Using a Conversational Agent to Facilitate Non-native Speaker's Active Participation in Conversation

Zixuan Guo
University of Tsukuba
Ibaraki, 3058550, Japan
s1721704@s.tsukuba.ac.jp

Tomoo Inoue
University of Tsukuba
Ibaraki, 3058550, Japan
inoue@slis.tsukuba.ac.jp

ABSTRACT

When a non-native speaker talks with a native speaker, he/she sometimes feels hard to take speaking turns due to language proficiency. The resulting conversation between a non-native speaker and a native speaker is not always productive. In this paper, we propose a conversational agent to support a non-native speaker in his/her second language conversation. The agent joins the conversation and makes intervention by a simple script based on turn-taking rules for taking the agent's turn, and gives the next turn to the non-native speaker for prompting him/her to speak. Evaluation of the proposed agent suggested that it successfully facilitated the non-native speaker's participation over 30% of the agent's interventions, and significantly increased the frequency of turn-taking.

1 INTRODUCTION

When a non-native speaker (NNS) uses a second language (L2) to talk with a native speaker (NS), the NNS has a disadvantage in language proficiency and sometimes has difficulty in presenting fully what he/she would like to say. Previous research shows that even when NSs and NNSs were presumed to have equal content knowledge, the NSs emerged as the talkers as they did when they were the relative experts [1].

KEYWORDS

conversational agent; turn-taking; second language communication; non-native speaker

CHI'19 Extended Abstracts, May 4-9, 2019, Glasgow, Scotland, UK.

ACM ISBN 978-1-4503-5971-9/19/05. DOI: https://doi.org/10.1145/3290607.3313075

In other words, NSs often control or take over conversations with NNSs, and this shows the possibility that NNSs have difficulty taking turns to speak. Zhu et al. examined interaction and feedback in mixed peer response groups by inspecting participants' turn-taking behaviors, and indicated that the NNSs as a group took fewer turns and produced fewer language functions during an oral discussion of writing [2]. Thus, in the L2 conversation, a mechanism that gives NNSs more chances to participate in conversations is desired in order to achieve more cooperative and productive communication between NSs and NNSs.

In this paper, we propose a conversational agent system to support NNSs in L2 conversation. Based on the existing turn-taking rules, the agent intervenes the conversation to prompt NNS's utterance.

2 RELATED WORK

2.1 Supporting Second Language Conversation

Various methods have been studied to support NNSs in L2 conversation. Using automatic speech recognition (ASR) to convert utterance into text and present it to NNS in real time is an example. Highlighting keywords in ASR transcripts was reported to reduce the cognitive load of NNSs and help NNSs catching up in audio conferences [3]. A system for NSs to input keywords during the conversation and share it with NNSs was proposed. It was then verified that mutual understanding and knowledge sharing between two speakers increased [4]. A speech-rate-awareness system was developed, which recognized the speech rate of NSs in real time and notified when the speech rate got too fast for the NNSs [5].

These studies try to support NNS's listening or understanding of conversation in L2 conversation. In contrast, we try to facilitate NNS's speaking by a conversational agent, because supporting NNS's participation has not been studied despite that NNSs take fewer turns in L2 conversation [2].

2.2 Agents for Supporting Conversation

Studies of using conversational agents or robots in multi-party conversation have been increasing. These agents or robots actively communicate with participants and play a role in achieving collaborative multi-party conversation.

An example is a conversational robot system that estimated each participant's conversational dominance and participation role, produced attentional behaviors, and intervened in the conversation in proper timing [6]. However, most studies intend to coordinate conversation among homogeneous participants where all the participants were assumed to have the same language proficiency. In contrast, the agent in this paper intends to coordinate L2 conversation where language proficiency is different between the two participants.

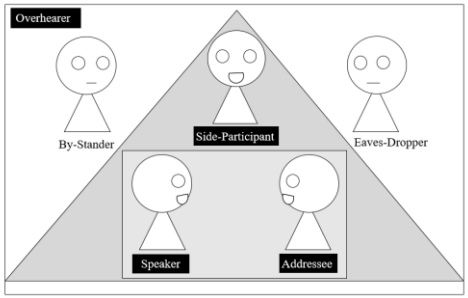

Figure 1: Participation role in multi-party conversation

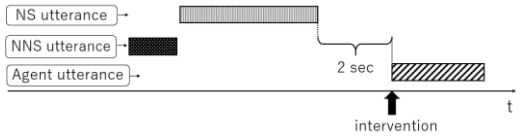

Figure 2: Intervention timing by the agent

Table 1: Sentence list for our agent

Next-speaker Selection Technique	*Sentence*
(a)	[NNS's last name], what do you think?
(a)	[NNS's last name], how about that?
(c)	[NNS's last name], don't you think so?
(c)	[NNS's last name] thinks so too, don't you?

2.3 Participation Role of Face-to-face Conversation

Goffman pointed out participation role [7] as shown in Fig. 1. In a face-to-face conversation among three or more participants, the participant who makes utterance is called a speaker, the listener is referred to as an addressee, and the participant who does not receive the speaker's speech is regarded as a side-participant.

Turn-taking occurs among the speaker and the addressee, but the side-participant can also join the conversation. In this way, it is possible for a third participant to moderate and control the conversation. Therefore, in a conversation where the turn is not allocated successfully between a NS and a NNS, if there is a third participant who plays a role of adjusting the imbalanced turn-taking between them, he/she may be able to actively improve their participation.

3 PROPOSED AGENT

3.1 Overview

To better balance NSs and NNSs in L2 conversation, a conversational agent is proposed who is a third participant and plays a role of adjusting the imbalanced turn-taking between NSs and NNSs [8]. The agent takes turns in the conversation and passes the next turn to NNSs for prompting NNSs utterances based on the turn-taking rules. Specifically, the implemented agent detects the speech of the two speakers in real time and decides whether or not to intervene in the conversation. If both the NS and the NNS were silent for a certain period of time, the agent takes its turn by speaking pre-determined sentences that prompt the NNS to take the next turn.

3.2 Method

To decide the timing for intervention, we consider the turn-taking rule [9]. The rule is that if "current speaker selects next" technique is not used and unless another "self-selects" (means taking the opportunity to start talking), the current speaker may continue. Because the agent does not want to disturb other participants talking, it takes its turn by start talking only when none of the participants is talking. More than two seconds of silence is regarded as the boundary of turns in conversation [10], as shown in Fig. 2.

To make sure that the agent can select the NNS as the person to speak next, we used the "current speaker selects next" technique, that is, the Next-speaker Selection Techniques as in the following (a) to (d) [9].

(a) Select a participant as the next speaker by asking a question and give the turn to him/her. (b) Repeat some parts of the prior speaker's utterance with a 'one-word question', e.g. "What?", "Who?". (c) Add an addressed question to the end of the turn, which is called 'tag question'. e.g. "Don't you agree?", "You know?". (d) Employ social identities to specify the next speaker.

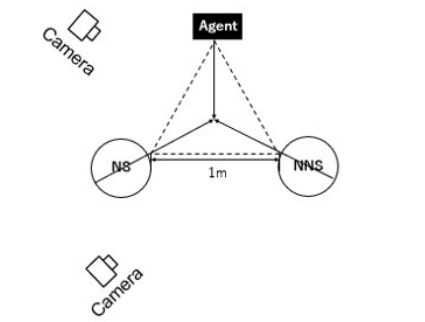

Figure 3: Setup of the experiment

Table 2: Topic list

1. Lesson, Seminar, and Research	7. Overseas Experience
2. Extracurricular Activity	8. Food
3. Hobby	9. Daily Life and Habit
4. Part-time Job	10. Clothes
5. Birthplace	11.Health
6. Travel	12. Festival

For (b) and (d), it is difficult to predefine a sentence that can specify the NNS as the next speaker for sure. In this study, we designed a sentence list using (a) and (c). All of the statements start with the NNS's last name. To avoid rudeness and abruptness in the Japanese situation, which may affect the participants' conversation experience and make the NNS nervous, we used euphemistic expressions to ask for the NNS's opinion. Table. 1 shows the sentence list for our agent. The agent uses one of these four sentences at random to speak to the NNS.

3.3 Implementation

We built a conversational agent using the MMD Agent [11], which is an open-source toolkit and has a basic function to perform language and non-verbal interaction. By adding a socket plug-in to MMD Agent, we were able to control the agent with Python program. For speech detection, the PyAudio [12], which is an audio I/O library using Python, was used to acquire real-time speech data in 16-bit PCM (Pulse Code Modulation) format.

4 EXPERIMENT

4.1 Participants and Environment

The experiment participants were 36 people of Japanese NSs and the same number of Japanese NNSs. The NNSs were international students from China who passed the Japanese Language Proficiency Test N2 Level [13]. We recruited them by using a Google Form at the University of Tsukuba. A pair of a NS and a NNS was the same gender who met each other for the first time.

The experiment was conducted in a laboratory setting. The environmental noise was 40 db. The distance between the two speakers was 1 m. The setup of the experiment is shown in Fig. 3.

4.2 Conditions and Procedure

To investigate how the intervention of the agent affects the conversation, we used a counterbalanced design with 3 within-subject conditions as follows.

Intervention Condition: The agent intervenes in the conversation.
Non-intervention Condition: The agent does not speak and intervene but only displayed.
Control Condition: The LCD screen does not display the agent.

For each condition, the two participants made conversation, whose video and voice were collected. The details were as follows. (1) The participants sat on chairs and wore pin microphones. They were explained the procedure of the experiment, and filled out the consent forms. (2) They selected a topic for conversation from the list (Table. 2) before starting the conversation. (3) They had a free conversation for 8 minutes. (4) After the conversation, they filled out the questionnaires.

Table 3: An example of DS

Sample 4	*02:10*	*Club*
NS: It is quite normal to get along well with students of different grades ...		
NNS: Oh—		
NS: ... Certainly.		
AGENT: [NNS's name] thinks so too, don't you?		
NNS: Yes.		

Table 4: An example of IS

Sample 11	*02:08*	*World Cup*
NS: How about this year? Is China strong? In football.		
NNS: Uh-huh...s, stro...		
AGENT: [NNS's name], how about that?		
NS: Wow, how about that?		
NNS: China ...is... not... strong...		

Table 5: An example of RF

Sample 13	*05:44*	*Vegetable*
NS: Is there any vegetable you saw for the first time since you came to Japan from China?		
NNS: Vegetables I saw for the first time		
AGENT: [NNS's name], how about that?		
NS: Huh?		
NNS: Giggle.		
NS: It will support us.		
NNS: That's amazing. Vegetables I saw it for the first time ... Well ... I cannot remember.		

5 RESULTS AND DISCUSSION

Seven minutes which excluded the first one minute from 8-minute data was used for the analysis, because the first one minute was not very usual including many greetings. Hence the analyzed data was 7 (min.) * 18 (persons) = 126 minutes per a NS or a NNS in each condition. The effective intervention rate and the frequency of turn-taking were obtained.

5.1 Effective Intervention Rate

70 interventions by the agent were found in the conversation in the intervention condition. These interventions were categorized as follows based on the responses of the participants.

Direct Success in turn-allocation (DS): The agent was directly successful in NNS's taking turn. This applied to the case when the turn was passed directly from the agent to the NNS. Table. 3 shows an example.

Indirect Success in turn-allocation (IS): The agent was indirectly successful in the NNS's taking turn. This applied to the case when the NS responded to the agent's utterance but the response directly introduced the NNS's taking turn. Table. 4 shows an example.

Responded Failure in turn-allocation (RF): The agent failed in the NNS's taking turn but was successful in drawing response. This applied to the case when the NS responded. Table. 5 shows an example.

Non-responded Failure in turn-allocation (NF): The agent was ignored and failed in the NNS's taking turn.

The rates of these inventions are shown in Table. 6. Over 30% of intervention was successful in passing turns to the NNS. NSs sometimes asked opinions of NNSs right after the agent utterances, which was categorized as IS. This behavior, seemed to be considerate of NNSs, was introduced by the agent.

5.2 Frequency of Turn-Taking

The frequency of turn-taking was calculated, as shown in Fig. 4. Results of normality test are $p = .068$, $p = .027$, $p = .066$ for the 3 conditions respectively. Results of a Friedman Test showed that the effect of our agent was significant, $\chi 2\ [2] = 7.000$, $p = .030$. Post hoc analyses using the Bonferroni Correction indicated that the frequency of turn-taking was significantly higher in the Intervention Condition ($M = 7.63$, $SD = 2.12$) than in the Control Condition ($M = 6.32$, $SD = 2.33$), $p = .023$. This suggests that the intervention was effective in making the conversation more active.

The pairwise comparison of the Non-intervention Condition with the Control Condition was non-significant. This suggests that the change from the two-person conversation to the three-person conversation and the changes of the participation role of the 2 participants did not affect the outcome.

Table 6: Rates of interventions

	DS	*IS*	*RF*	*NF*
Number of interventions	12	10	28	20
Rate	17%	14%	40%	29%

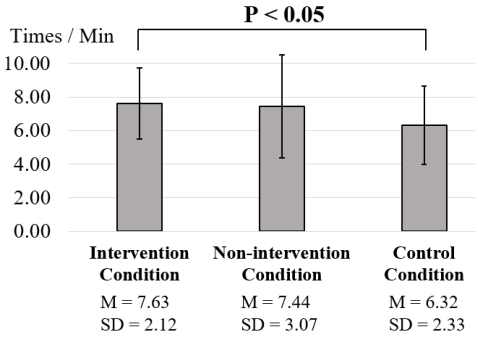

Figure 4: Frequency of turn-taking

6 CONCLUSIONS

In this paper, we proposed an agent to prompt NNS to speak in L2 conversation based on the participation role and the turn-taking rules in multi-party conversation. From the experiment, it was revealed that the agent using (even very simple) conversation rules could make effective interventions to help an NNS in L2 conversation, and could make conversation more active.

REFERENCES

[1] Zuengler Jane, 1993. Encouraging learners' conversational participation: The effect of content knowledge. In the Language Learning Journal Vol 43 No 3, 403-432. https://doi.org/10.1111/j.1467-1770.1993.tb00620.x

[2] Zhu Wei, 2001. Interaction and feedback in mixed peer response groups. In Journal of second language writing Vol 10 No 4, 251-276. https://doi.org/10.1016/S1060-3743(01)00043-1

[3] Hautasaari Ari, and Naomi Yamashita, 2014. Catching up in audio conferences: highlighting keywords in ASR transcripts for non-native speakers. In *Proceedings of the 5th ACM international conference on Collaboration across boundaries: culture, distance & technology*. ACM, New York, NY, USA, 107-110. https://doi.org/10.1145/2631488.2634064

[4] Hanawa Hiromi, Xiaoyu Song, and Tomoo Inoue, 2016. Key-Typing on Teleconference: Collaborative Effort on Cross-Cultural Discussion. In Collaboration Technologies and Social Computing, CCIS vol 647, Springer, 74-88. https://doi.org/10.1007/978-981-10-2618-8_7

[5] Jing Ye, and Tomoo Inoue, 2016. A Speech Speed Awareness System for Non-Native Speakers. In *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion*. ACM, New York, NY, USA, 49-52. https://doi.org/10.1145/2818052.2874322

[6] Yutaka Takase, Takashi Yoshino, and Yukiko I. Nakano, 2017. Conversational Robot with Conversation Coordination and Intervention Functionality in Multi-party Conversations. In IPSJ Journal Vol 58 No 5, 967-980.

[7] E. Goffman, 1981. Forms of Talk. University of Pennsylvania Press.

[8] Tomoo Inoue, Zixuan Guo, 2018. A conversational agent that uses turn-taking rules for supporting a non-native speaker. *In* Collaborative Technologies and Data Science in Smart City Applications , Logos, 1-9.

[9] Sacks Harvey, Emanuel A. Schegloff, and Gail Jefferson, 1974. A Simplest Systematics for the Organization of Turn-Taking for Conversation. In Language Vol 50 No 4, 696-735. https://doi.org/10.2307/412243

[10] Nakano Yukiko, and Yuki Fukuhara, 2012. Estimating conversational dominance in multiparty interaction. In *Proceedings of the 14th ACM international conference on Multimodal interaction*. ACM, New York, NY, USA, 77-84. https://doi.org/10.1145/2388676.2388699

[11] MMDAgent Homepage, http://www.mmdagent.jp/, last accessed 2018/02/01.

[12] Pyaudio Homepage, https://people.csail.mit.edu/hubert/pyaudio/docs/, last accessed 2018/02/01.

[13] Japanese-Language Proficiency Test, https://www.jlpt.jp/e/, last accessed 2019/02/13.