Exploring How to Display Referential Action to Support Remote Group Discussion

Tzu-Yang Wang
Yuki Noaki
wang.tzuyang@gmail.com
yknoaki@gmail.com
University of Tsukuba
Tsukuba, Ibaraki, Japan

Hideaki Kuzuoka
The University of Tokyo
Tokyo, Japan
kuzuoka@cyber.t.u-tokyo.ac.jp

ABSTRACT

Nowadays, remote conferences are widely seen in many situations. Although there is little problem to conduct one-on-one remote conferences, remote group discussion still contains many challenges. One of the main issue is that the remote participants often unnotice the referential action performed by the local participants. To solve such problem, we proposed to develop a function that automatically detects the presence of referential action and displays to the remote participant. One of the issues for this function is about what kind of information of the referential action should be provided and how the information should be displayed. In this research, we conducted a lab study to compare two displaying methods: Picture-in-Picture (PiP) and auto-pilot and compare two displaying contents: object being referred and person performing the referential action. The result shows that the PiP method had higher usability and remote participant had higher opportunity to join the conversation with the PiP method. On the other hand, displaying object being referred had higher usability than displaying the person performing the referential action.

CCS CONCEPTS

• **Human-centered computing** → **Laboratory experiments**; **Collaborative interaction**; User interface design.

KEYWORDS

PiP, referential action, remote communication, group discussion, WoZ

ACM Reference Format:
Tzu-Yang Wang, Yuki Noaki, and Hideaki Kuzuoka. 2021. Exploring How to Display Referential Action to Support Remote Group Discussion. In *Asian CHI Symposium 2021 (Asian CHI Symposium 2021), May 8–13,2021, Yokohama, Japan.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3429360.3468188

1 INTRODUCTION

Remote conference is becoming popular due to the COVID-19 pandemic. People who live far away from each other use teleconference tools to communicate and discuss things. Although most teleconference tools effectively work in the case of one-on-one conversation, some research has pointed out that the remote communication had a low communication quality in remote group discussion.

Among all types of remote group discussion, this research focused on the remote group discussion with physical referential actions. A referential action is a action that a speaker refers to objects or people during conversation. The speaker uses both verbal cues (e.g. naming or describing) and nonverbal cues (e.g. pointing, tapping, etc) to make sure that listeners pay attention on the objects or people [3, 14]. In a remote group discussion with physical referential actions, people do not always face at the camera and communicate, but sometimes move around, interacting with physical objects, and refer the physical objects in the discussion. For example, in a business meeting, many business plans and documents are placed on the whiteboards and tables. People walk toward different whiteboards and point at different documents and plans during the meeting.

However, during the remote group discussion, referential communication usually failed. The remote participants often unnotice the referential actions [7, 13] due to narrow field of view (FOV) [5] and incomplete nonverbal cues. Modern webcams have less then 120 degree of FOV which is much narrower than the FOV of human's eyes. On the other hand, many nonverbal cues, such as gazes, cannot be successfully transferred to remote participants. In a face-to-face group discussion, a participant A can easily notice that another participant B who is outside of his/her FOV is performing a referential action and can promptly turn to the participant B. This procedure requires many nonverbal information. For example, participant A observes other participants gazes to find out the location of participant B or judges the location of participant B with the spatial orientation of the voice. However, in a remote group discussion, the gaze direction cannot be correctly transferred on a 2D-display and the spatial orientation of the voice cannot be transferred through a 2.0 stereo speaker or headset. Thus, the remote participant has difficulty noticing the referential actions and locating the referential actions. Furthermore, missing the referential actions makes the remote participants having fewer chances to join the discussion and diversity of the opinions reduces. Finally, the quality of the remote group discussion becomes poorer.

Asian CHI Symposium 2021 , May 8–13, 2021, Yokohama, Japan

ACM ISBN 978-1-4503-8203-8/21/05…$15.00
https://doi.org/10.1145/3429360.3468188

Therefore, we considered that a function that automatically detects the presence of referential actions and shows to the remote participants can effectively improve the quality of the remote group discussion. For such function, in addition to the part of automatic detection, it is important to know the appropriate method to display the information and the appropriate information of the referential action that can support a better remote group discussion.

In this research, we conducted a lab study to address the issue. WoZ was used to replace the automatic detection of referential actions. We compared two possible methods to display the presence of referential actions: Picture-in-Picture (PiP) method and autopilot method, and compared whether showing the person who is performing referential action or showing the object being referred is more useful.

2 RELATED WORK

2.1 Method of Displaying Information Outside of View

Displaying the information outside of view is an important issue. One obvious method is to expand the FoV of users [1]. However, providing a wide FoV video on traditional displays often make the image too small and the users are hard to see detailed information. This might cause a detrimental effect on remote group discussion. Some other research tried to provide explicit hints. Lin et al. developed two assistance functions (auto-pilot and visual guidance) to navigate users' focus in 360-degree videos [8]. Auto-pilot method automatically directly changes users' views to the target which was outside of view; visual guidance method does not change the users' view but provided arrows in the screen to guide users' attention. It was found that auto-pilot method improved the feeling of presence better than visual guidance while observing a 360-degree sport videos.

Instead of providing the arrows to guide users' attention, Lin developed a Picture-in-Picture (PiP) method that displays the information outside of view in a smaller window that attached to the user's view [9]. The evaluation experiment showed that participants perceived better spatial information than arrow-based guidance while observing 360-degree videos.

However, there is no research comparing whether the auto-pilot method or PiP method is more suitable for displaying the information outside of view. Besides, since the past literature focused on 360-degree video, it is unclear how the two methods works in a remote group discussion. Thus, in this research, we adapted the two methods to display the presence of referential actions.

2.2 Component of Referential Action

Another important issue is what kind of information is needed to be provided to a remote participant. Referential actions include two main components: the person who performs referential action and the object being referred [6]. The distance between the person and the object can be quiet far. Thus, it is difficult to display both the person and the object to the remote participant. Displaying the person transfers the local participant's nonverbal behavior to the remote participant [12]; displaying the object supports remote participant to understand the spoken content more. Thus, it is necessary to investigate which information is more significant and should be displayed.

3 PROPOSED TELECONFERENCE SYSTEM

To explore an appropriate way for displaying the presence of referential actions, we developed our own teleconference system. For the hardware, we used a teleconference robot Kubi (Fig. 1 left). Kubi is a teleconference robot developed by Revolve Robotics. Tablets can be placed on the arm, and the remote users can freely control the rotation of the Kubi with software to change the viewpoint. Kubi can twist 300 degrees and tilt 90 degree, so the remote user can observe a wide area of the local environment.

As for the software part, there were two software: one is for the remote participant (Fig. 1 middle) and the other is for the observer (Fig. 1 right). Regarding the software for the remote participant, a real time video stream captured by the tablet's camera on the Kubi is shown on the screen. Besides, the software contains buttons for rotation so that remote participants can freely rotate the Kubi to change their views. Since this research used the WoZ method to replace the automatic detection, we developed a software for the observer. In this software, a real time video stream which is captured by an extra webcam is shown on the screen. With the video stream, the observer can completely see the whole local environment. When a local participant performs a referential action, the observer clicks the referred object or the person on the video stream. Then, it triggers the function of displaying the referential action.

We developed two different methods to display the referential action to the remote participant: auto-pilot method and PiP method (Fig. 2). Regarding the auto-pilot method, once the observer clicks on the software, the Kubi automatically rotates to the corresponding direction. Regarding the PiP method, once the observer clicks on the software, the referential action is captured by another webcams and shown in a small window attached to the main window of the software for the remote participant (see the red window in Fig. 1). The remote participant could see both video captured by the tablet on the Kubi and video showing the information of the referential action at the same time.

4 EXPERIMENT

4.1 Hypothesis

In this experiment, we compared two different displaying methods and two different displaying contents and set up two hypotheses (Fig. 2):

- Comparing with the auto-pilot method (Auto-pilot), displaying the referential action information to the remote participants with PiP method (PiP) results in a better discussion quality in the remote group discussion.
- Displaying the object being referred (Object) improves discussion quality better than displaying the person performing the referential action (Person).

4.2 Task and Apparatus

A travel planning task was used as our experiment task to simulate a group discussion situation. There were five travel places (Fig. 3

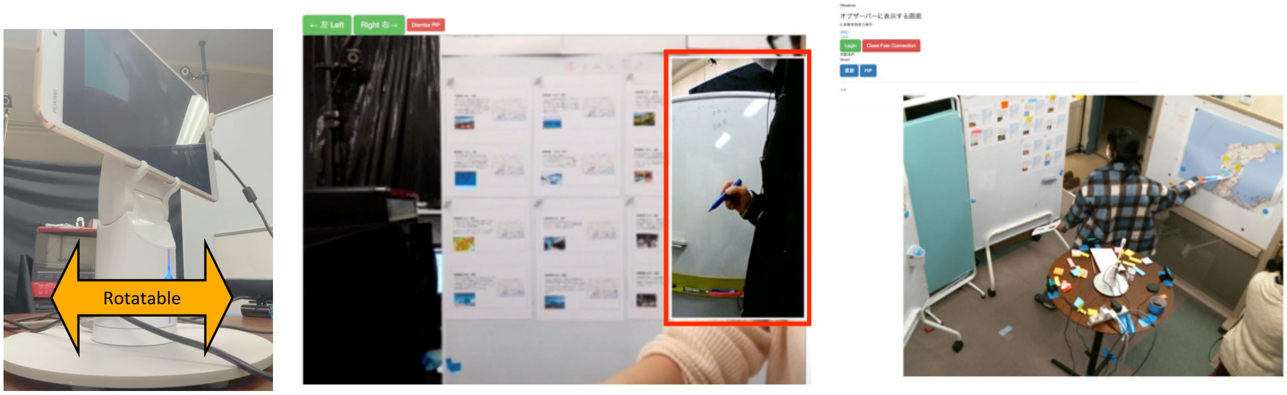


Figure 1: Left: Kubi can twist 300 degrees to allow users to change views; middle: a screenshot of the software for the remote participant; right: a screenshot of the software for the observer.

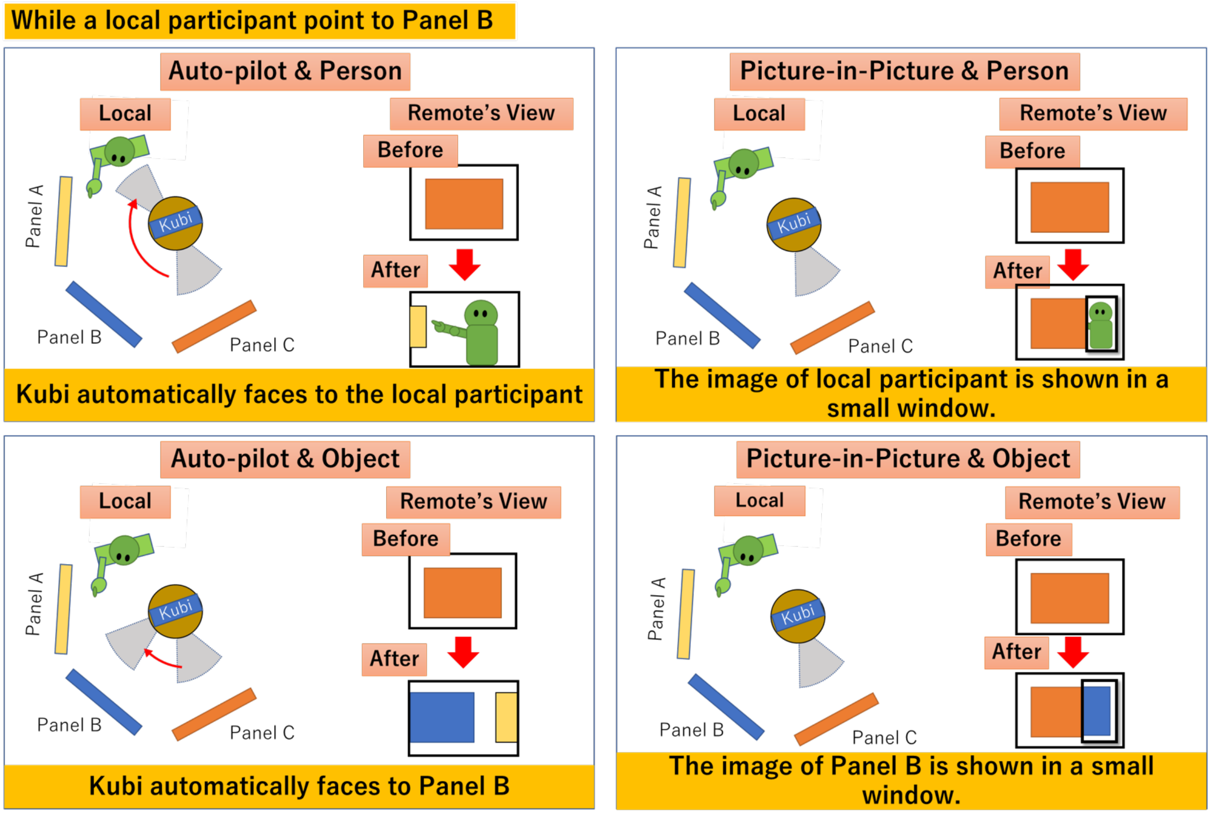
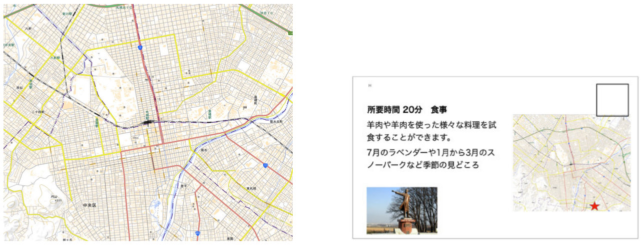

While a local participant point to Panel B
Auto-pilot & Person
Local
Remote's View
Before
After
Panel A
Panel B
Panel C
Kubi
Kubi automatically faces to the local participant
Picture-in-Picture & Person
Local
Remote's View
Before
After
Panel A
Panel B
Panel C
Kubi
The image of local participant is shown in a small window.
Auto-pilot & Object
Local
Remote's View
Before
After
Panel A
Panel B
Panel C
Kubi
Kubi automatically faces to Panel B
Picture-in-Picture & Object
Local
Remote's View
Before
After
Panel A
Panel B
Panel C
Kubi
The image of Panel B is shown in a small window.

Figure 2: A sample of four conditions compared in the experiment

left) for a practice session and four main sessions with four different conditions (described in section 4.4). For each travel place, there were 10–15 candidates of tourist attractions (Fig. 3 right). For each tourist attraction, the needed time, type, pictures, and short description were written on a paper board. Note that the name of tourist attraction was not written so that the participants might rely more referential gestures to refer objects during the discussion. For each task, the participants were asked to discuss with each other, selecting tourist attractions, considering the route, and plan a 10-hour trip.

Besides, to facilitate the conversation, each participant received three individual missions (e.g select at least two nature-related tourist attraction). While planning the trip, the participants are also required to complete the received missions. Note that the participants were asked not to explicitly tell their own missions to other participants.


所要時間 20分　食事

羊肉や羊肉を使った様々な料理を試食することができます。

7月のラベンダーや1月から3月のスノーパークなど季節の見どころ

Figure 3: Left: map of a travel place; right: tourist attraction paper board including needed time, type, pictures, and short description.

4.3 Experiment Environment

The trip planning task was conducted in a laboratory. The laboratory was separated to two area: local area and remote area. In the local area, a Kubi and a tablet were placed on a circle table among four whiteboard panels (Fig. 4). The distance between the panels and the Kubi was 1.5m. The paper boards of tourist attractions were placed on Panel B. The half of the map of the travel place was placed on Panel A; the another half of the map was placed on Panel C. A magnet ruler was placed on the panel that the participants could use it to measure the travel time between attractions. Panel D is the panel for participants to write memo on.

In the remote area, a PC with the proposed software was placed on a table, and the remote participants can operate the software with a mouse. Additionally, since the resolution of the streaming video captured by the tablet camera was low, it might be hard for the remote participant to clearly see the words on the paper boards of tourist attractions. Thus, in the remote area, a copy of tourist attraction description was prepared.

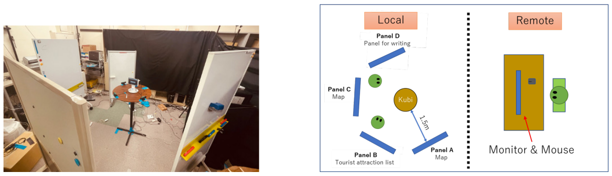


Figure 4: left: picture of the local area; right: bird view of the local area and remote area.

4.4 Experimental Conditions

The experiment was conducted in a within-participant design with two independent variables: displaying method and displaying content. As we described in section 3, for the displaying method, there were PiP (showing the information of referential action in a small window attaching to the main window of the remote software) and Auto-pilot (automatically changing the direction of the Kubi to display the information of referential action directly). For the displaying content, there were Object (object being referred) and Person (person performing the referential action). In total, there were four conditions (Fig. 2).

4.5 Participants

24 participants (13 males and 11 females) in the University of Tsukuba were recruited to participate the experiment and eight groups were formed. The average age was 21.2 yr and standard deviation was 1.83. In addition, one 23-year-old male participant was recruited as a observer for all eight experiments.

4.6 Procedure

Every experiment contained three participants. Two of them were local participant and stayed in the local area; the other participant was remote participant and stayed in the remote area. After signing the consent form, we explained the experiment procedure to the participants. For each session, a list of individual missions was given to each participant. After the preparation, the participants started the travel planning task; the remote participant discussed with the local participants through the Kubi and the proposed software. Meanwhile, the observer started to observe the group discussion, monitoring the referential action and operate the software. In the condition with Object, the observer clicked a object on the software when the object was referred; in the condition with Person, the observer clicked a local participant when the participant performed a referential action. After 15 minutes, the task ended no matter if the task was successfully conducted. The participants were asked to answer the questionnaire (described in section 4.7). After answering the questionnaire, the participants had a 10-minute break before the next session began. The experiment ended after all four sessions were finished. Note that the order of the four sessions with four conditions was counterbalanced to reduce the order effect.

4.7 Measurement

To measure the quality of the discussion, we focused on three parts: system usability, communication quality, and outcome quality (Table 1). For the system usability, we employed the system usability scale (SUS) as the questionnaire in this study. The SUS was a unidimensional scale with 10 items designed by Brooke [2] (Q1 – Q10), and has demonstrated high reliability and validity [11]. For the communication quality, we used the scale of quality of communication experience (QCE) developed by Liu et al [10]. The scale consisted of 15 items and consisted of 3 factors: "Clarity" (Q11–Q15), "Responsiveness" (Q16–Q20), and "Comfort" (Q21–Q25). For the outcome quality, we adopted the scale developed by DeStephen et al. [4]. The scale measured the small group consensus and contained 5 factors. 2 factors which were not associated with the experiment were removed and the 3 factors were "Feelings of agreement, satisfaction, and commitment (Satisfaction)" (Q26–Q30), "Feelings about the effectiveness of the individual participation (Effectiveness)" (Q31–Q33), and "Feelings about individual opportunity to participate (Opportunity)" (Q34–Q36). The whole questionnaire contained 36 items and it was a 7-point Likert scale.

5 RESULT

The result of questionnaire of remote participants and local participants was analyzed separately. For each factor, the score of reversed

Table 1: Subjective Questionnaire

Usability
1. I think that I would like to use this system frequently.
2. I found the system unnecessarily complex. (R)
3. I thought the system was easy to use.
4. I think that I would need the support of a technical person to be able to use this system. (R)
5. I found the various functions in this system were well integrated.
6. I thought there was too much inconsistency in this system. (R)
7. I would imagine that most people would learn to use this system very quickly.
8. I found the system very cumbersome to use. (R)
9. I felt very confident using the system.
10. I needed to learn a lot of things before I could get going with this system. (R)

Clarity
11. I understood what the other side was saying.
12. I understood what was important to the other side.
13. We clarified the meaning if there was a confusion of the messages exchanged.
14. I think the other side understood me clearly.
15. The messages exchanged were easy to understand.

Responsiveness
16. The other side responded to my questions and requests quickly during the interaction.
17. The conversation ran smoothly without any uncomfortable silent moments or I did not notice any uncomfortable silent moments.
18. I was willing to listen to the other side's perspectives.
19. When the other side raised questions or concerns, I tried to address them immediately.
20. One or both of us kept silent from time to time. (R)

Comfort
21. I was nervous talking to the other side. (R)
22. I felt the other side trusted me.
23. I felt the other side was trustworthy.
24. I felt comfortable interacting with the other side.
25. The other side seemed comfortable talking with me.

Feelings of agreement, satisfaction, and commitment toward the group's decision
26. The group reached the right decision.
27. I believe that our group's decision is appropriate.
28. I support the final group decision.
29. I believe we selected the best alternative available.
30. I would be willing to put my best effort into carrying out the group's final decision.

Feelings about the effectiveness of the individual participation
31. I believe I contributed important ideas during the decision-making process.
32. I believe I had a lot of influence on the group's decision-making.
33. I contributed important information during the group's decision-making process.

Feeling regarding individual opportunity to participate
34. During group meetings, I got to participate whenever I wanted to.
35. I believe that the other members of the group liked me.
36. Other members of the group really listened to what I had say.

Note: item with (R) is the reversed item.

item was reversed and all items were summed up. Later, two-way repeated measure ANOVAs were used to examine the effect of displaying method and displaying content on the score of each factor (Fig. 5).

5.1 Local Participant

As for the SUS score, the score of the PiP was marginally significantly higher than the score of the Auto-pilot ($F(1, 15) = 3.264$, $p = .091$). There was a significant interaction between the two independent variables ($F(1, 15) = 6.155$, $p = .025$). Simple effect tests were followed up and alpha level was adjusted to 0.25 based on the Bonferroni correction. The result showed that the condition of PiP and Person had a higher SUS score than the condition of Auto-pilot and Person ($F(1, 15) = 7.408$, $p = .016$).

For other factors, there were no significant difference between the different displaying methods and different displaying contents.

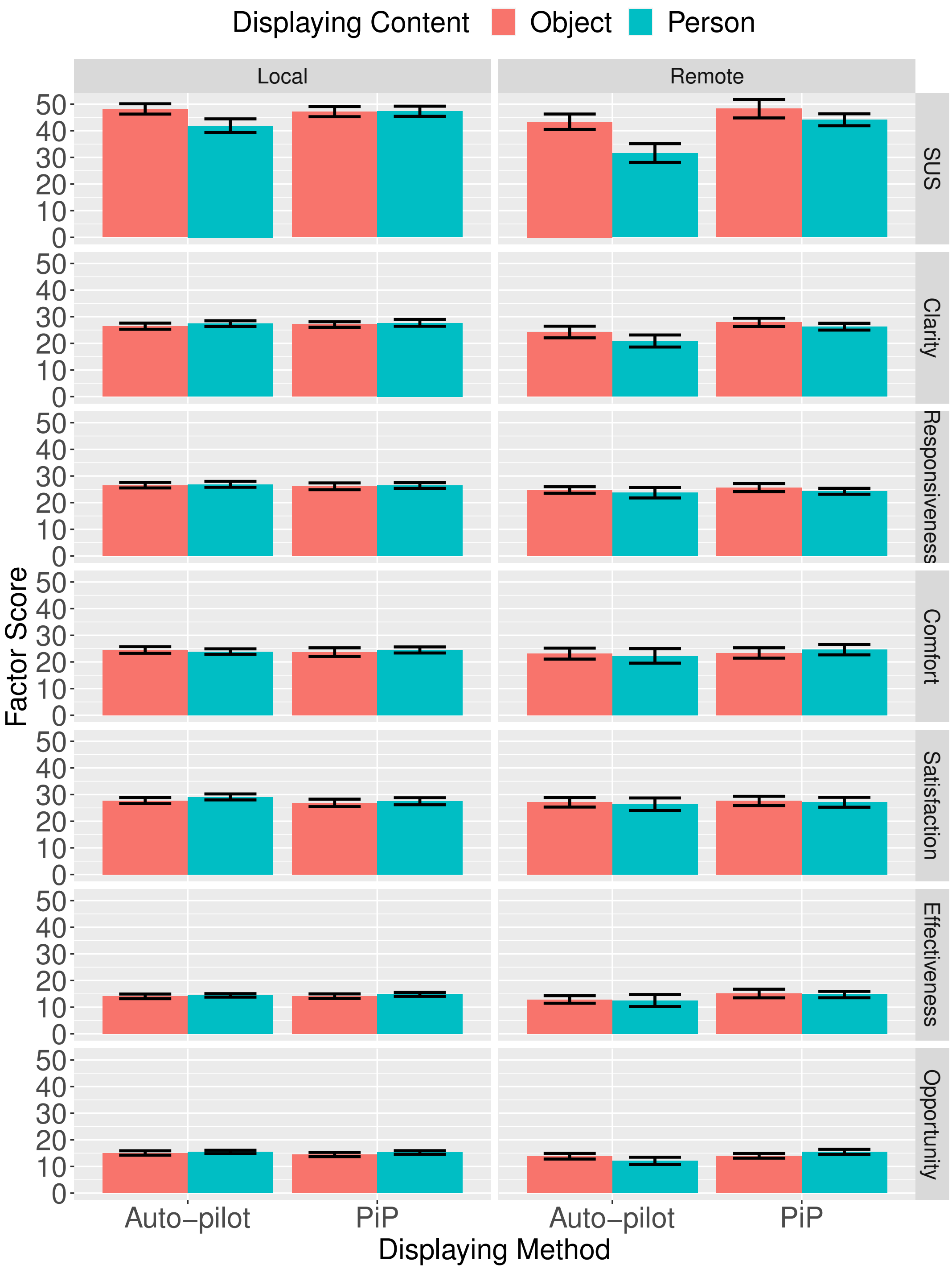


Figure 5: Result of the questionnaire

5.2 Remote Participant

As for the SUS score, the score of the PiP was significantly higher than the score of the Auto-pilot method ($F(1, 7) = 8.504$, $p = .023$). The SUS score of the Object was significantly higher than the score of Person ($F(1, 7) = 9.171$, $p = .019$). In addition, there was a significant interaction between the two independent variables ($F(1, 7) = 9.752$,

p = .017). Simple effect tests with Bonferroni correction showed that the condition of Auto-pilot and Object had a larger SUS score than the condition of Auto-pilot and Person (F(1, 7) = 10.8, p = .013). Besides, the condition of PiP and Person had a larger SUS score than the condition of Auto-pilot and Person (F(1, 7) = 11.58, p = .011).

Regarding the factor "Clarity", the score of PiP was marginally significantly higher than the score of Auto-pilot (F(1, 7) = 5.532, p = .051). The score of Object was marginally significantly higher than the score of Person (F(1, 7) = 4.268, p = .078).

Regarding the factor "Opportunity", the score of PiP was marginally significantly higher than the score of Auto-pilot (F(1, 7) = 3.611, p = .1). Besides, there was a significant interaction between the two independent variables (F(1, 7) = 6.464, p = .039). Simple effect tests with Bonferroni showed that the condition of PiP and Person had a higher score than the condition of Auto-pilot and Person (F(1, 7) = 10.65, p = .014).

For other factors, there were no significant difference between the different displaying methods and different displaying contents.

6 DISCUSSION

6.1 Local Participant

While there were almost no significance between different conditions for the local participants, the only major difference was the score of usability. The result showed that the condition of PiP and Person had higher usability score than the condition of Auto-pilot and Person. Based on the video analysis, we found that in the case of automatically facing to the person, the remote participants often controlled the Kubi by themselves to face to the objects being referred after the Kubi was faced to the participant performing referential actions. Thus, the local participants and the remote participants paused the conversation and restarted the conversation after the remote participants finished controlling the Kubi. The extra waiting time and the interruption are possible reasons causing the low usability.

6.2 Remote Participant

As for the remote participant, PiP had a higher usability score than Auto-pilot. Based on the observation and video analysis, we found that a possible reason is that the unintentional movement of the Kubi caused the remote participants hard to see the information on the white board. The remote participant should spend extra effort and time to figure out the information and get back to the discussion. This also caused the remote participant hard to understand the conversation topic, so that the score of Clarity was marginally lower in the condition with auto-pilot methods. Furthermore, the remote participants felt low opportunity to participate in the discussion.

Besides, the result of usability also showed that displaying objects had higher usability. The reason might be same as the reason described above that both local participants and remote participant had to wait until the remote participant adjust the direction of the Kubi to face to the objects.

6.3 Comparing Remote Participant and Local Participant

Another interesting finding is that the difference between conditions were significant for remote participants but not local participants. A possible reason is because the issue remote participants encountered might not be a major issue for the local participants. Our video analysis showed that in most groups, the local participants dominate the discussion. The two local participants still conducted fluent discussion between each other, and paid rather less attention to the remote participants. An evidence was that the percentage of time that the remote participants spoke in the remote discussion was 17.90% which was lower than the ideal percentage 33.3%. Thus, the two local participant might not feel that there was problem with the discussion.

On the other hand, the remote participants spent larger effort to actively joined in the discussion. Proposed functions directly influenced the easiness for the remote participants to join in the discussion. Thus, the difference between functions was more significant.

7 CONCLUSION

In this paper, we conducted an experiment to investigate an appropriate way to display information of referential action to remote participant to reach a higher quality of remote group discussion. The result showed that directly rotating the Kubi to face to the referential action created extra effort to the remote participant. Instead, displaying the information of the referential action with PiP method improved the quality of the remote group discussion. In addition, displaying the person performing the referential action was not useful for the remote participants since the remote participants often rotated the Kubi by themselves to faced to the object being referred.

ACKNOWLEDGMENTS

This research was funded by Tateisi Science and Technology Foundation Research Program C.

REFERENCES

[1] Jérôme Ardouin, Anatole Lécuyer, Maud Marchal, Clément Riant, and Eric Marchand. 2012. FlyVIZ: a novel display device to provide humans with 360 vision by coupling catadioptric camera with hmd. In *Proceedings of the 18th ACM symposium on Virtual reality software and technology.* 41–44.

[2] John Brooke. 1996. SUS: A quick and dirty usability scale.

[3] Herbert H Clark. 2003. Pointing and placing. *Pointing: Where language, culture, and cognition meet* 243 (2003), 268.

[4] Rolaynes DeStephen and Randy Y Hirokawa. 1988. Small group consensus: Stability of group support of the decision, task process, and group relationships. *Small Group Behavior* 19, 2 (1988), 227–239.

[5] William W Gaver. 1992. The affordances of media spaces for collaboration. In *Proceedings of the 1992 ACM conference on Computer-supported cooperative work.* 17–24.

[6] Charles Goodwin. 2003. Pointing as situated practice. *Pointing: Where language, culture and cognition meet* 217 (2003), 241.

[7] Christian Heath and Paul Luff. 1991. Disembodied conduct: Communication through video in a multi-media office environment. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* 99–103.

[8] Yen-Chen Lin, Yung-Ju Chang, Hou-Ning Hu, Hsien-Tzu Cheng, Chi-Wen Huang, and Min Sun. 2017. Tell me where to look: Investigating ways for assisting focus in 360 video. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems.* 2535–2545.

[9] Yung-Ta Lin, Yi-Chi Liao, Shan-Yuan Teng, Yi-Ju Chung, Liwei Chan, and Bing-Yu Chen. 2017. Outside-in: Visualizing out-of-sight regions-of-interest in a 360 video

using spatial picture-in-picture previews. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*. 255–265.

[10] Leigh Anne Liu, Chei Hwee Chua, and Günter K Stahl. 2010. Quality of communication experience: Definition, measurement, and implications for intercultural negotiations. *Journal of Applied Psychology* 95, 3 (2010), 469.

[11] Jeff Sauro and James R. Lewis. 2012. *Quantifying the User Experience: Practical Statistics for User Research* (1st ed.). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

[12] Tzu-Yang Wang, Yuji Sato, Mai Otsuki, Hideaki Kuzuoka, and Yusuke Suzuki. 2020. Effect of Body Representation Level of an Avatar on Quality of AR-Based Remote Instruction. *Multimodal Technologies and Interaction* 4, 1 (2020), 3.

[13] Naomi Yamashita, Katsuhiko Kaji, Hideaki Kuzuoka, and Keiji Hirata. 2011. Improving visibility of remote gestures in distributed tabletop collaboration. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work*. 95–104.

[14] George Yule. 1997. *Referential communication tasks*. Routledge.