Similar to the previous studies, the author collected subjective feedback (questionnaires and interviews), activity logs of remote participants, and video recordings of each screen. The questionnaire included twelve Likert-scale rating questions for local participants and fourteen questions for the remote users (see Table 7.2).

Table 7.2: Questions in the questionnaire after each session (the questions 3 and 4, highlighted in grey, were only for the remote user).

Q1	I enjoyed assembling a model
Q2	I was able to focus on assembling a model.
Q3	Using the interface distracted me from collaborating with my partner.
Q4	I was able to see the work environment properly.
Q5	I was able to express my idea properly.
Q6	I easily understood what my partner tried to do.
Q7	I felt I was interrupting my partner during collaboration (for a remote user) or My partner was interrupting me during collaboration (for a local user).
Q8	I felt my partner was focusing on my explanation (for a remote user) or I was able to focus on my partner's explanation (for a local user)
Q9	I felt connected with my partner. (connected: have aligned thoughts or feeling with your partner)
Q10	I felt we were together. (being together: staying in the same environment)
Q11	The communication with my partner was mentally stressful.
Q12	We focused on the same piece or area while collaborating.
Q13	My partner was aware of where I was drawing (for a remote user). My partner was aware of how I manipulated pieces (for a local user).
Q14	I felt we collaborated well.

For the questions, the author also considered main factors of remote collaboration (both users who encode and decode messages, channels, and the situation model) except for channels like other studies. Since the situation model is understanding current states of a task and a collaboration partner, the author had three questions (Q2, Q4, and Q12) about understanding the task and two questions (Q9 and Q10) about connectedness between users. The author prepared four questions (Q2, Q4, and Q10) about users' experience of decoding messages. For the experience of sending (encoding) message question 5 was prepared. To see how the different decoding experiences had effect on overall remote collaboration, the author also prepared questions 1, 3, 11 and 14.

## 7.3 Results

The author recruited twenty-four participants (in pairs) who were friends or family and who had experience of using videoconferencing regularly (more than once a month). There were 20 males and 4 females with ages ranging from 20 to 35 years old ( $M = 27.6$ ;  $SD = 4.4$ ).

The key results from this user study were:

- The independent view significantly improved the collaboration in terms of understanding the workspace, communication, and collaboration in both types of tasks.
- The freedom and stability of the view were the main benefits of independent view that may have led to better communication.
- Remote participants mostly had the fully zoomed-out view during the collaboration.
- The remote expert task had mostly one-way communication which was simpler compared to the parallel experience task which required bidirectional communication.
- Even though the one-way communication was easy to focus for collaboration, it led local participants to be passive and reduced satisfaction by reducing their contribution for collaboration.

The author analyzed the participants' answers to the Likert-scale questions, remote participants' activity logs, recorded videos, and user preference. To analyze the Likert-scale ratings between four conditions, the author used the Aligned Rank Transform (ART) test for non-parametric factorial analyses using repeated measure ANOVA procedures ( $\alpha = .05$ ) proposed by Wobbrock et al. (2011). This test is designed for non-parametric factorial data analysis and capable of analyzing data from a 2x2 factorial experimental design with within-subject measure.

### 7.3.1 Focusing and understanding the task

For the questions (Q2, Q12, and Q4 – only for remote participants) about focusing, figures 7.5 and 7.7 show the results comparing the three conditions

using the Likert scale ratings.

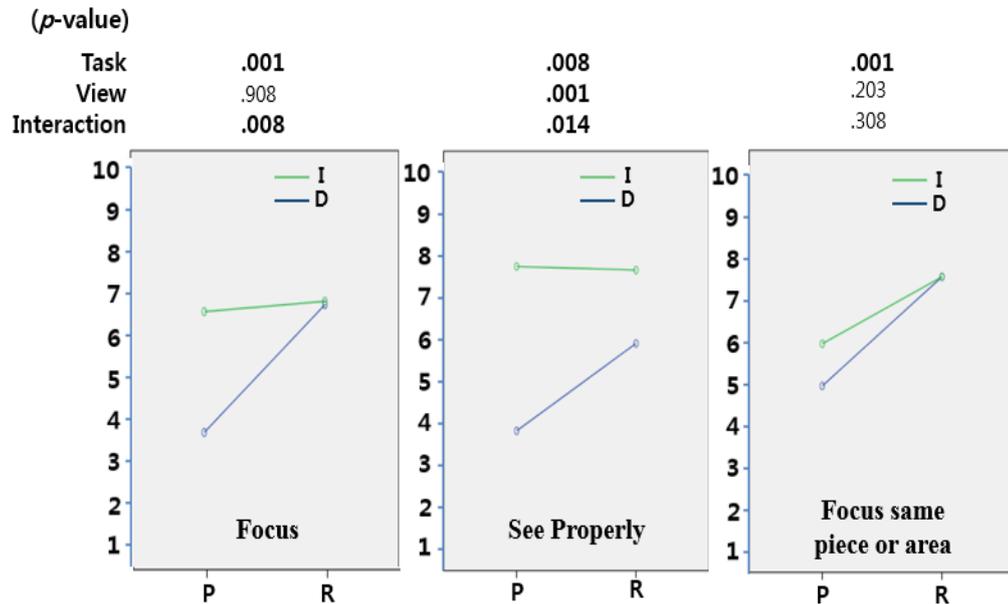


Figure 7.5: Results of Likert-scale ratings of Q2, Q4, and Q12 from remote participants (0: strongly disagree ~ 10: strongly agree; I: independent view, D: dependent view, P: parallel experience, R: remote expert)

Remote participants gave significantly higher points for remote expert collaboration than parallel experience collaboration for focusing on the task (Q2,  $F(1,12)=28.937$ ,  $p<.001$ ), seeing the workspace properly (Q4,  $F(1,12)=10.226$ ,  $p=.008$ ), and focusing on the same piece or area with partner (Q12,  $F(1,12)=21.764$ ,  $p=.001$ ). This would mean that the remote participants had a better level of focus on the task, saw the workspace properly, and felt that they were focusing on the same piece or area with their local partner.

For comparison between views, the remote participants gave significantly higher scores to the independent view than the dependent view in seeing the workspace properly (Q4,  $F(1,12)=80.084$ ,  $p<.001$ ). However, the ratings did not show significant difference between the views in focusing on the task (Q2,  $F(1,12)= 0.014$ ,  $p=.908$ ), and focusing on the same piece or area (Q12,  $F(1,12)=1.144$ ,  $p=.308$ ). This would mean that the independent view helped remote participants to see the workspace more accurately, but it did not have an effect on focusing on the task.

Interestingly, a significant interaction between task and view was found in the aspects of focusing on the task (Q2,  $F(1,12)=10.520$ ,  $p=.008$ ), and seeing the workspace properly (Q4,  $F(1,12)=8.574$ ,  $p=.014$ ). For these two aspects, the difference between independent and dependent views is more obvious in the parallel experience than in the remote expert. This means that the independent view is more critical in the parallel experience than the remote expert for these two aspects.

From the log data the author found that all remote participants except R4 (the remote participant in group 4 who stayed mostly in the fifty six centimeter away view) stayed in the fully zoomed-out view (sixty centimeter away view) for most of the time. The fully zoomed-out view (large field of view) had the benefit of the independent view. R1 mentioned that having a zoomed out view was helpful for seeing the workspace, and R3 and R2 said that they zoomed out all the time to see more area. On the contrary R4 and R6 complained about having a narrower view with the dependent view (where zooming was not available).

Since the prototype provided two ways of navigation with a keyboard and a mouse, the author compared the amount of navigation by keyboard and mouse in the two types of tasks. Figure 7.6 shows that remote participants navigated significantly more in the parallel experience than in the remote assistance task ( $F(1,12)=24.799$ ,  $p<.001$ ). To calculate the amount of navigation with a mouse, the author simply measured how much the mouse cursor was moved (in pixels) on the screen of the shared view while navigating. For keyboard navigation, the author calculated how much the shared view was panned in pixels. With the calculation, the author found that the mouse was used significantly more than the keyboard for navigation ( $F(1,12)=62.748$ ,  $p<.001$ ). Moreover, it shows that the difference between the remote assistance and the parallel experience in the amount of navigation was more obvious with the mouse interface than with the keyboard interface ( $F(1,12)=27.107$ ,  $p<.001$ ).

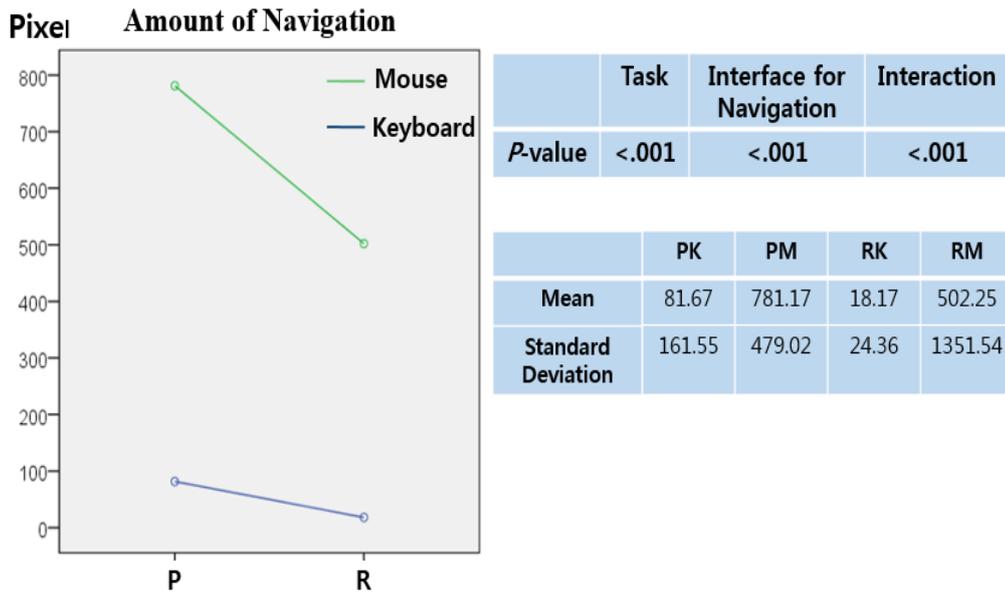


Figure 7.6: The amount of navigation by a keyboard and a mouse (PK: keyboard navigation in parallel experience, PM: mouse navigation in parallel experience, RK: keyboard navigation in remote assistance, RI: mouse navigation in remote assistance, P: parallel experience, R: remote assistance)

Several participants did not like the dependency of their view because remote participants did not have freedom of their view and could not focus on the task. R3 mentioned that *“her head movement was important to what I can see, so I cannot see everything”* and R8 reported that *“I cannot control the view to see what I want”*. On local end, L7 (the local user in group 7) reported that *“I felt like forcing him to focus where I was seeing”*. Some of the remote participants disliked that they could not keep focusing on an area or a piece when the local participants changed their view. R5 mentioned that *“if he moved fast, I could not focus on the task because he already changed the view”* and R3 reported that *“When he moved, my view was changed unexpectedly and I lost my focus on the task”*.

When the remote participants had independent view, the stability of the view (not shaking due to the local users’ head movement) was mentioned as one of the benefits. R7 mentioned that *“Independent view showed a stable shared view”* and similar comments were found in the interviews with R1 and R9. This stability might lead to an increase in the level of focusing on task. R4

commented, *“I can focus on where I want to see (with independent view)”* and R9 mentioned that *“I can hold the view of mine, so that I can keep in thinking”*.

Having freedom to look outside of the local participants' view was reported as another benefit of the independent view. R1, R8, and R10 noted that they had freedom from their partner's view. Since they could navigate the workspace as they wanted, they would feel the freedom. R2 mentioned that *“I preferred the independent view since I had access outside the other person's viewing area and I could always return back to the red viewing zone”*. R12 said that he liked being able to choose where he looked at while helping his partner. R1, R3, R4, R5, R6, R7 and R12 also reported the benefit of a navigable independent view. Interestingly, R1 mentioned that it was like having his own workspace. On the local end, L4 said *“I don't need to worry about providing my partner with a good view”*.

Figure 7.7 shows the local participants' ratings for questions 2 and 12. The local participants gave significantly higher scores for the remote assistance collaboration than for the parallel experience collaboration in the question about focusing on the same piece or area with partner (Q12,  $F(1,12)=12.645$ ,  $p=.005$ ) but not in the question about focusing on the assembling task (Q2,  $F(1,12)=.093$ ,  $p=.766$ ). The author could not find a significant difference between the views and there was not main interaction between the two factors, views and collaboration types.

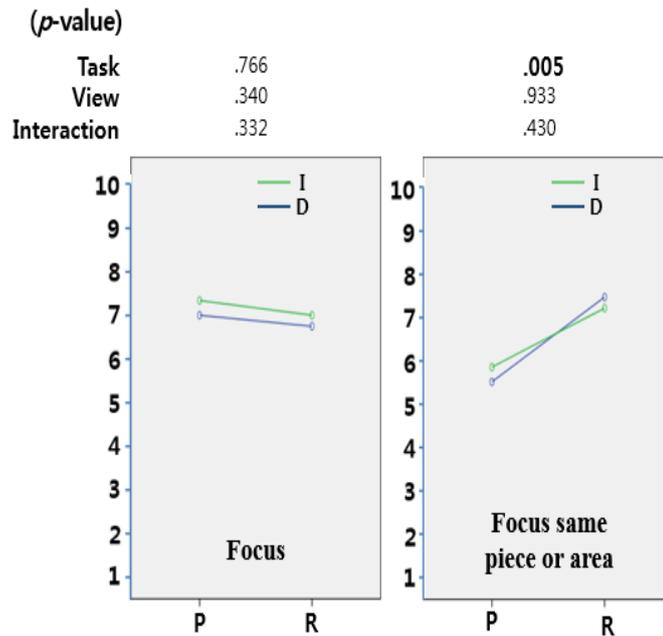


Figure 7.7: Results of Likert-scale ratings of Q2 and Q12 from local participants (0: strongly disagree ~ 10: strongly agree; I: independent view, D: dependent view, P: parallel experience, R: remote expert)

Participants reported that the remote assistance task made it easier to focus on collaboration and highlighted that only the remote participants have the solution of the task. R1, R2, and R3 mentioned that their partners focused on their directions because they were the only one knowing the answers. R10 commented that “*mostly one-way communication is so less distracting*”. Local participants also gave similar comments. L1 mentioned that “*I only focused on what my partner explained; he was my eyes and I was his hands*”.

### 7.3.2 Receiving/Decoding Messages

For the questions (Q6, Q7, Q8, and Q13) about focusing, figures 7.8 and 7.9 show the results comparing the three conditions using the Likert scale ratings.

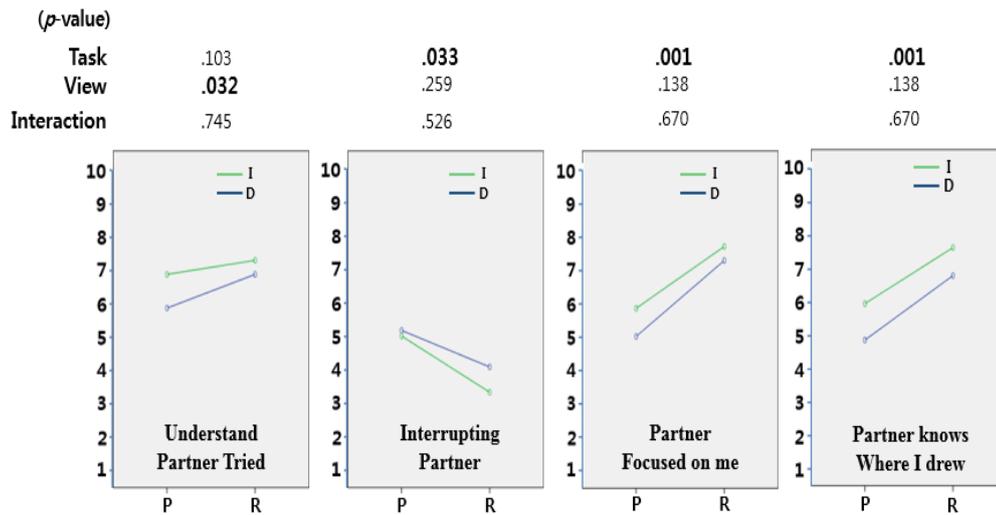


Figure 7.8: Results of Likert-scale ratings of Q6, Q7, Q8, and Q13 from the remote participants (0: strongly disagree ~ 10: strongly agree; I: independent view, D: dependent view, P: parallel experience, R: remote expert)

Remote participants gave significantly higher points for remote expert collaboration than parallel experience collaboration for being less interruptive for their partner (Q7,  $F(1,12)=5.909$ ,  $p=.033$ ), the partner focusing on their explanation (Q8,  $F(1,12)=33.113$ ,  $p<.001$ ), and partner being aware of where the remote participant was drawing (Q13,  $F(1,12)=12.051$ ,  $p=.005$ ). However, the remote user answers did not show any significant difference between the remote assistance and the parallel experience collaborations in response to the questions about understanding their partner (Q6,  $F(1,12)=3.160$ ,  $p=.103$ ). This would mean that remote participants felt that their local partners focused on them and their drawings more in the remote expert collaboration than in the parallel experience collaboration. However, they did not feel the difference in the level of understanding their partners.

For comparison between views, the remote participants gave significantly higher scores to the independent view than the dependent view in understanding what their partner was trying to do (Q6,  $F(1,12)=6.011$ ,  $p=.032$ ), but not in other aspects. These results are aligned with the remote participants understanding the workspace better with the independent view than with the dependent view. There was no main interaction between two

factors, views and collaboration types.

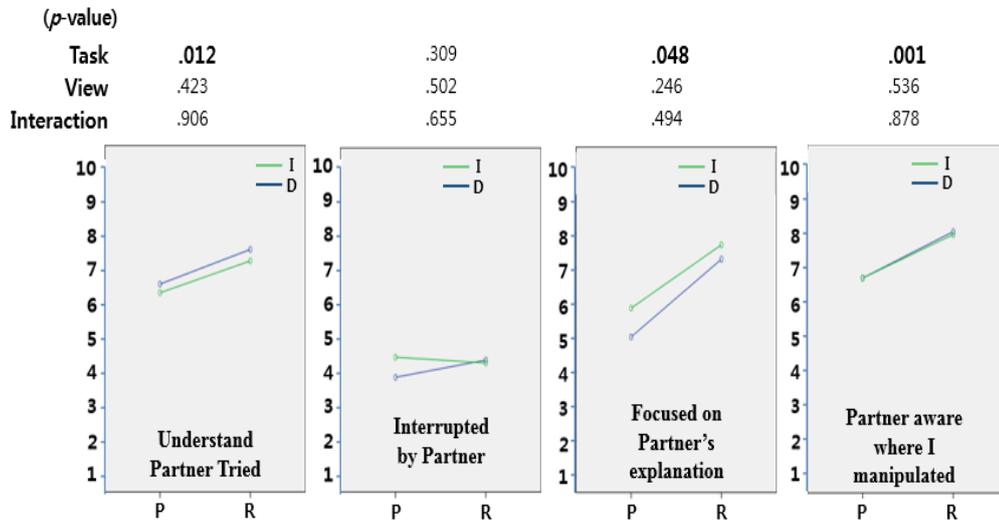


Figure 7.9: Results of Likert-scale ratings of Q6, Q7, Q8, and Q13 from the local participants (0: strongly disagree ~ 10: strongly agree; I: independent view, D: dependent view, P: parallel experience, R: remote expert)

Local participants gave significantly higher points for remote expert collaboration than parallel experience collaboration for understanding what their partner was trying to do (Q6,  $F(1,12)= 9.061, p=.012$ ), focusing on partner's explanation (Q8,  $F(1,12)=4.929, p=.048$ ), and partner being aware of where local partner was manipulating (Q13,  $F(1,12)=19.084, p=.001$ ). However, the local participants' answers did not show any significant difference between the remote assistance and the parallel experience collaborations in response to the questions about being interrupted by their partner (Q7,  $F(1,12)=.211, p=.655$ ). Moreover, there was no significant difference between the views for the aspects, and no main interaction between the view and the collaboration types.

### 7.3.3 Sending/Encoding Messages

Figure 7.10 shows the results comparing the three conditions using the Likert scale ratings for question 5; expressing ideas.

Remote participants gave significantly higher points for remote assistance collaboration than parallel experience collaboration for expressing ideas

properly (Q5,  $F(1,12)=7.781$ ,  $p=.018$ ) but this was not the case in the ratings from the local participants ( $F(1,12)=1.282$ ,  $p=.282$ ). This would mean that the remote participants felt better expressing ideas with remote assistance collaboration than with parallel experience collaboration, but local participants did not feel any significant difference in the feeling of expressing ideas between the collaboration types. Similarly, the remote participants gave significantly higher scores to the independent view than the dependent view in expressing ideas properly ( $F(1,12)=13.626$ ,  $p=.004$ ), but the local participants did not.

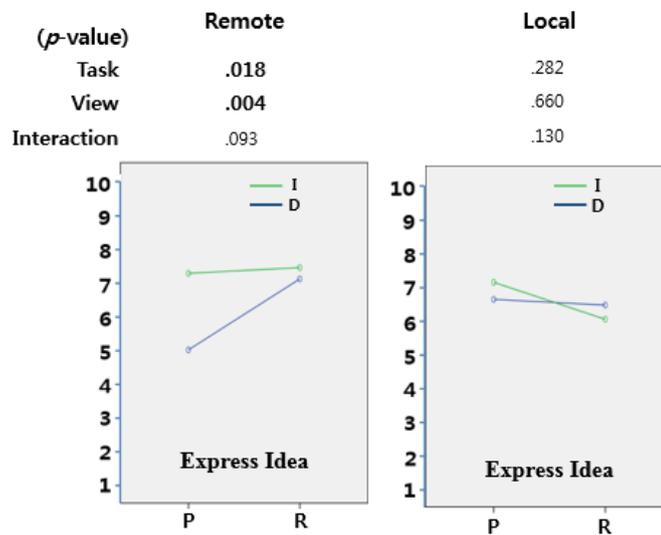


Figure 7.10: Results of Likert-scale ratings of Q5 from the remote and local participants (0: strongly disagree ~ 10: strongly agree; I: independent view, D: dependent view, P: parallel experience, R: remote expert)

The log data also shows corresponding results with participants' ratings (see Figure 7.11). Remote participants had significantly more drawing annotations in the remote assistance task than in the parallel experience ( $F(1,12)=32.728$ ,  $p<.001$ ), and significantly more drawings in the independent view than in the dependent view ( $F(1,12)=24.135$ ,  $p<.001$ ).

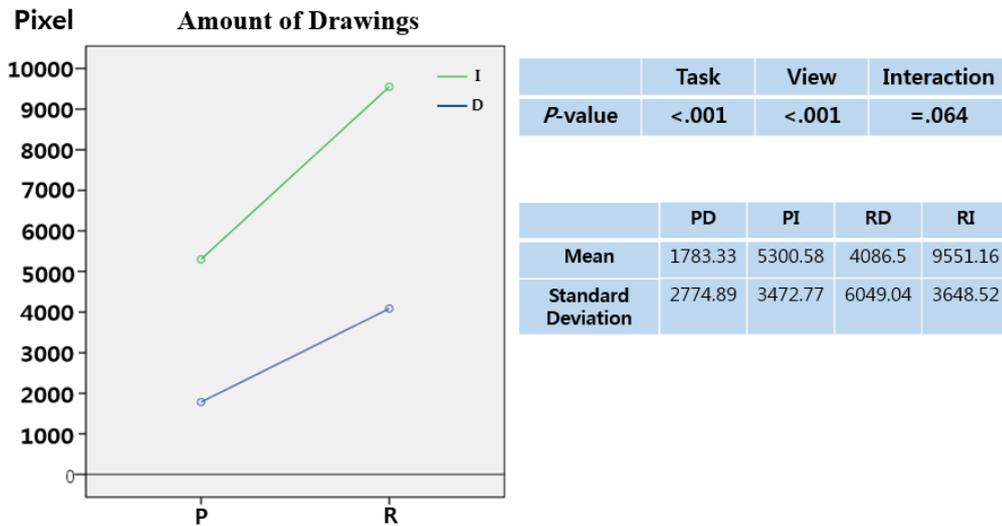


Figure 7.11: The number of drawings in four conditions (P: parallel experience, R: remote assistance, I: independent view, D: dependent view)

The author also compared the number of strokes they made for drawing annotations between the conditions (see Figure 7.12). Remote participants had significantly more strokes in the remote assistance task than in the parallel experience task ( $F(1,12)=8.731, p=.013$ ). They also made more strokes in the independent view than the dependent view ( $F(1,12)=11.271, p=.006$ ). The two independent variables (views and tasks) did not have a significant interaction ( $F(1,12) =.447, p=.518$ ).

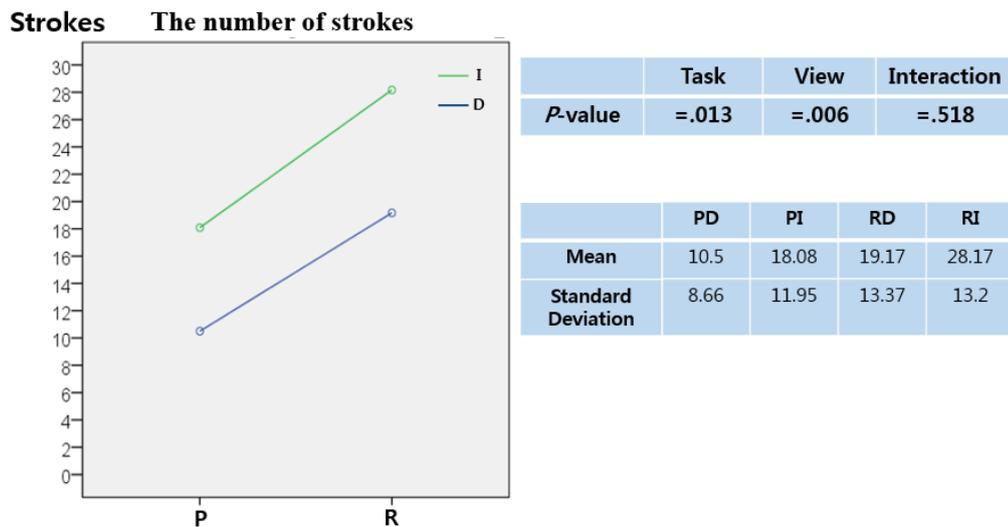


Figure 7.12: Number of strokes in four conditions (P: parallel experience, R: remote assistance, I: independent view, D: dependent view)

Several participants concerned ‘one-way communication’ for remote expert collaboration. R1 and R4 mentioned that “*he (the local partner) was passive and just waited for my next drawing*”, and “*it was more like teaching (instructing) rather than cooperating*” respectively. This issue was also revealed in comments of the local participants. L4 and L5 mentioned that they were just their partner’s hands and they kept quiet and only listened to the instructions. L2 said that the assembling was completely dependent on his partner, and L1 commented that “*I was not sure he made a mistake or not, just only follow*”.

For the parallel experience collaboration, the remote participants mentioned that they became less active with the parallel experience task than with the remote assistance collaboration. R4 mentioned that “*In the parallel experience, I had less work to do.*” R2, R3, and R7 gave similar comments. R10 mentioned that “*Having bidirectional communication caused being more emotional and waiting for my partner*”, and R11 reported that “*I felt like my input was second (low priority) to her*”. Moreover, R12 commented that “*I relied on my partner more because I trusted him to make the right decisions more than what I can offer him*”.

#### **7.3.4 Connectedness & Being Together**

Figure 7.13 shows the results comparing the three conditions using the Likert scale ratings for the questions about level of connectedness and being together.

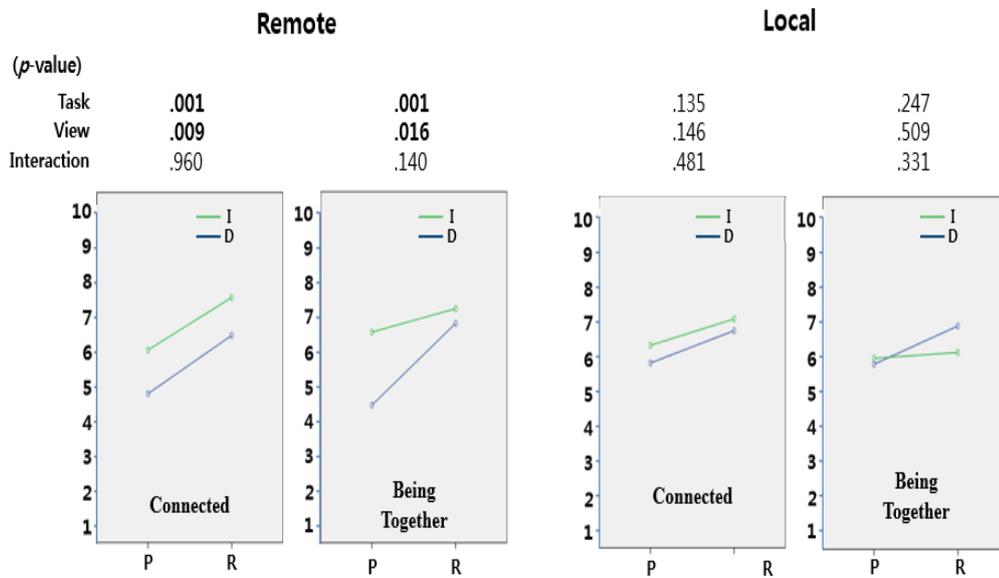


Figure 7.13: Results of Likert-scale ratings of Q9 and Q10 from the remote and local participants (0: strongly disagree ~ 10: strongly agree; I: independent view, D: dependent view, P: parallel experience, R: remote expert)

Remote participants gave significantly higher points for remote expert collaboration than parallel experience collaboration for being connected with their partner (Q9,  $F(1,12)=21.695$ ,  $p=.001$ ), and being together with their partner (Q10,  $F(1,12) = 18.817$ ,  $p=.001$ ). However, the local participants did not for the same aspect (being connected with their partners:  $F(1,12)=2.609$ ,  $p=.135$ , and being together with their partners :  $F(1,12)=1.495$ ,  $p=.247$ ). These would mean that remote participants felt more connected and being together with remote expert collaboration than with parallel experience collaboration, but local participants did not.

From the interview some remote participants commented why they felt more connectedness with remote expert collaboration than with parallel experience collaboration. The fact that remote participants had a solution and gave instruction to local participants may increase the feeling of connectedness. According to R1, the local partner was only focusing on where he was drawing, leading to a greater feeling of being together. R2 and R3 mentioned that “*I felt more connected because it was driven by one person*”, and “*I felt more connected because I could understand what she was doing*”

respectively.

For comparison between views, the remote participants gave significantly higher scores to the independent view than the dependent view in being connected with their partner (Q9,  $F(1,12)=10.003$ ,  $p=.009$ ), and being together with their partner (Q10,  $F(1,12)=8.015$ ,  $p=.016$ ). However, the local participants did not for the same aspect. These would mean that remote participants felt more connected and being together with independent view than with dependent view but local participants did not. There was no main interaction between two factors, views and collaboration types in both remote and local participants' ratings.

### 7.3.5 Overall Collaboration & User Preference

To see how the different collaboration types and views have effect on overall remote collaboration, the author also prepared questions 1, 3, 11 and 14. Figure 7.14 shows the results comparing the three conditions using the Likert scale ratings from the remote participants.

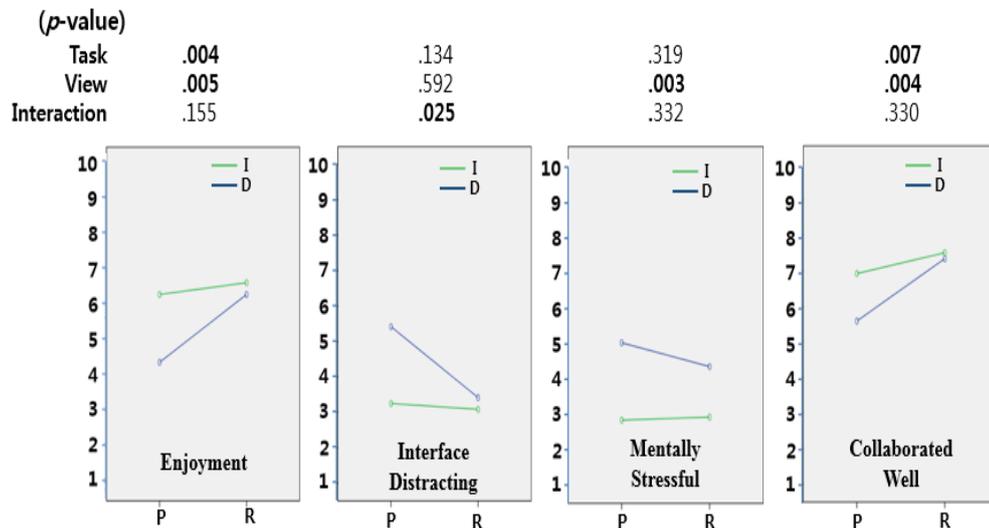


Figure 7.14: Results of Likert-scale ratings of Q1, Q3, Q11 and Q14 from remote participants (0: strongly disagree ~ 10: strongly agree; I: independent view, D: dependent view, P: parallel experience, R: remote expert)

Remote participants gave significantly higher points for remote assistance collaboration than parallel experience collaboration for enjoying the assembly

task (Q1,  $F(1,12)=12.787$ ,  $p=.004$ ), and collaborating well (Q14,  $F(1,12)=10.786$ ,  $p=.007$ ). This would mean that remote participants felt a higher level of enjoyment and collaborating well with remote expert collaboration than with parallel experience collaboration. However, the remote user answers did not show any significant difference between the remote assistance and the parallel experience collaborations in response to the questions about distraction (Q3,  $F(1,12)=2.617$ ,  $p=.134$ ), and mentally stressful communication (Q11,  $F(1,12)=1.091$ ,  $p=.319$ ).

For comparison between views, the remote participants gave significantly higher scores to the independent view than the dependent view in enjoying the assembling task (Q1,  $F(1,12)=12.343$ ,  $p=.005$ ), communication being mentally less stressful (Q11) ( $F(1,12)=14.175$ ,  $p=.003$ ), and collaborating well (Q4) ( $F(1,12)=12.733$ ,  $p=.004$ ). This would mean that remote participants felt a higher level of enjoyment and collaborating well with low level of mental stress with the independent view than with the dependent view.

Interestingly, remote participants reported a higher level of satisfaction with the remote assistance collaboration than parallel experience. R11 mentioned that *“I had a solution and was directing her so the task was more interesting for me”*. R10 also said that *“I was more active and I knew better to complete the modeling”*. On the local end, L2 and L4 said that their partners were more confident. In the interview with R12, the author found that R12 had become ignorant of the relationship with his local partner. R12 said that *“I did not need to see where he was looking at because I knew he was going to listen to me”*.

Remote participants felt that the dependent view was less effective in the communication. R5 stated *“I was too much dependent on his view, which allowed me to explain other things that he was doing. This handicapped very much on how much help I could provide in addition to what he was doing.”* A similar comment was found in the interview of R11: *“I was not able to see exactly where I wanted to see, so I was not able to draw as often as I liked”*.

R12 also reported that he did not draw as often as he wanted because drawing on the workspace was dependent on where his partner was looking.

Figure 7.15 shows the results comparing the three conditions using the Likert scale ratings from the local participants for the question 1, 11, and 14.

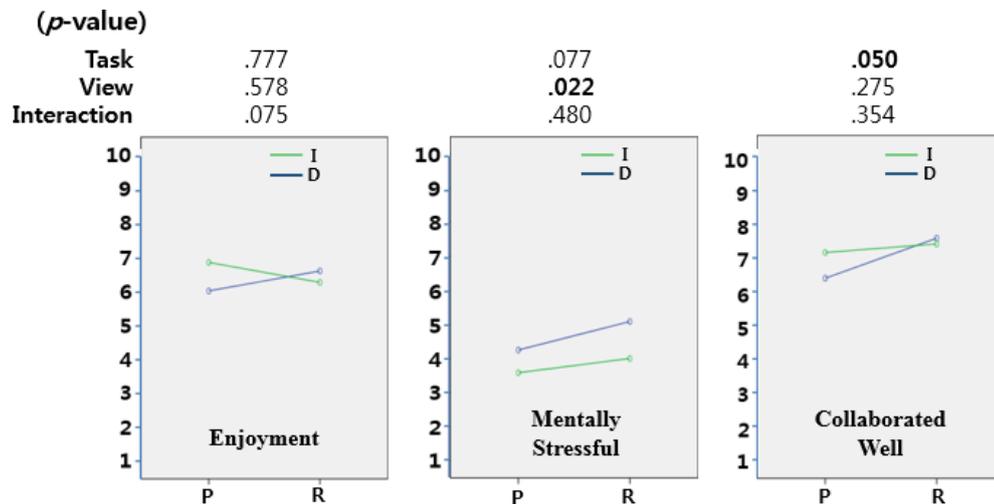


Figure 7.15: Results of Likert-scale ratings of Q1, Q11, and Q14 from local participants (0: strongly disagree ~ 10: strongly agree; I: independent view, D: dependent view, P: parallel experience, R: remote expert)

The local participants gave significantly higher scores for the remote assistance collaboration than for the parallel experience collaboration in response to the question about collaborating well (Q14,  $F(1,12)=4.827$ ,  $p=.050$ ) but not in enjoying the assembling task ( $F(1,12)=.084$ ,  $p=.777$ ), and communication was mentally stressful ( $F(1,12)=3.801$ ,  $p=.077$ ). These mean that local participants felt they were more collaborative with remote expert collaboration than with parallel experience collaboration. When comparing the views, the local participants felt a higher level of mental stress with the dependent view than with the independent view.

Since the performance (task completion time) of the parallel experience (where no one is given the solution for the task) is highly dependent on their ideas, the author did not collect nor compare the performance of two conditions in the parallel experience. The author compared the performances of two different views in the remote expert task (see Figure 7.16) using paired

$t$ -test ( $\alpha = .05$ ), and the result showed that participants completed the task significantly faster with independent view than with dependent view ( $t(11)=3.041, p=.011$ ).

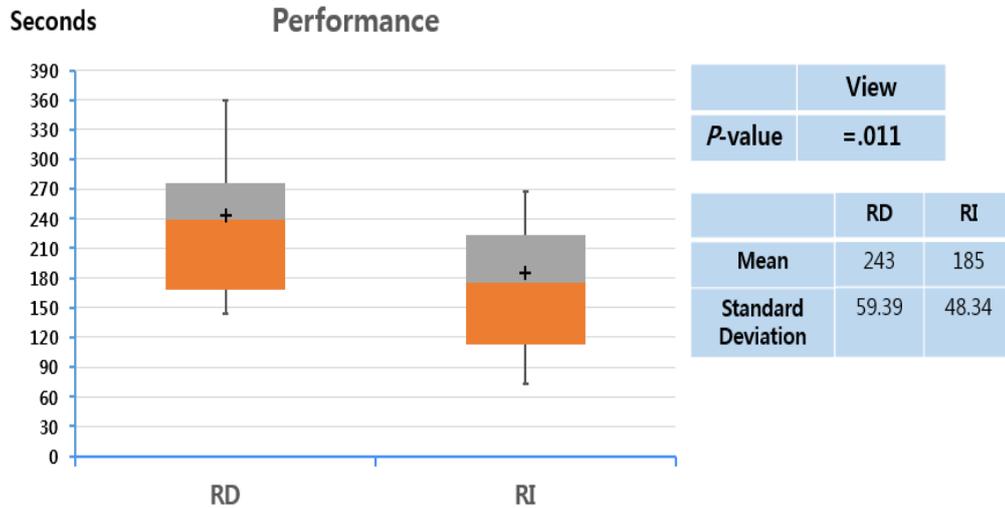


Figure 7.16: User performance with the remote assistance task

While the author did not ask their preference between the tasks (because the type of tasks is not subject to user's choice but is dependent on people's needs in a given situations), the author did ask participants' preference between the views for each task after they tried all conditions (see Figure 7.17). The author analyzed the remote ( $N=12$ ) and local ( $N=12$ ) participants' preferences separately.

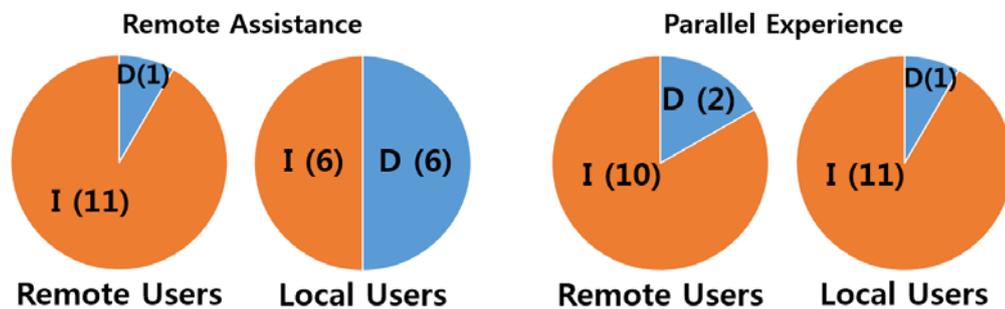


Figure 7.17: Participants' preference for each type of tasks (I: independent view, D: dependent view)

For the remote expert task, eleven remote participants (91.7%) preferred the independent view while one remote participant preferred the dependent

view (8.3%). Half of the local participants (50%) preferred the independent view while the other half (50%) preferred the dependent view. For parallel experience, ten remote participants (83.3%) preferred the independent view while the other two (16.7%) preferred the dependent view. Eleven local participants (91.7%) preferred the independent view while the other one (8.3%) preferred the dependent view.

The Wilcoxon Signed-Rank Test ( $\alpha = .05$ ) was used to analyze their preferences. For the remote expert task, the independent view was significantly more preferred than the dependent view ( $Z=-2.887, p=.004$ ) by the remote participants, while no significant difference in preference was found for the local end ( $Z=0, p=1.0$ ). For the parallel experience task, the independent view was significantly more preferred than the dependent view by both remote ( $Z=-2.309, p=.021$ ) and local participants ( $Z=-2.887, p=.004$ ).

## **7.4 Discussion**

Through the experiment, the author found that the benefit of independent view was revealed in both types of tasks, and the benefit was more pronounced in the parallel experience than in the remote expert task. Local participants were more passive in the remote expert task than in the parallel experience task, and the passive role in collaboration may lead to less head movement in the remote expert, as described in the interview. This would be an example of how the user state or the relationship between the users can have an effect on the user behaviour, such as having less head movement (Carey 1989).

The dependent view has the benefit of sharing exactly the same view, but it did not lead to participants' preferring the dependent view. Compared to the dependent view, the independent view has the benefit of stability and freedom of view and participants may take account of these advantages of independent view more than the advantage of the dependent view for their preference. This would be because stability and freedom of the view are more fundamental

requirements than watching the same area (sharing the exactly same view). With their natural eyes, people have both stability and freedom of view.

However, if the workspace is larger than the one used in this study (e.g. a room), the freedom of the view may become an issue affecting the ability to collaborate. Users need to understand what their partners are doing which could require coupling the remote and local user's views. Having an independent view in a larger workspace may cause difficulty in coupling both user's views.

According to Helmholtz's outflow theory (1867), people know they will turn their head or eyes before they do it. The brain sends a signal to their eyes to prepare the proper reaction for turning, so their eyes can release their focus. In the case of the dependent view, local users have already prepared before changing their view, but remote users cannot anticipate the change, and so feel disoriented when it happens. This could be a reason why remote participants were not comfortable with the dependent view.

Using a mouse for navigation interaction had the benefit of using only one device for all user interactions: navigation and drawing. In this study, remote participants used the mouse interaction significantly more for navigation than the keyboard. With this result, the author deduced that the benefit of using a mouse was more considerable than that of dividing the workload between two devices because the remote participants did not need to draw annotation and navigate the shared view simultaneously, and the workload for two interactions was not an issue.

## **7.5 Conclusion**

In this chapter the author reported a user study comparing dependent and independent views in two different types of remote collaborations; remote expert and parallel experience collaborations. The user study was conducted with four conditions (2x2, one view and one task as one condition). The results showed that the independent view significantly improved the

collaboration in terms of watching and understanding the workspace, communication, and collaboration in both types of tasks. The freedom and stability of the view were the main benefit of independent view. In addition, remote participants usually had the most zoomed-out view during the collaboration.

Participants generally gave higher points to the remote expert task than the parallel experience task because they could have better communication and collaboration. Even though they had better communication and collaboration in remote expert, it led local participants to be passive and ignored their ability as they simply followed the remote partners' instruction.

## **8 Discussion**

In this chapter, the author discusses on the overall results of the whole user studies.

### **8.1 Drawing Annotation Interface**

During the experiment the author observed that the drawing interaction was very intuitive. All of the remote participants found it easy to draw with a mouse because they were already familiar with it. None of the participants mentioned the incongruity between the virtual drawings and the real world scene. The robust tracking method used enabled virtual annotations to appear as if they were attached to the surface of real objects. However, the colour and shading of the virtual drawings made them stand out from the real world background and be easily distinguishing between real world objects and virtual annotations.

### **8.2 Remote Expert & Parallel Experience**

Since the previous studies (Fussell et al. 2004; Kirk et al. 2007; Alem et al. 2011; Gauglitz et al. 2012; Gauglitz et al. 2014b; Sodhi et al. 2013) have mainly focused on the remote expert collaboration, this author explored the interfaces of the visual communication cue for the parallel experience collaboration. The main difference between the remote expert collaboration and the parallel experience collaboration is whether a remote user has a solution or not for the shared task.

The main difference between the two types of collaboration may result in different preferences between a pointer and drawing annotation interface. Since the remote user has a solution for the local end task in remote expert collaboration, his/her main requirement is transferring the solution to the local user. This suggests that in remote expert collaboration, the richer communication cue which can transfer more information would be more effective, hence preferable.

In parallel experience collaboration, the remote user does not have a solution for the local end shared task; the main requirement of the remote collaboration being to exchange ideas and find the solutions together. Since the local user is in the shared environment and can share his/her ideas while conducting the task as they do in the real world, sharing ideas is not an issue. The remote user is required to use a given interface to share their ideas and it would be less effective and slower than the local partner. Being less effective and slower in discussion with a local partner may require quicker and easier visual communication cues, so speed and ease of visual communication cues would be more crucial factors than transferring more information.

Another difference between two types of collaboration is the user state during the collaboration. In our remote expert collaboration experiment, the remote user was very active but the local user was mostly passive while following the remote user's instruction. Since the remote user has a solution, he/she mainly focuses on transferring the solution to the local partner and the local partner mainly focuses on receiving the message rather than sharing possible solutions. In parallel experience collaboration, both remote and local users can be either active or passive because both users do not have a solution. Being active or passive in parallel experience collaboration depends on several factors; the characteristics of the users, having better ideas to solve the task, and the level of understanding of the current task.

Users may feel more complexity in parallel experience collaboration than in remote expert collaboration. In remote expert collaboration users mainly focus on transferring information from a remote user to a local user. In parallel experience collaboration the users also need to transfer information from a remote user to a local user, but they additionally need to understand the current state of the shared task and have ideas to share.

### **8.3 Connectedness**

Connectedness is how well users 1) understand each other and 2) have a

similar level of understanding about the current state of the task. Through the four experiments, the author found that connectedness was important for better collaboration. In communication, if one user understands the other user well, he/she can encode or decode messages according to the other user's needs. For example, if a local user tries to find a piece and the remote partner understands the local user well, the remote partner helps the local user to find the piece more effectively. This would lead to better collaboration. Furthermore, a similar level of understanding about the task would lead to better collaboration. For example, a local user understands the task and finds that the next step of the task is positioning a small triangle piece. In this case, if the remote user has a similar level of understanding, there is more chance that they will focus on the same next step, which would also lead to better collaboration.

In the third study, the author explored notification methods to increase the awareness of the annotation and as a result increasing connectedness (i.e. the local user knows when the remote user is drawing). The author compared non-notification, red box, and both freeze conditions. Both freeze condition forced the local user to watch the remote user's drawing and this may be the best way for transmitting the remote user's idea to the local user. However, the both freeze did not increase the level of connectedness or lead to better collaboration with the better information transmitting; it was less preferred than the other conditions and most participants in the study commented that it was interruptive. From this, the author deduces that an interface with a high level of interruption to the other user would not increase connectedness even if it has the benefit of effective information transfer.

The participants in the fourth experiment (Chapter 7) had a higher level of connectedness in remote expert tasks than in the parallel experience task. In a remote expert task, a local user is in one of two main states, (1) decoding a message and (2) manipulating a piece. In a parallel experience task, the local user can be in one of three states, (1) decoding a message, (2) manipulating a

piece, and (3) thinking about ideas. Having fewer local user states in the remote expert task could mean that there is more chance that the local user is in (1) decoding a message state where the local user focuses on the remote user's message and also mean that the remote participants can more easily predict which state the local user is in (understanding the local user better).

In the fourth study, the author compared dependent and independent views in two different types of tasks; remote expert and parallel experience. The dependent view forced a remote user to see the same area with a local user. Since both users had the same view, they would know where the other user was seeing (better understanding the other user). However, the dependent view did not increase the connectedness or lead to better collaboration. On the other hand, the independent view provided stability of the view and had better connectedness. From this, the author deduces that stability of the view should be achieved to increase the connectedness.

#### **8.4 Parallel Experience Collaboration Model**

Figure 8.1 is the collaboration model for the parallel experience collaboration. In parallel experience collaboration, neither user has solutions for the problem at the local end, so each should have a process of generating ideas coming from their understanding of the workspace or through discussion. The communication is bilateral when both users discuss. The remote user discusses verbally while showing spatial information through visual communication cue. The local user also discusses verbally while showing object manipulation. Every communication includes encoding or decoding processes, and the encoding and decoding include semantic noise.

Interestingly, the local user's object manipulation (encoding a message) and remote user's understanding of the shared workspace (understanding the local user manipulation as decoding message) can play a role of communication from local to remote user when the local user manipulates objects to show an idea (noting that manipulating an object cannot be a communication channel

when used for other purposes, e.g. following the remote users' drawing). As the processes play a role of communication, these also include semantic noise.

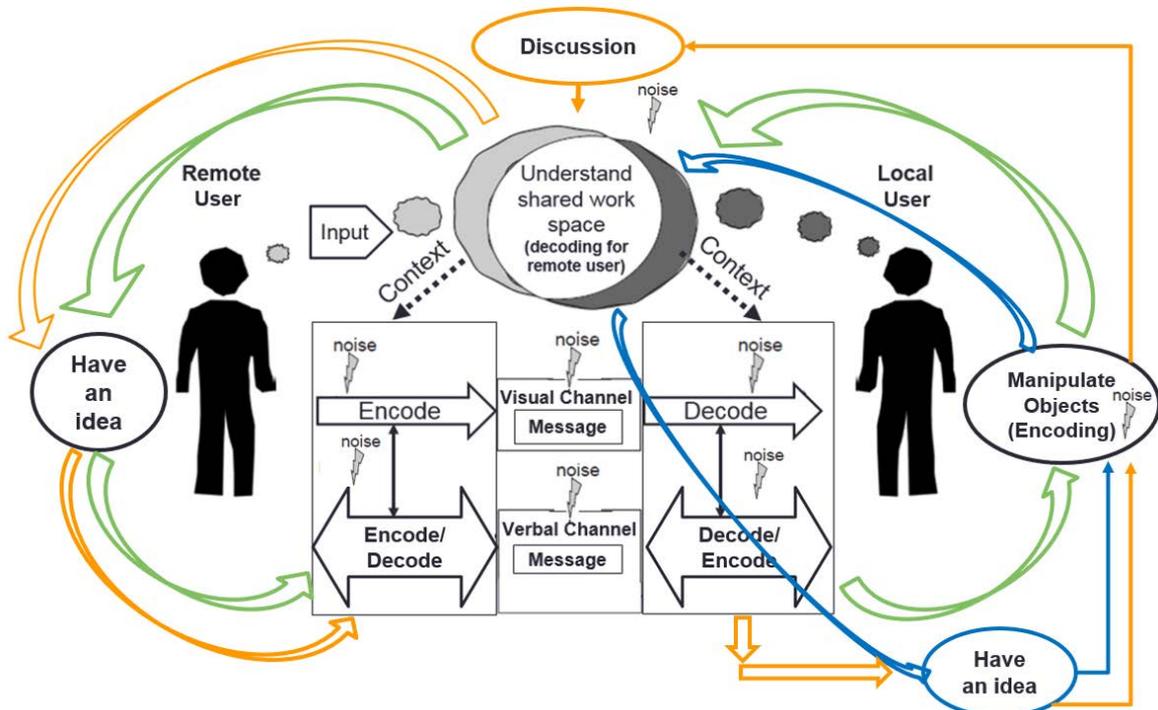


Figure 8.1: The communication model for parallel experience. The blue circulation depicts the communication when a local user and a remote user fall in *do-it-alone* state and *passive* state respectively, while the green circulation depicts the communication when the local user and the remote user falls in *passive* state and *do-it-alone* state respectively. The orange circle represents the communication when both users are in *collaborating together* state.

In the first user study, the author found the four user states, *collaborating together*, *playing in parallel*, *do it alone* and *passive* states. If the collaboration only had the processes represented with the blue circulation, the local user falls into the *do it alone* state and the remote user falls into a *passive* state. If the collaboration only had the processes represented with the green circulation, the local user falls into a *passive* state and the remote user falls into the *do it alone* state. The promising collaboration, *collaborating together* state, is represented with the orange circulation. If there is no communication while the local user manipulates an object with his or her own idea and the remote user draws an annotation with his or her own idea, they

fall into *playing in parallel* state.

## 8.5 Limitation of the studies

The studies in this Ph.D. mainly used the Tangram as the experimental task. The task space was relatively small (to reduce the effect from moving view by the local user's head movement in the first three studies) and the Tangram pieces were light and flat (to reduce the effect from the action of holding and placing objects). The study results would be different according to the size of the task space and the target objects in the task.

The visual communication cues, a pointer and drawing annotation, are difficult to include 3D spatial information in. Since the visual communication cues are attached to real world objects, it is not easy to describe 3D spatial information with them. The 3D annotation or hand gesture may be used for the 3D spatial information.

The prototype system did not support real time panorama view construction. To support this effectively, high quality live video feeds is required and this can affect the local user's movement.

In the user studies, the participants watched the live video from a Logitech C920 camera which provides a relatively small field of view (FOV) (78 degree<sup>4</sup>) compared to natural eye FOV (almost 180-degree<sup>5</sup>). However, the task space was relatively small and people's foveal vision covers only two degrees<sup>6</sup> of FOV.

The questionnaires used in the studies included several standalone questions, but did not focus on one core concept. Since parallel experience collaboration with visual communication cue was not explored well, the author focused on the parallel experience collaboration itself rather than

---

<sup>4</sup> [http://support.logitech.com/en\\_us/product/hd-pro-webcam-c920](http://support.logitech.com/en_us/product/hd-pro-webcam-c920)

<sup>5</sup> [https://en.wikipedia.org/wiki/Field\\_of\\_view](https://en.wikipedia.org/wiki/Field_of_view)

<sup>6</sup> [https://en.wikipedia.org/wiki/Fovea\\_centralis](https://en.wikipedia.org/wiki/Fovea_centralis)

focusing on a specific concept of it. To supplement the questionnaire, the author conducted intensive interviews with participants throughout the studies.

The remote expert collaboration and the parallel experience collaboration is not mutually exclusive. Remote collaboration can be remote expert collaboration then being changed to parallel experience collaboration according to users' need. However the previous studies only explored the use of visual communication cues in remote expert collaboration, so their studies did not cover the use of visual communication cues in parallel experience collaboration. Moreover, as Carey (1989) observed, if a local user merely unpacks and accepts the message from a remote user (as they do in remote expert collaboration), then the remote user is not only sharing ideas but also controlling the local user. This may mean that the type of remote collaboration is not easily changed from remote expert collaboration to parallel experience collaboration because the remote user controls the local partner.

## **9 Conclusion**

This chapter concludes the dissertation. The author summarizes the presented work and describes directions for future research. The main goal of this Ph.D. was introducing interfaces to help users have better parallel experience collaboration. The main goal was achieved by the studies listed below:

- 1) Study the use of visual communication cues in parallel experience (in the first study described in Chapter 5)
- 2) Suggest better interfaces to solve the weak points of visual communication cues for better parallel experience collaboration (in the second and third studies described in Chapter 6)
- 3) Study the difference in the use of independent views between parallel experience and remote expert collaborations (in the fourth study described in Chapter 7)

The main results of each study are described in section 10.1 and potential future works are introduced in section 10.2.

### **9.1 Summary**

In this PhD, the author conducted two pilot studies and four main user studies. In the pilot studies (described in chapter 3), the author found that drawing annotation is more efficient as a visual communication cue than pointers for the remote expert collaboration. However, with a more complicated task, a local user sometimes got confused with the overlapped annotations that led to less effective collaboration.

The author conducted four user studies in a parallel experience task in which both users did not know the solution, hence needing to discuss and share their ideas. The first study (described in chapter 5) compared using a pointer and drawing annotation in a parallel experience task. The users preferred the pointer interface rather than drawing annotation, and the author found that the pointer interface could increase connectedness with frequent communication between the users.

In the second study, the author designed an interface (auto freeze condition) that encouraged quick and easy communication. Comparing auto freeze, manual freeze and no freeze conditions, the auto freeze condition was the most preferred. The auto freeze supported quick and easy drawing annotations by reducing the processes of drawing annotation.

The third study explored two notification methods. The first notification was the red box condition that showed a thick red outline around the edges of the local user's screen when the remote user draws an annotation. The second notification was the both freeze condition that shows the same frozen view on both ends. The red box notification was most preferred in the user study with the benefit that the local user easily knows when the remote user draws an annotation without significant interruption.

In the fourth study, the author compared independent and dependent views in remote expert and parallel experience tasks. Participants preferred the independent view in both tasks because the independent view provided stability and freedom of the view to the remote users. Moreover, independency of the remote user's view was more important in the parallel experience collaboration because the local user made more head movements (that led to more viewpoint movements) in the parallel experience task than in the remote expert task.

Additionally, the author introduces a communication model for parallel experience collaboration including verbal and visual communication cues.

Overall, the interface for drawing annotation in parallel experience collaboration would be better to 1) support quick and easy use of drawing annotation interface, 2) include appropriate notification that helps a local user to know when a remote user draws an annotation and 3) support an independent view 4) with the function that a remote user can start drawing while toggling from the dependent view to independent views.

## 9.2 Future Research

This thesis described the interfaces sharing task space for parallel experience collaboration using the visual communication cues *pointers* and *drawing annotation*. In the first study (Chapter 5), the author found that the pointer cue was more suitable than the drawing annotation cue in the parallel experience collaboration. In the second study (Chapter 6), the author explored a better interface for drawing annotation cue, auto freeze interface. The first next study may be to compare the pointer cue and annotation cue with auto freeze interface.

The author mainly focused on the Tangram task and applied similar experimental design and analysis in all four user studies. The next study would employ different experimental tasks such as the task having a larger workspace, and use different analysis such as looking at communication patterns and conversational analysis.

The author described a number of related works, and previous research mainly focused on three visual communication cues: a pointer, drawing annotations, and hand gesture. Since the author's studies only explored the use of the pointer and drawing annotation cues, future study will expand this to hand gesture interaction for parallel experience.

In this research the author were mostly focused on explicit conscious cues, such as drawing annotation or pointer indication for object manipulations. In the future the author would like to explore sharing emotional cues between users when both users focus on other tasks and share the task space live video with augmented virtual objects. Sharing the facial expression would be a good emotional cue because people are already familiar with this.

## References

- Adamczyk, P.D. & Bailey, B.P., 2004. If not now when?: the effects of interruption at different moments within task execution. *Proceedings of the SIGCHI conference on Human factors in computing systems*, 6(1), pp.271–278. Available at: <http://portal.acm.org/citation.cfm?id=985692.985727>.
- Alem, L., Tecchia, F. & Huang, W., 2011. Remote Tele-expert System for Maintenance Operators in Mines. *2011 Underground Coal Operators' Conference The*, pp.171–177. Available at: <http://ro.uow.edu.au/coal/355/>.
- Almeida, I.D.S. et al., 2012. AR-based video-mediated communication: A social presence enhancing experience. *Proceedings - 2012 14th Symposium on Virtual and Augmented Reality, SVR 2012*, pp.125–130.
- Azuma, R. & Azuma, R., 1997. A survey of augmented reality. *Presence: Teleoperators and Virtual Environments*, 6(4), pp.355–385. Available at: <http://scholar.google.com/scholar?q=intitle:A+Survey+of+Augmented+Reality#0>.
- Bell A. G. (1876). Improvement in telegraphy. U.S. Patent No. 174,465, filed February 14, 1876, granted March 7, 1876. Washington, DC: U.S. Patent and Trademark Office.
- Billinghurst, M. & Kato, H., 2002. Collaborative augmented reality. *Communications of the ACM*, 45(7).
- Brubaker, J.R., Venolia, G. & Tang, J.C., 2012. Focusing on Shared Experiences : Moving Beyond the Camera in Video Communication. *DIS '12: Proceedings of the Designing Interactive Systems Conference*, pp.96–105.
- Comport, A.I. et al., 2006. Real-time markerless tracking for augmented reality: The virtual visual servoing framework. *IEEE Transactions on Visualization and Computer Graphics*, 12(4), pp.615–628.
- DeVito, J. A., *The interpersonal communication book*, 8th edition, New York: Longman, 1998

- Edwards, M.B. & Gronlund, S.D., 1998. Task interruption and its effects on memory. *Memory (Hove, England)*, 6(6), pp.665–687.
- Fussell, S. et al., 2004. Gestures Over Video Streams to Support Remote Collaboration on Physical Tasks. *Human-Computer Interaction*, 19(3), pp.273–309.
- Fussell, S.R., Setlock, L.D. & Kraut, R.E., 2003. Effects of head-mounted and scene-oriented video systems on remote collaboration on physical tasks. *Proceedings of the conference on Human factors in computing systems - CHI '03*, (5), p.513. Available at: <http://portal.acm.org/citation.cfm?doid=642611.642701>.
- Gauglitz, S. et al., 2014a. In touch with the remote world. *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology - VRST '14*, pp.197–205. Available at: <http://dl.acm.org/citation.cfm?id=2671015.2671016>.
- Gauglitz, S., Lee, C., Turk, M., et al., 2012. Integrating the physical environment into mobile remote collaboration. *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services - MobileHCI '12*, pp.241–250. Available at: <http://dl.acm.org/citation.cfm?id=2371574.2371610>.
- Gauglitz, S., Lee, C., Sweeney, C., et al., 2012. Interactive Remote Collaboration using Augmented Reality Interactive Remote Collaboration using Augmented Reality. , (September), p.2012.
- Gauglitz, S. et al., 2014b. World-stabilized annotations and virtual scene navigation for remote collaboration. *Proceedings of the 27th annual ACM symposium on User interface software and technology - UIST '14*, pp.449–459. Available at: <http://dl.acm.org/citation.cfm?id=2642918.2647372>.
- Gaver, W.W. et al., 1993. One is not enough: multiple views in a media space. *Proceedings of SIGCHI Conference on Human Factors in Computing Systems (CHI)*, pp.335–341.
- Gillie, T. & Broadbent, D., 1989. What makes interruptions disruptive? A study of length, similarity, and complexity. *Psychological Research*, 50(4), pp.243–250.
- Griffin, E. A.: 2005, A first look at communication theory, 5th edn, McGrawHill

- Gouran, D., Wiethoff, W. E. and Doelger, J. A.: 1994, *Mastering communication*, Allyn and Bacon (Boston)
- Gupta, K., "Wearable Tools for Affective Remote Collaboration", Master Thesis, University of Canterbury, Christchurch, New Zealand, 2015
- Hameed, S. et al., 2009. Using informative peripheral visual and tactile cues to support task and interruption management. *Human factors*, 51(2), pp.126–135.
- Hauber, J., "Understanding Remote Collaboration in VideoCollaborative Virtual Environments." Doctoral Thesis, University of Canterbury, Christchurch, New Zealand, 2008
- Huang, W. & Alem, L., 2011. Supporting hand gestures in mobile remote collaboration: a usability evaluation. *Proceedings of the 25th BCS Conference on Human-Computer Interaction*, pp.211–216. Available at: <http://dl.acm.org/citation.cfm?id=2305356>.
- Inkpen, K. et al., 2013. Experiences2Go : Sharing Kids ' Activities Outside the Home with Remote Family Members. *Cscw'13*, (February 23–27), pp.1329–1339.
- Inkpen, K. et al., 2012. Video kids: augmenting close friendships with asynchronous video conversations in videopal. *CHI '12: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp.2387–2396. Available at: <http://dl.acm.org/citation.cfm?doid=2207676.2208400&npapers3://publication/doi/10.1145/2207676.2208400>.
- Isaacs, E. a & Tang, J.C., 1994. What video can and cannot do for collaboration: A case study. *Multimedia Systems*, 2(2), pp.63–73. Available at: <http://link.springer.com/10.1007/BF01274181>.
- Ishii, H., Kobayashi, M. & Arita, K., 1994. Iterative design of seamless collaboration media. *Communications of the ACM*, 37(8), pp.83–97.
- Ishii, H., Kobayashi, M. & Grudin, J., 1992. Integration of Interpersonal Space and Shared Workspace - ClearBoard Design and Experiments. *Proceedings of the International Conference on Computer Supported Cooperative Work (CSCW'92)*, 11(4), pp.349–375.
- Kasahara, S. & Rekimoto, J., 2014. JackIn: integrating first-person view with out-of-body vision generation for human-human augmentation. *Proceedings of the 5th Augmented Human ...*, pp.1–8. Available at: <http://dl.acm.org/citation.cfm?id=2582097>.

- Kato, H. & Billinghurst, M., 1999. Marker tracking and HMD calibration for a video-based augmented reality conferencing system. *Proceedings 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR'99)*.
- Kim, S., Junuzovic, S. & Inkpen, K., 2014. The Nomad and the Couch Potato. In *Proceedings of the 18th International Conference on Supporting Group Work - GROUP '14*. New York, New York, USA: ACM Press, pp. 167–177. Available at: <http://dl.acm.org/citation.cfm?doid=2660398.2660409>.
- Kirk, D., Rodden, T. & Stanton Fraser, D., 2007. Turn it this way. , pp.1039–1048. Available at: <http://dx.doi.org/10.1145/1240624.1240782>.
- Kurata, T. et al., 2004. Remote collaboration using a shoulder-worn active camera/laser. *Wearable Computers, 2004. ISWC 2004. Eighth International Symposium on*, 1, pp.62–69.
- Kuznetsova, A. et al., 2015. Expanding Object Detector ' s H ORIZON : Incremental Learning Framework for Object Detection in Videos. *Cvpr*, pp.28–36.
- Kuzuoka, H., Ishimoda, G., et al., 1994. Can the GestureCam be a surrogate? *Media*, pp.181–196. Available at: <http://dl.acm.org/citation.cfm?id=1241958.1241970>.
- Kuzuoka, H., Kosuge, T. & Tanaka, M., 1994. GestureCam: A video communication system for sympathetic remote collaboration. *Proceedings of the 1994 ACM conference on Computer supported cooperative work*, pp.35–43. Available at: <http://dl.acm.org/citation.cfm?id=192866>.
- Latorella, K.A., 1999. Investigating interruptions: Implications for flightdeck performance. , (October).
- Li, S.Y.W., Magrabi, F. & Coiera, E., 2012. A systematic review of the psychological literature on interruption and its patient safety implications. *Journal of the American Medical Informatics Association*, 19(1), pp.6–12.
- Lu, S. a. et al., 2013. Supporting Interruption Management and Multimodal Interface Design: Three Meta-Analyses of Task Performance as a Function of Interrupting Task Modality. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, p.0018720813476298–. Available at: <http://hfs.sagepub.com/cgi/content/long/0018720813476298v1>.

- O'Hara, K., Black, A. & Lipson, M., 2006. Everyday practices with mobile video telephony. *Proceedings of the SIGCHI conference on Human Factors in computing systems - CHI '06*, pp.871–880. Available at: <http://portal.acm.org/citation.cfm?doid=1124772.1124900>.
- Poelman, R. et al., 2012. As if being there: mediated reality for crime scene investigation. *Proceedings of the ACM ...*, (5), pp.1267–1276. Available at: <http://dl.acm.org/citation.cfm?id=2145394>.
- Posner, M.I., 1980. Orienting of attention. *The Quarterly journal of experimental psychology*, 32(1), pp.3–25.
- Reichherzer, C., Nassani, a & Billinghamurst, M., 2014. [Poster] Social panoramas using wearable computers. *Mixed and Augmented Reality (ISMAR), 2014 IEEE International Symposium on*, (September), pp.303–304.
- Rekimoto, J. & Nagao, K., 1995. The World through the Computer: Computer Augmented Interaction with Real World Environments. *Proc 8th Ann ACM Symp User Interface and Software Technology UIST ACM Press*, pages, pp.29–36. Available at: <http://portal.acm.org/citation.cfm?doid=215585.215639>.
- Roszak, T.: 1986, *The cult of information: the folklore of computers and the true art of thinking.*, University of California Press
- Sakata, N. et al., 2003. WACL: supporting telecommunications using - wearable active camera with laser pointer. *Seventh IEEE International Symposium on Wearable Computers, 2003. Proceedings.*, pp.53–56. Available at: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1241393>.
- Schnaars, S. & Wymbs, C., 2004. On the persistence of lackluster demand - the history of the video telephone. *Technological Forecasting and Social Change*, 71(3), pp.197–216.
- Seo, D. et al., 2014. Real-time panoramic video streaming system with overlaid interface concept for social media. *Multimedia Systems*, 20(6), pp.707–719.
- Shannon, C.E. and Weaver, W., *The Mathematical Theory of Communication*, The University of Illinois Press, 1949
- Sodhi, R.S. et al., 2013. BeThere: 3D Mobile Collaboration with Spatial Input. *Proceedings of the SIGCHI Conference on Human Factors in Computing*

*Systems - CHI '13*, pp.179–188. Available at:  
<http://dl.acm.org/citation.cfm?doid=2470654.2470679>.

Suzuki, T.; Taniguchi, H. & Takada, H. (1986), A Real-Time Electronic Conferencing System based on Distributed UNIX., in 'USENIX Summer' , USENIX Association, , pp. 189-199 .

Tan, W. et al., 2013. Robust monocular SLAM in dynamic environments. *2013 IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2013*, pp.209–218.

Tang, a, 2004. Embodiments and VideoArms in mixed presence groupware. Available at:  
<http://grouplab.cpsc.ucalgary.ca/grouplab/uploads/Publications/Publications/2004-VideoArms.Report2004-741-06.pdf>.

Tang, J.C. & Minneman, S.L., 1991. Videodraw: a video interface for collaborative drawing. *ACM Transactions on Information Systems*, 9(2), pp.170–184.

Wickens, C. et al., 2011. Auditory-Visual Redundancy in Vehicle Control Interruptions: Two Meta-analyses. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 55(1), pp.1155–1159.

Wobbrock, J.O. et al., 2011. The aligned rank transform for nonparametric factorial analyses using only ANOVA procedures. *CHI '11 Conference on Human Factors in Computing Systems*, pp.143–146.

Woods, D. D. (1995). The alarm problem and directed attention in dynamic fault management. *Ergonomics*, 38, 2371–2393

Yarosh, S., Inkpen, K.M. & Brush, a. J.B., 2010. Video playdate. *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10*, p.1251. Available at:  
<http://portal.acm.org/citation.cfm?id=1753514>  
<http://portal.acm.org/citation.cfm?doid=1753326.1753514>.

## Appendices

### Appendix A : Questionnaire for Demographic Information

#### Questionnaire # 1

**Group:**    \_\_\_

Please fill out section A before you start tasks. If you have any questions, please feel free to ask the experiment conductor.

#### Section A (Before Start Tasks)

1. Please choose your gender below      Male                      Female
2. How old are you?                      \_\_\_\_\_
3. Relationship with you partner ?      \_\_\_\_\_
4. Have you ever used a video conferencing system such as Skype or Google Hangout?  
  
YES   NO  
If yes, how often do you use the video conferencing system?  
\_\_\_\_\_ e.g.) Weekly, Daily or Monthly
5. Which of you usually make a decision?      Me                      My partner
6. Whom do you usually have a video chatting with?  
  
\_\_\_\_\_

7. If you did any special activity during the video chatting, please describe it.

---

---

## **Appendix B : Material for Study about the Use of Visual Communication cues**

Human Interface Lab (HIT Lab) NZ  
Telephone: +64 3 364 2349  
Email: seungwon.kim@pg.canterbury.ac.nz



### **Information Sheet**

**Title: Pointing and Annotation interface for Remote Collaboration**

#### **Invitation**

You are invited to participate as a subject in the research project "Pointing and Annotation interface in Remote Collaboration". This information sheet provides you with the research study details. This investigation will take about one and half hour to finish. The study will be conducted by Seungwon Kim, a PhD candidate at Human Interface Technology Laboratory New Zealand (HITLab NZ). Seungwon Kim will be available to answer your questions and provide further explanations. This study is under the supervision of Professor Mark Billingham, header of HITLab NZ, and Post-Doctoral fellow Gun Lee. They can be contacted at (03) 364-2349 or email: {seungwon.kim, gun.lee, mark.billingham }@canterbury.ac.nz. They will be pleased to discuss any concerns you may have about participation in the study. If you agree to take part in the research study, you will be asked to sign a consent form. Your participation in this study is voluntary. You are free to choose whether or not you will take part in the study.

#### **1. 'What is the purpose of this study?'**

The purpose of this study is to explore the role of pointing and annotation cues with 2D and 3D tasks in a computer interface for remote collaboration.

#### **2. 'Why have I been invited to participate in this study?'**

You are eligible to participate in this study because you have average skill in

using a mouse and you can hold and move small pieces as you want. Also, your ability of speaking English is enough to describe the orientation and position of the pieces with pointing or annotation interaction on still image and live video.

### **3. 'What does this study involve?'**

At the beginning, the instructor will ask you and your partner to enter a normal room at HITLab NZ. In the room, you will be asked to sit on a normal chair and a normal desk will be in front of you. The desk is the area you will conduct this experiment with a personal computer, head mounted display(HMD) system or a tablet. Your partner and you cannot see each other and your partner will be asked to sit at the opposite side of you.

There is a one questionnaire that is needed to be filled out before you start using our remote collaboration system. The questionnaire includes several questions asking about your experience of Skype or Google Talk, your age, and your gender.

Next, the instructor will show you how to give commands with one of our systems. The commands are achieved with normal mouse interaction. Then, you will have time to practice with the system, and carry out a task that involves verbal communication with your partner and changing the position of pointers or annotating on the view. If you participate in this investigation as an remote user, then you will discuss with your partner for assembling a model by verbal communication , a pointer or annotation. If you participate in this investigation as a local user, you will wear HMD or hold a Window Surface Pro and discuss with your partner by verbal communication or hand gestures. After finishing the task, you will be asked to fill out a questionnaire that includes several questions about your experience with one of our systems.

There are three different types of systems, and you will be asked to conduct one task for each types of systems as described above. For finishing one task, it will take about 4~6 minutes, and the experiment will be finished in an hour.

During the experiment, two videos will be recorded for observing users' interaction. The first will record the screen of PC and the second will record the view of the tablet user with a small camera attached on the cap that the tablet user wears on. All inputs of users will be recorded into a log file. There is a questionnaire that contains three sections, before experiment, after each task, and after experiment. Then, final interview will be conducted.

### **4. Are there any possible risks from participation in this study?**

There are no risks anticipated with participation in this study. Participants can

withdraw from the project at any time with no penalty, including withdrawal of any information provided.

**5. Are there any possible invasion of privacy from participation in this study?**

There will be two video recordings, one for PC screen and the other one for tablet user's view. During the experiment, there is no chance to record participants' appearance, but your verbal communication will be recorded. However, all the data we get from you will be treated in a confidential manner. Only an experiment conductor, Mr. Seungwon Kim, and two supervisors, Dr. Gun Lee and Prof. Mark Billinghamurst, have the authority to access to the video recordings.

The results of the project may be published, but you may be assured of the complete confidentiality of data gathered in this investigation. To ensure anonymity and confidentiality, you are not required to give any personal information during this study without your age and your gender. The measured data will be used only by Mr. Seungwon Kim, Dr. Gun Lee and Prof. Mark Billinghamurst. Your data will be kept for a maximum of five years before being destroyed.

**6. Are there any possible benefits from participation in this study?**

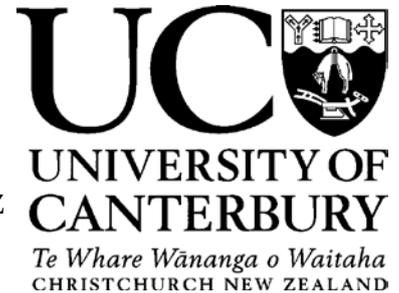
Benefits to participating in a study include:

- The satisfaction of helping others by contributing to the research on remote collaboration systems, or helping to identify a possible new interface for remote collaboration systems.
- The opportunity to have an experience of using pointing and annotation communication cues with video chatting program similar with Skype and Google Talk.

**Thank you for taking the time to consider this study.  
If you wish to take part in it, please sign the attached  
consent form.**

**This information sheet is for you to keep.**

Human Interface Lab (HIT Lab) NZ  
Telephone: +64 3 364 2349  
Email: seungwon.kim@pg.canterbury.ac.nz



## Consent Form

### **Title: Pointing and Annotation interface for Remote Collaboration**

1. I have read and understood the 'Information Sheet' for this project.
2. The nature and possible effects of the study have been explained to me.
3. I understand that the study involves mouse interaction and wearing head mounted system for remote collaboration.
4. I understand that participation involves the risk(s) that is not beyond those that occur in daily life. Participants will be volunteers and can withdraw from the project at any time with no penalty.
5. I understand that all research data will be securely stored on the HITLab NZ for 5 years, and will then be destroyed by administration staff via secure shredding.
6. Any questions that I have asked have been answered to my satisfaction.

7. I agree that research data gathered from me for the study may be published and I cannot be identified as a participant in the publication.
8. I understand that the researchers will maintain my identity confidential and that any information I supply to the researcher(s) will be used only for the purposes of the research.
9. I agree to participate in this investigation which approximately takes one and half hour and understand that any data I have supplied will be withdrawn at any time in accordance with my wish.

Name of Participant:

Signature:

Date:

### Statement by Investigator

I have explained the project & the implications of participation in it to this volunteer and I believe that the consent is informed and that he/she understands the implications of participation

If the Investigator has not had an opportunity to talk to participants prior to them participating, the following must be ticked.

The participant has received the Information Sheet where my details have been provided so participants have the opportunity to contact me prior to consenting to participate in this project.

Name of Investigator    Mr Seungwon Kim

Signature                      of  
Investigator

Name of investigator : Mr Seungwon Kim

Signature of investigator \_\_\_\_\_

Date \_\_\_\_\_

## Questionnaire (Remote) Condition Group:

Below are statements that describe your experience. Please circle on a number to indicate your level of agreement or disagreement with each statement (0 - Strongly Disagree, 10 - Strongly Agree)

1. I enjoyed assembling a model.

0      1      2      3      4      5      6      7      8      9      10

2. I felt connected with my partner.

0      1      2      3      4      5      6      7      8      9      10

3. I was able to focus on the activity (building models)

0      1      2      3      4      5      6      7      8      9      10

4. I felt I was there with my partner

0      1      2      3      4      5      6      7      8      9      10

5. I was able to express my idea properly

0      1      2      3      4      5      6      7      8      9      10

6. I easily understood what my partner express.

0      1      2      3      4      5      6      7      8      9      10

7. (Short Interview)

1) What do you like most and dislike most with this condition?

**Questionnaire (Local) Condition \_\_\_\_\_ Group:**

—

Below are statements that describe your experience. Please choose a number to indicate your level of agreement or disagreement with each statement (0 - Strongly Disagree, 10 - Strongly Agree)

1. I enjoyed assembling a model.

0      1      2      3      4      5      6      7      8      9      10

2. I felt connected with my partner.

0      1      2      3      4      5      6      7      8      9      10

3. I was able to focus on the activity (building models)

0      1      2      3      4      5      6      7      8      9      10

4. I felt my partner was here with me.

0      1      2      3      4      5      6      7      8      9      10

5. I was able to express my idea properly

0      1      2      3      4      5      6      7      8      9      10

6. I easily understood what my partner express.

0      1      2      3      4      5      6      7      8      9      10

7. (Short Interview) What do you like most and dislike most with this condition

## Questionnaire (R) After three conditions

**Group:**   —

Please rank the condition according to your experience. (1 - best, 2 - second best, 3 - worst)

1. Rank the conditions according to *the level of you enjoyed*.

None additional visual cue (        )     Annotation    (        )

Pointers (        )

2. Rank the conditions according to *the level of you felt connected with your partner*.

None additional visual cue (        )     Annotation    (        )

Pointers (        )

3. Rank the conditions according to *the level of you focused on assembling a model*.

None additional visual cue (        )     Annotation    (        )

Pointers (        )

4. Rank the conditions according to *the level of you felt being at the work space with your partner*.

None additional visual cue (        )     Annotation    (        )

Pointers (        )

5. Rank the conditions according to *how much you were able to properly express your idea*.

None additional visual cue (       )     Annotation    (       )  
Pointers (       )

6. Rank the conditions according to *how easily you understood your partner's expression*.

None additional visual cue (       )     Annotation    (       )  
Pointers (       )

7. (Short Interview)

You mostly choose (       ) condition as the best. Could you explain why?

You mostly choose (       ) condition as the worst. Could you explain why?

## Questionnaire (L) After three conditions

**Group:** —

Please rank the condition according to your experience. (1 - best, 2 - second best, 3 - worst)

1. Rank the conditions according to *the level of you enjoyed*.

None additional visual cue ( )      Annotation ( )  
Pointers ( )

2. Rank the conditions according to *the level of you felt connected with your partner*.

None additional visual cue ( )      Annotation ( )  
Pointers ( )

3. Rank the conditions according to *the level of you focused on assembling a model*.

None additional visual cue ( )      Annotation ( )  
Pointers ( )

4. Rank the conditions according to *the level of you felt your partner being with you*.

None additional visual cue ( )      Annotation ( )  
Pointers ( )

5. Rank the conditions according to *how much you were able to properly express your idea.*

None additional visual cue (       )     Annotation    (       )  
Pointers (       )

6. Rank the conditions according to *how easily you understood your partner's expression.*

None additional visual cue (       )     Annotation    (       )  
Pointers (       )

7. (Short Interview)

You mostly choose (       ) condition as the best. Could you explain why?

You mostly choose (       ) condition as the worst. Could you explain why?

## Appendix C : Material for the Study about Freeze Frame

Human Interface Lab (HIT Lab) NZ

Telephone: +64 3 364 2349

Email: seungwon.kim@pg.canterbury.ac.nz

### Information Sheet



**Title: Freeze function while using annotation interface for Remote Collaboration**

#### Invitation

You are invited to participate as a subject in the research project "Freeze function while using annotation interface in Remote Collaboration". This information sheet provides you with the research study details. This investigation will take about one and half hour to finish. The study will be conducted by Seungwon Kim, a PhD candidate at Human Interface Technology Laboratory New Zealand (HITLab NZ). Seungwon Kim will be available to answer your questions and provide further explanations. This study is under the supervision of Professor Mark Billinghurst, header of HITLab NZ, and Post-Doctoral fellow Gun Lee. They can be contacted at (03) 364-2349 or email: {seungwon.kim, gun.lee, mark.billinghurst}@canterbury.ac.nz. They will be pleased to discuss any concerns you may have about participation in the study. If you agree to take part in the research study, you will be asked to sign a consent form. Your participation in this study is voluntary. You are free to choose whether or not you will take part in the study.

#### 1. 'What is the purpose of this study?'

The purpose of this study is to explore the role of freeze function on the user of annotation cues with 2D and 3D tasks in a computer interface for remote collaboration.

#### 2. 'Why have I been invited to participate in this study?'

You are eligible to participate in this study because you have average skill in using a mouse and you can hold and move small pieces as you want. Also, your ability of speaking English is enough to describe the orientation and

position of the pieces with annotation interaction on still image and live video.

### **3. 'What does this study involve?'**

At the beginning, the instructor will ask you and your partner to enter a normal room at HITLab NZ. In the room, you will be asked to sit on a normal chair and a normal desk will be in front of you. The desk is the area you will conduct this experiment with a personal computer or head mounted display(HMD) system. Your partner and you cannot see each other and your partner will be asked to sit at the opposite side of you.

There is a one questionnaire that is needed to be filled out before you start using our remote collaboration system. The questionnaire includes several questions asking about your experience of Skype or Google Talk, your age, and your gender.

Next, the instructor will show you how to give commands with one of our systems. The commands are achieved with normal mouse interaction. Then, you will have time to practice with the system, and carry out a task that involves verbal communication with your partner and annotating on the view. If you participate in this investigation as an remote user, then you will discuss with your partner for assembling a model by verbal communication and by drawing annotation. If you participate in this investigation as a local user, you will wear HMD and discuss with your partner by verbal communication or hand gestures. After finishing the task, you will be asked to fill out a questionnaire that includes several questions about your experience with one of our systems.

There are three different types of systems, and you will be asked to conduct one task for each types of systems as described above. For finishing one task, it will take about 4~6 minutes, and the experiment will be finished in one and half hour.

During the experiment, two videos will be recorded for observing users' interaction. The first will record the screen of PC and the second will record the view of the tablet user with a small camera attached on the cap that the tablet user wears on. All inputs of users will be recorded into a log file. There is a questionnaire that contains three sections, before experiment, after each task, and after experiment. Then, final interview will be conducted.

### **4. Are there any possible risks from participation in this study?**

There are no risks anticipated with participation in this study. Participants can withdraw from the project at any time with no penalty, including withdrawal of any information provided.

**5. Are there any possible invasion of privacy from participation in this study?**

There will be two video recordings, one for PC screen and the other one for tablet user's view. During the experiment, there is no chance to record participants' appearance, but your verbal communication will be recorded. However, all the data we get from you will be treated in a confidential manner. Only an experiment conductor, Mr. Seungwon Kim, and two supervisors, Dr. Gun Lee and Prof. Mark Billinghurst, have the authority to access to the video recordings.

The results of the project may be published, but you may be assured of the complete confidentiality of data gathered in this investigation. To ensure anonymity and confidentiality, you are not required to give any personal information during this study without your age and your gender. The measured data will be used only by Mr. Seungwon Kim, Dr. Gun Lee and Prof. Mark Billinghurst. Your data will be kept for a maximum of five years before being destroyed.

**6. Are there any possible benefits from participation in this study?**

Benefits to participating in a study include:

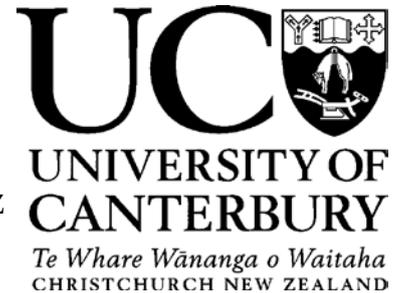
- The satisfaction of helping others by contributing to the research on remote collaboration systems, or helping to identify a possible new interface for remote collaboration systems.
- The opportunity to have an experience of using pointing and annotation communication cues with video chatting program similar with Skype and Google Talk.

**Thank you for taking the time to consider this study.**

**If you wish to take part in it, please sign the attached consent form.**

**This information sheet is for you to keep.**

Human Interface Lab (HIT Lab) NZ  
Telephone: +64 3 364 2349  
Email: seungwon.kim@pg.canterbury.ac.nz



## Consent Form

### **Title: Freeze function while using annotation interface for Remote Collaboration**

1. I have read and understood the 'Information Sheet' for this project.
2. The nature and possible effects of the study have been explained to me.
3. I understand that the study involves mouse interaction and wearing head mounted system for remote collaboration.
4. I understand that participation involves the risk(s) that is not beyond those that occur in daily life. Participants will be volunteers and can withdraw from the project at any time with no penalty.
5. I understand that all research data will be securely stored on the HITLab NZ for 5 years, and will then be destroyed by administration staff via secure shredding.
6. Any questions that I have asked have been answered to my satisfaction.

7. I agree that research data gathered from me for the study may be published and I cannot be identified as a participant in the publication.
8. I understand that the researchers will maintain my identity confidential and that any information I supply to the researcher(s) will be used only for the purposes of the research.
9. I agree to participate in this investigation which approximately takes one and half hour and understand that any data I have supplied will be withdrawn at any time in accordance with my wish.

Name of Participant:

Signature:

Date:

### Statement by Investigator

I have explained the project & the implications of participation in it to this volunteer and I believe that the consent is informed and that he/she understands the implications of participation

If the Investigator has not had an opportunity to talk to participants prior to them participating, the following must be ticked.

The participant has received the Information Sheet where my details have been provided so participants have the opportunity to contact me prior to consenting to participate in this project.

Name of Investigator Mr Seungwon Kim

Signature of Investigator

Name of investigator : Mr Seungwon Kim

Signature of investigator \_\_\_\_\_

Date \_\_\_\_\_

## Questionnaire (Remote) Condition \_\_\_\_\_

**Group:** \_\_\_\_\_

Below are statements that describe your experience. Please circle on a number to indicate your level of agreement or disagreement with each statement

(0 - Strongly Disagree, 10 - Strongly Agree)

1. I was able to focus on the activity (building models)

0      1      2      3      4      5      6      7      8      9      10

2. I was able to express my idea properly

0      1      2      3      4      5      6      7      8      9      10

3. It was easy to draw/annotate on the remote view.

0      1      2      3      4      5      6      7      8      9      10

4. I was able to draw annotation on the remote view as soon as I wanted to.

0      1      2      3      4      5      6      7      8      9      10

5. I easily understand what my partner was trying to do and explaining

0      1      2      3      4      5      6      7      8      9      10

6. I felt we collaborated well

0      1      2      3      4      5      6      7      8      9      10

7. The interface was *mentally stressful* to use

0      1      2      3      4      5      6      7      8      9      10

8. I felt time pressure while drawing on the remote view.

0      1      2      3      4      5      6      7      8      9      10

9. I had to be careful while drawing on the remote view.

0      1      2      3      4      5      6      7      8      9      10

10. The interface was *physically stressful* to use

0      1      2      3      4      5      6      7      8      9      10

11. What do you like most and dislike most with this condition?

(Short Interview will be followed)

## Questionnaire (Local) Condition \_\_\_\_\_

**Group:** \_\_\_\_\_

Below are statements that describe your experience. Please circle on a number to indicate your level of agreement or disagreement with each statement

(0 - Strongly Disagree, 10 - Strongly Agree)

1. I was able to focus on the activity (building models)

0      1      2      3      4      5      6      7      8      9      10

2. I was able to express my idea properly

0      1      2      3      4      5      6      7      8      9      10

3. I easily understand what my partner was trying to do and explaining

0      1      2      3      4      5      6      7      8      9      10

4. I felt we collaborated well

0      1      2      3      4      5      6      7      8      9      10

5. The interface was *mentally stressful* to use

0      1      2      3      4      5      6      7      8      9      10

6. The interface was *physically stressful* to use

0      1      2      3      4      5      6      7      8      9      10

7. What do you like most and dislike most with this condition?

(Short Interview will be followed)

## Appendix D : Material for the Study about Notificatioion

Human Interface Lab (HIT Lab) NZ  
Telephone: +64 3 364 2349  
Email: seungwon.kim@pg.canterbury.ac.nz



# Information Sheet

**Title: Visual Notification while using annotation interface for Remote Collaboration**

## Invitation

You are invited to participate as a subject in the research project "Visual Notification while using annotation interface in Remote Collaboration". This information sheets provides you with the research study details. This investigation will take about one and half hour to finish. The study will be conducted by Seungwon Kim, a PhD candidate at Human Interface Technology Laboratory New Zealand (HITLab NZ). Seungwon Kim will be available to answer your questions and provide further explanations. This study is under the supervision of Professor Mark Billingham, header of HITLab NZ, and Post-Doctoral fellow Gun Lee. They can be contacted at (03) 364-2349 or email: {seungwon.kim, gun.lee, mark.billinghurst }@canterbury.ac.nz. They will be pleased to discuss any concerns you may have about participation in the study. If you agree to take part in the research study, you will be asked to sign a consent form. Your participation in this study is voluntary. You are free to choose whether or not you will take part in the study.

### 1. 'What is the purpose of this study?'

The purpose of this study is to explore the role of visual notification on the use of annotation cues with 2D and 3D tasks in a computer interface for remote collaboration.

## **2. ‘Why have I been invited to participate in this study?’**

You are eligible to participate in this study because you have average skill in using a mouse and you can hold and move small pieces as you want. Also, your ability of speaking English is enough to describe the orientation and position of the pieces with annotation interaction on still images and a live video.

## **3. ‘What does this study involve?’**

At the beginning, the instructor will ask you and your partner to enter a normal room at HITLab NZ. In the room, you will be asked to sit on a normal chair and a normal desk will be in front of you. The desk is the area you will conduct this experiment with a personal computer or head mounted display(HMD) system. Your partner and you cannot see each other and your partner will be asked to sit at the opposite side of you.

There is a one questionnaire that is needed to be filled out before you start using our remote collaboration system. The questionnaire includes several questions asking about your experience of Skype or Google Talk, your age, and your gender.

Next, the instructor will show you how to give commands with one of our systems. The commands are achieved with normal mouse and keyboard interaction. Then, you will have time to practice with the system, and carry out a task that involves verbal communication with your partner and annotating on the view. If you participate in this investigation as a remote user, then you will discuss with your partner for assembling a model by verbal communication and by drawing annotation. If you participate in this investigation as a local user, you will wear HMD and discuss with your partner by verbal communication or hand gestures. After finishing the each task, you will be asked to fill out a questionnaire that includes several questions about your experience with one of our systems.

There are three different types of systems, and you will be asked to conduct one task for each types of systems as described above. For finishing one task, it will take about 4~6 minutes, and the experiment will be finished in one and half hour.

During the experiment, two videos will be recorded for observing users’ interaction. The first will record the screen of PC and the second will record the view of the tablet user with a small camera attached on the cap that the tablet user wears on. All inputs

of users will be recorded into a log file. There is a questionnaire that contains three sections, before experiment, after each task, and after experiment. Then, final interview will be conducted.

**4. Are there any possible risks from participation in this study?**

There are no risks anticipated with participation in this study. Participants can withdraw from the project at any time with no penalty, including withdrawal of any information provided.

**5. Are there any possible invasion of privacy from participation in this study?**

There will be two video recordings, one for PC screen and the other one for tablet user's view. During the experiment, there is no chance to record participants' appearance, but your verbal communication will be recorded. However, all the data we get from you will be treated in a confidential manner. Only an experiment conductor, Mr. Seungwon Kim, and two supervisors, Dr. Gun Lee and Prof. Mark Billinghurst, have the authority to access to the video recordings.

The results of the project may be published, but you may be assured of the complete confidentiality of data gathered in this investigation. To ensure anonymity and confidentiality, you are not required to give any personal information during this study without your age and your gender. The measured data will be used only by Mr. Seungwon Kim, Dr. Gun Lee and Prof. Mark Billinghurst. Your data will be kept for a maximum of five years before being destroyed.

**6. Are there any possible benefits from participation in this study?**

Benefits to participating in a study include:

- The satisfaction of helping others by contributing to the research on remote collaboration systems, or helping to identify a possible new interface for remote collaboration systems.
- The opportunity to have an experience of using pointing and annotation communication cues with video chatting program similar with Skype and Google Talk.

**Thank you for taking the time to consider this study.**

**If you wish to take part in it, please sign the attached consent form.**

**This information sheet is for you to keep.**

## **Appendix E : Material for the Study about Independent view**

Human Interface Lab (HIT Lab) NZ  
Telephone: +64 3 364 2349  
Email: seungwon.kim@pg.canterbury.ac.nz



# **Consent Form**

**Title: Visual Notification while using annotation interface for Remote Collaboration**

1. I have read and understood the 'Information Sheet' for this project.
2. The nature and possible effects of the study have been explained to me.
3. I understand that the study involves mouse interaction and wearing head mounted system for remote collaboration.
4. I understand that participation involves the risk(s) that is not beyond those that occur in daily life. Participants will be volunteers and can withdraw from the project at any time with no penalty.
5. I understand that all research data will be securely stored on the HITLab NZ for 5 years, and will then be destroyed by administration staff via secure shredding.

6. Any questions that I have asked have been answered to my satisfaction.
7. I agree that research data gathered from me for the study may be published and I cannot be identified as a participant in the publication.
8. I understand that the researchers will maintain my identity confidential and that any information I supply to the researcher(s) will be used only for the purposes of the research.

9. I agree to participate in this investigation which approximately takes one and half hour and understand that any data I have supplied will be withdrawn at any time in accordance with my wish.

Name of Participant:

---

Signature:

Date:

---

### Statement by Investigator

I have explained the project & the implications of participation in it to this volunteer and I believe that the consent is informed and that he/she understands the implications of participation

If the Investigator has not had an opportunity to talk to participants prior to them participating, the following must be ticked.

The participant has received the Information Sheet where my details have been provided so participants have the opportunity to contact me prior to consenting to participate in this project.

Name of Investigator    Mr Seungwon Kim

Signature                    of  
Investigator

Name of investigator : Mr Seungwon Kim

Signature of investigator \_\_\_\_\_

Date \_\_\_\_\_

## Questionnaire (Remote) Condition \_\_\_\_\_

**Group:** \_\_\_\_\_

Below are statements that describe your experience. Please circle on a number to indicate your level of agreement or disagreement with each statement

(0 - Strongly Disagree, 10 - Strongly Agree)

1. I was able to focus on the activity (building models)

0      1      2      3      4      5      6      7      8      9      10

2. I felt interrupted when my partner was drawing (or I felt I was interrupting my partner when I was drawing)

0      1      2      3      4      5      6      7      8      9      10

3. I was able to express my idea properly

0      1      2      3      4      5      6      7      8      9      10

4. I easily understood what my partner was trying to do and explaining.

0      1      2      3      4      5      6      7      8      9      10

5. I was able to focus on my partner's explanation. (I felt my partner was focusing on my explain)

0      1      2      3      4      5      6      7      8      9      10

6. I felt we collaborated well

0      1      2      3      4      5      6      7      8      9      10

7. The communication with my partner was mentally stressful

0      1      2      3      4      5      6      7      8      9      10

8. We focused on the same piece or area while collaborating

0      1      2      3      4      5      6      7      8      9      10

9. I was aware of where my partner was drawing (or My partner was aware of where I was drawing)

0      1      2      3      4      5      6      7      8      9      10

10. What do you like most and dislike most with this condition?

(Short Interview will be followed)

**Questionnaire (Local) Condition \_\_\_\_\_**

**Group: \_\_\_\_\_**

Below are statements that describe your experience. Please circle on a number to indicate your level of agreement or disagreement with each statement (0 - Strongly Disagree, 10 - Strongly Agree)

1. I was able to focus on the activity (building models)

0      1      2      3      4      5      6      7      8      9      10

2. I felt interrupted when my partner was drawing (or I felt I was interrupting my partner when I was drawing)

0      1      2      3      4      5      6      7      8      9      10

3. I was able to express my idea properly

0      1      2      3      4      5      6      7      8      9      10

4. I easily understood what my partner was trying to do and explaining.

0      1      2      3      4      5      6      7      8      9      10

5. I was able to focus on my partner's explanation. (I felt my partner was focusing on my explain)

0      1      2      3      4      5      6      7      8      9      10

6. I felt we collaborated well

0      1      2      3      4      5      6      7      8      9      10

7. I knew the moment when my partner wanted to explain his idea

0 1 2 3 4 5 6 7 8 9 10

8. The communication with my partner was mentally stressful

0 1 2 3 4 5 6 7 8 9 10

9. We focused on the same piece or area while collaborating

0 1 2 3 4 5 6 7 8 9 10

10. I was aware of where my partner was drawing (or My partner was aware of where I was drawing)

0 1 2 3 4 5 6 7 8 9 10

11. What do you like most and dislike most with this condition?

(Short Interview will be followed)

# The fourth user study (described in Chapter 7)

Human Interface Lab (HIT Lab) NZ  
Telephone: +64 3 364 2349  
Email: [seungwon.kim@pg.canterbury.ac.nz](mailto:seungwon.kim@pg.canterbury.ac.nz)



## Information Sheet

**Title: The use of drawing annotation with different shared views and in different types of remote collaboration tasks**

### Invitation

You are invited to participate in the research project "The use of drawing annotation with different shared views and different types of remote collaboration task". This information sheets provides the research study details. This investigation will take about 70 minutes to finish. The study will be carried out as a requirement for Ph.D thesis of Seungwon Kim, a PhD candidate at Human Interface Technology Laboratory New Zealand (HITLab NZ). This study is under the supervision of Associate Professor Christoph Bartner, Professor Mark Billingham, and Dr. Gun Lee. Seungwon Kim will be available to answer your questions and provide further explanations. Seungwon Kim and his supervision team can be contacted at (03) 364-2349 or email: [seungwon.kim@pg.canterbury.ac.nz](mailto:seungwon.kim@pg.canterbury.ac.nz) and [{gun.lee, christoph.bartneck}@canterbury.ac.nz](mailto:{gun.lee, christoph.bartneck}@canterbury.ac.nz). They will be pleased to discuss any concerns you may have about participation in the study. If you agree to take

part in the research study, you will be asked to sign a consent form. Your participation in this study is voluntary. You are free to choose whether or not you will take part in the study.

## **2. ‘What is the purpose of this study?’**

The purpose of this study is to explore how users collaborate differently according to the different shared views and different types of remote collaboration tasks. People start using video conferencing system for collaboration beyond face-to-face communication such as remotely helping for fixing a bike, and sharing their experience by showing what they are watching. Previous researchers mostly studied remote expert task where their focus was how effectively a remote user instructs a local user to solve an issue. In this study, we extend our focus to parallel experience where a local user not only receives instructions but also actively shares his/her ideas with a remote user with equal roles. Especially we will study the difference between remote expert and parallel experience when a remote user has a dependent or an independent view from the one for a local user.

## **2. ‘Why have I been invited to participate in this study?’**

You are eligible to participate in this study because you have average skill in using a mouse and keyboard, and you can hold and move small Tangram pieces as you want. Also, your ability of speaking English is enough to describe the orientation and position of the pieces with drawing annotation on a shared live video.

## **3. ‘What does this study involve?’**

At the beginning, the instructor will ask you and your partner to enter a room at HITLab NZ. In the room, you will be asked to sit on a chair in front of a desk. The desk is the area you will conduct this experiment with a personal computer or head mounted display(HMD) system. Your partner and you cannot see each other and your partner will be asked to sit at the opposite side of you.

Before using our remote collaboration system, a questionnaire, asking about your experience of using video conferencing software (such as Skype or Google Talk), your age, and your gender, is required to be filled out.

Next, the instructor will show you how to use the prototypes.

If you participate in this investigation as a remote user, you will discuss with your partner for assembling a model with Tangram blocks by verbal communication while you draw annotation. If you participate in this

investigation as a local user, you will wear HMD and discuss with your partner by verbal communication and hand gestures. After finishing the task, you will be asked to fill out a questionnaire and short interviews will be followed.

As a remote user, you can draw annotation with a mouse and virtually navigate live video with keyboard interaction. You will have time to practice with the system, and carry out a task that involves verbal communication with your partner and drawing annotating on the shared view. There will be two types of shared view (dependent and independent views) and two types of remote collaboration tasks.

At the end of experiment, you will be given the opportunity to check your answers in questionnaires and written transcript during interviews.

There are four different conditions, and you will be asked to conduct one task for each condition as described above. For finishing one task, it will take about 4~6 minutes, and the experiment will be finished in about 70 minutes.

You may receive a copy of the project results by contacting the researcher at the conclusion of the project.

#### **4. Are there any possible risks from participation in this study?**

Participation does not involve any risk(s) beyond the one occurred in normal life. Participants will be volunteers and can withdraw from the study at any time with no penalty. Withdrawal of participation will also include the withdrawal of any information you have provided.

#### **5. Are there any possible invasion of privacy from participation in this study?**

There will be two video recordings, one for the remote user's PC screen and the other one for the local user's view. The video recording will not record your appearance, but your verbal communication when you use our prototype for collaboration with your partner in this experiment. The experiment conductor will also write your answers during the interview on the back pages of questionnaires you filled out. At the end of experiment, you will have the opportunity to check your answers in questionnaires and the written transcript during interviews. In any case, all the data we get from you will be treated in a confidential manner. Only the experiment conductor, Mr. Seungwon Kim, and two supervisors, Dr. Gun Lee and Prof. Christoph Bartner, will have the authority to access to the video recordings.

The results of the project may be published, but you may be assured of the complete confidentiality of data gathered in this investigation: your identity will not be disclosed at any time. To ensure anonymity and confidentiality, you are not required to give any personal information during this study aside your age and your gender. The only identifying data, which is the consent form, will never be disclosed after stored in a locked cabinet, and it will be treated separately from other information collected. Non-identifying data, such as questionnaire and transcript of interview, will be documented as electronic file during data analyze and be kept in a locked cabinet. The electronic file will be stored in a password-protected computer.

The raw measured data will be accessed only by Mr. Seungwon Kim, Dr. Gun Lee, Associate Prof. Christoph Bartner and Prof. Mark Billinghurst. Your data will be kept for a maximum of ten years before being destroyed.

#### **6. Are there any possible benefits from participation in this study?**

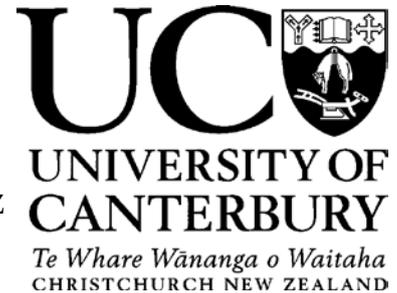
Benefits to participating in a study include:

- The satisfaction of helping others by contributing to the research on remote collaboration systems, or helping to identify a possible new interface for remote collaboration systems.
- The opportunity to have an experience of using a visual communication cue with video conferencing system.
- The final results will be reported via your e-mail if you agree to receive an e-mail (reporting only combined results and never reporting individual ones).

This project has been reviewed and approved by the University of Canterbury Human Ethics Committee Low Risk process, and participants should address any complaints to the Chair of Human Ethics Committee, University of Canterbury, Private Bag 4800, Christchurch (human-ethics@canterbury.ac.nz).

**If you agree to participate, you are asked to complete the consent form**

Human Interface Lab (HIT Lab) NZ  
Telephone: +64 3 364 2349  
Email: seungwon.kim@pg.canterbury.ac.nz



## Consent Form

**Title: The use of drawing annotation with different shared views and different types of remote collaboration tasks**

1. I have been given full explanation of this project and had the opportunity to ask questions.
2. I understand what is required of me (mouse and keyboard interaction and wearing head mounted system for remote collaboration) if I agree to take part in the research.
3. I understand that participation is voluntary and I may withdraw at any time without penalty. Withdrawal of participation will also include the withdrawal of any information I have provided.
4. I understand that any information or opinions I provide will be kept confidential to the researcher and the administrators of the research project and that any published or reported results will not identify the

participants. I understand that a thesis is a public document and will be available through the UC Library.

5. I understand that all the recordings and data collected for the study will be kept in locked and secure facilities and/or in password protected electronic form and will be destroyed after ten years. (Securely stored in one of the cabinet of the HITLab NZ for 10 years, and will then be destroyed by administration staff via secure shredding)
6. I understand participation does not involve any risk(s) beyond the one occurred in normal life. .
7. I understand that I am able to receive a report on the findings of the study by contacting the researcher at the conclusion of the project.
8. Any questions that I have asked have been answered to my satisfaction.
9. I agree to participate in this investigation which takes approximately 70 minutes and understand that any data I have supplied will be withdrawn at any time in accordance with my wish.

I understand that I can contact the researcher (Seungwon Kim, seungwon.kim@pg.canterbury.ac.nz)

or supervisor (Christoph Bartneck, christoph.bartneck@canterbury.ac.nz) for further information. If

I have any complaints, I can contact the Chair of the University of Canterbury Human Ethics

Committee, Private Bag 4800, Christchurch  
(humanethics@canterbury.ac.nz)

Please check left box if you agree to receive the final results of this study via e-mail and

write your e-mail address below (Only combined results will be e-mailed and any individual result will not be reported)

By signing below, I agree to participate in this research project.

Name of Participant:

---

Signature:

Date:

---

### **Statement by Investigator**

I have explained the project & the implications of participation in it to this volunteer and I believe that the consent is informed and that he/she understands the implications of participation

The participant has received the Information Sheet where my details have been provided so participants have the opportunity to contact me prior to consenting to participate in this project.

Name of Investigator:

---

Signature:

Date:

---

## Questionnaire (Remote) Condition \_\_\_\_\_

**Group:** \_\_\_\_\_

Below are statements that describe your experience. Please circle on a number to indicate your level of agreement or disagreement with each statement

(0: Strongly Disagree ~ 5: Neither agree nor disagree ~ 10: Strongly Agree)

1. I enjoyed assembling a model.

0      1      2      3      4      5      6      7      8      9      10

2. I was able to focus on assembling a model.

0      1      2      3      4      5      6      7      8      9      10

3. Using the interface distracted me from collaborating with my partner.

0      1      2      3      4      5      6      7      8      9      10

4. I was able to see the work environment properly.

0      1      2      3      4      5      6      7      8      9      10

5. I was able to express my idea properly.

0      1      2      3      4      5      6      7      8      9      10

6. I easily understood what my partner tried to do.

0      1      2      3      4      5      6      7      8      9      10

7. I felt I was interrupting my partner during collaboration.

0 1 2 3 4 5 6 7 8 9 10

8. I felt my partner was focusing on my explanation.

0 1 2 3 4 5 6 7 8 9 10

9. I felt connected with my partner.

0 1 2 3 4 5 6 7 8 9 10

10. I felt we were together.

0 1 2 3 4 5 6 7 8 9 10

11. The communication with my partner was *mentally stressful*.

0 1 2 3 4 5 6 7 8 9 10

12. We focused on the same piece or area while collaborating.

0 1 2 3 4 5 6 7 8 9 10

13. My partner was aware of where I was drawing.

0 1 2 3 4 5 6 7 8 9 10

14. I felt we collaborated well.

0 1 2 3 4 5 6 7 8 9 10

15. What do you like most and dislike most with this condition?

# Questionnaire (Local) Condition \_\_\_\_\_

**Group:** \_\_\_\_\_

Below are statements that describe your experience. Please choose a number to indicate your level of agreement or disagreement with each statement  
(0: Strongly Disagree ~ 5: Neither agree nor disagree ~ 10: Strongly Agree)

1. I enjoyed assembling a model.

0      1      2      3      4      5      6      7      8      9      10

2. I was able to focus on assembling a model.

0      1      2      3      4      5      6      7      8      9      10

3. I was able to express my idea properly.

0      1      2      3      4      5      6      7      8      9      10

4. I easily understood what my partner tried to do.

0      1      2      3      4      5      6      7      8      9      10

5. My partner was interrupting me during collaboration.

0      1      2      3      4      5      6      7      8      9      10

6. I was able to focus on my partner's explanation.

0      1      2      3      4      5      6      7      8      9      10

7. I felt connected with my partner.

0 1 2 3 4 5 6 7 8 9 10

8. I felt we were together.

0 1 2 3 4 5 6 7 8 9 10

9. The communication with my partner was *mentally stressful*.

0 1 2 3 4 5 6 7 8 9 10

10. We focused on the same piece or area while collaborated.

0 1 2 3 4 5 6 7 8 9 10

11. My partner was aware of how I manipulated pieces.

0 1 2 3 4 5 6 7 8 9 10

12. I felt we collaborated well.

0 1 2 3 4 5 6 7 8 9 10

13. (Short Interview) What do you like most and dislike most with this condition