Coordinating Turn-Taking and Talking in Multi-Party Conversations by Controlling a Robot's Eye-Gaze

Ryo Sato[1] Yugo Takeuchi[1]

[1] Graduate School of Informatics, Shizuoka University, Japan

Abstract: This study realizes smooth turn-taking and discussion in multi-party conversations by designing the physical behavior of a robot. In this paper, humans estimated a robot's utterances by looking at the direction of its eye-gazes. In addition, we investigated how the robot's behavior as a listener who expressed his eye-gazes influenced the next speaker's comments and context. Humans adjusted their utterances by looking at the eye-gaze of the robot as a listener. We believe our study will effectively produce desirable discussions in multi-party conversations.

1 Introduction

As more and more communication robots are being developed, they must be adapted to the face-to-face dialogs of humans to realize communication between humans and robots and humans through robots. Since face-to-face dialogs are the most natural communication form for humans, it is difficult to realize smooth turn-taking between humans and robots [1]. Many interactive robots have been developed that address this problem. For example, ROBITA can deal with human utterances based on eye-gazes to understand them [1]. This robot presumes suitable timing to start his utterances to naturally participate in conversations.

These robots adaptively participate in conversations by passively presuming interactive environments. But their development isn 't based on the assumption of aggressively coordinating conversations. Multi-party conversations may deteriorate without following implicit turn-taking rules. Furthermore, for decision-making or consensus building, dialog directionality can be lost, and the dialog 's quality can decline if opinions based on various viewpoints are not exchanged. One method to solve these problems is to endow chairpersons with the authority and the ability to determine the next speaker and to restrict what that speaker can say. The existence of a chairperson with such ability is important to realize smooth multi-party conversations; but no such robots have been developed.

In human face-to-face communication, much verbal/nonverbal information is conveyed among participants. The eye-gazes of others are used to guess to whom utterances are being addressed. Furthermore, the eye-gazes of others convey intentions about the next speaker when others turns his eyes to another person [2]. We can identify the speaker and influence what the speaker will say by designing eye-gaze for communication robots.

In this study, we investigate whether designing eye-gazes for such robots can coordinate turn-taking and talking in multi-party conversations. Our study is expected to effectively encourage desirable dialogs in such multi-party conversations as collaborative learning scenes. We conducted two experiments. One investigated how humans estimate the robot's utterances by watching its eye-gazes. The other investigated whether a robot 's eye-gaze as a listener can provide a context for the next speaker's utterances.

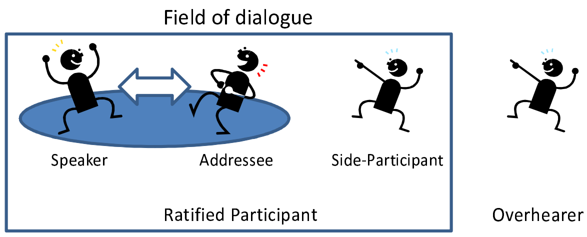

Figure 1: Participation roles

2 Participation-Role and Turn-Taking

Generally, dialog participants are roughly divided into two types: speakers and listeners. For one-on-one dialogs, the listener automatically becomes the next speaker, and the current speaker becomes the listener. But for multi-party conversations, the next speaker is not clear because there are more than two listener candidates.

In multi-party conversations, the participants are roughly divided into two types, ratified participants and overhearers; ratified participants are further divided into three types: speakers, addressees, and side-participants [3][4] (Fig. 1). Ratified participants are recognized as the conversation participants by other participants. In this study, we defined a dialog field as the place where participants interact with verbal/nonverbal information about each other to understand the intentions and relationships with others. This field consists of ratified participants.

Turn-taking means that the current speaker becomes the listener, who then becomes the next speaker. We usually talk by repeating turn-taking, which consists of four kinds of phenomena: continuation, overlap, silence, and change. They happen due to the utterances/non-utterances of speakers and listeners [5]. Generally, two types of phenomena, such as silence and overlap, do not happen without deciding the order of utterances in advance [6]. Participation-role-taking happen during the timing of turn-taking, and such effects of nonverbal information as eye-gaze are so great that turn-taking is adjusted [7][8][9]. The eye-gazes of others are used to infer to whom the utterances are being addressed and can convey the intention to turn to face someone [2]. We consider the relations between eye-gaze and turn-taking in Section 3.

3 Eye-Gaze and Turn-Taking

In face-to-face communication, humans share the same time and space and can get not just verbal information but also nonverbal information through visual, auditory, and touch senses. With such nonverbal information as position, utterance timing, eye-gazes, and gestures, humans can convey intention, emotion, and address utterances to each other. Conveying nonverbal information plays an important role to realize smooth conversations. In contrast, conversations through media might suffer from insufficient nonverbal information, and the available channels are more limited than in face-to-face communication. Interaction is easily fragmented. Since these nonverbal information effects are expected to be adapted for human-robot interaction, previous study has tried to realize smooth conversations by designing such information for robots [1].

3.1 Eye-Gaze and Joint Attention

In communication with others, others' eye-gaze is an important key to infer their visual attention. Humans usually turn their eyes and concentrate to get visual information of their targets. So humans can guess the interests of others by focusing their own attention on the attention of others. Relating each other's visual action with a mental state by focusing on the attention of others is called joint attention [10]. We can realize a conversation based on intuitive recognition by communicating information that corresponds to each other's environments. We can also realize instructions through sympathy by joint attention with others. Therefore, implementing joint attention with humans is an important assignment for interactive robots. Previous study has tried to realize joint attention between humans and robots by implementing eye-gaze for robots [11].

3.2 Eye-Gaze and Addressing

Eye-gaze can reveal not only the interests of others but it can also address utterances. Kendon argued that turn-taking is realized when a speaker closely observes the listener for a turn-taking signal and a listener closely observes the speaker as acceptance of this signal [8]. Enomoto argued that dialogs consist of three persons and investigated whether nonverbal information functions as a selection method for the next speaker in multi-party dialogs. A participant who has turned his eyes to the speaker becomes the next speaker when all participants turn their eyes to the speaker [12].

Other than a function that starts the conversation, eye-gaze also includes the following three functions [8]:

- Monitoring: speaker checks whether she should continue to talk by looking at the listeners' eye-gazes.
- Regulation: speaker adjusts her opinion by checking for favorable impressions from listeners by looking at their eye-gazes.
- Expressive: listener gives the speaker her impressions of the speaker's comments.

In addition, the speaker often selects the next speaker by checking his interest in the conversation by turning his own eyes. Listeners can predict the next desirable speaker by looking at the speaker's eye-gaze.

Such eye-gaze effects are not limited to human-human communication. In human-robot interaction, humans can guess the intention of a robot's utterances by looking at the robot's eye-gaze even if it is artificial. The robot's eye-gaze can select the next speaker from among the participants by planning his eye-gaze. We revitalize conversations by giving all members a fair chance to join.

Until now in this paper, we have considered how robot's eye-gaze can realize smooth turn-taking in multi-party conversations. Next in Section 4 we experimentally investigated whether our hypothesis is correct. Our experiment 's aim is to reveal how humans judge whether they are the target of the robot's utterances by looking at the robot's eye-gaze.

4 Experiment 1

4.1 Participants

All 15 Japanese participants were informatics students at Shizuoka University.

4.2 Experimental Environment

For meeting the robot, a participant and an experimenter were separated by a partition to prevent the participant from judging the target of the robot's utterances without information gleaned by looking at

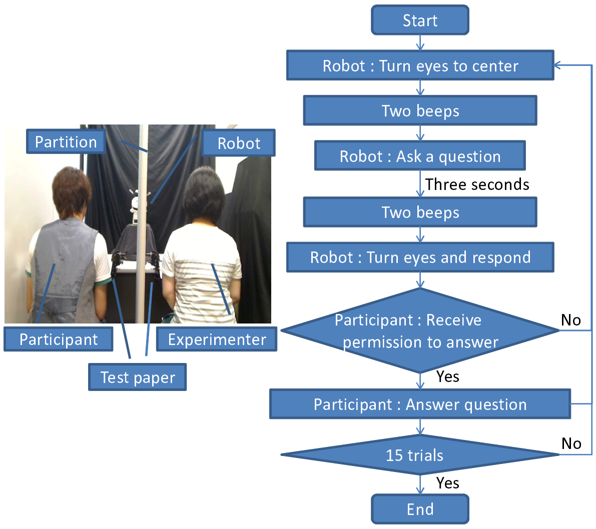

Figure 2: Experimental environment and procedure

the robot's behavior. The participant and experimenter positions were randomly changed every participant. The participants were told that the experimenter would also take part in this experiment. Identical test papers, on which five numbers were written in different colors, were put in front of them. Our robot was a cow-puppet placed over a dome-shaped camera. Its eye-gaze was expressed by rotating the camera inside of the puppet.

4.3 Subject

The participant and the experimenter had to solve calculation problems, which were numbers written on a piece of paper as operands, to determine whether they received permission to answer based on watching the robot's behavior. For example, the robot said colors as numbers: "What does red plus yellow equal? "

We conducted this experiment in the following procedures (Fig. procedure1). First, the robot turns his eyes to the space between the participant and the experimenter and asks a question after two beeps. Participant and experimenter are given three seconds for calculation. Three seconds later, two beeps sounded again to signal the end of the calculation time, and the robot turns his eyes in the direction shown in the experimental condition (Table 4.4) and says " yes. " By watching the robot's behavior, participants determined whether they were given permission to answer. In contrast, the experimenter explained in the participant's presence that he should answer using fingers without talking so that the participants can judge without observing whether the experimenter answered.

Table 1: Experimental conditions

		Gaze before utterances				
		Self	Self-Others	Others-Self	Others	Center
Gaze with utterances	Self	S-S	SO-S	OS-S	O-S	C-S
	Others	S-O	SO-O	OS-O	O-O	C-O
	Center	S-C	SO-C	OS-C	O-C	C-C

4.4 The Experimental Condition

This experiment was conducted in a two-way within-subjects design (Table 4.4). The first factor, called the " gaze before utterances " and denotes the direction in which the robot turns his eyes before making his utterance, consists of five levels: self, others-self, self-others, others, and center. On each level, the robot turns his eyes in the following five directions before making his utterance:

- Self: only turns to participant
- Others-self: turns to experimenter and then to participant
- Self-others: turns to participant and then to experimenter
- Others: only turns to experimenter
- Center: looks between them

The second factor, which is called the " gaze with utterances "and denotes the direction in which the robot turns his eyes with his utterance, consists of three levels: self, others, and center. On each level, the robot turns his eyes in the following three directions with his utterance:

- Self: turns to participant
- Others: turns to experimenter
- Center: looks between them

Each condition was randomly assigned to every participant and 15 trials were performed (one trial for each condition).

4.5 Analysis

We observed the following two points.

- Did the participants answer?
- What was the interval time between the end of the robot's utterance and the start of the participant's answer when the participant answered?

4.6 Hypotheses and Prediction

The following are our hypotheses and predictions:

- First Hypothesis

  Humans make different judgments about being addressed by the robot's utterances by watching the direction in which the robot turns his eyes before it makes its utterances. The quicker the robot turns his eyes to the human than the other, the more the human will judge that the robot is talking to the human.

- First Prediction

  By different levels of the " gaze before utterance " factor, the tendency that the participants answer will be strong:

  - $self > self - others > others - self > center > others$

- Second Hypothesis

  Humans make different judgments about being addressed by the robot's utterances by watching the direction in which the robot turns his eyes and utters. The more the robot turns his eyes to the humans, the more they judge that the robot is talking to the humans. The more the robot turns his eyes to the others, the more the humans judge that the robot is talking to the others.

- Second Prediction

  By different levels of the " gaze with utterance " factor, the tendency where the participants answered the questions will be strong:

  - $self > center > others$

4.7 Results

First, we analyzed the number of times that the participants answered the " gaze before utterance " factor by a chi-square test. We observed no significant difference ($x^2_{(4)} = 1.353, n.s.$) We also analyzed the number of times that the participants answered the " gaze with utterance " factor by a chi-square test and found a significant difference ($x^2_{(2)} = 38.482, p < .01$). By assessing with Ryan's multiple comparison method, the participants who answered increased in the order of levels self, center, and others ($p < .05$) (Fig. 3).

Furthermore, we analyzed the interval time on the self level of the " gaze with utterance " factor by a one-way ANOVA. We analyzed the 12 participants who answered all the levels of the " gaze before utterance " factor and found a main effect ($F_{(4,44)} = 3.224, p < .05$). By assessing with the LSD multiple comparison method, the interval time on the others level was longer than the other levels ($p < .05$) (Fig. 4).

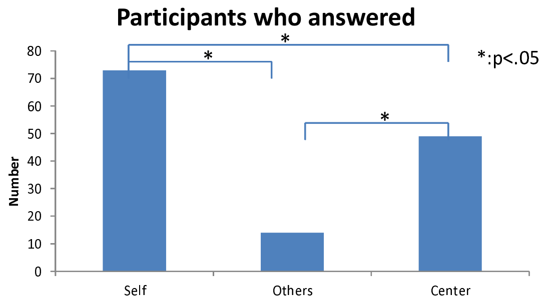

Figure 3: Number of participants who answered

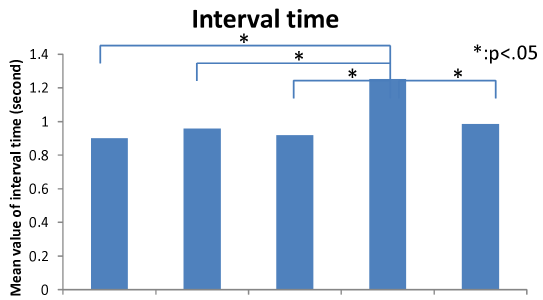

Figure 4: Interval time

4.8 Consideration

In Section 4.7, we found no significant difference for the number of participants who answered in the " gaze before utterance " factor. This result means that our first hypothesis was not supported. In contrast, we observed a main effect for the number of participants who answered in the " gaze with utterance " factor and found that they answered when they turned to the robot's eyes but they did not answer when the robot turned his eyes to the experimenter. This supports our second hypothesis. Based on the interval time result, the participants to whom the robot's eyes were turned spent more time when the robot only turned his eyes to the experimenter than in the other conditions. Even though the participants predicted that the robot would talk to others because it turned its eyes to the others before talking, the participants to whom the robot's eyes were turned talked. Participants probably had difficulty judging whether they had been given permission to answer. These results show that the robot's eye-gaze before utterances is not an important factor to determine whether humans utter, but it suggests that this might change the degree of difficulty for such judgments.

We found that humans judge whether they have the right to answer by looking for a robot's eye-gaze that simultaneously occurs with an utterance. But previous research has not focused on the utterance

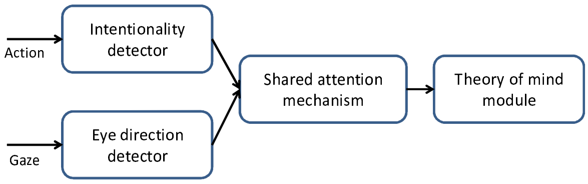

Figure 5: Intention understanding model

context of humans who have received the right to answer. On the other hand, others' eye-gazes cannot just address utterances but they can also suggest the intention that they should turn his eyes [2]. Speakers might be able to control the context of their utterances by looking the robot's eye-gaze. We consider the relation between eye-gaze and intention in Section 5.

5 Eye-gaze and Intention, Context

Human eye-gazes can tell others not only where they are looking and on what they are concentrating but also the intention underlying where they turned their eyes. Baron-Cohen found that eye-gaze is important to understand intention in a study on autism [13] (Fig. 5). Their model consists of four modules: an Intentionality Detector (ID), an Eye Direction Detector (EDD), a Shared Attention Mechanism (SAM), and a Theory of Mind Module (ToMM). ID estimates the inner states of others by looking at their face expressions. EDD estimates the direction of the eye-gazes of others. SAM detects the object in which others are interested by following their eyes. ToMM is a function for understanding the inner states of others by building a general model that integrates these three modules. Baron-Cohen concluded that humans infer the intentions of others by relating the object in their eyes and their focus of attention by these four modules [13].

Tomasello argued that joint attention includes the primitive awareness about others of infants who have the visual ability to pay attention to an object with their eyes. Their intention is different from infant intention [10]. According to Johnson, infants only monitor the eye-gazes of others when they regard the others as " intentional agents, " whose existence seems to have a relationship to the infants themselves [14]. In other words, some of the characteristics of intention intervene so that the person follows the eyes of others. The Baron-Cohen model, which is recognized widely in the domain of understanding eyes and mind reading, is expected to be applied to human-robot interactions. Some studies shared the intentions of humans and robots by trying that the robot follows their eyes [15].

Humans have also been found to have different understanding of the behaviors of others and different impressions due to the variations of the context of the behaviors of others. Humans understand such behaviors differently even if such behaviors are the same. This phenomenon is called the " Construct Accessibility Effect " or the " Priming Effect. " In addition, with nonverbal information, human expression influences the inferences of others about their inner states based on context information before and after nonverbal information such as eye-gaze; this is called the " Kuleshov Effect. "

Based on these researches, a robot's eye-gaze can reveal its intentions to others due to different understandings based on its before and after context information. Since humans who look at a robot's eyes might have different understandings, they can adjust their utterances. If these effects could be realized, we might be able to effectively encourage desirable dialogs in multi-party conversations. To realize such effects, we focused on the robot's behavior as a listener by eye-gaze. The robot's eye-gaze is not only a natural listener behavior but it can also express different degrees of interest [2]. The robot's eye-gaze can adjust the content of multi-party conversations. For example, if the robot agrees/disagrees with the speaker, the next speaker may agree/disagree with the previous speaker to satisfy the robot. In Section 6, we describe our experiment that investigated whether a robot's behavior as a listener who expresses eye-gaze influences the next speaker 's utterances.

6 Experiment 2

6.1 Participants

All 15 Japanese participants were informatics students at Shizuoka University.

6.2 Experimental Environment

A participant and an experimenter (a cow-puppet with an audio speaker) were arranged to meet our robot called robovie-mR2 [16]. Their positions are randomly changed every participant. A display unit behind the robot shows the conversation theme.

6.3 Subject

The participant and the experimenter were told to give advice about the theme to help the robot reach a decision. The participant gets points when his suggestions are chosen by the robot. The theme is called " The Desert Survival Task. " The participants must choose four things from eight choices to survive under one of three situations: desert, uninhabited island, and mountains.

We conducted our experiment in the following procedure (Fig. 6). First, an audio speaker located above the participants announces the theme and asks, " what do you think? " The robot turns his eyes to the experimenter who gives his opinion about his four choices of things: " I think I 'll take some matches, a watch,

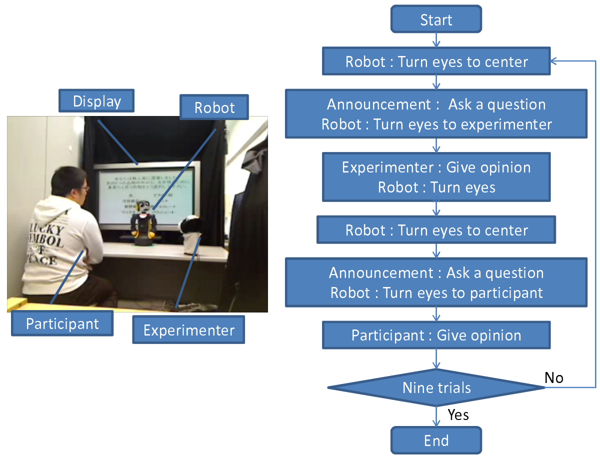

Figure 6: Experimental environment and procedure

Table 2: Experimental condition

		Gaze in first half		
		Speaker	Listener	Center
Gaze in latter half	Speaker	S-S	L-S	C-S
	Listener	S-L	L-L	C-L
	Center	S-C	L-C	C-C

a compass, and a map of an aerial photograph. "Next the robot turns his eyes in the direction shown in the experimental condition (Table 2). The announcement asks again, "what do you think?", and the robot turns his eyes to the participant who gives his opinion that also consists of four things. Each round of things and situations is randomly changed every participant.

6.4 The Experimental Condition

Our experiment was conducted under a two-way within-subjects design (Table 2). The first factor is called the "gaze in first half" and denotes the direction in which the robot turns his eyes while the experimenter gives his opinion in the first half: the time from beginning to say the first thing from saying the second thing. This factor consists of three levels: speaker, listener, and center. The second factor is called "gaze in latter half" and denotes the direction in which the robot turns his eyes while the experimenter gives his opinion in the latter half: the time from beginning to say the third thing to saying the fourth thing. This factor also consists of three levels: speaker, listener, and center. Each condition was randomly assigned to every participant, and nine trials were performed (one trial for each condition).

6.5 Analysis

We observed the following three points:

- The number of things suggested by the experimenter was completely identical with the things the participant said.
- The number of things suggested by the experimenter in the first half agreed with the things the participant said.
- The number of things suggested by the experimenter in the latter half agreed with the things the participant said.

6.6 Hypotheses and Predictions

The following are our hypotheses and predictions:

- Hypothesis

  Humans who look at the robot's eye-gaze might have different understandings and adjust their utterances. If the robot turns his eyes to the speaker while listening to the speaker's utterances, humans who see this scene might believe that the robot has a positive impression about the speaker's opinion. Then the next speaker adapts his opinions of those of the speaker. In contrast, if the robot turns his eyes to the listener while listening to the speaker's utterances, humans who see this scene might believe that the robot has a negative impression of the speaker's opinion. Then the next speaker will not change his opinion to match the speaker.

- Prediction

  The amount of things suggested by the experimenter will agree with the things the participants suggested and will increase as follows:

  - $S-S > S-C, C-S > S-L, L-S, C-C > L-C, C-L > L-L$

6.7 Results

The amount of things suggested by the experimenter entirely agreed with the things suggested by the participants (Fig. 7). By a two-way ANOVA analysis, we found no effect of interaction ($F_{(4.56)} = 0.36, n.s.$) But we did find a marginally significant effect of the "gaze in first half" factor ($F_{(2,28)} = 2.71, p < .10$) and a significant effect of the "gaze in latter half" factor ($F_{(2,28)} = 5.88, p < .01$). By assessing with the LSD multiple comparison method, the number in the speaker level was larger than the listener level ($p < .05$) in the "gaze in first half" factor (Fig. 8). The number in the speaker level was also larger than the listener level ($p < .05$) in the "gaze in latter half" factor (Fig. 9).

We analyzed the amount of things suggested by the experimenter in the first half that agreed with the things suggested by the participants in a two-way ANOVA. We found no interactive or main effect of any factor ($F_{(4,56)} = 1.28, n.s.$, $F_{(2,28)} = 1.00, n.s.$, $F_{(2,28)} = 0.49, n.s.$). We analyzed the amount of

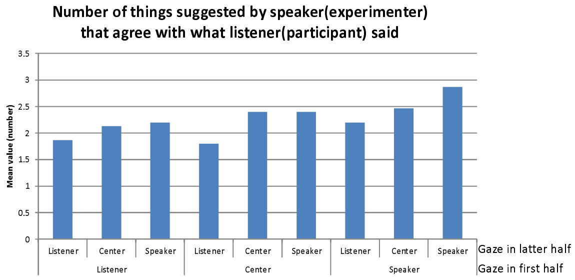

Figure 7: Number of complete agreements

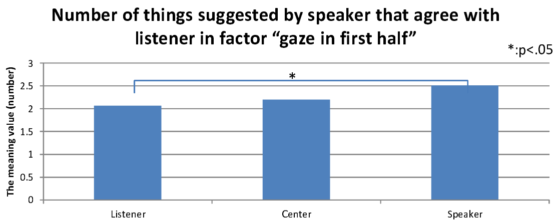

Figure 8: Number of agreements in first half

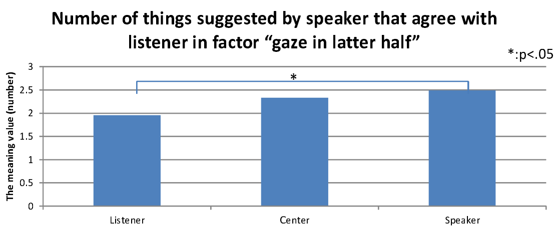

Figure 9: Number of agreements in latter half

things the experimenter said in the latter half that agreed with the things said by the participants by a two-way ANOVA. We found no interactive or main effect of any factor ($F_{(4,56)} = 2.17, n.s.$, $F_{(1.64)} = 1.00, n.s.$, $F_{(2,28)} = 1.05, n.s.$).

6.8 Consideration

According to Section 6.7, the participants succeeded in the context of opinion when the experimenter answered when the robot turned its eyes to the experimenter while listening to his utterances. In contrast, the participants failed when the robot turned his eyes to them while listening to the experimenter. These results support our hypothesis. Based on the questionnaire results, the participants thought that the robot had a positive impression, such as interesting and clear, when it turned his eyes to the experimenter while listening to his opinion. Under such circumstances, the participant might think that it was desirable to give his opinion in the successful context of the experimenter's opinion. But when the robot turned his eyes to the participant while listening to the experimenter's opinion, the participants thought that the robot held such a negative impression as uninteresting or unclear. The participants might think that it was desirable to give their opinions in the unsuccessful context of the experimenter's opinion.

The degree of these effects changes by the timing when the robot turned his eyes because the result showed significant effects of the " gaze in latter half " factor in contrast with just a marginally significant effect of the " gaze in first half " factor. These effects were greater when the robot turned his eyes in the late timing than in the early timing. Perhaps the participants thought that it was natural when the robot turned his eyes to the speaker as one of the listeners in the first half and that it was an intentional behavior in the latter half. The participants especially controlled the contents of their utterances by looking at the robot's eye-gaze in the latter half.

Based on these results, we found that the next speaker can adjust his utterances by controlling the robot 's eye-gaze as a listener. In our experiment, the contents of the speaker's utterances don't tell a story, and these details of the story/utterances do not change by altering their order. But conversations often consist of utterances that do tell a story. Humans can have different impressions of a robot 's eye-gaze as listeners if the speaker is narrating a story. This problem is future work.

7 Conclusion

Many communication robots have been developed. Such robots must adapt themselves to human face-to-face dialogs to realize communication between humans and robots and between humans through robots because face-to-face dialogs are the most natural communication form for humans. It is difficult to realize smooth turn-taking between humans and robots for spoken dialogs [1]. Some studies focused on this problem, where robots adaptively participate in conversations by passively presuming interactive environments. But they failed to assume that interactive environments are aggressively coordinated. In this study, we investigated how the effects of designing eye-gazes for robots determine turn-taking and talking exchanged in multi-party conversations.

We also investigated how humans estimate the target of utterances based on where a robot turned its eye-gaze. Humans estimate a robot's utterance by looking at the direction of its eye-gaze. Then we investigated whether the robot's behavior as a listener expressing his eye-gaze influenced the next speaker's conversation and context. Humans adjusted their utterances by looking at the robot 's eye-gaze as listeners. In the future, we expect this study to effectively lead conversations in desirable discussions in such multi-party conversations as collaborative learning scenes.

But it remains unclear whether these effects apply to ordinary conversations because our experiments were conducted in a laboratory. In ordinary multi-party conversations, much verbal/nonverbal information that is too difficult to control in laboratory settings is conveyed between the participants. We must not only build models of a robot's eye-gaze in a laboratory experiment but also verify them in a field experiment.

References

[1] Yosuke Matsusaka, Tsuyoshi Tojo, Tetsunori Kobayashi: Conversation Robot Participating in Group Conversation, *Trans.IEICE*, Vol. E86-D, No. 1, pp. 26–36 (2003)

[2] Norihiro Hagita, Atsushi Fukayama, Takehiko Ohno, Naoki Mukawa, Minako Sawaki: Messages Embedded in gaze of Interface Agents - Impression management with agent's gaze -, *Proceedings of the ACM Conference on Human Factors in Computing Systems, CHI2002*, pp. 41–49 (2002)

[3] Goffman, E.: Replies and responses, *Language in Society*, Vol. 5, pp. 257–313 (1976)

[4] Clark, H., Carlson, T. B.: Hearers and speech acts, *Language*, Vol. 58, pp. 332–373 (1982)

[5] Hanae Koiso, Yasuharu Den: Towards a precise model of turn-taking for conversation: A quantitative analysis of overlapped utterances, *Proceedings of the DiSS-LPSS Joint Workshop 2010 (The 5th Workshop on Disfluency in Spontaneous Speech and the 2nd International Symposium on Linguistic Patterns in Spontaneous Speech)*, pp. 55–58 (2010)

[6] Kanayo Ogura, Kazushi Nishimoto: Is a Face-to-Face Conversation Model Applicable to Chat Conversations?, *Proceedings of the 18th PRICAI2004 Workshop on "Language Sense on Computer"*, pp. 26–31 (2004)

[7] Duncan, Jr., S.: Some signals and rules for taking speaking turns in conversation, *Journal of Personality and Social Psychology*, Vol. 23, pp. 283–292 (1972)

[8] Kendon, A.: Some functions of gaze direction in social interaction, *Acta Psychologica*, Vol. 26, pp. 22–63 (1967)

[9] Duncan, Jr., S., Fiske, D. W.: Face-to-Face interaction, *Research, methods and theory*, Hillsdale, NJ: Lawrence Erlbaum

[10] M. Tomasello: Joint Attention as Social Cognition, C. moore and P. J. Dunham(eds.), *Joint Attention, Its Origins and Role in Development*, pp. 103–130, LEA (1995)

[11] Akira Ito, Shunsuke Hayakawa, Kouji Mototaka, Kazunori Terada: Can we feel a gaze pressure from a robot? - Development of an eye-contact robot -, *The 12th IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN 2003)*, pp. 103–106 (2003)

[12] Yosuke Matsusaka, Mika Enomoto, Yasuharu Den: Simultaneous prediction of dialog acts and address types in three-party conversations, *Proceedings of the 9th International Conference on Multimodal Interfaces*, pp. 66–73 (2007)

[13] S. Baron-Cohen: Mindblindness - An Essay on Autism and Theory of Mind, *MIT Press* (1995)

[14] Johnson, S., Slaughter, V., Carey, S.: Whose gaze will infants follow ? - The elicitation of gaze-following in 12-month-olds - , *Science*, Vol. 1, pp. 233–238 (1998)

[15] B. Scassellati: Theory of mindfor humanoid robots, *Int. Conf. On Humanoid Robotics* (2000)

[16] ATR: robovie - mR2, http://www.irc.atr.jp/ reo/mr2/ (2012/8/28 the latest access)