Regular Paper

Visual Appearance is not Always Useful: Voice is Preferred as a Mediating Conversational Agent to Support Second Language Conversation

Tomoo Inoue,* Kumiko Kawai,** Boyu Sun**, and Kosuke Sasaki**

*Faculty of Library, Information and Media Science, University of Tsukuba, Japan
**Graduate School of Library, Information and Media Studies, University of Tsukuba, Japan
*inoue @slis.tsukuba.ac.jp

Abstract – Two representations of a conversational agent to support second language communication were compared, one of which was a CG character agent and the other was a voice agent. Except for the appearance, these agents were identical in terms of the function to intervene in the conversation between the native speaker and the non-native speaker and to pass the next speaking turn to the non-native speaker. A few distinctions were found between the two in the influence on the second language conversation. The voice agent was more effective in intervention, and was perceived more useful by the participants.

Keywords: Conversational agent, Second language communication, Voice user interface, Non-native speaker, Turn taking

1 INTRODUCTION

In recent years, globalization has increased the opportunities for conversation in the second language among people from different countries. Cross-cultural communication has disincentives such as differences in language and culture. Especially when a non-native speaker (hereinafter NNS) conducts conversation in a second language with a native speaker (hereinafter NS), he/she often face problems in listening comprehension of NS speech as well as in speaking, because his/her oral skills and knowledge in the second language are limited and thus difficult to make a story smoothly. As a result, there are differences in the amount of speech between NS and NNS [1], and the tendency of NS to take the initiative in conversation [2].

To improve these biases, methods and tools to support more participation of NNS in conversation have been proposed [3]-[6]. However, most of those are not very easy to use in everyday life in terms of portability and equipment cost, as they require a PC with a display or the installation of a robot, and so on [7]. More convenient form of a support tool is desired with the expansion of opportunities for second language conversation. Audio support without video can be considered along this line. In this research, we compared the voice agent with the existing CG character agent which works in the same way.[1]

2 RELATED WORK

2.1 Second Language Conversation Support Systems

Various studies on supporting second language conversation have been conducted recently. Majority of them are on presenting visual information such as text and images. For example, a face-to-face cross-cultural communication support system was proposed that displayed related information on the nouns appeared in a conversation [3]. A keyword sharing system to promote mutual understanding in a remote conversation between NS and NNS has been studied where keywords in the voice conversation are entered by NS for accurate and efficient content summarization and they are shown in each screen for better understanding and participation of NNS [4]. In the speech speed awareness system, the speech speed of the speaker is constantly detected and when it becomes too fast for NNS, the system notifies it through a screen [5]. Guo et al. has developed a CG character agent with voice to support second language conversation [6]. When a long silence occurs in the conversation between NS and NNS, the agent takes the speaking turn and talks to the NNS to pass the next speaking turn to him/her. It aims to encourage NNS utterance for better participation, which would result in better productivity in conversation.

All of these second language conversation support systems require a screen and a PC, and are not very portable. Considering that the second language conversation takes place in various places, it is desirable to develop a small portable system.

2.2 CG Character Agents and Robots for Intervention in Conversation

A CG character agent and a robot are often introduced to support conversation. Huang et al. proposed a CG character agent that participated in human conversation and presented information during the conversation [8]. From a WoZ experiment of intervention to conversation, the following four types of interventions was shown to be effective. No.1 was "provide-topic" that provides new topic to the conversation.

[1] This is an extended version of the paper presented at IWIN 2019.

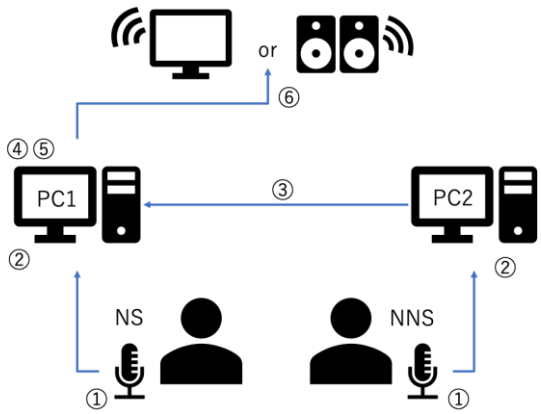

Figure 1: Configuration of the agent system.

No.2 was “more-information” that provides additional information to the topic of the conversation. No.3 was “recall-support” that supplements the information that the user has forgotten. No.4 was “discussion-support” that organizes the content of the conversation so far. In particular, the rating of “provide-topic” was high, which indicated the demand for keeping conversation.

Akiba et al. developed a robot agent to be a facilitator in order to eliminate social imbalances in multi-party conversation [9]. The robot was life-size and intervened in the conversation between the two participants who were leading it when there was a less involved participant.

Those agents and robots also have visual appearances and need bulky equipment.

2.3 Representation of Agent and Its Psychological Influence

Research has also been conducted on the psychological effects of the form of an agent on users. Naito et al. [10] compared with a robot, a CG character, and a voice agent on the user's attitude toward information acceptance. A robot significantly increased the user's acceptance attitude, and there was no difference between a CG character and a voice agent on the information acceptance attitude. There have been other experiments to improve affinity and communication by making the agent look and act like human. According to Takeuchi et al., a human-like agent could promote relaxed dialogue with little psychological burden, but at the same time, could cause excessive expectations for its behavior [11]. Hara et al. conducted a study asking if it is possible to give more feeling of being watched by making the appearance of a camera robot eyeball-shaped. It was shown that the eyeball-shaped camera robot could make people feel more nervous and afraid than the normal camera robot. As a result, it was shown that the presence of the eye could make people feel more nervous and afraid than the camera [12].

Those studies indicate representation of an agent has psychological effects and influences on human behavior. In this research, we are interested in the influence of appearance of an agent that facilitates second language conversation between NS and NNS.

Table 1: Agent intervention patterns.

Last speaker	Intervention pattern after 2 seconds of silence
NS	Ask a specific participant for an opinion ・What do you think, NNS?
	Ask for a new opinion ・Do you have any opinion, NNS?
NNS	Ask for an utterance that supplements own opinion ・I see NNS. Could you tell me more?
	Ask for a new opinion ・Do you have any other opinion, NNS?

3 CONVERSATIONAL AGENT

The agent used in this study is a second language conversation support agent based on the turn taking rule developed by Guo et al. [6]. It participates in conversation and try to pass the speaking turn to NNS according to the next speaker selection technique and the turn taking rule in conversation between NS and NNS.

Figure 1 shows the system configuration. The voices of NS and NNS are by taken by microphones (1). The volume of each voice is detected by the voice detection module in PCs (2). The voice detection module notifies to the silence detection module whether received sound exceeds pre-determined threshold value (3). The silence detection module gathers information from the voice detection modules to know silence, and notifies to the Wizard, an experimenter, when silence continues for more than 2 seconds, because it is regarded as another turn even if the former speaker speaks again [13] (4). The Wizard selects an intervention speech which is most appropriate in a next turn (5). The agent starts performing intervention by the selected speech (6).

As the agent's intervention, Guo et al. applied the adjacent pair pattern "Question-Response" (e.g. "What do you think, NNS?") and the additional question (e.g. “You think so, don’t you, NNS?”) based on the next speaker selection technique [14]. However, it might be unnatural when the agent intervenes after the NNS utterance. Therefore, in this research, the intervention patterns based on the “Question-Response” shown in Table 1 were used with reference to the intervention by Yoshino et al. [15]. In this study, we used the WoZ method to manually determine the intervention content, because it was decided based on the last speaker before the silence.

4 EXPERIMENT

We investigated the influence of representation of our conversational agent that facilitates second language conversation between NS and NNS. Specifically, it was whether the voice agent and the CG agent provided the same support and the same psychological effect on the participants.

(a) CG condition

(b) Voice condition

Figure 2: Scenes from the experiment.

4.1 Participants

The participants were a total of 48 people of 24 pairs each consisting of a native Japanese speaker and a non-native Japanese speaker. JLPT N2 was required as the language skill of a non-native Japanese speaker to meet our target who could speak Japanese but not fluently like native speakers. JLPT is a standardized criterion-referenced test to evaluate and certify Japanese language proficiency for non-native speakers [16]. N2 level certifies the ability to understand Japanese used in everyday situations, and in a variety of circumstances to a certain degree [17]. They were recruited from a Web bulletin board, and were undergraduate or graduate university students. Each pair was new to each other and of the same sex in this experiment.

4.2 Design

In order to investigate the influence of agent representation, we conducted within-subject experiments under the following three conditions.

1) CG condition

The CG character agent displayed on the screen intervenes in the conversation.

2) Voice condition

The voice agent from the loudspeaker intervenes in the conversation.

3) FTF condition

Participants talk face-to-face without the agent.

Figure 2 shows the scenes from the experiment.

The CG agent was implemented using the MMD agent, which is an open-source toolkit for building voice interaction systems with a CG character [18]. An agent built by the MMD agent has a basic function to perform language and non-verbal interaction with people. The appearance of the agent is shown in Fig. 2(a). For our purpose, specific motion of the agent including facial expression was not used. What was used in this study was its default behavior and facial expression, which was to sit still with slight swaying and normal facial expression to avoid giving a strange impression. By adding a socket plugin-in to the MMD Agent, the agent was able to be controlled with a Python program. For speech detection, one of audio I/O libraries PyAudio was used [19]. For speech synthesis, a Japanese text-to-speech system Open JTalk was used [20]. It provides the voice of a female speaker model called the "May" character. The voice agent was the same with the CG agent but only used its voice part.

Each pair had free conversation in 5 minutes for each condition. The topic was different each time. The order of participation in the conditions was balanced by the Latin square method.

4.3 Data

1) Conversational behavior

The video and audio data were collected from three cameras, the panoramic camera that captures all of NS, NNS, and the agent, and the two cameras that capture each speaker from the front. From this video and audio, we analyzed the conversational behavior of NS, NNS and the agent. The total length of data used for analysis was 3 conditions * 24 pairs * 5 minutes = 360 minutes.

From the collected video data, speech segments were extracted under the condition of minimum silent interval of 400 milliseconds [21] using the segmentation function of the annotation tool ELAN. The part where the utterances overlapped and the part where the voice was low were segmented manually to avoid incorrect segmentation. Then the speech segments excluding non-verbal speech segments such as laughter, cough and sigh were classified as NS speech, NNS speech, and agent speech.

2) Questionnaire

The questionnaire consisted of the following items. They were the items such as naturalness of communication, ease of speech, for evaluating overall impression of communication and conversation [3], [22]-[24], items regarding the appropriateness of the agent intervention and usefulness, items for measuring stress in conversation, and items regarding the agent's impression for evaluating the influence of the agent representation on the participants [11], [25]-[26]. Items related to an agent were asked to the CG condition and the voice condition, and not to the FTF condition. All items were measured in 7-point Likert scale where 1 corresponded to strongly

Table 2: Categories of intervention.

	Effective	Ineffective	Total
CG (#)	17	9	26
CG (%)	65.4	34.6	100
Voice (#)	37	4	41
Voice (%)	90.2	9.8	100

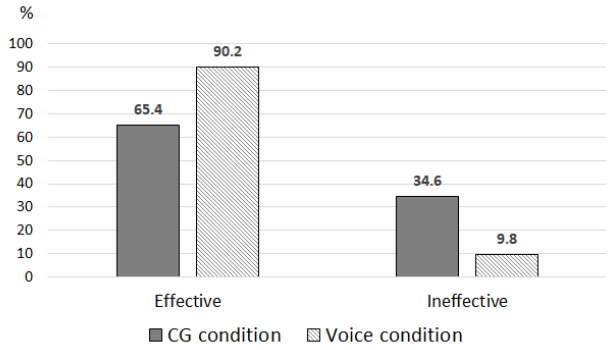

Figure 3: Ratio of interventions by the agent.

disagree to 7 corresponded to strongly agree. It is better evaluated in the item #1 to #18 and #22 to #23 as having higher scores, while in the item #19 to #21 as having lower scores.

5 RESULT

5.1 Number of Intervention by the Agent

We found a mistake in the operation of the wizard in a pair, whose data was removed.

The total number of agent interventions in the 23 pairs was 70, of which 29 times in the CG condition and 41 times in the voice condition. Paired t-test found a marginally significant difference between the conditions (t(23)=-1.96, p = .06).

5.2 Categories of Intervention by the Agent

Interventions were categorized into 2 types of effective intervention and ineffective intervention, where NNS took the next speaking turn in effective intervention while he/she did not take it in ineffective intervention.

We removed 3 interventions, all of which were in the CG condition, from the analysis. For one, the system stopped right after the intervention due to malfunction. For other 2, the interventions were the end of the sessions.

Table 2 and Fig. 3 show the result. The effective intervention was 65.4% in the CG condition and 90.2% in the voice condition.

A two-way ANOVA was used to compare whether there are differences in the number of interventions between the conditions of CG and voice, and between the categories of effective intervention and ineffective intervention. The result showed a marginally significant difference for the main effect between conditions (F(1,88) = 3.13, p = .08) and a significant difference for the main effect between categories (F(1,88) = 23.4, p <.01). A significant difference was also found for interactions (F(1,88) = 8.70, p = .004), but since there was a sufficient difference in the number of interventions between the categories, simple main effects were examined by a Bonferroni-corrected t-test using the number of interventions in each category.

As a result, significant differences were found between the number of effective interventions in the voice condition and the following numbers: the effective interventions in the CG condition (t(22) = -3.66, p = .008), the ineffective interventions in the CG condition (t(22) = -5.18, p = .0002), the ineffective interventions in the voice condition (t(22) = 5.13, p = .0002).

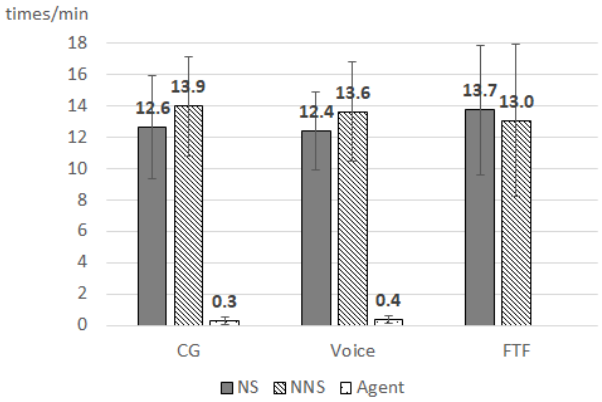

Figure 4: Speech frequency.

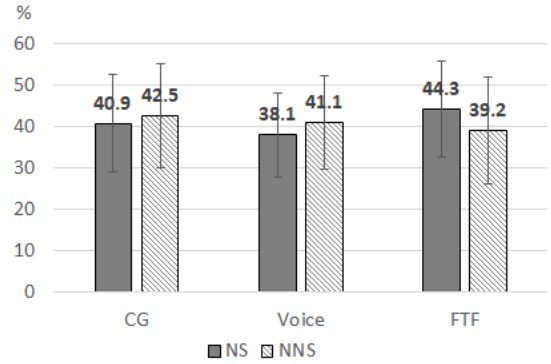

Figure 5: Speech time ratio.

5.3 Participants' Speech

The speech frequency is shown in Fig. 4. A two-way ANOVA comparing between the speakers of NS and NNS, and between the conditions of CG, voice, and FTF revealed no significant differences in the main effect and interaction (Inter-speaker main effect, F(1,132)=0.98, p=.32; inter-condition main effect, F(2,132)=0.14, p=.87; interaction, F(2,132)=1.09, p=.34).

The ratio of the time each speaker was speaking during the conversation (speech time / conversation time * 100) is defined as the speech time ratio. Figure 5 shows the speech time ratio of each NS and NNS for each condition. A two-way

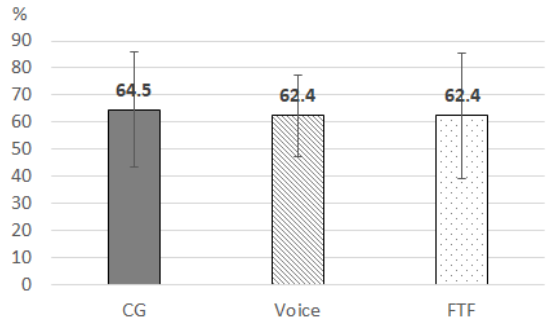

Figure 6: Speech balance.

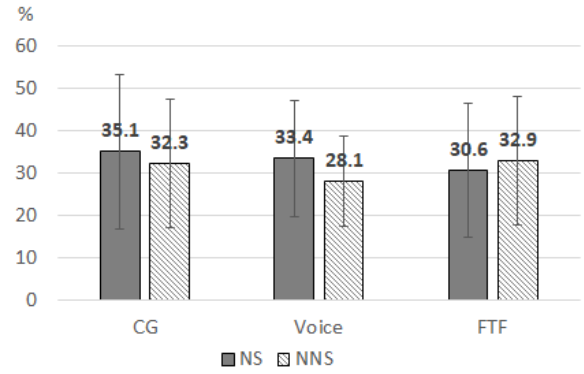

Figure 7: Back channeling rate.

ANOVA comparing between the speakers of NS and NNS, and between the conditions of CG, voice, and FTF revealed no significant differences in the main effect and interaction (Inter-speaker main effect, $F(1,132)= 0.006$, $p=.94$; inter-condition main effect, $F(2,132)= 0.62$, $p=.62$; interaction, $F(2,132)= 1.49$, $p=.23$).

Figure 6 shows the results of the speech balance, which is defined as (the number of utterances of the less talkative speaker) / (the number of utterances of the more talkative speaker) * 100 (%) [27]. A one-way ANOVA comparing between the conditions of CG, voice, and FTF revealed no significant difference ($F(2,69)=0.09$, $p=.92$).

Back channeling is a short expression sent by the listener while the speaker is using the right to speak [28]. To compare the degree that the participant plays the role of the listener, the back channeling rate (the number of back channeling / the number of utterance * 100) was calculated. Back channeling was judged according to the definition of the back channeling expression by Yoshida et al. [29]. The results are shown in Fig. 7. A two-way ANOVA comparing between the speakers of NS and NNS, and between the conditions of CG, voice, and FTF revealed no significant differences in the main effect and interaction (Inter-speaker main effect, $F(1,132)=0.52$, $p=.47$; inter-condition main effect, $F(2,132)=0.44$, $p=.64$; interaction, $F(2,132)= 0.74$, $p=.48$).

5.4 Questionnaire

Figure 8 shows the items for impression of communication and conversation. Figure 9 shows the items for the usefulness of the agent. Figure 10 shows the items for the impression of the agent. Figure 11 shows the items for the stress in conversation.

To examine whether there were differences in the results of the questionnaire between conditions, Friedman tests were performed for questions #1 to #9 and #19 to #23, which asked participants in all three conditions, and Wilcoxon's signed-rank tests were performed for questions #10 to #18, which asked participants in the CG and voice conditions. As a result, we found that the scores of NNS in the following items were significantly or marginally significantly lower in the CG condition than in the voice condition. They are the item #10 "I felt the agent was involved in the conversation" ($Z = -2.07$, $p =.04$), the item #13 "I could speak a lot because of the agent intervention." ($Z = -2.03$, $p = .04$), and the item #14 "I felt the agent's utterance timing was appropriate." ($Z = -1.75$, $p = .08$).

6 DISCUSSION

6.1 Influence of the Agent Representations on Conversation

It was surprising that the influence of the agent representation seemed differ in some aspects.

The tendency of more number of interventions in the voice condition in 5.1 indicates more tendency of long silence occurred at the same time. Although the differences were not statistically significant, it seems to be consistent with the fact that the voice condition had smaller values than the CG condition in all of Fig. 4, 5, and 7 which related to participants' conversation.

Figure 7 shows that the intervention was relatively effective in both the voice and CG conditions, and that the intervention resulted in a higher proportion of NNS speech in the voice condition, which might not be meet our expectation.

Again, although the difference was not statistically significant, the following can be read from Fig. 4, 5 and 7. The speech frequency of NNS seemed slightly higher than that of NS in the CG condition and the voice condition, while that of NNS seemed lower than that of NS in the FTF condition. Also, the speech time ratio of NNS seemed slightly higher than that of NS in the CG condition and the voice condition, while that of NNS seemed lower than that of NS in the FTF condition. Moreover, the back channeling rate of NNS seemed slightly lower than that of NS in the CG condition and the voice condition, while that of NNS seemed higher than that of NS in the FTF condition. It is interesting that there is such consistency in the trend of conversation change. NNS seemed to be more of a listener in the FTF condition. But agent intervention seemed to have reversed it.

Overall, the result could assure the feasibility of a voice agent.

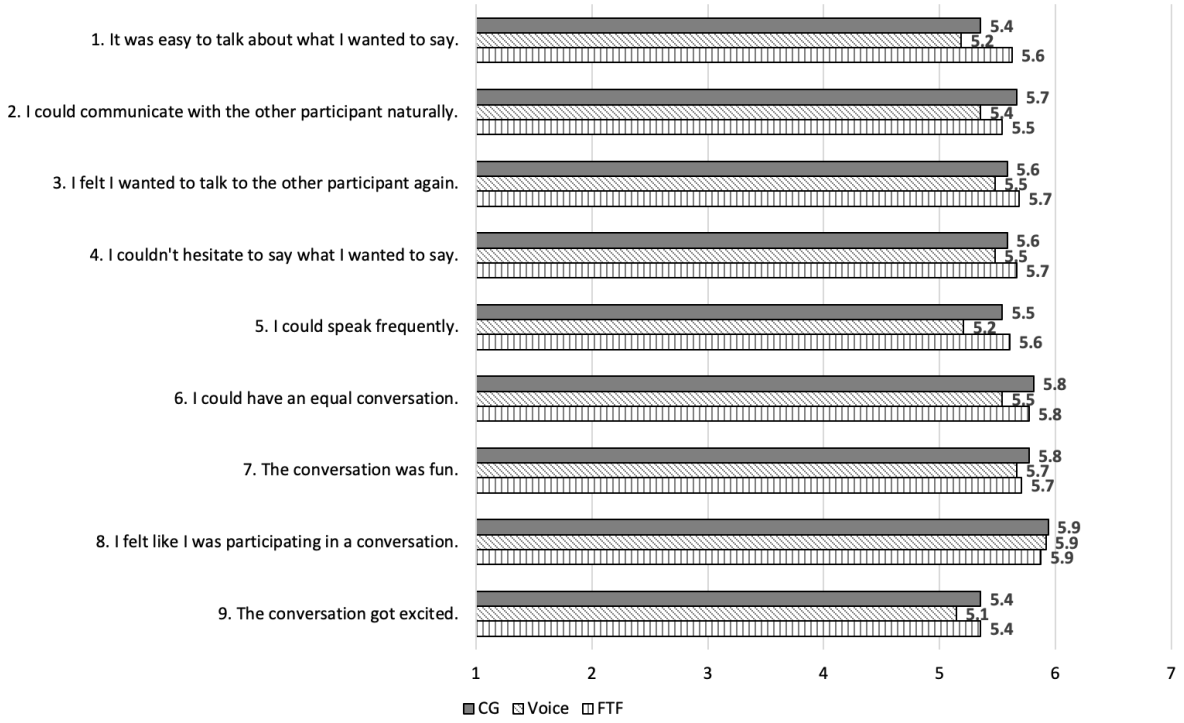

Figure 8: Questionnaire result for conversation.

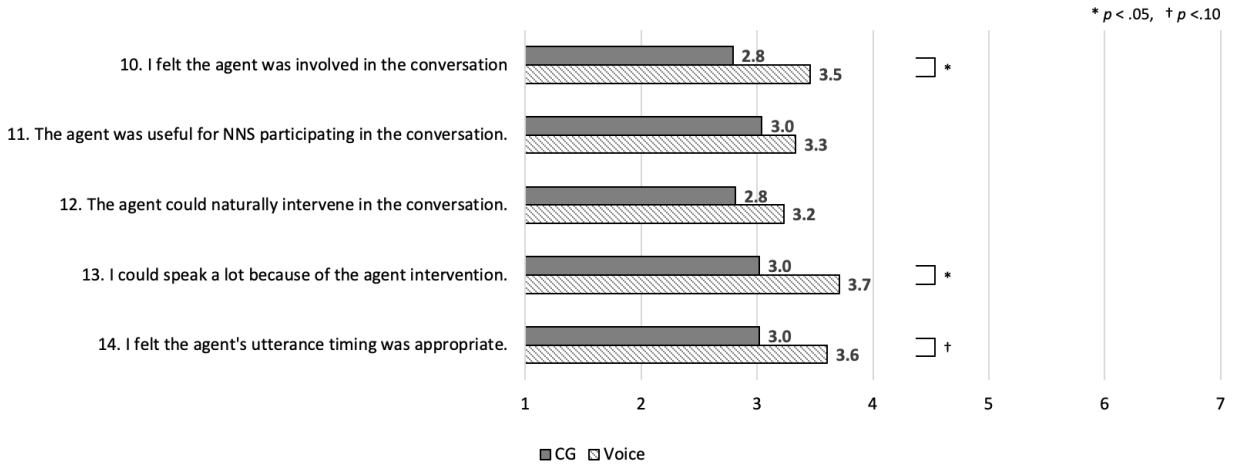

Figure 9: Questionnaire result for the agent usefulness.

6.2 Influence of the Agent Representations on Subjective Impression

From the questionnaire, the voice agent was better accepted than the CG agent in some question items, which related to the agent usefulness. Nevertheless, the average scores for the voice agent were lower than neutral, so they were not highly rated; the both agents were rated low, meaning that CG agents were rated even lower.

It might be possible that these lower ratings for the agents, especially for the CG character agent, were the reflection of the over expectation for an agent with very limited capability mentioned in Section 2.3.

For other question items, the significant difference was not found between the conditions. One of the possible reasons may be that the CG character appearance does not play any role in this conversational agent system. Actually, only the voice element of the agent system was used for intervening in the conversation, which could make the visual element not useful.

There were no significant differences between any of the agent conditions in comparison to the FTF. In other words, the agents did not have a subjectively noticeable effect on

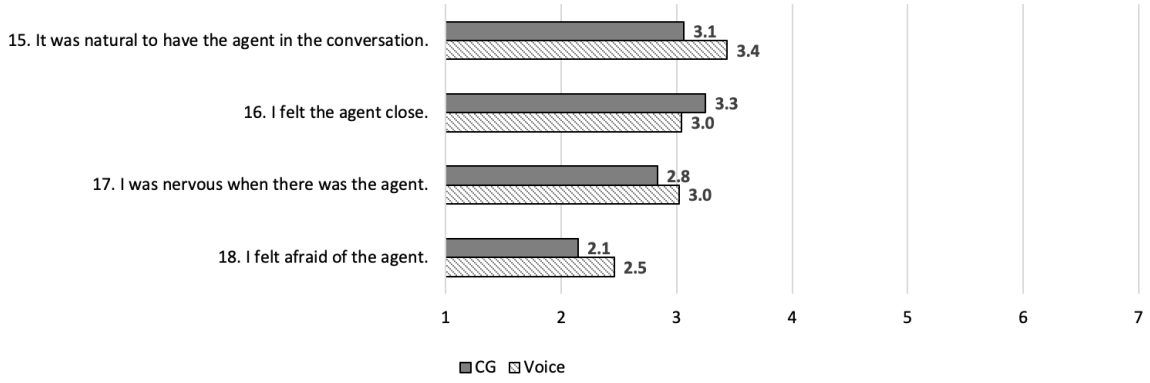

Figure 10: Questionnaire result for the agent impression.

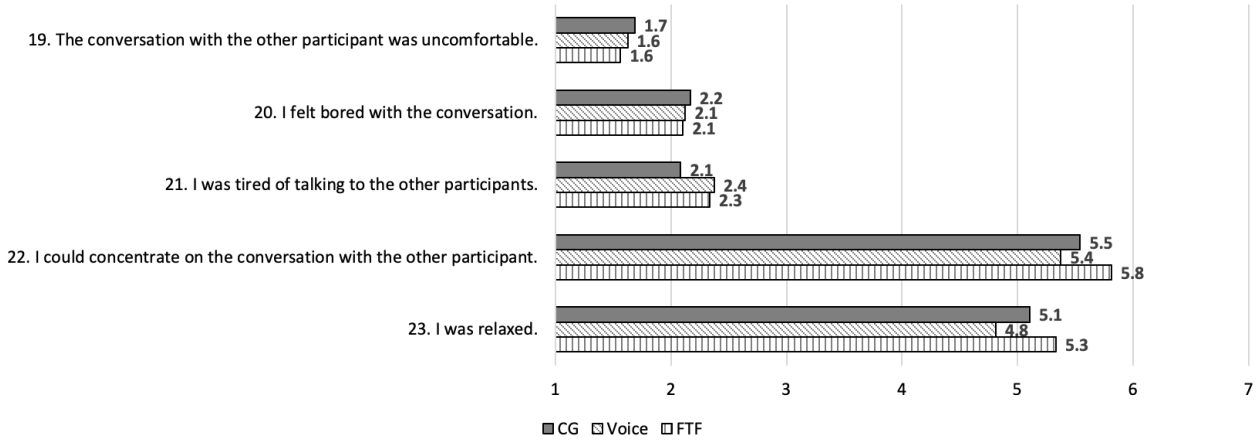

Figure 11: Questionnaire result for stress.

conversation or stress. The objective data showed that the agents worked in a certain way on the conversation, but were not very aware of it by the participants.

Another possibility that the agents did not get noticeable difference might be the positions where the agents were placed. The agents were placed on the side of the participants as in Fig. 2. This setting could certainly affect the magnitude of an agent's impact, in particular the magnitude of the visual agent's impact, and the results. Thus the agent placement is the information needed to reproduce the results.

6.3 A Mediating Conversational Agent to Support Second Language Conversation

A summary of the findings of this study are as follows.

1) In many cases, NNS actually acquires a speaking turn and starts the next utterance by an agent who participates in the conversation between NNS and NS and takes a simple action of giving the right to speak to NNS.

2) The voice agent is more effective than the CG agent in transferring the right to speak, and is also perceived to be more useful.

3) The addition of the agent (who intervenes only during long silences of more than 2 seconds) does not affect the conversation very much, nor does the participants feel any effect.

In this study, as a second language conversation support system, we investigated the influence of the representation of the agent who intervenes in the conversation, based on the related research that the agent representation has a psychological effect. It was possible that the influence of CG was not seen from the similarity with [10], the results were somewhat surprising and the CG agent performed worse. As a study with similar direction of the result, in classical mediated communication experiments of problem-solving tasks, it is known that audio is an essential medium and video has little impact [30]. It can be said that the usefulness of media such as audio and video depend largely on what is done. CG was not only unnecessary in the agent's role in this experiment, but it may also have undercut expectations of competence for human-like CG.

Of course, the results of this experiment may also depend on the appearance and behavior of the CG agent used in this experiment. The visual impact may be increased by using an agent that gives a stronger impression, or an agent that has

facial expressions or gestures. The same applies to the agent placement described in 6.2. This is the limitation of this study.

How can we use such an agent? The good news is that a voice agent would be sufficient for this purpose, as you don't have to make any visual parts (if it's not a special one). If you just want to output sound, you only need a loudspeaker, which makes it more portable. For example, you don't have to sit down at a dedicated conference table where a CG agent shows up to start a conversation. You can use it for a little standing talk by running it on a smartwatch. Since conversations occur everywhere, it's nice to have the system easier to use. If it becomes easier to use, it is conceivable to improve the practicality by working on a method of voice intervention in the future.

7 CONCLUSION

In this research, we compare a CG character agent and a voice agent to support second language conversation. In the evaluation by 24 pairs of participants in the experiment, the intervention in conversation by the agent generally resulted in passing the speaking turn to NNS successfully, but more when the agent was voice-only. From the results of questionnaire survey, the voice agent gave an impression that it was more useful than the CG character agent. This finding indicates that visuals are not always important when designing agents to intervene with people, but rather demonstrates the availability of agents in environments where providing sufficient visuals is not easy, such as mobile environments.

REFERENCES

[1] W. Zhu, “Interaction and Feedback in Mixed Peer Response Groups”, Journal of Second Language Writing, Vol.10, No.4, pp.251-276 (2001).

[2] T. Hifumi, “Relationship of Linguistic Processing and Psychological Processing in Conversations Between Native Speakers of Japanese and Nonnative Speakers”, Japanese Journal of Educational Psychology, Vol.47, pp.490-500 (1999).

[3] K. Okamoto, and T. Yoshino, “Development of Face-to-face Intercultural Communication Support System using Visualized Keyword of Conversation”, Proceedings of Forum on Information Technology, Vol.8, No.3, pp.393-396 (2009).

[4] H. Hanawa, X. Song, and T. Inoue, “Communication Support by a Native Speaker’s Entering Keywords on Discussion with a Non-native Speaker - Effects of Entering Keywords by a Native Speaker on Audio Conference -”, IEICE Technical Report, Vol.116, No.31, pp.139-144 (2016).

[5] J. Ye, and T. Inoue, “A Speech Speed Awareness System for Non-Native Speakers”, Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion, pp.49-52 (2016).

[6] Z. Guo, M. Tang, and T. Inoue, “A Conversational Agent for a Non-native Speaker to Talk with a Native Speaker”, IEICE Technical Report, Vol.117, No.509, pp.107-112 (2018).

[7] S. Yamada, “Are Robots Necessary for Human-Agent Interaction?”, Proceedings of the 23rd Annual Conference of the Japanese Society for Artificial Intelligence, Vol.23, No.1G2-OS3-11 (2009).

[8] H. Huang, S. Otogi, R. Horita, and K. Kawagoe, “Toward a Guide Agent Who Actively Provides Information in Multi-user Conversation – A Trial in Finding the Timings for Intervention”, Transactions of Japanese Society for Artificial Intelligence, Vol.31, No.1, 10.1527/tjsai.DSF-514 (2016).

[9] I. Akiba, Y. Matsuyama, and T. Kobayashi, “Procedures of Obtaining Initiatives for Multiparty Conversation Facilitation Robots”, IPSJ SIG Technical Reports, Vol.2013-SLP-97, No.10 (2013).

[10] H. Naito, S. Kawamura, and Y. Takeuchi, “An Empirical Study of Indicating Reliable Messages in Real World Interaction”, The Transactions of the IEICE. A, Vol.92, No.11, pp. 840-851 (2009).

[11] Y. Takeuchi, and Y. Katagiri, “Inducing Social Agreement Responses Toward Interface Agent”, IPSJ Journal, Vol.41, No.5, pp.1257-1266 (2000).

[12] T. Hara, and K. Terada, “Gankyu robot ni yoru hishikan ga doutokusei koujou ni ataeru eikyo”, IPSJ SIG Technical Reports, Vol.2015-HCI-164, No.1 (2015).

[13] J. Rubin, “A Review of Second Language Listening Comprehension Research”, The Modern Language Journal, Vol.78, No.2, pp.199-221 (1994).

[14] H. Sacks, E. A. Schegloff, and G. Jefferson, “A Simplest Systematics for the Organization of Turn-Taking for Conversation”, Language, Vol.50, No.4, Part1, pp.696-735 (1974).

[15] T. Yoshino, M. Yatsushiro, Y. Takase, and Y. Nakano, “Intervening in Multiparty Conversations by Conversational Agents based on Dominance Estimation”, Proceedings of the 29th Annual Conference of the Japanese Society for Artificial Intelligence, Vol.29, No.1I4-4, pp.1-3 (2015).

[16] “Objectives and History | JLPT Japanese-Language Proficiency Test”, http://www.jlpt.jp/e/about/purpose.html, last accessed 2020/09/11.

[17] “Competence Required for Each Level | JLPT Japanese-Language Proficiency Test”, https://www.jlpt.jp/e/about/levelsummary.html, last accessed 2020/09/11.

[18] “mmdagent.jp”, http://www.mmdagent.jp/, last accessed 2020/09/11.

[19] “PyAudio Documentation”, https://people.csail.mit.edu/hubert/pyaudio/docs/, last accessed 2020/09/11.

[20] “Open JTalk - HMM-based Text-to-Speech System”, http://open-jtalk.sp.nitech.ac.jp/, last accessed 2020/09/11.

[21] B. F. Freed, N. Segalowitz, and D. P. Dewey. “CONTEXT OF LEARNING AND SECOND LANGUAGE FLUENCY IN FRENCH: Comparing Regular Classroom, Study Abroad, and Intensive Domestic Immersion Programs”, Studies in Second Language Acquisition, Vol.26, No.2, pp.275-301 (2004).

[22] R. Liu, and T. Inoue, “Application of an Anthropomorphic Dining Agent to Idea Generation”, Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication, pp.607-612, (2014).
[23] T. Inoue, “Naturalistic Control of Conversation by Meal: Induction of Attentive Listening Attitude through Uneven Meal Distribution in Co-dining”, Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication, pp.601-606 (2014).
[24] T. Takahashi, M. Hamasaki, and H. Takeda, “Web Community System Mediated by Avatar-like Agents”, Proceedings of the 15th Annual Conference of the Japanese Society for Artificial Intelligence, 1F1-10 (2001).
[25] Y. Hayashi, and K. Miwa, “Cognitive and Emotional Characteristics in Human-Human / Human-Agent Interaction”, The Transactions of Human Interface Society, Vol.10, No.4, pp.445-455 (2008).
[26] Y. Hayashi, E. Cooper, V. Kryssanov, A. Urao, and H. Ogawa, “Psychological Characteristics on Communication with a Conversational Agent - A Study on Schema and Embodiment -”, Transactions of Japan Society of Kansei Engineering, Vol.11, No.3, pp.459-467 (2012).
[27] A. Shamekhi, Q. V. Liao, D. Wang, R. K. E. Bellamy, and T. Erickson, “Face Value? Exploring the Effects of Embodiment for a Group Facilitation Agent”, Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, pp.391-403 (2018).
[28] S. K. Maynard, Discourse Analysis, Kuroshio Shuppan (1993).
[29] N. Yoshida, K. Takanashi, and Y. Den, “Taiwa ni okeru aiduchi hyougen no nintei to sono mondaiten ni tsuite”, Proceedings of the Annual Meeting of the Association for Natural Language Processing, Vol.15, pp.430-433 (2009).
[30] A. Chapanis, “Interactive Human Communication”, Scientific American, Vol.232, pp.36-42 (1975).

(Received November 12, 2019)
(Revised April 6, 2020)

Tomoo Inoue is Professor of the Faculty of Library, Information and Media Science of University of Tsukuba, Japan. His research interests include Computer-Supported Cooperative Work, Human-Computer Interaction and Educational Technology. He received his Ph.D. in Engineering from Keio University in 1998. He is a recipient of awards including Outstanding Paper Award, Activity Contribution Award and SIG Research Award from Information Processing Society of Japan (IPSJ). He has served a number of academic committees, including IPSJ Journal Group Editor-in-Chief, IPSJ Transactions on Digital Content Editor-in-Chief, IEICE Technical Committee on Human Communication Science Chair, IPSJ SIG Groupware and Network Services Steering Committee Secretary, ACM CSCW Papers Associate Chair, CollabTech Steering Committee, and IEEE TC CSCWD. He has co-authored and edited "Idea Generation Methods and Collaboration Technologies (Kyoritsu Shuppan)(in Japanese)," and "Communication and Collaboration Support Systems (IOS Press)" among others.

Kumiko Kawai received her M.Sc. in Informatics from University of Tsukuba in 2019. She is currently with Ateam Inc. Her research interests include human interface and interaction.

Boyu Sun received his B.Sc. in Engineering from Huaqiao University in 2018. He is currently a graduate student in University of Tsukuba, Japan. His research interests include human-computer interaction.

Kosuke Sasaki received his M.Sc. in Informatics from University of Tsukuba in 2016. He is currently a Ph.D student in University of Tsukuba, Japan. His research interests include Computer-Supported Cooperative Work.