Gaze and Conversation Dominance in Multiparty Interaction

Yuki Fukuhara
Graduate School of Science and Technology
Seikei University
Musashino-shi, Tokyo 180-8633 Japan
dm106221@cc.seikei.ac.jp

Yukiko Nakano
Dept. of Computer and Information Science
Seikei University
Musashino-shi, Tokyo 180-8633 Japan
y.nakano@st.seikei.ac.jp

ABSTRACT
With the goal of designing conversational agents that can join and manage conversations with multiple participants, in this paper, we conduct an experiment to collect multiparty conversations with a virtual agent, and recognize head direction as each participant's focus of attention. Then, we analyze how gaze and mutual gaze affect floor management and conversation dominance; we assumed that the most dominant participant may control the participation framework and lead the conversation to make a decision. Based on the analysis, we found that turn-releasing success ratio is different depending on the levels of conversation dominance. We also found that the frequency of gaze and mutual gaze are different depending on the participation roles and conversation dominance. These results suggest that gaze behaviors are strongly related to conversation dominance, and may become a good predictor of dominance in multiparty conversations.

Author Keywords
Empirical study, eye-gaze behavior, conversation dominance, Wizard-of-Oz experiment.

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION
In public information kiosk systems which are used in museums, shopping malls, and entertainments, it usually happens that a group of people come to the system and operate it while talking with each other. Therefore, to build conversational agents that can serve as public information kiosk, such systems are required to be capable of managing multiparty conversations. However, little has been studied about interaction between a group of users and an agent. Thus, for the purpose of building conversational agents that can use bodily expressions, computational models of nonverbal behaviors in multiparty conversations are necessary.

2nd Workshop on Eye Gaze in Intelligent Human Machine Interaction.
February 13, 2011, Palo Alto, California, USA.

Figure 1. Three people joining a multiparty conversation

Figure 1 shows a snapshot of a multiparty conversation by three people. Based on the face direction of each participant, we can assume that person A and person B are talking with each other, and person C is not a speaker or a hearer. If this interaction continues, it may happen that the person C cannot get enough chance to contribute to the conversation. It may also happen that one person leads and dominates the conversation, and other participants follow the leader.

These typical phenomena in multiparty conversations may be useful in managing multiparty conversation by a system or an agent. For example, if the system needs to finish the conversation, it would be effective to persuade the conversational leader to make a decision and finish this conversation as soon as possible. On the contrary, to help an inferior person get a chance to tell her/his opinion, the agent may give a turn to her/him.

With the goal of designing conversational agents that can join and manage conversations with multiple participants, in this paper, first, we conduct an experiment to collect multiparty conversations with a virtual agent, analyze head direction as each participant's focus of attention, and discuss how gaze and mutual gaze relate to floor management and conversation dominance.

RELATED WORK

Goffman [1] discussed the concept of participation in conversations, and based on his discussion, Clark [2] distinguished participants from overhearers. In his definition, participant includes "speaker" and "addressee (hearer)" as well as others taking part in the conversation but not currently being addressed. The non-addressed participant is "side participant". All other listeners are "overhearers", who have no rights or responsibilities in a given conversation. The role of each conversational participant is automatically assigned once a speaker starts speaking. At this moment, the addressee is chosen by the speaker, and the rest of the participants become side participants.

Previous studies showed that the participant's gaze behaviors are different depending on her/his role. People gaze more while listening (75%) than while speaking (41%) in dyad conversations [3]. Analyzing multiparty conversations, [4] reported similar results that about 1.6 times more gaze occurred while listening than while speaking. However, when a person starts speaking to all other participants, her/his gaze is distributed to all of them. In this condition, the total percentage of gaze (while speaking) rises to 59% of the time. Then, [4] applied these findings to a design of mediated communication.

The relationship between dominance and conversational behaviors was studied in [5], and they found that in dyad conversations, less dominant person is more likely to be the first to break the mutual gaze. The higher status person gaze roughly the same percentage of the time while listening and speaking, whereas the lower status person gazes relatively more while listening [6]. Moreover, dominant person who is in higher status and has more power initiates speech more and controls the whole interaction [7]. However, the notion of dominance in these studies indicates a stable social relationship, and not exactly the same as conversation dominance focused in this study where there is no power or status difference among the participants and the dominance is temporally evoked to perform the experimental task.

In studies of multiparty conversations with conversational agents, interactions between one user and multiple agents were mainly studied. [8] developed a conversational presentation system that generates conversations between two agents that answer the user's questions. [9] proposed a multiparty conversational system that can manage negotiation conversation among three parties: two virtual agents and a user.

Nagao et al [10] proposed a plan-based multiparty conversation system for two users. More recently, Bohus et al [11] proposed an open-world dialogue system where multiple people with different and varying intentions enter and leave, and communicate and coordinate with each other and with interactive systems. They also proposed a method to predict user's intention of engagement in conversation in open-world context [12]. Huang et al [13] proposed a quiz agent where a group of users join the game, and the system estimates whether the users are excited about talking with the agent. However, the previous work has not exploited gaze information to interpret multiparty participation framework, and not applied a multiparty gaze model to conversational agents.

Figure 2. Animation agent used in the experiment

EXPERIMENT

We conducted an experiment to collect multiparty conversations with a virtual agent in a Wizard-of-Oz setting, and investigate how participation framework is dynamically changed, and what types of verbal/nonverbal behaviors may be useful as a predictor of conversation dominance.

Task

A group of three people participated in the experiment as a subject group. The task of the subject group was to discuss and decide where they would go out to enjoy a weekend. To collect information about the visiting places, they could ask questions to the agent displayed in front of them. The agent answered the questions, and also recommended some places to visit. The picture of the agent is shown in Figure 2. To experimentally control the participation attitude towards the conversation, in each session, one of the subjects played a role of person who cannot go out that weekend. We expected that this person would participate in the conversation less dominantly and positively than the others, and has more possibility of becoming a side participant.

Moreover, the other two subjects were assigned different goals in selecting the places. For example, one subject was instructed to go to a shopping place, and the other was instructed to go to a scenery place. Thus, their task was to discuss and choose two places that satisfy each member's requirement as much as possible. The roles of the subjects were changed every session. The subject who could not go out was also different for each session. To motivate the subjects, we instructed them that they can get more rewards if they chose good places that satisfied all the requirements.

Procedure

We actually had another experimental condition that four people had a conversation. So, in experimental procedure, a group of subjects had 4 conversations for each condition

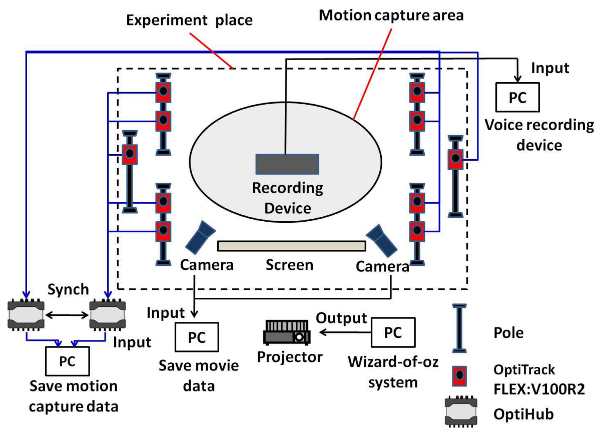


Figure 3. Experimental setting

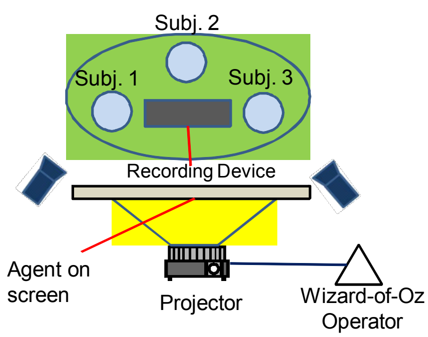


Figure 4. Proxemics of the subjects

and 8 conversations in all. The experimental contents are different in all the conversations. So, we had 8 kinds of experimental contents, and they were randomly assigned to each condition. Note that we only analyzed the corpus for three party conversations in the rest of this paper.

Experimental Setting

Figure 3 shows the setting of the equipments used in this experiment. We used 10 cameras for OptiTrack motion capture system that were mounted on 6 poles. The animated agent was displayed on a 100 inch screen in front of the subjects. When a subject asked a question to the agent, an experimenter (wizard) understood that question and chose one of the answer menus or typed in some texts to answer the question. Then, the spoken response was produced through TTS. The agent did not display any facial expressions or gestures, but her lip movement was synchronized with speech. The interaction was video-taped using two Hi-vision cameras. Each subject wore a sweater with markers for motion capturing (see Figure 1). So, the motion capture system measured upper body motions for each subject. Speech sound for each subject was recorded through individual microphones. The proxemics of subjects and coordination of audio and video equipments are shown in Figure 4. As shown in the figure, the subjects were instructed to stand around the audio device in front of the screen, so that they can see each other and the agent.

Data

From the experiment described above, we collected the following data;

- Video data shot from right front and left front of the subjects
- Speech sound for each subject
- Upper body motion data measured by a motion capture system
- Score for conversation dominance. After the experiment, we showed the video data to 5 people who did not join the experiment, and asked to rate each subject in terms of conversation dominance by asking like “Who is leading the conversation”, or “who is the leader of the conversation?” Then, we calculated the average of rating scores, and used the values as the conversation dominance score for each subject.

ANNOTATION

Participation role

In multiparty conversations, to specify the participation framework, it is necessary to specify the hearer as well as the side participant. Therefore, to analyze the structure of multiparty interaction, and relationship between the subjects and their participation roles, we annotated the participation framework for each utterance.

First, we identified the time period of each utterance using Praat, and read the Praat data into Anvil annotation tool [14] to annotate utterances. Then, looking at the video, we annotated the participation roles: speaker, hearer, and side participant. When the speaker talked to the other two subjects, both of them were annotated as hearer. In addition to the participation roles, we annotated the utterance type: speak to someone, respond to someone, and soliloquy. In this paper, since we focus on human multiparty conversations, we did not analyze interaction between the user(s) and the agent.

In addition to the utterance annotation, we identified and annotated turn boundaries. A sequence of utterances whose speaker was identical and a pause between the utterances was less than 2 seconds was identified as one turn.

Automatic Annotation of Gaze

Previous studies revealed that gaze distribution is deeply related to turn taking/releasing, and conversational coordination [15, 16]. Therefore, we think that gaze data is indispensable in analyzing multiparty conversations. However, manually annotating gaze by looking at video is very time consuming. Thus, in this study, we estimated the gaze direction from the head motion data obtained from a motion capture. Decision tree algorithm implemented in Weka (J48) was applied to the head motion data: x, y, z

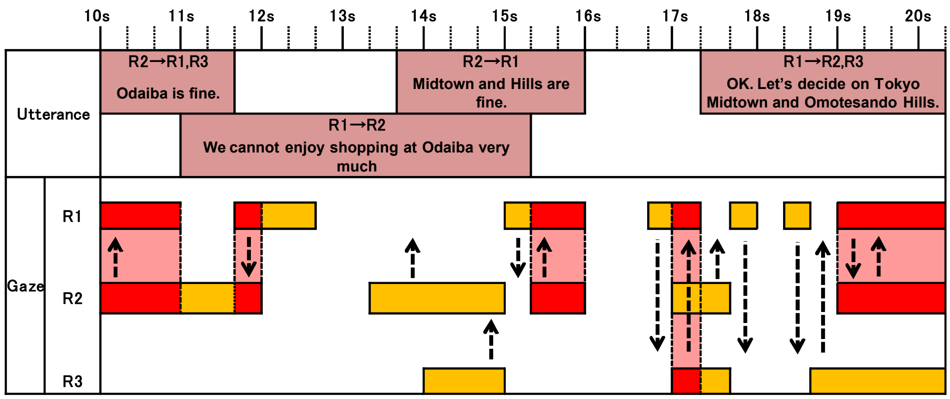


Figure 5: Example of interaction

position and rotation. We prepared training data by manually annotating gaze behaviors for each subject for 5 sessions. The decision tree judges whether the subject looks at the subject standing at the left position, the right position, or the center position towards the screen. By employing the leave-one-out method, four sessions were used for training and one for testing. We examined all the combinations of training and test sets. The averages of precision, recall, and F-measure are shown in Table 1. Since the accuracy of the decision tree was sufficient, we decided to use this model to automatically annotate subject's gaze directions.

Table 1. Accuracy of estimating gaze direction for each subject

Subject (Position)	Precision	Recall	F-Measure
Subj.1 (right)	0.872	0.914	0.893
Subj.2 (center)	0.927	0.922	0.925
Subj.3 (left)	0.954	0.95	0.952

ANALYSIS

In this section, we analyze the participant's behaviors with respect to the order of dominance in conversation. The order was determined based on the score of the conversation domination, which was described in the previous section. We call a conversational participant who got the highest score "rank 1 participant (R1)", a participant who got the second highest score "rank 2 participant (R2)", and a participant who got the lowest score "rank 3 participant (R3)". The averages of conversation domination scores are 1.16, 2.08, 2.75, for R1, R2, R3 respectively. The scores indicate clear difference between ranks. To check the validity of experimental control, we calculated how many people playing as a person who cannot go out were actually ranked as R3. As a result, in 5 out of 8 sessions, subjects who cannot go out were ranked as R3. For the rest of three sessions, they were ranked as R2. Subjects who cannot go out have never been ranked as R1. Thus, we believe that the experimental instruction successfully controlled the subjects' interaction.

Example of eye gaze exchanges in collected corpus

An example of multiparty interaction is shown in Figure 5. The upper track shows subject's utterances, and the lower track shows gaze behaviors for R1, R2, and R3 subject respectively. For instance, from 10s to 11s, R1 and R2 established mutual gaze (R1 looks at R2, and R2 looks at R1) while R2 was speaking. At 13.3s to 15s, R2 was looking at and talking to R1. At the same time (14s to 15s), R3 was looking at R2 as the speaker, but did not get R2's attention. Overall, it seems that R1 get the highest attention among three participants, and mutual gaze was more frequently established between R1 and R2, while R3 had less chance to establish mutual gaze. In the rest of this paper, we will more closely investigate such gaze behaviors.

Distribution of participation role

First, we investigated whether the distribution of participation role was different depending on the order of

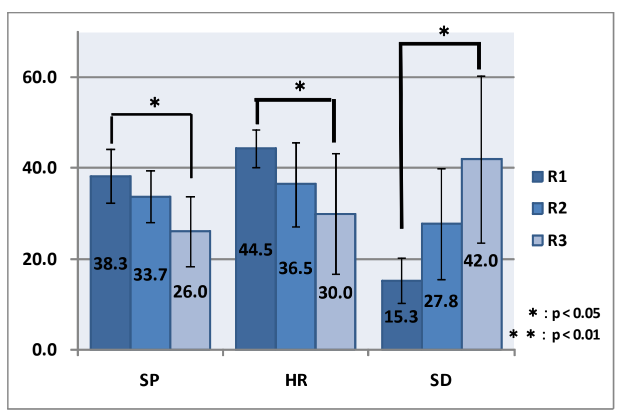


Figure 6. Distribution of participation role

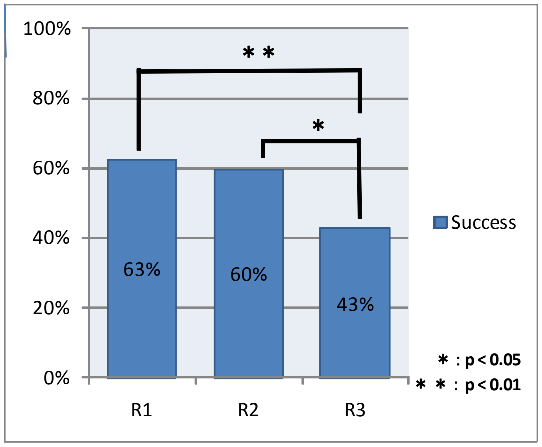


Figure 7. Turn-releasing success ratio

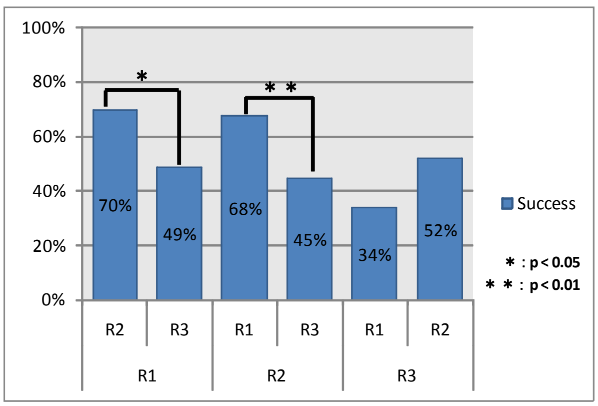
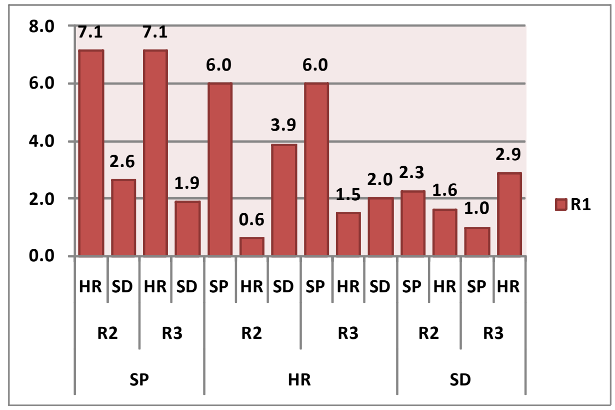


Figure 8. Turn-releasing success ratio WRT the addressee of the gaze signal

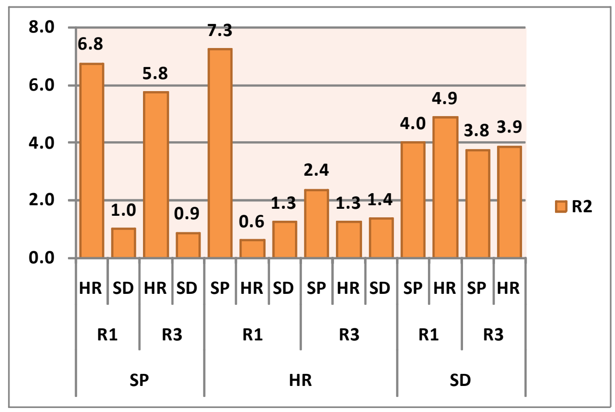
dominance in conversation. Figure 6 shows the average number of times for each participation role. R1 more frequently served as the speaker (SP) then the other two, and less frequently served as the side participant (SD). As the result of examining the difference of average by ANOVA, the average difference is statistically significant for all the participation roles (SP: $F(2,21)=6.338$ $p < .001$, Scheffe's post hoc test; R1 vs R3 $p < .05$, HR: $F(2,21)=4.481$ $p < .005$, post hoc test; R1 vs R3 $p < .05$, SD: $F(2,21)=7.302$ $p < .001$, post hoc test; R1 vs R3 $p < .05$) In the following sections, we investigate whether such difference in the distribution of participation roles is related to eye gaze.

Choosing the next speaker using gaze

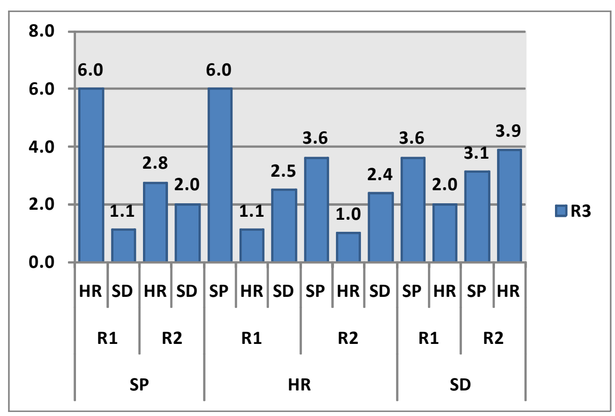
In previous studies [15, 16], it was found that the speaker chooses the next speaker using gaze. The next speaker, who receives the turn-releasing gaze signal by establishing a mutual gaze with the current speaker, looks away for a very short period of time, and then starts speaking. This is a typical gaze exchange for turn taking. We investigated whether this turn taking process is different depending on the dominance of a participant in a given conversation. We



(a)



(b)



(c)

Figure 9. Frequency of gaze

calculated the turn-releasing success ratio for each participant. For each utterance, we identified who was gazed at by the speaker for the last one second of the utterance. If the person who received the turn-releasing signal actually started speaking within 2 seconds, we counted this as the success of choosing the next speaker.

Figure 7 shows the turn-releasing success ratio with respect to the order of conversation dominance. The success ratio for rank 1 (R1) and rank 2 (R2) participants is over 60%. On the contrary, that for rank 3 (R3) participant is about 40%. As the result of statistical test for difference of ratio, the turn-releasing success ratio is different depending on the order of conversational dominance (χ^2=10.7785, p<0.0045). As the result of Post-hoc test using Ryan method, the difference between R1 and R3, and between R2 and R3 were statistically significant. These results suggest that R3 was less successful in choosing the next speaker using gaze.

Moreover, we analyzed whether the turn-releasing success ratio is different depending on who releases the turn to whom. As shown in Figure 8, the turn-releasing success ratios between R1 and R2 (R1 gives a turn to R2, and R2 gives a turn to R1) was about 70%. On the contrary, in releasing a turn to R3 or receiving a turn from R3, the success ratio became much lower than 30%. Therefore, these results suggest that the turn taking was quite successful between R1 and R2, but not for R3.

Getting attention from others

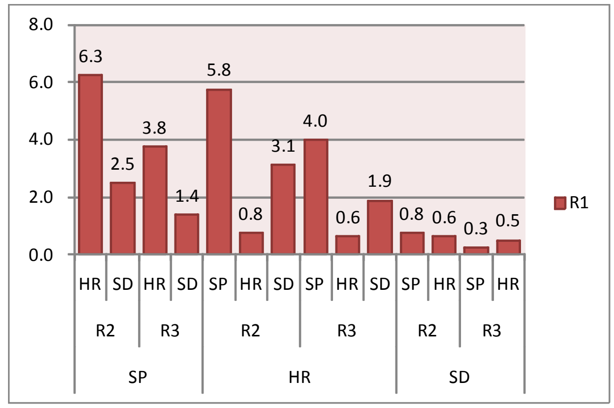
Even when the participants are not speaking, they are looking at someone. So, we analyzed participants gaze with respect to the participation role and the conversation dominance. Figure 9 (a), (b), (c) show the average number of eye gaze per conversation. When gaze shifts to other participants or elsewhere, one unit of gaze is completed. In the current analysis, we did not consider the duration of eye gaze in counting gaze behaviors. As shown in Figure 9 (a), when R1 was the speaker, the frequency of looking at R2 as the hearer and that of looking at R3 were the same (7.1). Likewise, in Figure 9 (b), when R2 was the speaker, R2 almost equally looked at R1 (6.8) and R3 (5.8). On the contrary, in Figure 9 (c), when R3 was the speaker, R3 more frequently looked at R1 (6.0) than R2 (2.8) as the hearer. This suggests that R3 more respected R1 than R2 as the hearer.

In analyzing the gaze behaviors as the hearer, when R1 was the hearer, s/he equally looked at R2 (6.0) and R3 (6.0) as the speaker. However, when R2 was the hearer, s/he more respected R1 (7.3) than R3 (2.4) as the speaker. Similarly, when R3 was the hearer, s/he more respected R1 (6.0) than R2 (3.6) as the speaker. We did not find clear trends for the distribution of gaze behaviors for side participants.

These results suggest that R1 got more attention from other participants as the speaker and as the hearer. On the other hand, while R1 equally looked at R2 and R3, gaze behaviors of R2 and R3 were not balanced (more looking at R1).

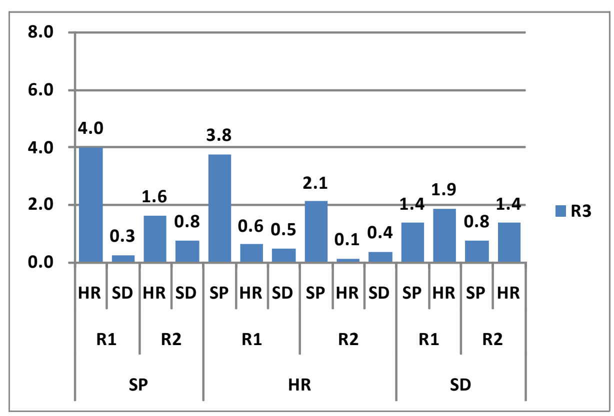
Mutual gaze

In addition to gaze, mutual gaze is also very important in investigating the interaction. Thus, we automatically identified mutual gaze as the gaze overlap between two participants. Figure 10 (a), (b), (c) show the average number of mutual gaze for each participant per conversation. As shown in Figure 10 (a), when R1 was the speaker, mutual gaze was more frequently established with R2 as the hearer (6.3) than with R3 (3.8). In Figure 10 (b), when R2 was the speaker, mutual gaze was more frequently



(a)

(b)



(c)

Figure 10. Frequency of mutual gaze

established with R1 as the hearer (5.8) than with R3 (2.1). When R3 was the speaker, mutual gaze was more frequently established with R1 as the hearer (4.0) than with R2 (1.6). These results suggest that when R1 or R2 is the speaker, mutual gaze is more frequently established between them (not with R3). It is interesting that even though R1 and R2 equally look at other participants (shown in Figure 9), mutual gaze is more frequently exchanged between R1 and R2.

From the hearer's perspective, it is obvious that the hearer mainly looks at the speaker, but the frequency of mutual gaze is different depending on the conversation dominance. When R1 was the hearer, mutual gaze was more frequently established with R2 (5.8) than with R3 (4.0) as the speaker. When R2 was the hearer, mutual gaze was more frequently established with R1 (6.3) than with R3 (1.6) as the speaker. When R3 was the hearer, mutual gaze was more frequently established with R1 (3.8) than with R2 (2.1) as the speaker. Again, these results suggest that when R1 or R2 is the hearer, they spend less attention to R3 as the speaker even though R3 respects R1 as the speaker. More interestingly, when R1 was the hearer, the frequency of mutual gaze with R2 as the side participant (3.1) was close to that with R3 as the speaker (4.0). This suggests that R1 respects R2 even if R2 is the side participant.

In the analysis of side participant, since R1 rarely became the side participant, the frequencies were very low for all the cases. An interesting finding is that when R2 was the side participant, mutual gaze with R1 was more frequent than with R3 regardless of the participation roles.

These results suggest that both participation roles and conversation dominance affect the participants' gaze behaviors and mutual gaze between the participants. It was found that R1 got highest respect from other participants, and R2 was respected by R1. However, R3 did not get attention from other participants.

CONCLUSION AND FUTURE WORK

This paper presented our empirical study of multiparty conversations. By analyzing turn taking, gaze, and mutual gaze, we found that these nonverbal behaviors were affected by participation roles and conversation dominance. As the next step, we are aiming at estimating the participation framework based on gaze and mutual gaze behaviors, and then estimating the conversation domination.

We admit that there are many other factors that affect the dominance of conversation, such as personal relationship, and personality of each participant. Our research goal is to recognize what is happening in a given conversation, but not estimating static social relationship or personality. In our experiment, we control the relationship between the subjects by assigning a role to each subject.

Although this study focused on human-human interaction, in future work, we will analyze gaze and mutual gaze in human-agent multiparty conversation, and compare to the results found in this study.

ACKNOWLEDGMENTS

This work is partially funded by JSPS under a Grant-in-Aid for Scientific Research in Priority Areas "i-explosion" (21013042), and MEXT Grant-in-Aid for Building Strategic Research Infrastructures.

REFERENCES

1. Goffman, E., *Forms of Talk*. 1981, Philadelphia, PA: University of Pennsylvania Press.
2. Clark, H.H., *Using Language*. 1996, Cambridge: Cambridge University Press.
3. Argyle, M. and M. Cook, *Gaze and Mutual Gaze*. 1976, Cambridge: Cambridge University Press.
4. Vertegaal, R. *The GAZE Groupware System: Mediating Joint Attention in Multiparty Communication and Collaboration*. in *CHI 1999*. 1999.
5. Dovidio, J.F. and S.L. Ellyson, *Patterns of visual dominance behavior in humans*, in *Power, dominance, and nonverbal behavior*, S.L. Ellyson and J.F. Dovidio, Editors. 1985, New York: Springer-Verlag. p. 129-149.
6. Knapp, M.L. and J.A. Hall, *Nonverbl Communication in Human Interaction*. 2010: Wadsworth.
7. Burgoon, J.K. and N.E. Dunbar, *Nonverbal expressions of dominance and power in human relationships*, in *The Sage handbook of nonverbal communication*, V. Manusov and M.L. Patterson, Editors. 2006, Sage publications. p. 279-297.
8. Eichner, T., et al. *Attentive Presentation Agents*. in *The 7th International Conference on Intelligent Virtual Agents (IVA)*. 2007.
9. Traum, D., et al. *Multi-party, Multi-issue, Multi-strategy Negotiation for Multi-modal Virtual Agents*. in *the 8th International Conference on Intelligent Virtual Agents (IVA08)*. 2008.
10. Nagao, K. and A. Takeuchi. *Social Interaction: Multimodal Conversation with Social Agents*. in *Twelfth National Conference on Artificial Intelligence*. 1994. Menlo Park, CA: AAAI Press.
11. Bohus, D. and E. Horvitz. *Open-World Dialog: Challenges, Directions, and Prototype*. in *IJCAI2009 Workshop on Knowledge and Reasoning in Practical Dialogue Systems*. 2009.
12. Bohus, D. and E. Horvitz. *Learning to Predict Engagement with a Spoken Dialog System in Open-World Settings*. in *SIGdial'09*. 2009. London, UK.
13. Huang, H.-H., et al. *How multiple concurrent users react to a quiz agent attentive to the dynamics of their game participation*. in *AAMAS*. 2010.
14. Kipp, M. *Anvil - A Generic Annotation Tool for Multimodal Dialogue*. in *the 7th European Conference on Speech Communication and Technology*. 2001.

15. Duncan, S., Some signals and rules for taking speaking turns in conversations. Journal of Personality and Social Psychology, 1972. 23(2): p. 283-292.

16. Kendon, A., Some Functions of Gaze Direction in Social Interaction. Acta Psychologica, 1967. 26: p. 22-63.