CHAT ROBOT COUPLING MACHINE RESPONSES AND SOCIAL MEDIA COMMENTS FOR CONTINUOUS CONVERSATION

Hidekazu Minami[1], *Hiromichi Kawanami*[1], *Masayuki Kanbara*[1], *Norihiro Hagita*[1,2*]

[1]Nara Institute of Science and Technology, Graduate School of Information Science
8916-5 Takayama-cho, Ikoma, Nara, 630–0101, Japan
[2]ATR Institute International, Intelligent Robotics and Communication Laboratories
2-2-2 Hikaridai Seika-cho, Soraku-gun, Kyoto, 619–0288, Japan
{minami.hidekazu.ls7, kawanami, kanbara}@is.naist.jp, hagita@atr.jp

ABSTRACT

In this paper we propose a communicative robot that facilitates talk with a user who lacks a chance of verbal communication with others for any reasons (e.g. the elderly who live alone). As our first trial, this system is designed to support a user to talk while watching a TV program. It features two conversation techniques: realizing natural timing of simultaneous response and providing interesting utterances from social media networks related to the TV program the user is watching. To achieve natural response timing, the proposed system includes three response functions: backchannel, repetition and machine answering. While the system keeps talking using the three kinds of responses, it searches human text comments about the TV program from social network services (SNS). When a comment is found, the robot outputs the comment by synthetic speech. A preliminary experiment is conducted to evaluate how the proposed response technique encourages a user to speak to the chat robot. The results show that average number of user utterances with the proposed robot using social media networks and the above three functions is significantly higher than that when the chat robot only outputs social media comments.

***Index Terms*—** Human-computer interaction, chat robot, spoken dialogue system, communication assistance

1. INTRODUCTION

The loss of oral communication especially among the elderly has been focused on as a social problem. In 2014, the Japanese government reported that about 28% and 22% of elderly men and women (age 65 and more) living alone have only one (or less) opportunity to talk to another every few days in average[1]. The report also drew attention to the risk of resulting medical problems such as cognitive impairment and depression. In the modern society, this problem is not limited to the elderly but also a problem for a young person.

One solution is to increase the number of volunteers with skills such as active listening[2], but this requires time and money. Another solution is to introduce ICT (information and communication technology) into their lives. One example is spontaneous interaction through the internet society network service via a mobile device or a PC. But it requires much effort to familiarize with such devices especially for elderly people.

Thus, a communicative robot (chat robot) has been focused on as a substitute for an active listening volunteer and ICT, Many chat robots have been developed and some are provided as consumer products in Japan. For example, [3] developed a robotic pet which delivers daily information to a senior from a distant place such as local government office or welfare institution. The system also monitors the senior. Some other systems employ eye contact, body gestures such as nodding in addition to speech interface. However, the quality of the conversation still needs to be improved.

Miyazawa et al.[4] investigate factors that encourage humans to continue dialogue. This research reports importance of sociality and humor in conversation in addition to naturalness in response (timing and content) From these viewpoints, we developed a novel chat robot that encourages a user to have chances to speak in a daily life.

2. CHAT ROBOT FOR NATURAL TALK

2.1. Preceding research

In the field of communication robots, Matsubara et al.[5] developed an assistant active listening robot that nods to the user at appropriate times as they talk, encouraging them to talk about whatever is on their mind. However, the robot itself did not speak to the user.

As part of their research on two-way conversation, NTT DOCOMO[6] developed a chat engine named *Shabette*

*This work was partly supported by Stragetic Information and Communications R&D Promotion Programme (SCOPE) of the Ministry of Internal Affairs and Communications of Japan (13230713) and JSPS KAKENHI Grant Number 24500115.

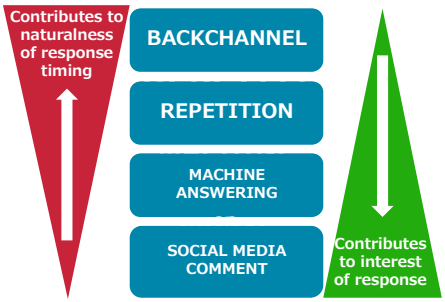

Fig. 1. Contribution of response functions

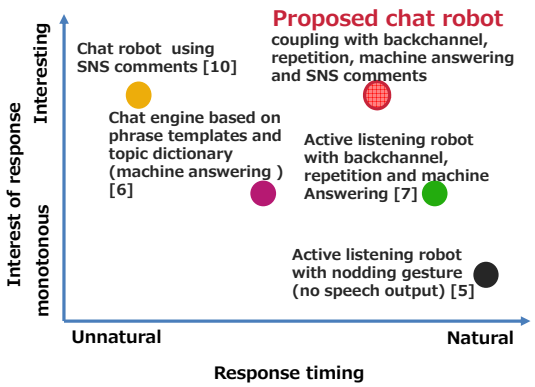

Fig. 2. Characteristics of the preceding research

Concier (concierge by speaking). The system outputs an answer to a user question using phrase templates and a topic dictionary. The system employed an agent character and enabled to speech input and speech output. However, its response text tended to be monotonous as it used phrase templates and its response timing was unnatural.

Kobayashi et al.[7] developed an active listening robot designing to be a substitute for a human volunteer. It had the oral functions of backchannel (called *aizuchi* in Japanese). and repetition, in addition to conventional responses using phrase templates and a topic dictionary. It had appropriate response timing although its responses were monotonous.

To overcome the problem of monotonous responses, response generation using text Corpus gathered from the World Wide Web are widely reported (e.g. [8, 9]). Takahashi et al.[10] proposes a chat robot that employs a social media network service (SNS) as a resource for interesting lively responses. The chat robot is designed to use while its user watching a TV program. It gathers social comments related to the TV program from the internet. However, their system employs Wizard-of-Oz strategy, where the tester manually transcribes the subject's utterance and writes it to the SNS manually to obtain others' responses. This system also has a problem that the robot cannot output anything until someone writes a comment on the SNS.

2.2. Response functions

As mentioned before, the preceding research introduced four response functions: backchannel[5, 7], *repetition*[7], *machine answering*[6, 7] and *social media comments*[10]. They can be characterized by contribution to natural response timing and the interest of the response as shown in Fig. 1. In this paper, the automatic response generation functions; backchannel, repetition, machine answering are referred to as machine responses.

Figure 2 summarizes the previous research on a two-dimensional plane; the horizontal axis represents naturalness of the response timing and the vertical axis represents interest of the response. The figure shows that a chat robot that realizes both natural response timing and interesting utterances is not yet available. In this paper, we propose a chat robot that employs both social media comments and machine responses.

3. CHAT ROBOT COUPLING SOCIAL MEDIA NETWORKS AND MACHINE RESPONSES

3.1. System structure

Figure 3 illustrates the process flow of the proposed robot. When the system starts, it accesses a social network service correspond to the user's interests and gathers comments in real time. If a comment is available, the system outputs the comment by synthesized speech. When the user speaks, the system responds with backchannel and repetition with natural timing. If no new social media comments are found, the machine answering process is activated.

The following functions are realized by the robot.

- Backchannel
  This makes a user feel that the system is paying attention to and listening to them. Although actual backchannel is performed in the middle of speaking, in this system, it is only generated when the system detects the end of a user utterance. The voice activity is detected using the zero-cross count and waveform amplitude installed in the free speech decoder Julius[11]. When detected, the system outputs backchannel such as *Hai* ("I see" or "yes") after a fixed pause[12].

- Repetition
  This allows a user to know that the system has correctly heard what they said. The user utterance is first transcribed to text using Julius then the text is sent to API keyword extraction software [13]. If preset keywords are found in a user utterance, the system generates a template-based response including the keyword.

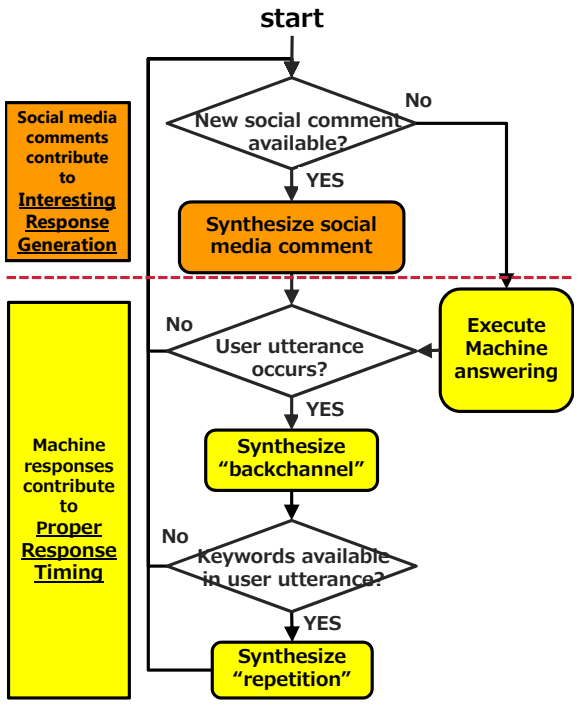

Fig. 3. Process flow of the proposed chat robot

- Machine answering
  In this system, the chat engine API provided by NTT DOCOMO[14] is used. This function is activated when a user utterance ends but no new social comments are available.

The above three functions mainly contribute to natural response timing.

- Social media comment
  This function provides interesting responses written by humans interested in the topic. The system gathers social media comments related to the user's interests such as TV program the user is watching or a keyword spoken by the user. In this system, Twitter[15] is used as a resource.

3.2. Example of dialogue

Figure 4 shows an example of dialogue with our initial chat robot, while a user (a graduate school student) was watching sumo wrestling on TV. It illustrates the transcription with time stamps and which response function generates the utterances. It can be observed that the backchannel and repetition have natural timing (utterance id's 2, 3 and 7) and the machine answering compensates for the absence of social media comments (utterance id 8). Moreover, the system utterances based on social media comments (utterance id's 1 and 5) and machine answering (utterance id 8) successfully lead to subsequent user utterances.

4. EXPERIMENT

4.1. Method

We investigated how our proposed robot encourages users to speak. As we mentioned before, our goal is to encourage the elderly to speak, however, four graduate school students participated in this experiment instead of the elderly as a preliminary investigation. The subjects were asked to talk freely with two different chat robots while watching a TV program. One robot (Robot A) generated responses only using social media comments. The other (Robot B, proposed) generated responses using machine responses backchannel, repetition, machine answering) and social media comments.

4.2. Criteria

4.2.1. Number of utterances

The number of user utterances was used as an objective criterion. To investigate the effect of the chat robot clearly, only user utterances triggered by a robot utterance were counted.

4.2.2. Subjective scores

The subjects were also asked to evaluate the robots by answering two questions with scores 1 to 5 (1 for the lowest evaluation and 5 for the highest). Q1 is *"Was the dialogue natural?"*, which evaluates naturalness of the dialogue. Q2 is *"Would you like to talk with the robot again?"*. This evaluates whether the robot encourage a user to talk continuously or not.

4.3. Experimental conditions

Each subject was asked to watch pre-recorded TV program with both Robot A and Robot B. Each session was 15 minute long. The TV program in this experiment were presented using pre-recorded video files with synchronized SNS comments gathered while the TV programs were first-aired.

Table 1. Experimental tools

Chat robot	Unazuki-Kabochan[16]
Robot control unit	Raspberry Pi Model B+ (Memory: 512MB)
SNS	Twitter
Keyword extractor	Yahoo! JAPAN, Text analyzer
Chat engine	NTT DOCOMO, Zatsudan-Taiwa
Speech recognition	Julius 4.4.1
Speech synthesis	Voicetext[17]

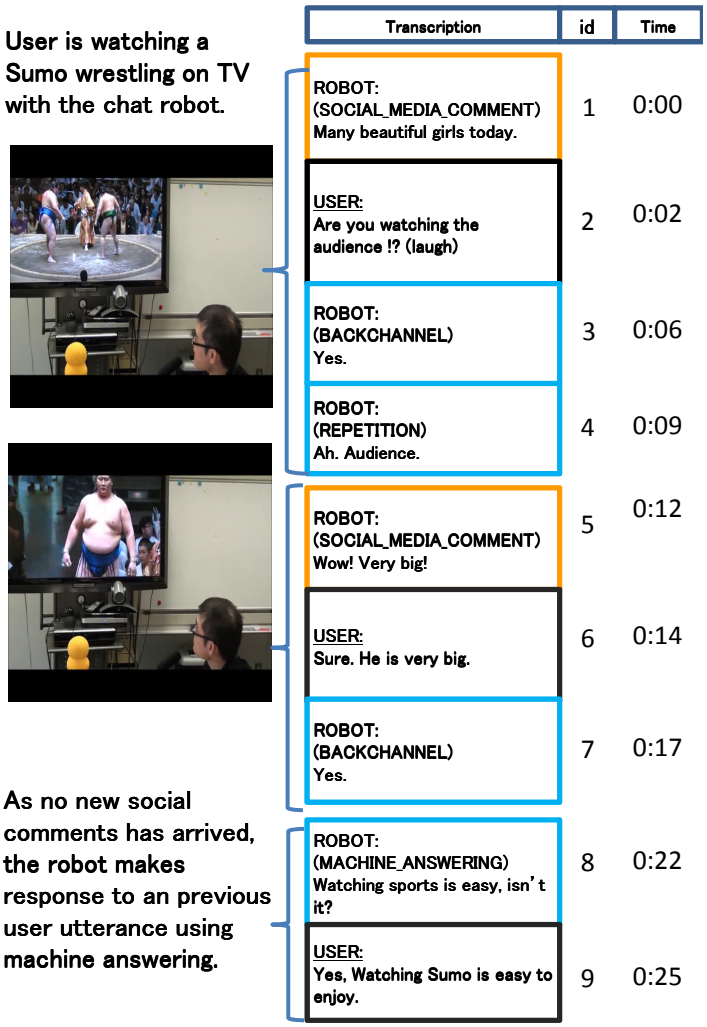

Fig. 4. Dialogue example of the proposed chat robot

In the experiment using Robot B, pre-stored society media comments were partly omitted from the database to simulate a condition that no social media comments are uploaded. This intends to investigate that machine response can substitute for social media comments, Silent periods are inserted at intervals of two minutes during a session and it continues one minute each.

The software and tools used in this experiment are listed in Table 1. As the robot body, "Unazuki-kabochan (Nodding Kabo)"[16] was used. The robot has a childlike body and its arms are controlled to move up and down repeatedly while it speaks. The mechanics of the robot were controlled by Raspberry Pi, and the other software executed on a laptop PC. The acoustic model and language model used in the speech recognition were standard speaker-independent models included in the Julius package.

The experimental setting is shown in Fig. 5. A video camera was installed behind a subject to record the experiment for future analysis.

5. RESULTS

5.1. Number of utterances

Figure 6 shows the average frequency of user utterances triggered by a chat robot per minute. The number of utterances when using Robot B (proposed) was significantly higher than that when using Robot A ($t(28) = 1.9 \cdot 10^{-4}$, $p < 0.05$ in T test).

Takahashi et al.[10] previously reported that a chat robot only using social media comments (Robot A in this experiment) encourages user utterances. This result shows that coupling social media comments and machine responses (backchannel, repetition and machine answering) further pro-

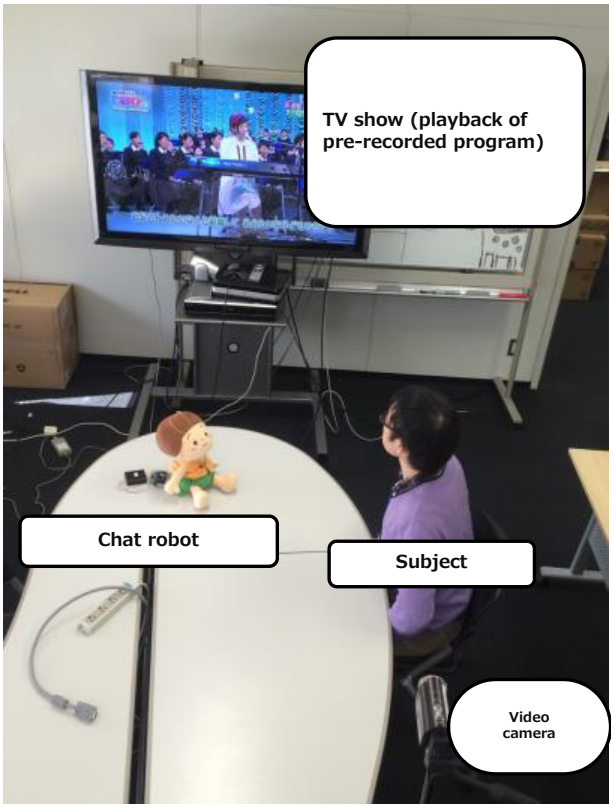

Fig. 5. Experimental setting

motes user utterances.

5.2. Subjective score

Result of question 1 is shown in Fig. 7. The average scores for Robot A and Robot B (proposed) were 1.25 and 3.75, respectively. A significant difference between the average scores are indicated by T-test ($t(6) = 0.4 \cdot 10^{-4}$, $p < 0.05$). Result of question 2 is shown in Fig. 8. There are also significant difference on the average scores ($t(6) = 0.01$, $p < 0.05$).

In addition, the subjects were interviewed by the tester after the experiment. One subject pointed out that although Robot A outputted interesting utterances, they were unnatural as part of a dialogue and that dialogue was discouraged since Robot A did not respond to the user.

6. CONCLUSION AND FUTURE WORK

In this paper, we proposed a chat robot that couples machine responses and comments from social media networks. From the experimental results using frequency of user utterances and the subjective score for the naturalness of dialogue, it is shown that the proposed chat robot is effective for prompting users to speak.

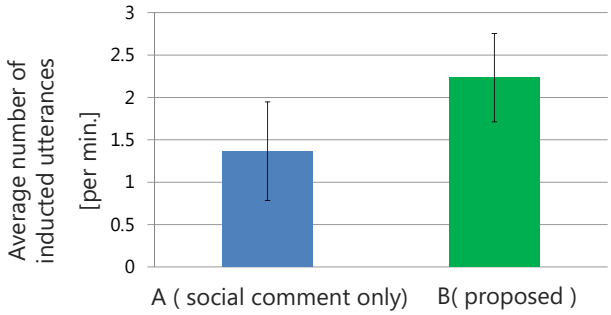

Fig. 6. Average numbers of utterances

In the future, we plan to carry out experiments to confirm our proposed system support a user involving elderly people to talk to others through SNS a real environment.

We will also improve the system to be suitable for the elderly. It is supposed that it is hard for the elderly to catch beginning of synthetic speech while concentrating on a TV. To solve this we plan to introduce a mechanism to attract a user's attention before the system speaks. Kendon revealed that importance of eye-contacting between speakers for turn-taking [18]. Based on this, turning a robot head before speaking mechanism will be introduced to the system.

In addition, we will investigate how our system can build and change relationship with its user through long-term operation.

7. REFERENCES

[1] Cabinet Office, Government of Japan, Annual Report on the Aging Society, 2014. `http://www8.cao.go.jp/kourei/whitepaper/w-2014/zenbun/26pdf_index.html` (in Japanese)

[2] Y. Nakanishi, H. Sugisawa and H. Ishikawa, "User Evaluation of Listening Volunteer for Housebound Elderly People: Examination in the Case of Interviewer," *Bulletin of Institute of Sociology and Social Work*, Meiji Gakuin University, No. 39, pp. 85–96, 2009. (in Japanese)

[3] H. Yamamoto, H. Miyazaki, T. Tsuzaki and Y. Kojima, "A Spoken Dialogue robot, Named Wonder, to Aid Senior Citizens Who Living Alone with Communication," *Journal of Robotics and Mechatronics*, vol.14, no.1, pp. 54–59, 2002.

[4] K. Miyazawa, T. Tokoyo, Y. Masui, N. Matsuo and H. Kikuchi., "Factors of Interaction in the Spoken Dialogue System with High Desire of Sustainability," *Trans. IEICE (A)*, Vol. 95, No. 1, pp. 27–36, 2012. (in Japanese)

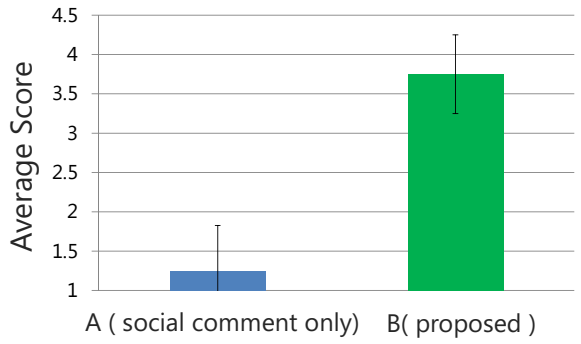

Fig. 7. Subjective scores for naturalness of dialogue timing

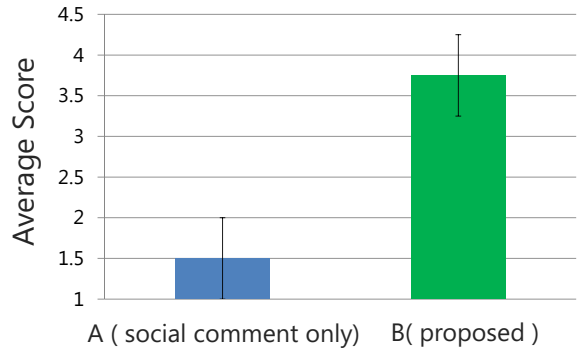

Fig. 8. Subjective scores for encouragement of dialogue

[5] D. Matsubara and H. Ueda., "A Robot Listening to Users' Grumble," *IEICE Technical report*, MVE, Vol. 111, No. 38, pp. 45–50, 2011. (in Japanese)

[6] NTT DOCOMO, "DOCOMO to Launch Shabette Concier? Voice-agent Application", `https://www.nttdocomo.co.jp/english/info/media_center/pr/2012/001580.html`.

[7] Y. Kobayashi, D. Yamamoto and S. Yokoyama., "Design Targeting Voice Interface Robot Capable of Active Listening," Proc. *HRI '10*, pp. 161–162, 2010.

[8] P. Dybala, M Ptaszynski, R. Rzepka and K. Araki, "Activating Humans with Humor – A Dialogue System That Users Want to Interact with," *IEICE Trans. Inf. & Syst.*, vol.E92-D, no.12, pp. 2394–2401, 2009.

[9] A. Ritter, C. Cherry, B. Dolan, "Data-Driven Response Generation in Social Media," *Proc. 2011 Conference on Empiricial Methods in Natural Language Processing (EMNLP)*, pp. 583–593, 2011.

[10] T. Takahashi, M. Kanbara and N. Hagita., "A Social Media Mediation Robot to Increase an Opportunity of Conversation for Elderly : Mediation Experiments Using Single or Multiple Robots," *IEICE Technical report*, CNR, Vol. 113, No. 84, pp. 31–36, 2013. (in Japanese)

[11] A. Lee and T. Kawahara., "Recent Development of Open-Source Speech Recognition Engine Julius," *Proc. APSIPA ASC*, pp. 131–137, 2009.

[12] H. Nishi and J. Kojima., "Automatic Telephone Answering System Based on Keyword Network Estimation Method," *IEICE Technical report*, Vol. 88, No. 92, pp. 71–77, 1988. (in Japanese)

[13] Yahoo! JAPAN Developer Network, Text Analysis, `http://developer.yahoo.co.jp/webapi/jlp` (in Japanese)

[14] DOCOMO Developer Support, Zatsudan-Taiwa, `https://dev.smt.docomo.ne.jp/?p=docs.api.page&api_docs_id=5` (in Japanese)

[15] "Welcome to twitter." `https://twitter.com/`

[16] PIP Co., Ltd., Unazuki-Kabochan (Chat robot), `http://www.pip-club.com/kabo/index.html` (in Japanese)

[17] HOYA Service Corporation, Voicetext, `http://voicetext.jp/` (in Japanese, English voice sample available)

[18] Adam Kendon, "Some functions of gaze-direction in social interaction," *Acta psychologica*, Vol. 26, pp. 22–63, 1967.