Starting a Conversation by Multi-robot Cooperative Behavior

Takamasa Iio(✉), Yuichiro Yoshikawa, and Hiroshi Ishiguro

JST ERATO, Osaka University, Toyonaka, Osaka 5608531, Japan
iio@irl.sys.es.osaka-u.ac.jp

Abstract. In a human-robot conversation, it is difficult for the robot to start the conversation just when the user is ready to listen to the robot, due to recognition technology issues. This paper proposes a novel approach to starting a conversation smoothly by using the cooperative behavior of two robots. In this approach, the two robots try to fill the blank time until the person is ready to listen by showing an interaction between the robots to attract the person's attention. To evaluate the effectiveness of the approach, we conducted an experiment, which compared the following three methods of starting a conversation: early timing, late timing by one robot, and the proposed method. The results showed that participants almost ready to listen and not feel awkward when interacting with two robots with the proposed method, compared to one robot with early and late timing.

Keywords: Human robot interaction · Multi-robots · Starting of conversation

1 Introduction

Recently, in the research field of human-robot interaction, availability of using multiple robots have got attentions of researchers. According to their works [1–6], robots could give users positive effects by coordinating appropriately; for example, it is reported that such robots are more attractable [3] and easier to talk [1] and keep users to a conversation [5]. We are studying how multiple robots should behave with each other to improve the quality of human-robot interaction. This paper focus on using multiple robots in a situation that robots start a conversation with a user. Starting a conversation naturally is important because if a robot showed strange behaviors at the start, the user would remain feeling a strange impression during the conversation. This situation prevents the user from paying an attention to the contents of the conversation. Therefore, we need to consider how a robot should start a conversation.

However, it is not easy for a robot to start a conversation naturally. Let's consider a case of human-human interaction. Goodwin [13] has shown systematic procedures in which speakers obtain the attention of a user. Although he indicated characteristic phenomena in speaker's utterance such as restarts, pauses and hesitations, in a simple case that a speaker attempts to start a conversation with an user, the speaker's behaviors tend to proceed as follows: (1) The person probably offers a short greeting term like "Hey" to attract the user's attention. (2) The person waits for the user to respond by being ready to listen. (3) The person then offers a main topic for discussion.

A. Kheddar et al. (Eds.): ICSR 2017, LNAI 10652, pp. 739–748, 2017.
https://doi.org/10.1007/978-3-319-70022-9_73

For a robot to achieve these behaviors, the robot needs to recognize whether the user is ready to listen as in the second step above. This recognition will be composed of a variety of sensing mechanisms such as face [7] and gaze direction [8–10], a distance and a positional relation with a user [11–13]. For example, Kuzuoka et al. [11] revealed that, in an elderly care home, a gaze direction of a caregiver let an elderly know whether they could start a conversation or not, and provided suggestions of the design policy of the starting of a conversation by a robot. Nakano et al. [8] analyzed a user's gaze behavior and found that patterns of gaze transition correlated with human participation or observational judgement of a user's engagement in a conversation. They proposed an engagement estimation method by judging a gaze direction and showed that the method improves the impression of the user in human-agent interaction. Shi et al. showed that controlling the distance and the positional relation improves a quality of a conversation [12]. Those studies have tried to model a human behavior in starting a conversation and apply the model to a robot. If the robot could recognized a gaze direction, a distance and a positional relation with a user, their methods would work well. However, such recognitions sometimes go wrong in real environment. If a recognition failure happened, the methods would not perform well.

Recognition failure of user's acknowledgement prevents a robot from starting a conversation naturally because the robot cannot deal with the second step of the above process. Figure 1 shows such a failure case, in which a false positive recognition event happened. In the second scene, the robot failed to judge as the user was ready to interact, but in fact, they were not. Due to the recognition, the robot started to speak on a main topic, even though the user is not ready to listen. Such behaviors would not only make the user awkward but also result in a failure of communicating the main topic, as shown in the third scene. To avoid such an unpleasant situation, we can tighten the threshold for the recognition to reduce the likelihood of a false positive. However, this change increases the risk of a false negative recognition. In other words, even though the user was ready, the robot does not speak on a main topic for an extended time.

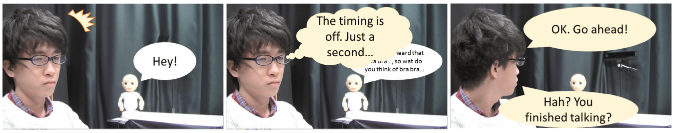

Hey!
The timing is off. Just a second...
heard that bra bra..., so wat do you think of bra bra...
OK. Go ahead!
Hah? You finished talking?

Fig. 1. The failure case of starting a conversation by a recognition failure. In the first scene, the robot says "Hey" to attract the user's attention. Then, the robot waits for the user and is ready to listen, but the robot misrecognized the user's state as being ready to listen and started to speak on a main topic. As a result, the user missed what the robot said.

In this paper, upon accepting the risk of false negative recognition, we consider how to alleviate the user's uncomfortable feelings to a start of a conversation by a robot. Our idea about this is simple but novel; it is that two robots show a user a collaboration interaction between them to fill an unnatural blank time before starting a conversation. This idea is inspired by studies on using multi-robots in a human-robot conversation [1–6]. These studies have suggested potential merits of a social context generated by multi-robots in conversation. For example, Sakamoto et al. [3] reported that, in a field trial in a station, people were more likely to stop to listen to a conversation between two robots than to listen to a single robot. Their results suggest that the social context generated by two robots attracted user's attention more than a speech by one robots. Arimoto et al. [1] investigated the effect of using multi-robots in a human-robot conversation. They compared the impressions of a conversation with a robot and with two robots, and found that participants who talked with two robots felt the robots were less ignored than one robot. Furthermore, Iio et al. [5] developed a turn-taking pattern in which two robots behave according to a prescheduled scenario to avoid the robot's verbal responses sometimes sounding incoherent in the context of the conversations. They proved that participants who talked to two robots using the turn-taking pattern felt the robot's responses to be more coherent than those who talked to a single robot not using it. Arimoto et al. [1] and Iio et al. [5] pointed out that showing a participant a collaborative responses between two robots may influence the participant to improve his or her sense of conversation.

Inspired by those studies [1, 5], this paper propose a behavior design of a collaborative response between two robots for starting a conversation, even though the robots misrecognized an user's cue for the starting of a conversation. This paper is organized as follows. The next section described a behavior design for a collaborative response before a conversation. The Sect. 3 explains an experiment that we conducted to verify the effectiveness of the design. We show the results of the experiment in the Sect. 4 and discuss the effectiveness of the proposed approach in the Sect. 5. Finally, the Sect. 6 concludes our study.

2 Behavior Design to Start a Conversation

If a robot speaks on a main topic before a user is ready (i.e. Fig. 1), the user would be not only awkward but also, in the worst case, miss to listen to the main topic. Therefore, we consider it is better to delay starting a main topic a little longer, even though it would make the user awkward. In our idea, two robots show a user a short interaction until starting a conversation to alleviate the feelings of awkwardness in the user.

Based on this idea, we designed a behavior pattern between two robots as shown in Fig. 2. The following steps correspond to the scenes of Fig. 2.

1. One of robots, called a speaker robot, faces a user and speaks a short greeting term like "Hey" to attract their attention. At that time, another robot, called a bystander robot, turns to them in several hundreds of milliseconds (e.g. 0.5 s).
2. After a few seconds (e.g. 2.0 s), the bystander robot looks back at the speaker robot and says something to fill the empty time from the short greeting term to the main topic. In Fig. 2, the bystander robot says, "What's going on?" The speaker also turns back to the bystander robot when the bystander robot starts to speak.
3. The speaker robot responds to the utterance of the bystander robot.
4. Finally, after the response, the speaker robot turns to the user again and speaks on the main topic. The bystander robot also looks at them during this statement.

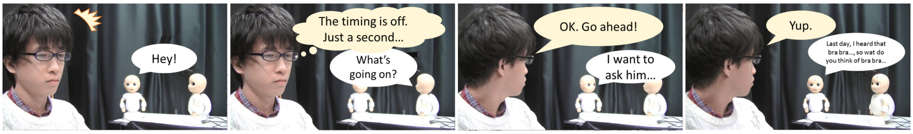

Hey!
The timing is off. Just a second...
What's going on?
OK. Go ahead!
I want to ask him...
Yup.
Last day, I heard that bra bra..., so wat do you think of bra bra...

Fig. 2. The proposed behavior pattern between two robots before starting a conversation.

The point of this pattern is that two robots try to reduce a user's aggravation by demonstrating their short conversation. The short conversation creates enough time for the user to be ready to listen. If the user gives their attention early, they observe the conversation. We suppose this situation will be more tolerable for them than a situation where the robot does not say anything for a few seconds. If the user's attention occurs later, they would listen to the main topic at an acceptable time.

3 Experiment

3.1 Hypothesis

Our hypothesis is as follows: When a robot does a call to start a conversation with a user, whether the user can be ready to listen quickly or late is depending on situations. Therefore, if the robot starts to say a main topic quickly after the call, the user will sometimes not be ready to listen. Conversely, if the robot delays to start the main topic, the user will sometimes wait a long time for the starting and feel awkward. On the other hand, if two robots show the user the interaction we proposed, the user will be almost ready to listen and not feel awkward. To verify this hypothesis, we developed a task called "Detective game on Skype" and conducted an experiment to verify this hypothesis.

3.2 Detective Game on Skype

The game is played by one participant who is assigned to a question master and two confederates who are assigned to detectives. These people are separated to different rooms and communicate on Skype with only sound. The participant is given a paper with a question and its answer like the following example:

- **Question:** Two men who are good friends met for the first time in a decade, but they did not say anything. They were not visually or hearing impaired, and the place where they met did not prohibit talking. Why did they not talk to each other?
- **Answer:** They were divers and met under the sea.

The participant is asked to give detectives the question. The detectives ask the participant to find the answer. The participant attempts to lead the detectives to the answer by responding to their questions appropriately. However, the participant must not say anything except for "Yes, that's right" or "No, that's wrong". Due to this restriction, the detective needs to consider better questions to quickly determine the answer. The following is an example:

- **Detective:** Did they quarrel at that time?
- **Participant:** No, that's wrong.

In this experiment, the detectives were confederate; therefore, they had behaved according a scenario we prepared to make to what each participant listens same among participants. In addition, to remove side-effects derived from confederates, we recorded their voices and played the voice on Skype in the experiment. That is to say, in the experiment, the confederates was actually not in that place.

This task has an advantage to be able to control a participant's state. As mentioned in Sect. 3.1, whether a participant is ready to listen quickly or late is depending on his or her situation when a robot did a call. If the participant is busy to listen to detective's asking, he or she will not be ready to listen to the robot. On the other hand, if the participant is free because the detective has not ask him or her yet, he or she will be ready to listen. In this task, to make a robot do a call in various timing, we manipulated both the timings of detective's asking (playing voices) and robot's call.

3.3 Environment, Apparatus and Procedure

Figure 3 shows an experimental environment and apparatus. Robots were put on a table against a wall near a poster on the wall, and a laptop computer for Skype was set on another table about 1.5 m away from the robots. We employed CommU (Fig. 3, right), which is a conversation robot developed by VSTONE. This robot has three DOFs for its west, three DOFs for its neck, and two DOFs for each eye so it controls its gaze in a flexible manner. This flexible head and gaze control is important for humanlike social behaviors such as turn-taking in a conversation, establishing engagement via eye contact, and expressing attention. The participant played the detective game on Skype using the laptop on the table in front of them. This experiment was conducted by Wizard of Oz method; therefore, an operator played detective's asking and robot's behavior composed of utterances and gestures.

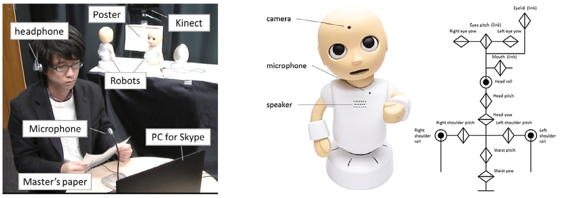


Fig. 3. Experimental environment and apparatus (left) and CommU (right).

In this experiment, a participant was given an instruction of an experimental procedure and a rule of the detective game at first. We asked the participant, "When a robot call to you, you tell the detectives to pause in the game and respond to the robot's main topic. After the response, you answer a questionnaire about impressions of the timing of the main topic. Then, you tell the detectives to restart the game." The participant played the detective game three times in different condition, respectively. In each game, a robot did a call and said a main topic four times in various timing.

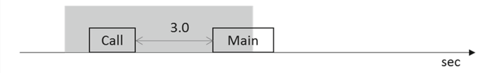
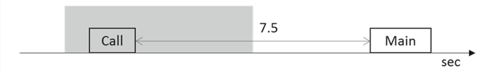
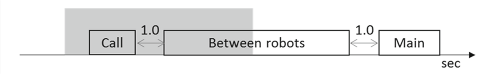
3.4 Condition

We developed three conditions, which are 1E, 1L and 2L. This experiment was a within-participant design, that is, the participants experienced every condition. The order of the experiences was counterbalanced.

- **1E:** One robot and early timing. The robot does a call to attract the participant's attention (e.g. "Hey"). After 3.0 s, the robot speaks on a main topic. The main topic was a simple question for the participant, such as, "Don't you feel cold?" The interval 3.0 s was decided from our preliminary trials.
- **1L:** One robot and late timing. A robot does a call. After 7.5 s, the robot then speaks on a main topic, the same as 2L, which is explained next.
- **2L:** Two robots and late timing. A robot does a call. Then, 1.0 s later, another robot says something to the first robot like "What's going on?" The first robot responds to it 1.0 s later. After that, 1.0 s later, the same robot speaks on a main topic. As a result, the timing of speaking on a main topic is after 7.5 s from a call.

The timeline of the robots' behaviors in each condition are described in Table 1. Each gray box illustrates a case of overlap of detective's question with robot's call and main topic.

Table 1. The timeline of robots' behaviors in each condition.

Condition	Timeline
1E	
1L	
2L	

3.5 Measurements

Participants filled in a questionnaire after every trial. We produced items from the following three aspects. These items were rated on a 5-point scale. Scores of one, three, and five mean disagreement, neutral and agreement, respectively.

- **Items about earliness of the timing:**
  (a) Did you feel that the timing of the robot's question was too early?
  (b) Were you ready to listen to the robot's question?
- **Items about lateness of the timing:**
  (c) Did you feel the timing of robot's question to be too late?
  (d) Did you feel awkward waiting for the robot to speak?

- **Item for general impressions:**
  (e) Did you feel the timing of the robot's question to be natural?
  (f) Did you feel the robot's behaviors asking you questions to be humanlike?

4 Results

Eighteen people (nine males and nine females, who were university students) participated in this experiment. Figure 4 shows the results of the questionnaire. We analyzed the results in one-way paired ANOVA and used Bonferroni method for multiple comparison.

(a) **Earliness of the timing of the question:** The average scores of 1E, 1L and 2L were 3.13 (SD = 0.97), 1.36 (SD = 0.49) and 1.74 (SD = 0.72). The significant difference among the conditions was found (F(2, 17) = 32.69, $p < .01$). The multiple analysis showed the significant difference between 1E and 1L ($p < .05$) and between 1E and 2L ($p < .05$). These results mean that the participants felt timings of a main topic too early in 1E, compared to the other conditions.

(b) **Readiness to listen to the question:** The average scores of 1E, 1L and 2L were 3.04 (SD = 0.99), 4.44 (SD = 0.73) and 4.29 (SD = 0.82). The significant difference among the conditions was found (F(2, 17) = 26.26, $p < .01$). The multiple analysis showed the significant difference between 1E and 1L ($p < .05$) and between 1E and 2L ($p < .05$). These results mean that the participants were not ready to listen to a main topic so more in 1E than the other conditions.

(c) **Lateness of the timing of the question:** The average scores of 1E, 1L and 2L were 1.65 (SD = 0.63), 3.19 (SD = 1.02) and 1.88 (SD = 0.74). The significant difference among the conditions was found (F(2, 17) = 23.40, $p < .01$). The multiple analysis showed the significant difference between 1E and 1L ($p < .05$) and between 1L and 2L ($p < .05$). These results mean that the participants felt timings of a main topic too late in 1L, compared to the other conditions.

(d) **Awkwardness of the waiting:** The average scores of 1E, 1L and 2L were 1.68 (SD = 0.70), 2.57 (SD = 0.89) and 1.61 (SD = 0.64). The significant difference among the conditions were found (F(2, 17) = 11.51, $p < .01$). The multiple analysis showed the significant difference between 1E and 1L ($p < .05$) and between 1L and 2L ($p < .05$). These results mean that the participants felt more awkward in 1L than the other conditions.

(e) **Naturalness of the timing of the question:** The average scores of 1E, 1L and 2L were 3.24 (SD = 0.86), 2.76 (SD = 1.05) and 3.71 (SD = 0.49). The significant difference among the conditions was found (F(2, 17) = 6.08, $p < .01$). The multiple analysis showed the significant difference between 1L and 2L ($p < .05$). These results mean that the participants felt unnatural in 1L more than 2L.

(f) **Robot's human-likeness:** The average scores of 1E, 1L and 2L were 2.82 (SD = 1.10), 2.78 (SD = 1.19) and 3.31 (SD = 0.90). The significant difference among the conditions was found (F(2, 17) = 3.66, $p < .05$), but the multiple comparison did not show the significant difference between any conditions.

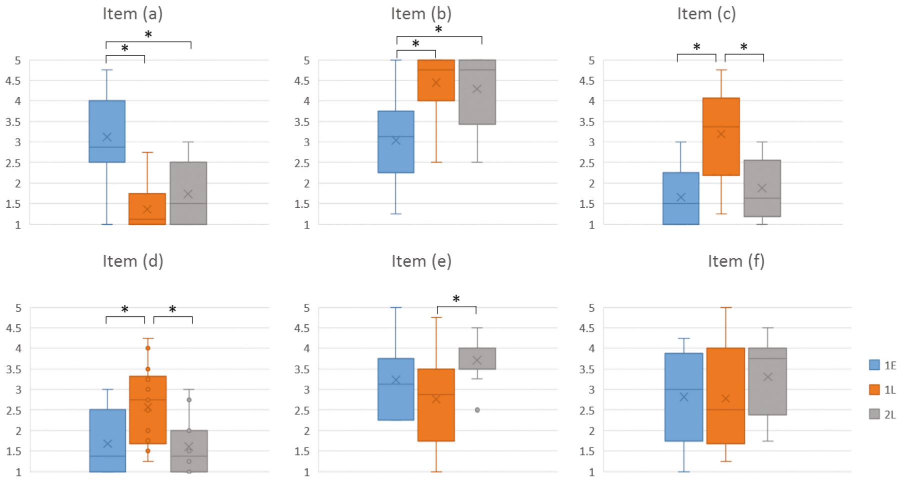


Fig. 4. Results of the questionnaire.

5 Discussion

5.1 Effectiveness of the Proposed Method

We proposed the method which is to show an interaction between two robots until starting a main topic in order to alleviate user's awkwardness. We discuss whether the method contributed to the purpose or not.

The average score of 2L in the earliness of the timing of the question (item (a)) was 1.73. The score was pretty low. It means that the participants did not feel that the robot started main topics too early. Moreover, the average score of 2L in the readiness to listen to the question (item (b)) was 4.29. The score was pretty high. It means that the participants were ready to listen to main topics. Thus, the proposed method would provide the participants an enough time to make ready to listen.

The average score of 2L in the lateness of the timing of the question (item (c)), the average score of 2L was 1.88. The score was as low as that of 1E. On the other hand, the average score of 1L was significantly higher than that of 2L. It means that in 2L the participants did not feel that the robot started main topics too late, even though the time from a call until starting a main topic was same in 1L and 2L. Moreover, in the item (d), awkwardness of the waiting, the average score of 2L was 1.61. The score was pretty low as same as that of 1E. The score was also significantly lower than that of 1L. It means that the participants felt less awkward in 2L than 1L. Those results indicate that the proposed method contributes to reducing participant's awkwardness.

The average score of 2L in the regarding naturalness of the timing of the question (item (e)) was significantly higher than that of 1L. It means that showing an interaction make the participants feel more natural than keeping waiting with silence. On the other hand, there was no significant difference between 1E and 2L. However, the variance of scores of 2L looks smaller than that of 1E (see the graph (e)). It suggests that the participants sometimes rated 1E as unnatural. This may be because there were some situations when the participants were so busy to listen to detective's asking that they could not be ready to listen a main topic. In contrast to this, our proposed method would work well in such a case because the participants could be ready to listen in most cases.

Finally, in the evaluation of robot's human-likeness (item (f)), there was significant difference among conditions but multiple comparison did not show the difference between each condition. This reason may be due to pretty large variances of each condition. If deepening a discussion of human-likeness, we need to design more real situation.

The above results and discussion support our hypothesis that if two robots show the user the interaction we proposed, the user will be almost ready to listen and not feel awkward. Thus, our proposed method was effective for starting a conversation.

5.2 Implication

Our proposed cooperative behavior patterns by multi-robots is not a perfect solution to start a conversation naturally but an interesting approach that has a strong effect on alleviating uncomfortable impressions in cases where false negative recognitions of an user's attention often occur.

In the future, these recognition technologies will continuously improve. However, it is difficult to recognize human status perfectly, as even humans sometimes also fail to do this. We believe that our study showed a novel approach for using multi-robot interaction to deal with the difficulties, and the experimental results suggested that the approach is effective.

5.3 Limitation

This experiment assumed that the participant always reacted to the robot's main topic. In a real environment, there are possible situations where people are not interested in or not aware of the robot. The experiment did not consider the effectiveness of our approach for such people or conditions.

We fixed the length of the human-robot interaction in this experiment because we intended to investigate the effectiveness of a basic situation. To apply our approach in a real environment, we need to use a recognition technique for the user's attention. This investigation will be the topic of future research.

6 Conclusion

This study proposed an approach for alleviating the negativity of people's impression of a robot starting a conversation. In the approach, after a robot speaks to attract a person's attention, the robot and another robot display for the person a short communication sequence between the robots until the first robot speaks on a main topic. Through an experiment to investigate the effectiveness of our approach, we found our proposed approach enabled the start of a conversation, by making an user ready to listen without causing an uncomfortable impression. The results suggested the effectiveness of our proposed approach using multi-robot cooperative behaviors, and it will contribute to HRI as a new approach to deal with starting a conversation naturally.

Acknowledgments. This work was supported by JST ERATO Grant Number JPMJER1401, Japan.

References

1. Arimoto, T., et al.: Cooperative use of multiple robots for enhancing sense of conversation without voice recognition. SIG-SLUD **B5**(2), 76–77 (2015). (In Japanese)
2. Takahashi, T., et al.: A social media mediation robot to increase an opportunity of conversation for elderly: mediation experiments using single or multiple robots. In: Technical Committee on Cloud Network robotics (CNR), vol. 113, No. 84, pp. 31–36 (2013). (In Japanese)
3. Sakamoto, D., et al.: Humanoid robots as a broadcasting communication medium in open public spaces. Int. J. Social Robot. **1**(2), 157–169 (2009)
4. Shiomi, M., et al.: Do synchronized multiple robots exert peer pressure? In: Proceedings of the Fourth International Conference on Human Agent Interaction, Biopolis, Singapore, pp. 27–33 (2016)
5. Iio, T., et al.: Pre-scheduled turn-taking between robots to make conversation coherent. In: Proceedings of the Fourth International Conference on Human Agent Interaction, Biopolis, Singapore, pp. 19–25 (2016)
6. Karatas, N., Yoshikawa, S., De Silva, P.R.S., Okada, M.: NAMIDA: multiparty conversation based driving agents in futuristic vehicle. In: Kurosu, M. (ed.) HCI 2015. LNCS, vol. 9171, pp. 198–207. Springer, Cham (2015). doi:10.1007/978-3-319-21006-3_20
7. Sinder, C.L., et al.: Where to look: a study of human-robot engagement. In: International Conference on Intelligent User Interfaces (IUI 2004), pp. 78–84 (2004)
8. Nakano, Y.I., et al.: Estimating user's engagement from eye-gaze behaviors in human-agent conversations. In: International Conference on Intelligent UserINterfaces, pp. 139–148 (2010)
9. Yamazaki, K., et al.: Prior-to-request and request behaviors within elderly day care: implications for developing service robots for use in multiparty settings. In: Bannon, L.J., Wagner, I., Gutwin, C., Harper, R.H.R., Schmidt, K. (eds.) ECSCW 2007, pp. 61–78. Springer, London (2007). doi:10.1007/978-1-84800-031-5_4
10. Bergstrom, N., et al.: Modeling of natural human-robot encounters. In: IEEE/RSJ International Conference on Intelligent Robots and System (IROS 2008), pp. 2623–2629 (2008)
11. Kuzuoka, H., et al.: Re-configuring spatial formation arrangement by robot body orientation. In: ACM/IEEE International Conference on Human-Robot Interaction (HRI 2010), pp. 285–292 (2010)
12. Shi, C., et al.: Spatial formation model for initiating conversation. In: Conference on Robotics: Science and Systems (RSS 2011) (2011). doi:10.15607/RSS.2011.VII.039
13. Goodwin, C.: Restarts, pauses, and the achievement of a state of mutual gaze at turn beginning. Sociol. Inq. **50**(3–4), 272–302 (1980)