2021 30th IEEE International Conference on Robot and
Human Interactive Communication (RO-MAN)
August 8 - 12, 2021 - Vancouver, BC, CA (Virtual Conference)

Proposing Remote Video Conversation System "PARAPPA": Delivering the Gesture and Body Posture with Rotary Screen*

Koki Ijuin[1], Kunihiro Ogata[1], Kentaro Watanabe[1], Hiroyasu Miwa[1] and Yoshinobu Yamamoto[2]

***Abstract*— Globalization and the effect of recent infectious disease are changing the remote conversations as a new normal in business meetings, social provision and casual chatting. Previous research show that the remote conversations have a difficulty on showing the presence or attention to the other interlocutors, especially in the situation where the majority of interlocutor share the same place and one or a few interlocutor participate from difference place. This paper proposed "Parappa", a remote video conversation system, for those unbalanced condition by utilizing both physical and virtual approaches to share the nonverbal behaviors of remotely-participating interlocutor. The proposed system is constructed with a rotatable screen which projects the life-size avatar of remote interlocutor. The rotation of the screen represents the body posture and the projected avatar show the gesture. The results of preliminary analyses of eye gaze activities and frequency of the screen rotation during conversation suggests the possibilities that proposed system enables to show the presence of remote interlocutors.**

I. INTRODUCTION

The number of remote conversations are increasing due to globalization and effect of COVID-19, because the workers are unable to meet face-to-face. Many companies have been turning to use videotelephony services in order to reduce time, money and risk of spreading infection. There are several studies which suggest that the remote conversation by video supports the interlocutors' satisfaction [1]. However, those who has experienced the remote conversations, to come up with the conflict of speech and less reaction in multiparty conversations. When we look back to the ordinary face-to-face conversations, we consciously or unconsciously utilize the nonverbal behavior in order to coordinate the conversational flows, as the many researches already reported. However, videotelephony services remain have some difficulty in usage of multimodal behavior compared to face-to-face conversations such as eye contact and body posture [2].

Several studies have aimed to further understand face-to-face communication and multimodal signaling of social interactions [3], [4], [5]. The nonverbal behaviors are used to signal their speaking intentions to the partner [6], to express intimacy [7] [8], and to coordinate conversations in typical face-to-face conversations[9]. In addition, the importance of nonverbal behaviors increase in multiparty conversations, where the conflict of speech, or turn taking, occurs more often than the dual conversations [10], [11]. From those reports, there is a possibility that utilizing interlocutors' nonverbal behaviors in remote conversations gives them more smooth and satisfied conversations.

The issue which often, and only, occurs in remote conversations is that the some of the interlocutors join to the conversation from the same place with a single computer, such as a business meeting between two different meeting rooms. The typical situation of such unbalanced remote conversations is that the only one interlocutor participate to the conversations from the different place while other interlocutors are in the same place and share the multimodal behaviors directly. It is obvious that the single remote interlocutor has many obstacles in order to participate to the conversation where the others can converse without any delay and deficit of speech and nonverbal behaviors. This is also a challenge for remote services which require close interactions among participants, such as remote workshops and training, and a virtual visit to senior people at a care facility.

From those situations we are facing, there is a strong need to understand how the interlocutors communicate in remote conversations in order to support remote conversations from the perspective of utilizing nonverbal behavior. We propose the new videotelephony system which assist the unbalanced remote conversation from the perspective of supporting to express the nonverbal behavior of remote-participating interlocutor. Proposed system *"Parappa"*(shown in Fig.1) is created in order to utilize both physical and virtual aspects where the remote-participating interlocutor's body posture is shown artificially by controlling screen direction and gestures is projected as the life-size avatar to the screen.

Fig. 1. Proposed System:Parappa

This paper aims to propose the possibilities that utilizing

*This study is partly supported by JSPS KAKENHI Grant Number JP19H04416.

[1] National Institute of Advanced Industrial Science and Technology, Human Augmentation Research Center. Kashiwa II Campus, University of Tokyo, 6-2-3 Kashiwanoha, Kashiwa, Chiba 277-0882 Japan koki-ijuuin@aist.go.jp

[2] The University of Fukuchiyama, 3370 Hori, Fukuchiyama. Kyoto 620-0886, Japan

nonverbal behavior enable to show the presence of remote-participating interlocutors. The proposed system *"Parappa"* delivers nonverbal behavior by combining both physical and virtual approaches. The purpose of this paper is to investigate how nonverbal behaviors such as eye gaze and body posture were used in the remote conversations using *Parappa*. The preliminary experiments and analyses were conducted in order to confirm how the remote conversations using *Parappa* were constructed.

This paper is structured as follows. We first introduce the researches related to the functions of nonverbal behavior in face-to-face and remote conversation. Then, we explain the structure of proposed system *"Parrapa"* and preliminary experimental setup and the method of annotation. We show and discuss results of preliminary analyses and conclude with the summary and future plan.

II. RELATED WORKS

A. Function of nonverbal behavior

In typical face-to-face conversations, the interlocutors use speech and nonverbal behavior for many purposes, as written in previous section. Many researches report that eye gaze is one of the strongest visual cues in face-to-face conversations, and it has a variety of functions, such as managing the attention of interlocutors, expressing intimacy, and coordinating turn management [6] [7] [8] [9]. The fundamental eye gaze related to turn management were discussed in Kendon [6], who stated that there are at least four such functions: (1) to provide visual feedback; (2) to organize the conversation's flow; (3) to interpret emotions and relationship; and (4) to concentrate on understanding the utterance by shutting out visual information. He also demonstrated that speakers look away at a turn's beginning and look back to their partners near the turn's ending. Those eye gaze activities such as gazing at or avoiding conversational partners might be used for some functions of turn management during two-person conversations. In contrast, Beattie [12] reported that there was no relation between eye gaze and floor apportionment under the experimental condition in which two participants play different social positions. Those results suggest that the eye gaze activity combines many functions, and the importance of those functions might changes by the condition of conversational setup.

B. Eye gaze activities in multiparty conversation

The aforementioned studies mainly dealt with dual conversations. However in multiparty conversation, there is a possibility that the features used for managing turn control may be different from dual conversations. As for triad conversations, Kalma [10] reported that the recipient of prolonged gaze, i.e., the participant who is gazed at by the speaker during the silence after his/her utterance, tended to take the floor. Lerner [11] also reported that the speaker anticipates the next speaker explicitly in many ways, including eye gaze. For turn management with eye gaze, the speaker signals the assumed next speaker by their gaze, requiring him/her to see that gaze, and the other participant also grasps the expectation that someone else will speak next. Vertegaal discussed the importance of gaze in multiparty conversations for signaling conversational attention [13], and Jokinen showed that the speaker's gaze is important for coordinating turn taking in multiparty conversations and that partners pay attention to the speaker's gaze behavior [14]. These results in face-to-face conversations show that the eye gaze activity plays an important role in conversations for coordinating turns, and the importance of those functions of eye gaze activity change in the situations such as number and social relation of interlocutors.

C. Nonverbal behavior in remote conversation

There are also several researches which discussed on the function of nonverbal behaviors in remote conversations. The previous research showed that when the remote interlocutors discuss about complex matters, video helps to provide information about them [1]. The other research showed that the interlocutors are more satisfied when communicating through video than audio-only, but the video does not affect the outcome performance of various tasks except for negotiation tasks [15]. Video enable to provide benefit of visual information to interlocutors in a way that they can read important cues from each other's face. Gaver [16] also mentioned that the gestures in remote conversations are mainly used as a signal to get attention. These researches suggests that the visual information in remote conversations assists to satisfy the interlocutors and to communicate well for both providing more contextual information and signaling important cues.

As mentioned above, it is not rare case to have remote conversations in which only one or few interlocutors participate from the different physical place. The previous research [17] suggested the problems occur in those unbalanced remote conversations for both in-room interlocutors and remote interlocutors. The problem for in-room interlocutors is that they tend to forget about remote interlocutors. For remote interlocutors, it is difficult to detect in-room turn change and to identify all in-room interlocutors due to the device restriction such as delay, audio quality and camera's view angle. The in-room interlocutors are able to share their nonverbal behaviors so that the smooth conversation is established just with in-room interlocutors and the remote interlocutors have a difficulty to follow the conversations and tend to be ignored if both interlocutors voluntarily encourage to participate all interlocutors. The previous research of assisting unbalanced remote conversations reported that using the flashlights to show the eye gaze direction of remote interlocutor enable to show the social presence of remote interlocutor where the in-room interlocutor paid more attention to the remote interlocutor [18].

From those studies, the nonverbal behaviors are important in human-human conversations for many reasons. However, the importance and usage of functions of nonverbal behaviors depend on the situation of the conversational environment. It is an important issue to examine how the nonverbal behaviors contribute to the unbalanced remote conversations which can often happen in many business situations, especially the

pandemic of COVID-19. The infection risk differs between cities so that some companies allow employees to come, but other companies in a different place can not.

III. Proposed System

To make the remote interlocutors feel like being in the same room, we propose a system *Parrapa* as shown in the Fig. 2.

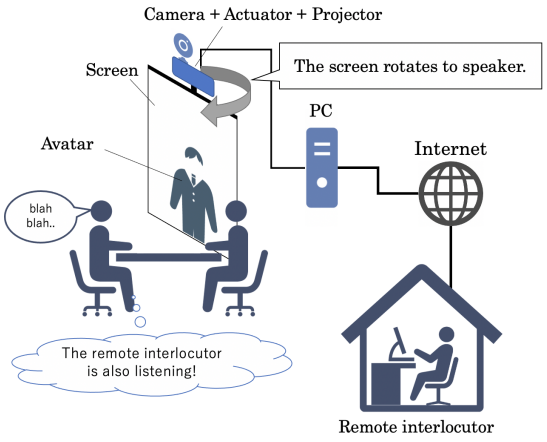

Fig. 2. Proposed System

Remote interlocutors can communicate the in-room interlocutors via avatars. The avatar is displayed onto the screen using a projector. These projector and screen have the camera and actuator, and the actuator rotates the screen automatically or manually to face the speakers. By rotating the screen, the in-room interlocutors can receive the intention that the remote interlocutors are listening and can recognize its existence. Remote interlocutors will become more likely to participate in conversations without feeling alienated from onsite members. Furthermore, the remote interlocutors can pay attention to the in-room interlocutor who want to talk, and it is expected that smooth communication can be realized by this system.

Here, in the screen rotation control, a method of automatically controlling based on information such as utterances and line of sight from the interlocutors, command operation by a remote interlocutors, or semi-automatic control by two inputs can be considered. It is considered necessary to use these properly according to the purpose and situation.

IV. Preliminary Experiment

We propose a system which allows remote interlocutors; to show their gestures with the life-size screen, and to show their attention by rotating the screen. We made two hypotheses that;

1) in-room interlocutors can capture the remote interlocutor's gesture and specify his/her attention with the direction of the screen so that in-room interlocutors can converse without forgetting remote interlocutors,
2) remote interlocutor is able to capture visual information widely by rotating the screen so that the difficulty of participating conversation will decrease.

In this paper, we conducted the preliminary experiments to confirm how the conversations constructed with *Parappa*. From the basic statistics of eye gaze and frequency of screen rotation, we roughly confirmed these hypotheses.

A. Experimental System

The experimental configuration of the remote communication support system used in this study is described in Fig. 3. The participants will be divided into two rooms to talk remotely. Room1 has a rotating projector and two participants, and Room2 has one participant. The PC in Room2 can interact through the rotating screen in Room1 via the teleconferencing system. In addition, the participants in Room2 can rotate the screen by wireless communication using the game controller.

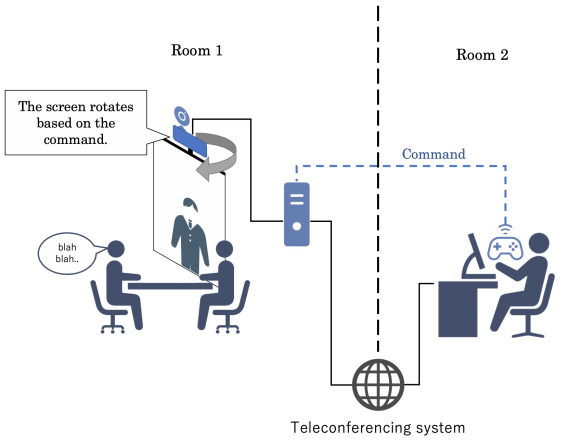

Fig. 3. Experimental System

As the game controller, we used the Nintendo Switch Joy-Con developed by Nintendo Co., Ltd. Joy-Con can be connected to a PC via BLE communication and can send commands wirelessly to the PC that controls the motor of the rotting projector. V-SERVO (VS-SV1150) manufactured by Vstone Co., Ltd. was used as the actuator of the rotating projector. By using the Vstone's communication cable, position control is possible by serial communication from a PC. In this research, it is possible to control the angle of the screen in 10 degree increments by inputing the Joy-Con.

As the avatar, we used the filter of Snap Camera which automatically convert the remote participant's face to anime-like. The all participants did not feel discomfort and they were able to perceive the changes of facial expressions with anime-like face.

B. Experimental Condition

In order to examine these hypotheses, we conducted a preliminary experiment. Table.I shows the condition of the preliminary experiment. We conducted two separate conversations of five minutes with a group of three participants. The participants work at the same office, and they know each other. Two different topics were given to the participants, free-flowing conversation and goal-oriented conversation. Free-flowing conversation is similar to casual conversations

in which they just talked along the given topic freely. Goal-oriented conversation requires the participants to decide one answer so that they need to cooperate and all the participants requires to give some idea or opinion.

Each participant wore Tobii Pro Glasses 2 in order to capture the eye gaze activities. The Omnidirectional camera is used to capture both gesture and audio information. Fig.4 shows experimental condition of Room1. In-room participants and proposed system were located around the table with the same distance, it was difficult to look both participants without moving head.

TABLE I
EXPERIMENTAL CONDITION

	Experimental condition
participants	3 males
relation of participants	workplace colleagues
# of conversation	2
duration of conversation	5 minutes
conversation topic	Free flowing conversation: ”like and dislike animals” Goal oriented conversation : ”one service which requires in their office”
equipments	Tobii Pro Glasses 2

Fig. 4. Experimental Setup

C. Annotation

After collecting the experimental data, we annotated the conversation. The annotation should be done for both utterances and nonverbal behaviors, although in this preliminary analysis, we focus on eye gaze activities. The eye gaze activities were annotated for each participant. For the remote participant, eye gaze toward the display of the computer was annotated because it was difficult to determine where exactly he/she gazed at in the display. For in-room participants, eye gaze toward the screen of the proposed system and toward the another in-room participants were annotated. Fig.5 and 6 are the snapshot of annotation software ”Tobii Pro Lab”. Red circle in these figures show the fixation point of participant.

V. PRELIMINARY ANALYSES

Preliminary analyses of eye gaze activities in unbalanced remote conversation were done. We first analyzed the ratio

Fig. 5. Annotation software Tobii Pro Lab for in-room participants

Fig. 6. Annotation software Tobii Pro Lab for remote participants

of duration to confirm how long each participant gazes at whom. The results for each participant for both free-flowing and goal-oriented conversations in Fig.7. The results showed that remote participant mostly gazed at the display during both conversations. In-room participants gazed at the screen and other in-room participants almost at the same amount. However, the participants gazed away or moving the gaze-target during the conversations. These results suggest that the remote participant kept his gaze toward display in order to grasp many visual information.

We also recorded how often the remote participant rotated the screen. Table.II shows the number of fixation toward each in-room participant. We define ”toward in-room participant” as the remote participant rotates the screen more than 10 degrees from the neutral position, where the screen vertically faces toward the middle of the table. Table.III shows the ratio of duration to confirm how long the screen faces to each in-room participant, and the average duration of staying in the same position. These results showed that the remote participant set the screen in the neutral position often compared to facing to in-room participants.

These results suggest that remote participant tried to grasp

overview of the conversation by facing toward the middle of in-room participants.

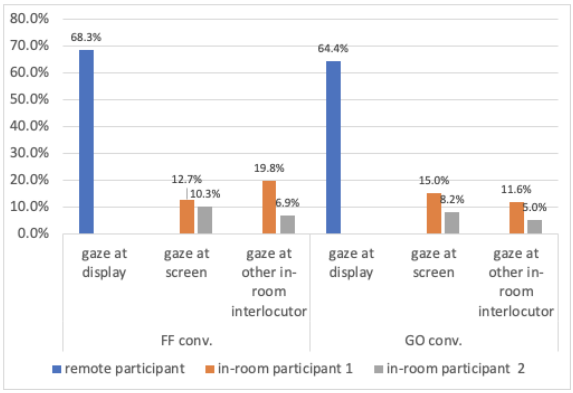

Fig. 7. Ratio of duration which the each interlocutor gazes at other interlocutor, system's screen or computer's display. Note that FF conv. stands for free-flowing conversations and GG conv. for goal-oriented conversation

TABLE II
NUMBER OF REMOTE PARTICIPANT ROTATE THE SCREEN TOWARD EACH IN-ROOM PARTICIPANTS.

	in-room participant 1	in-room participant 2	toward neutral
FF conv.	5	1	5
GO conv.	4	3	7

VI. DISCUSSION

We examined the eye gaze activities of participants and screen rotation from the preliminary experiments of two conversations.

The results of eye gaze activities by in-room participants showed that they gazed at the other in-room participant and the screen almost the same amount. The physical rotation of the screen might attract the attention of in-room participants, which is natural in face-to-face conversations by the body posture. The rotation of the screen enables to show his/her attention direction which can be perceived with both visual and audio information unconsciously because of the size of screen and small sound from the actuator. For this, intentional gazing toward partner or interruption with the speech is needed in ordinary videotelephony services. This might support our first hypothesis which the proposed system assists showing the presence of remote participant. The results also showed that the in-room participants gaze more at other things compared to at the screen and the other in-room participant. The in-room participants tend to stay their gazes around the middle of the screen and the other in-room participant, or tend to gaze around the other participants. This suggests the possibility that the proposed system attracts the attention of interlocutors too much so that the in-room participants were needed to prepare to the remote participant in order to react. However, the amount of eye gaze toward the interlocutors depends on the characteristic of the participants, the further analyses are needed to verify what causes these eye gaze patterns.

TABLE III
RATIO OF DURATION REMOTE PARTICIPANT ROTATE THE SCREEN TOWARD EACH IN-ROOM PARTICIPANTS.

	in-room participant 1	in-room participant 2	toward neutral	average duration[s]
FF conv.	32%	24%	44%	28s
GO conv.	21%	20%	60%	23s

The results also showed that remote participant mainly gazed at the display in order to capture the visual information of in-room participants. The remote participant rotated the screen almost every 25 seconds in average, and he/she kept the screen in the middle of the in-room participants often. These results suggested that the remote participant was able to capture the overview of in-room conversation and he/she can rotate the screen toward the in-room participant in order to focus on the specific in-room participant. This might weakly support the second hypothesis that the remote participant was able to capture the both overview and specific visual information by rotating the screen, although the further analyses of utterances and comparison between non-rotating condition is needed to examine the effect of participating to conversations.

VII. FUTURE WORKS

In this paper, we proposed the new remote conversation system which assists to participate the remote interlocutors in unbalanced remote conversations, where the majority of interlocutors share the same place. The preliminary experiments and analyses suggested the possibilities that the physical rotation of screen and the virtual life-size screen enable to show the presence of remote interlocutor, from the fact that the in-room interlocutors gaze at both participants evenly. The results also suggest that the remote interlocutor was able to grasp visual information widely by rotating the screen, and the rotation also show his attention to the in-room interlocutors. Although this paper argued the effectiveness of proposed system by the limited number of conversations, we still believe the preliminary analyses supported the possibility of new forms of remote conversations by combining physical aspect, rotating screen, and visual aspect, gestures in screen.

We also found several issues on both analyzing the effectiveness of the proposed system and improving the proposed system itself.

Quantitative and comparative analyses are needed to verify the hypotheses. In this paper, we set two types of conversations in order to consider the wide variety of situation using remote conversations. However, we were not able to discuss the difference between them. We are now planning to collect more data with different conditions. The wide variety of participants is needed to confirm the effectiveness of the system. We need to have detail analyses of each function, screen rotation and showing gesture, through the comparison with the normal video chat conversation.

We also used the humanlike avatar to express the nonverbal behaviors of the remote interlocutors. The avatar enables

to exaggerate the gestures and facial expression of remote interlocutors in order to assist both showing and signaling toward the interlocutors. To improve the proposed system, we are planning to create the nonverbal behavior amplifier which the specific behaviors of the remote interlocutor will be detected automatically and emphasized by the avatar's over-action. It is also necessary to shift command transmission to the rotating projector to packet communication via the network. The proposed system would be also applicable to the remote, interactive services. The value created through such usages in the service setting will be an interesting research issue to be investigated.

ACKNOWLEDGMENT

This study is partly supported by JSPS KAKENHI Grant Number JP19H04416.

REFERENCES

[1] E. Veinott, J. Olson, G. Olson, and X. Fu, "Video helps remote work: Speakers who need to negotiate common ground benefit from seeing each other," in *Proceedings of the Conference on Computer Human Interaction. CHI'99, ACM Press, PA, USA*, 1999, pp. 302–309.

[2] J. Sun and H. Regenbrecht, "Implementing three-party desktop video-conferencing," in *Proceedings of the 19th Australasian conference on Computer-Human Interaction: Entertaining User Interfaces*, 2007, pp. 95–102.

[3] A. Pentland, "Socially aware, computation and communication," *Computer*, vol. 38, no. 3, pp. 33–40, 2005.

[4] D. Gatica-Perez, "Automatic nonverbal analysis of social interaction in small groups: A review," *Image and Vision Computing*, vol. 27, no. 12, pp. 1775–1787, 2009.

[5] A. Vinciarelli, M. Pantic, and H. Bourlard, "Social signal processing: Survey of an emerging domain," *Image and vision computing*, vol. 27, no. 12, pp. 1743–1759, 2009.

[6] A. Kendon, "Some functions of gaze-direction in social interaction," *Acta Psychologica*, vol. 26, pp. 22–63, 1967.

[7] A. Mehrabian and M. Wiener, "Decoding of inconsistent communications," *Journal of Personality and Social Psychology*, vol. 6, no. 1, pp. 109–114, 1967.

[8] A. Mehrabian and S. R. Ferris, "Inference of attitudes from nonverbal communication in two channels," *Journal of Consulting Psychology*, vol. 31, no. 3, pp. 248–252, 1967.

[9] H. H. Clark, *Using Language*. Cambridge: Cambridge University Press, 1996.

[10] A. Kalma, "Gazing in triads: A powerful signal in floor apportionment," *British Journal of Social Psychology*, vol. 31, no. 1, pp. 21–39, 1992.

[11] G. H. Lerner, "Selecting next speaker: The context-sensitive operation of a context-free organization," *Language in Society*, vol. 32, no. 02, pp. 177–201, 2003.

[12] G. W. Beattie, "Floor apportionment and gaze in conversational dyads," *British Journal of Social and Clinical Psychology*, vol. 17, no. 1, pp. 7–15, 1978.

[13] R. Vertegaal, R. Slagter, G. Veer, and A. Nijholt, "Eye gaze patterns in conversations: there is more to conversational agents than meets the eyes," in *CHI '01 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2001, pp. 301–308.

[14] K. Jokinen, H. Furukawa, M. Nishida, and S. Yamamoto, "Gaze and turn-taking behavior in casual conversational interactions," *ACM Transactions on Interactive Intelligence Systems*, vol. 3, no. 2, pp. 12:1–30, 2013.

[15] J. Short, E. Williams, and B. Christie, *The social psychology of telecommunications*. Toronto; London; New York: Wiley, 1976.

[16] W. W. Gaver, "The affordances of media spaces for collaboration," in *Proceedings of the 1992 ACM conference on Computer-supported cooperative work*, 1992, pp. 17–24.

[17] N. Yankelovich, N. Simpson, J. Kaplan, and J. Provino, "Porta-person: Telepresence for the connected conference room," in *CHI'07 extended abstracts on Human factors in computing systems*, 2007, pp. 2789–2794.

[18] M. Shikida and A. Ratikan, "Assisting gaze awareness by using faint lights during a video conference between an unbalanced number of participants (in japanese)," *Journal of Information Processing Society of Japan*, vol. 58, no. 1, pp. 166–175, 2017.