User's Communication Behavior in a Pseudo Same-room Videoconferencing System BHS

Tomoo Inoue[*], Mamoun Nawahdah[**], and Yasuhito Noguchi[***]

[*]Faculty of Library, Information and Media Science, University of Tsukuba, Tsukuba, Japan
[**]Faculty of Engineering and Technology, Birzeit University, Birzeit, Palestine
[***]Graduate School of Library, Information and Media Studies, University of Tsukuba, Tsukuba, Japan
[*]inoue@slis.tsukuba.ac.jp, [**]mnawahdah@birzeit.edu, ***noguchi@slis.tsukuba.ac.jp

Abstract - This paper presents user's communication behavior in a pseudo same-room videoconferencing named "Being Here System," in comparison with a conventional videoconferencing. The system extracts the remote person's figure and superimposes it on the local site's front view in a large display in real-time. This method makes the local person feel as if the remote person was before him/her in his/her spatial environment. To investigate the influence of the system on user's communication, the recorded video of the system evaluation experiment was analyzed. This revealed that the system significantly affected user's communication behavior such as turn taking, speech overlap, and gaze directions [1].

Keywords: Videoconferencing, presence, video overlay, telecommunication, Kinect application

1 INTRODUCTION

Communication can be carried out in face-to-face (F2F) or through media. In F2F communication, the exchange of information, thoughts, and feelings is made when the participants exist in the same physical space at the same time. In this communication, nonverbal cues (e.g., eye contact, facial expression, body movement, interpersonal distance, etc.) may influence the way the message is interpreted by the receiver. In contrast, although mediated communication including videoconferencing provides people with many advantages given the increased globalization and the need for rapid knowledge transfer across borders and time zones, the absence of nonverbal cues may make communication difficult. Hence communication process is affected in mediated communication. A person may feel less presence of remote participants in mediated settings, and he/she may fail to interpret other people's behavior correctly and/or accurately. Therefore, one of the design goals of a videoconferencing system is to create a medium setup that is as close as possible to F2F.

Many studies have suggested that generating a life-sized view is likely to enhance the user's sense of presence [8, 12, 13, 24, 11, 16]. Here "presence" or "sense of presence" refers to the user's feeling of connection to the remote person with whom they are interacting [18]. The life-sized view makes it easy to read the other person's behavior that is essential for smooth communication.

Figure 1: A user talks to a remote user through BHS.

Large displays can be used to achieve a life-sized view. However, this means that a considerable region of the local person's front view is replaced by the remote site's background, which makes no integration or continuity in the local person's front view. This may decrease the user's sense of co-presence. Meanwhile, the remote site's background in some environments might be 'cluttered' with static or movable objects. This may either be a distraction or be more engaging, giving a greater sense of the other person's environment [5].

"Being Here System (BHS)" is a system to achieve pseudo same-room videoconferencing system using a large display [23]. The system provides the communication environment where the remote user's life-sized figure is visually situated in the local site (Fig. 1) and vice versa. The display shows the local site's front view, which would otherwise have been obstructed by the display, as a background. In this way, the user feels as if the remote user is present before him/her in the same room. In other words, the user feels co-presence of the other remote user. BHS was initially evaluated by a questionnaire filled by users after performing a videoconferencing experiment. The questionnaire results revealed that BHS achieved higher sense of co-presence of remote users than the conventional videoconferencing system.

In this paper, we further investigated the user's behaviors when communicating using BHS. The motivated question is whether BHS affects verbal and/or nonverbal communication structure. The considered verbal communication parameters in this study are turn taking, speech time, and speech overlap. Regarding nonverbal

[1] This research was partially supported by the JSPS Grant-in Aid for Scientific Research 26330218.

parameters, gaze direction is considered. The communication behavior analysis revealed that BHS has significantly affected users' conversational behaviors.

2 RELATED WORKS

2.1 Media Space Systems

There have been various studies done on remote communication and media spaces, and a host of systems have been developed over time. Many of these studies have been devoted to proposing and/or implementing methods aimed at enhancing the sense of presence in videoconferencing.

One early system called "Hydra" sought to enhance the sense of presence by supporting directional gaze cues and selective listening in 4-way videoconferencing [27]. A multi-party videoconferencing system called "MAJIC" was constructed by Okada et al. to support eye contact [24]. In this system, life-sized video images of participants were projected onto a large curved transparent display. Another line of research focused on the seating arrangement in video-mediated meetings, in order to enhance the sense of presence [12]. The system was designed for multiple participants so that the video image of any remote participant be always placed where a viewer need to make no effort to see it. A different approach to enhance the sense of presence was introduced by Morikawa et al. [20]. In this study, a system called "HyperMirror" was constructed, in which all participants were meant to feel as if they were sharing the same virtual space. To provide a greater sense of presence than had been achieved with conventional desktop videoconferencing, Gibbs et al. created the "TELEPORT" system [8], which was based on special rooms, called display rooms, in which one wall was a "view port" into a virtual extension. A side-by-side media space concept was proposed to enhance the presence feelings, which was suggested to be more appropriate for side-by-side style interactions such as collaborative writing and training [28]. Other effective attempts to enhance the presence feelings involved movable displays [22] and movable cameras [21].

It is natural to devote more attention to people present before one, since the felt presence of remote people is considerably weaker [32]. To overcome this inclination, robotic means have been employed to convey the sense of presence in videoconferencing, enhancing the remote people's felt presence. In this regard, a study by Sakamoto et al. investigated the effect of using a humanoid robot system as a telecommunication medium [26]. Another study, by Yankelovich et al., introduced a system called "Porta-Person" to enhance the sense of social presence for remote-meeting participants [32]. This goal was achieved by providing a high-fidelity audio connection and a remotely controlled telepresence display with video or animation. In the same manner, Venolia et al. developed a telepresence device, called "Embodied Social Proxy (ESP)", which represented a remote coworker at roughly human-scale [29]. In this system, they found that the physical presence of the ESP was a powerful reminder of the presence of the remote worker in the meetings.

The studies and implemented systems above focused primarily on creating a high-presence media space. Our study, in turn, makes its own contribution to this field. To mimic real situations, the remote person's figure should be presented locally, without his/her remote background. Typically, this can be achieved by using mixed-reality (MR) technology and special head-mounted display (HMD) equipment [11, 4, 15]. Using HMD for some people might be encumbering and uncomfortable. This setup is likely to decrease the sense of presence. In contrast, our proposed system can be easily implemented in both sites, allowing both participants to experience the same effects.

2.2 Commercial Videoconferencing

In commercial videoconferencing business firms, many solutions have been introduced under the name "Telepresence" technology for high presence feelings. Telepresence is defined as an illusion that a mediated experience is not mediated [18]. In videoconferencing experience, telepresence gives you the feeling as if the remote participants are in the same room with you. To create the same-room illusion, some commercial telepresence solutions use a combination of technology elements, such as utilizing large displays for life-sized dimensions and hidden high-definition cameras strategically placed to create the appearance of a direct eye contact, and environmental design, such as consistent furniture arrangements across locations. The life-sized dimensions allow participants to see facial expressions, make eye contact, and read body language. Such solutions are: Cisco TelePresence TX9000 Series [1] , Polycom® RealPresence™ Immersive [2] , TANDBERG[3], PeopleLink TelePresence[4] , etc.

In one hand these solutions simulate high presence meeting environments as if the other people were sitting across the table in the same room. But on the other hand these solutions are very expensive, require large-spaces, and have to be installed in a fixed environment with pre-installed matching furniture in both sides to achieve maximum telepresence feelings. In contrast, BHS can be implemented using an affordable equipments and can be installed easily almost anywhere.

2.3 Verbal and Nonverbal Communication Analysis

It's well known that in F2F communication, people switch speaking and listening by using a complicated mechanism of verbal and nonverbal cues [2]. A major nonverbal cue in speaking involves the use of eye contact [1]. In F2F communications, failure to maintain eye contact is commonly considered to be a sign of deception, and leads to feelings of mistrust [2]. Vertegaal et al. concluded that gaze is an excellent predictor of conversational attention in multiparty conversations [30]. A study by Karmer et al.

[1] http://www.cisco.com/en/US/products/ps12453/
[2] http://www.polycom.com/products/telepresence/_video/
[3] http://www.tandberg.com/
[4] http://www.peoplelink.in/telepresence.html

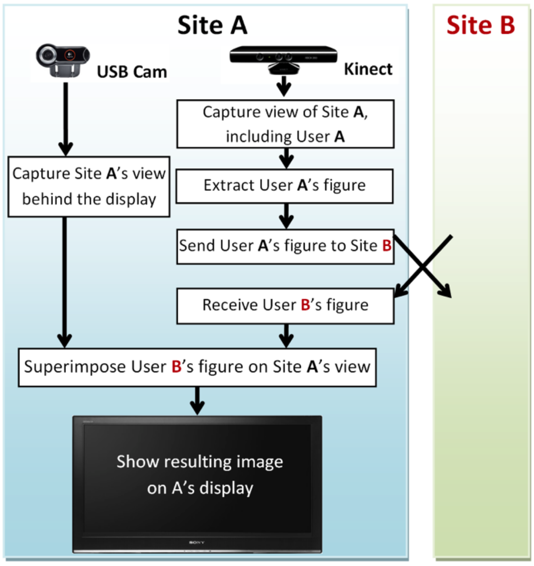


Figure 2: The process diagram of BHS.

proposed a method of measuring people's sense of presence in videoconferencing system based on linguistic features of their dialogues [17]. This study shows that 30% of the variance in self-reported presence can be accounted for by a small number of task-independent linguistic features.

The seating arrangements on group video communication affect participant's behaviors as well. A study by Inoue et al. presented a videoconferencing system "HERMES" that integrates F2F and video-mediated meetings [12]. In This study they observed that participants tended to pay much attention to the monitor when using lined-up seating arrangement. This problematic behavior solved by the combination of round seat arrangement and multiple monitors. Another study by Yamashita et al. revealed that seating arrangements affect speaker switches without verbal indication of the next speaker [31]. This study found that in some seating arrangement, the participants shared a higher sense of unity and reached a slightly better group solution.

Our study as well examined the proposed high-presence videoconferencing system for any verbal and/or nonverbal effects on communication comparing with a conventional videoconferencing system.

3 BEING HERE SYSTEM

A videoconferencing system "Being Here System (BHS)" was constructed to achieve pseudo same-room environment to the users. The two sites, 'Site A' and 'Site B', were connected over a local network to permit the exchange of live video. Each site was equipped with a display installed upright 70 cm above the floor, a USB camera, a Kinect™ RGB-D camera, a computer connected to the network, a speaker and a microphone, and a chair. The user was seated at 1.2 m distant from the display since this was considered to be appropriate distance for F2F meetings [10].

The process diagram of BHS is shown in Fig 2. We used the USB camera to capture the local site's front view, that is, the region concealed behind the display. The USB camera was placed behind the display in the center, and the camera's angle and zoom were calibrated so that the region behind the display was exclusively captured. This captured image (640 by 480 pixels) was used as a background for the display.

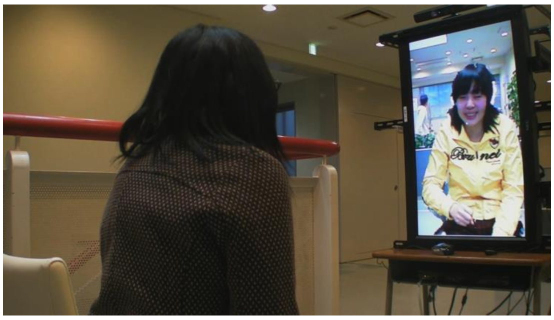


Figure 3: The 30 inch portrait display setup.

To capture the site view and extract the user's figure from it at run-time, the Kinect was used, which was placed centrally over the display and focused on the person's face. OpenNI API was used to analyze the Kinect image depth data by identifying the user in the scene and replacing the background with a transparent color. The resulted image was transmitted to the other site at 15 frames per second.

The final step in the process was to superimpose the received remote user figure onto the local front view. This was accomplished by merging the extracted user's figure and the background. Finally, the processed video was presented on the large display.

With this simple system architecture, the system is supposed to be expanded to the multi-point conferencing easily. This is a noteworthy feature that other existing systems have not achieved because of their limited spatial alignment and/or expensive customized devices.

4 SYSTEM EVALUATION

An experiment was conducted to evaluate BHS. The main objective of this experiment was to study the influence of BHS on user's communication behavior in comparison with conventional videoconferencing. Two remote sites, Site A and B, were constructed. In Site A, a large flat-panel display (46 inches) was used (Fig. 1), while in Site B, a 30-inch display was used. The 30-inch display was fixed in a vertical portrait position, presenting a life-sized image of an adult's upper body (Fig. 3). We used the portrait mode to study the effects of different background sizes.

4.1 Conditions

Two videoconferencing modes, "***Conventional mode***" where the remote site view was displayed in the local site's display and "***Superimpose mode***" where the remote person's figure was extracted and superimposed on the local site's front view, were established.

In the experiment, we considered the following videoconferencing conditions:

- Large Superimpose (***LS***): superimpose mode via large display.
- Portrait Superimpose (***PS***): superimpose mode via portrait display.

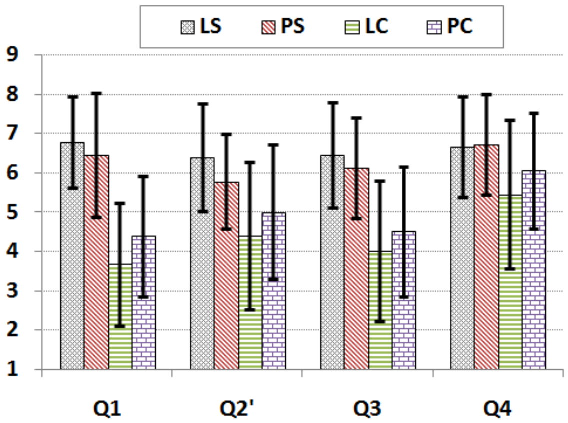


Figure 4: The participants' average sense of presence results *(Note: Q2' result is the positive form of the original Q2).*

- Large Conventional (***LC***): conventional mode via large display.
- Portrait Conventional (***PC***): conventional mode via portrait display.

4.2 Participants

Nine pairs, 7 females and 11 males whose ages ranged from 23 to 36 years old and who were familiar with each other, took part in the experiment. Among them, 17 participants had experience using videoconferencing systems. Most used the videoconferencing principally to talk to remote family members and/or remote close friends.

4.3 Procedure

One of the pair used the system at Site A, while the other used Site B. Before performing the videoconferencing tasks, the participants were asked to complete a basic demographic survey. After this, the experimenter introduced the system to the participants. The experiment began with a familiarization session. Each participant performed four videoconferencing sessions to test the conditions. In each session, participants were instructed to talk about a selected general topic for approximately 10 minutes. After that they were asked to complete the questionnaire about the system. The four general topics were:

Study life in X city: discuss with the other person the pros and cons of studying in X city; how long you have been in X city; why you choose X university, compare X city with other cities you have been in, etc.

Buying a new laptop: discuss the laptop's specifications; the suggested shops; prices; usage; etc.

Planning a trip: for the coming summer vacation, discuss the trip's options; where to go; locally or abroad; cost; weather; attraction; etc.

Plans after graduation: discuss with the other person your plans after graduation, the possibility of pursuing a higher degree; work options, etc.

The conditions orders were randomized to ensure that the order of the tested conditions would not affect the result.

4.4 Questionnaire

In the questionnaire, we asked participants to evaluate each of the statements according to the feeling they experienced during the videoconferencing session. To investigate the participants' sense of co-presence in each condition, the following statements were used [10]:

- ***Q1***: "I felt as if the other person existed in the same room."
- ***Q2***: "I didn't feel as if I were talking with the other person in the same room."
- ***Q3***: "I felt as if I were facing the other person in the same room."

The perceptual distance between the participants is an aspect of the sense of co-presence. To evaluate this, the following statements were used:

- ***Q4***: "I felt that the distance between me and the other person was comfortable for chatting."
- ***Q5***: "I felt that the distance between me and the other person was around:__________ "

All of these statements, except Q5, were rated on a 9-point Likert scale, where 1 = strongly disagree, 3 = disagree, 5 = neutral, 7 = agree, and 9 = strongly agree.

4.5 Videotaping

Two cameras were used to record the experiment sessions at HD 720 resolution (1280 by 720 pixels). The first camera was placed over the display facing the participant in order to capture his/her facial expressions, gestures, and postures. The second camera was installed upright 1 m above the floor beside participant in order to capture him/her from the side and the display content.

5 RESULTS

5.1 Questionnaire

Figure 4 shows the average results of the participants' sense of the other person's presence while videoconferencing, under the four conditions. A comparison was done using two-factor ANOVA test. The first factor is the videoconferencing mode (i.e. Conventional and Superimpose). The second factor is the used display (i.e. Large and Portrait). We found main effect of videoconferencing mode over the participants' sense of other person's presence as if in the same room (Q1: $F(1,68) = 55.26$, $p<0.01$, Q2': $F(1,68) = 14.08$, $p<0.05$, Q3: $F(1,68) = 31.71$, $p<0.01$). This indicates that the superimposed videoconferencing mode enhanced the presence feelings more than the conventional videoconferencing mode. On the other hand, the results shows no main effect of the used display (Q1: $F(1,68) = 0.31$, Q2': $F(1,68) = 0.0$, Q3: $F(1,68) = 0.06$). The result also shows that there is no interaction between the used mode and display (Q1: $F(1,68) = 2.31$, Q2': $F(1,68) = 2.73$, Q3: $F(1,68) = 1.34$).

Moreover, we found main effect of videoconferencing mode over the feeling of comfortable distance between the user and the other person (Q4: $F(1,68) = 7.14$, $p<0.01$), while no main effect of the used display (Q4: $F(1,68) =$

Figure 5: Screenshot of one of the ELAN's annotated video.

0.89). The result also shows there is no interaction between the used mode and display (Q4: $F(1,68) = 0.62$). This indicates that the superimposed videoconferencing mode enhanced the feeling of comfortable distance between the user and the other person. In addition, participants who used the superimpose videoconferencing mode were able to estimate the distance more accurately. The average estimated distances were as follows:

- LS: 1.3 m (s.d. = 0.6).
- PS: 1.3 m (s.d. = 0.6).
- LC: 2.2 m (s.d. = 1.2).
- PC: 1.8 m (s.d. = 1.0).

(The actual distance between the participant and the display was 1.2 m).

5.2 Communication Behavior

Communication behavior was analyzed quantitatively using the recorded video. ELAN[1] was used to annotate the video. A total of 36 recorded videos from 9 pairs by 4 conditions were annotated for user's communication behaviors such as speech and gaze. The middle 2 minutes of each session was analyzed, which resulted in a total of 72 minutes data. Figure 5 shows a screenshot of one of the ELAN's annotated video.

The following aspects were used for the analysis:

- ***Speech***: happens when a person speaks for at least 1.5 seconds [14].
- ***Turn taking***: is defined as the manner in which orderly conversation normally takes place. The principles of turn-taking were first described by sociologists Sacks et al. in [25]. In this study, we adopted the same turn definition with [27] as the person's number of continuous segment of speech between silent intervals for at least 1.5 seconds.
- ***Overlap***: is a simultaneous speech by two persons. This might happen when taking turns or when responding to other person's speech while talking.
- ***Gaze***: happens during a conversation when a person look at the other [1].
- ***Gaze aversion***: this term is defined for the analysis in this paper. It happens when a person averts his/her gaze from the other.

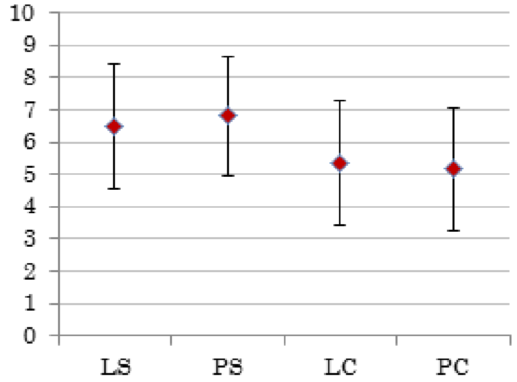


Figure 6: The participants' average number of turn taking per minute.

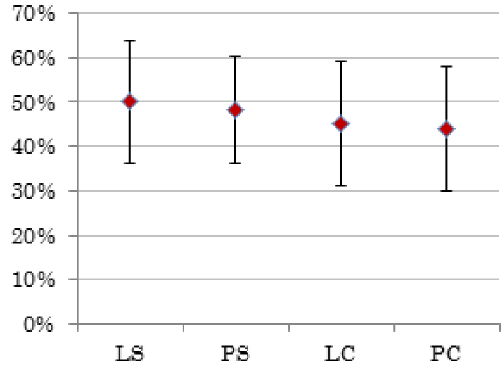


Figure 7: The participants' average speech rate.

5.3 Speech

Figure 6 shows the average results of the participants' number of turn taking per minute under the four videoconferencing conditions. A comparison among the conditions was done using one-way repeated-measures ANOVA test. We found significant difference in the number of turn taking ($F(3,51) = 6.49$, $p<0.05$). From Tukey's HSD post-hoc test, the superimpose conditions were significantly different from the conventional conditions.

Figure 7 shows the average results of participants' percentage of individual speech. We found no significant difference between the conditions ($F(3,51) = 0.48$).

Figure 8 shows the average results of the participants' percentage of speech overlap. We found significant difference in the percentage of speech overlap ($F(3,51) = 11.69$, $p<0.01$). The superimpose conditions were significantly different from the conventional conditions.

5.4 Gaze

Figure 9 shows the average results of the participants' number of gaze aversion per minute. We found a significant

[1] http://tla.mpi.nl/tools/tla-tools/elan/

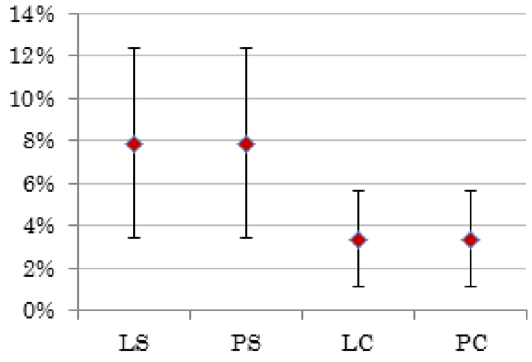


Figure 8: The participants' average percentage of speech overlap.

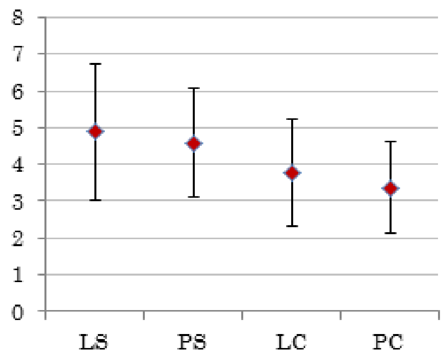


Figure 9: The participants' average number of gaze aversion per minute.

difference in the number of gaze aversion ($F(3,51) = 4.83$, $p<0.05$). The superimpose conditions were significantly different from the conventional conditions.

6 DISCUSSION

6.1 Display Size

We expected that the results could be different depending on the display sizes because of the different background sizes. One participant mentioned that the portrait display's frame concealed a relatively large area of the front view compared with the large display, which may be related to the study by Bi et al. on the effects of bezels of large tiled display that the bezels affected tunnel steering [3].

The results were, however, very similar between different display sizes. More regions are concealed in the large display conditions. It is expected that the superimpose mode can compensate this by displaying the front view as a background of the display. Actually the large superimpose condition and the portrait superimpose condition obtained the similar results. This indicates that the display size has no major effect as long as the displayed background is integrated with the actual front view.

In the conventional modes, the large display shows remote site more, which might decrease the sense of co-presence compared to the portrait display. This was not observed in the communication behaviors. The small differences of questionnaire results between the large conventional condition and the portrait conventional condition could be explained by this, although the differences are not statistically significant.

6.2 Communication Behavior

In this paper it was shown that the superimpose mode significantly affected participants' verbal and nonverbal communication behaviors. The superimpose conditions increased the number of turns by around 130% more than the conventional conditions (Fig. 6). The participants' average percentage of speech wasn't affected by the tested conditions (Fig. 7). This result is consistent with a related research by Sellen, which compared face-to-face and video-mediated conversations to find no difference in speech rate [27]. Each participant spoke 48% of the session time on average in our experiment. We found that the percentage of speech overlap in the superimpose conditions were twice more than the conventional conditions (Fig. 8). Because Cohen's study [7] and Sellen's study [27] found that face-to-face imposes more simultaneous speech compared with video conditions, this can be one of the evidences that our proposed superimpose mode could be closer to the F2F than the conventional mode.

Gaze is an important aspect of nonverbal communication [6, 9, 27]. We investigated the gaze directions and counted the number of gaze aversion in this study. The result shows that the participants tended to avert their gazes more when they used the superimposed conditions compared with the conventional conditions (Fig. 9). In F2F conversations, people use more gaze when they are further apart [1]. This means that the participants who used the superimposed conditions might feel closer to the other person.

6.3 Multipoint BHS

BHS was a point-to-point conferencing system. However, the system was supposed to be expanded to the multi-point conferencing easily considering its simple architecture, as noted in Section 3. Figure 10 shows a schematic diagram of the multi-point BHS which does not need a network, though obviously the system can be expanded to multi-point using a multicast network.

To examine its feasibility, the actual multipoint BHS was implemented (Fig. 11). It can be observed that the multipoint BHS provides higher sense of co-presence than the conventional multipoint videoconferencing (Fig. 12), even from the figures.

7 CONCLUSION

In BHS, the pseudo same-room effect is achieved by superimposing the remote person's figure, which is extracted from the remote site view using the Kinect RGB-D camera, with the local front view on a large display. BHS effectively reduced the psychological distance between the remote participants.

In this study, we investigated user's verbal and nonverbal communication behaviors while using BHS, in comparison with a conventional videoconferencing. The analysis of the

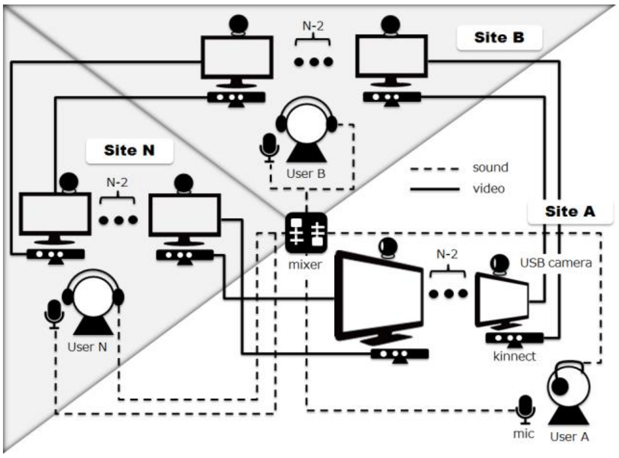


Figure 10: Design of Multipoint BHS.


Figure 11: Multipoint BHS.


Figure 12: Conventional multipoint videoconferencing.

recorded video revealed that using BHS significantly affected user's communication behavior. This result suggests that considering the local site front view as a background is one practical way to create the same-room illusion, which facilitates communication.

REFERENCES

[1] M. Argyle, Bodily Communication, 2nd Edn, Routledge, (1988).

[2] E. Bekkering, and J. Shim, “Trust in videoconferencing,” Commun. ACM, Vol. 49, No. 7, pp.103–107 (2006).

[3] X. Bi, S. Bae, and R. Balakrishnan, “Effects of interior bezels of tiled-monitor large displays on visual search, tunnel steering, and target selection,” In Proceedings of the 28th international conference on Human factors in computing systems, CHI’10, ACM, pp.65–74 (2010).

[4] M. Billinghurst, and H. Kato, “Out and about real world teleconferencing,” BT Technology Journal, Vol. 18, pp.80–82 (2000).

[5] D. Chatting, J. Galpin, and J. Donath, “Presence and portrayal: video for casual home dialogues,” In Proceedings of the 14th annual ACM international conference on Multimedia, MULTIMEDIA’06, ACM, pp.395–401 (2006).

[6] M. Chen, “Leveraging the asymmetric sensitivity of eye contact for videoconference,” In Proceedings of the SIGCHI conference on Human factors in computing systems: Changing our world, changing ourselves, CHI’02, ACM, pp.49–56 (2002).

[7] K. Cohen, “Speaker interaction: video teleconferences versus face-to-face meetings,” In Proceedings of Teleconferencing and Electronic Communications, pp.189–199 (1982).

[8] S. Gibbs, C. Arapis, and C. Breiteneder, “Teleport towards immersive copresence,” Multimedia Systems, Vol. 7, No. 3, pp.214–221 (1999).

[9] D. Grayson, and A. Monk, “Are you looking at me? eye contact and desktop video conferencing,” ACM Trans. Comput.-Hum. Interact, Vol. 10, No. 3, pp.221–243 (2003).

[10] Y. Ichikawa, K. Okada, G. Jeong, S. Tanaka, and Y. Matsushita, “Majic videoconferencing system: experiments, evaluation and improvement,” In Proceedings of the fourth conference on European Conference on Computer-Supported Cooperative Work, ECSCW’95, Kluwer Academic Publishers, pp.279–292 (1995).

[11] T. Inoue, “Mixed reality meeting system enabling user to keep and share interpersonal distance in the real world,” Journal of Information Processing Society of Japan, Vol. 50, No. 1, pp.246–253 (2009).

[12] T. Inoue, K. Okada, and Y. Matsushita, “Integration of face-to-face and video-mediated meetings: Hermes,” In Proceedings of the international ACM SIGGROUP conference on Supporting group work: the integration challenge, GROUP’97, ACM, pp.405–414 (1997).

[13] T. Ishida, A. Sakuraba, and Y. Shibata, “Proposal of high realistic sensation system using the large scale tiled display environment,” In Proceedings of Network-Based Information Systems (NBiS), pp.444–449 (2011).

[14] J. Jaffe, and S. Feldstein, “Rhythms of Dialogue,” Academic Press, New York, (1970).

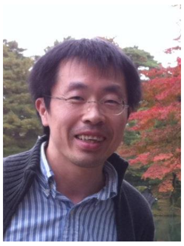
[15] T. Kantonen, C. Woodward, and N. Katz, “Mixed reality in virtual world teleconferencing,” In 2010 IEEE Virtual Reality Conference (VR), pp.179–182 (2010).

[16] E. Koh, “Conferencing room for telepresence with remote participants,” In Proceedings of the 16th ACM international conference on Supporting group work, GROUP’10, ACM, pp.309–310 (2010).

[17] A. Kramer, L. Oh, and S. Fussell, “Using linguistic features to measure presence in computer-mediated communication,” In Proceedings of the SIGCHI conference on Human Factors in computing systems, CHI’06, ACM, pp.913-916 (2006).

[18] M. Lombard, and T. Ditton, "At the heart of it all: The concept of presence," Computer-Mediated Communication, Vol. 3, No. 2, (1997).
[19] M. Mantei, R. Baecker, A. Sellen, W. Buxton, T. Milligan, and B. Wellman, "Experiences in the use of a media space," In Proceedings of the SIGCHI conference on Human factors in computing systems: Reaching through technology (CHI'91), pp.203–208 (1991).
[20] O. Morikawa, and T. Maesako, "Hypermirror: toward pleasant-to-use video mediated communication system," In Proceedings of the 1998 ACM conference on Computer supported cooperative work (CSCW'98), pp.149–158 (1998).
[21] H. Nakanishi, K. Kato, and H. Ishiguro, "Zoom cameras and movable displays enhance social telepresence," In Proceedings of the 2011 annual conference on Human factors in computing systems (CHI'11), pp.63–72 (2011).
[22] H. Nakanishi, Y. Murakami, and K. Kato, "Movable cameras enhance social telepresence in media spaces," In Proceedings of the 27th international conference on Human factors in computing systems (CHI'09), pp.433–442 (2009).
[23] M. Nawahdah, and T. Inoue, "Being Here: Enhancing the Presence of a Remote Person through Real-Time Display Integration of the Remote Figure and the Local Background," Transactions of the Virtual Reality Society of Japan, Vol. 17, No. 2, pp.101-109 (2012).
[24] K. Okada, F. Maeda, Y. Ichikawa, and Y. Matsushita, "Multiparty videoconferencing at virtual social distance: Majic design," In Proceedings of the 1994 ACM conference on Computer supported cooperative work (CSCW'94), pp.385–393 (1994).
[25] H. Sacks, E. A. Schegloff, and G. A. Jefferson, "Simplest Systematics for the Organization of Turn-Taking for Conversation," Language Vol. 50, No. 4, pp. 696–735 (1974).
[26] D. Sakamoto, T. Kanda, T. Ono, H. Ishiguro, and N. Hagita, "Android as a telecommunication medium with a human-like presence," In Proceedings of the ACM/IEEE international conference on Human-robot interaction (HRI'07), pp.193–200 (2007).
[27] A. J. Sellen, "Speech patterns in video-mediated conversations," In Proceedings of the SIGCHI conference on Human factors in computing systems (CHI'92), pp.49–59 (1992).
[28] P. Tanner, and V. Shah, "Improving remote collaboration through side-by-side telepresence," In Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems (CHI EA'10), pp.3493–3498 (2010).
[29] G. Venolia, J. Tang, R. Cervantes, S. Bly, G. Robertson, B. Lee, and K. Inkpen, "Embodied social proxy: mediating interpersonal connection in hub-and-satellite teams," In Proceedings of the 28th international conference on Human factors in computing systems (CHI'10), pp.1049–1058 (2010).
[30] R. Vertegaal, R. Slagter, G. van der Veer, and A. Nijholt, "Eye gaze patterns in conversations: there is more to conversational agents than meets the eyes," In Proceedings of the SIGCHI conference on Human factors in computing systems (CHI'01), pp.301–308 (2001).
[31] N. Yamashita, K. Hirata, S. Aoyagi, H. Kuzuoka, and Y. Harada, "Impact of seating positions on group video communication," In Proceedings of the 2008 ACM conference on Computer supported cooperative work (CSCW'08), pp.177–186 (2008).
[32] N. Yankelovich, N. Simpson, J. Kaplan, and J. Provino, "Porta-person: telepresence for the connected conference room," CHI'07 extended abstracts on Human factors in computing systems, pp.2789–2794 (2007).
[33] Y. Noguchi, T. Inoue, "Study on Placement and Presentation of Multiple Avatars in MR-based Distributed Meeting," Information Processing Society of Japan (IPSJ) Journal, Vol.48, No.1, pp54-62 (2007).

(Received December 18, 2013)



Tomoo Inoue is Professor of the Faculty of Library, Information and Media Science of University of Tsukuba. His research interests include HCI, CSCW, and Technology-enhanced learning. He received his Ph.D. in Engineering from Keio University in 1998. He is a recipient of awards including Best Paper Award, Activity Contribution Award and SIG Research Award from Information Processing Society of Japan (IPSJ). He has served a number of academic committees, currently including IEICE SIG Human Communication Science, IEICE SIG Multimedia on Cooking and Eating Activities, VRSJ SIG Cyberspace, IPSJ SIG Digital Contents Creation, IPSJ SIG Groupware and Network Services, APSCE CUMTEL SIG, and IEEE TC CSCWD.

Mamoun Nawahdah was born in Palestine in 1976. He received the B.Eng. degree in Computer System Engineering from the PPU, Palestine in 1997, the M.S. degree in Scientific Computing from the Birzeit University, Palestine in 2005, and the Ph.D. degree in Information Science from the University of Tsukuba, Japan in 2013. In 2006, he joined the Faculty of Engineering and Technology, Birzeit University, as a Lecturer, and in 2013 became an Assistant Professor. His current research interests include pair programming, human computer interaction (HCI), Mixed Reality, and remote communication.

Yasuhito Noguchi received his M.S. degree from the Graduate School of Library, Information and Media Studies at University of Tsukuba, Japan in 2008. He is currently working toward Ph.D. degree in the Graduate School of Library, Information and Media Studies at University of Tsukuba. His research interests include groupware.