Effect of Mutual Self-Disclosure in Spoken Dialog System on User Impression

Shunsuke Tada*, Yuya Chiba*, Takashi Nose* and Akinori Ito*
* Graduate School of Engineering, Tohoku University, Japan
E-mail: {shunsuke.tada.q4@dc, yuya@spcom.ecei, tnose@m, aito@spcom.ecei}.tohoku.ac.jp
Tel/Fax: +81-022-795-7085

***Abstract*—Many of current spoken dialog systems can conduct non-task-oriented dialog. The systems that can improve user impression are required for users to keep using them. This paper focuses on self-disclosure, that is a process that a person reveals information about herself/himself to an interlocutor in human-human conversation. It is known that the self-disclosure plays a vital role to develop an intimate relationship. However, it is still unclear how exchanging the self-disclosures affects the user impression in the human-machine dialog. In this paper, we conduct dialog experiments to investigate the effectiveness of mutual self-disclosures between the user and the system. To achieve this goal, we built a spoken dialog system which conducts the dialog that the user and the system disclose information about themselves alternately. The dialog experiments revealed that the proposed system can improve the user impression regarding satisfaction and friendliness.**

I. Introduction

In recent years, non-task-oriented spoken dialog systems have been studied actively [1]–[4]. Many non-task-oriented systems do not have a clear goal of the dialog, and try to keep a chat-like talk. It is known that such short talks are useful even in task-oriented dialog. The non-task-oriented systems are usually evaluated by subjective measures, such as user impression and satisfaction, because those systems have no clear objective measure to evaluate the performance.

Therefore, several studies focused on developing methods to improve the user impression. For example, Kobori *et al.* [5] reported that the introduction of the small talk during an interview improves the user impression. Besides, Miyashita *et al.* [6] conducted research that increases the user's intention to talk with the system by gradually increasing the behavior of a robot that expresses intimacy. Kim *et al.* [7] conducted experiments of human-robot interaction in Korean, and indicated that the robot is perceived more friendly when the robot call the user in the familiar form. Kageyama *et al.* [8] reported that the user impression improves by gradually changing form of the system's utterance. From these studies, the relationship between the system and the user seems to have a significant impact on the user impression.

This study focuses on self-disclosure to improve the user impression. Self-disclosure is defined as "an act of revealing personal information to others" [9] in the conversation. In the human-human dialog, the self-disclosure plays a vital role in the course of interpersonal intimacy. For example, Altman and Taylor regarded that individuals are getting more comfortable to talk about the private and personal matter as relationship become closer [10]. Besides, a person who discloses the information about herself/himself expects a recipient to return her/his information to the same degree [11], [12]. Therefore, exchanging the self-disclosures is expected to be useful to build the close relationship between the user and the system, and improve the user impression. The effects of self-disclosure in the human-machine dialog have been studied thus far [13], [14]. However, it is still unclear how exchanging self-disclosures affects the user impression.

In this paper, we aim to investigate the effect of the mutual self-disclosure in the human-machine dialog by dialog experiments. To achieve this goal, we built the system that conducts the dialog where the user and the system disclose themselves each other. The dialog system is constructed by combining the example-based approach and the template-based approach. In Sections II and III, we describe the details of an experimental system. Then, results of dialog experiments are reported in Section IV. Finally, the paper is concluded in Section V.

II. Dialog system for mutual self-disclosure

A. Overview of the experimental system

To construct the target dialog system, we separated the dialog conducted by the system into two sections. One is a system's self-disclosure section: the user asks the questions and the system discloses the information about itself. The other is a user's self-disclosure section: the system asks the questions and the user discloses the information about herself/himself. The example-based approach and the template-based approach are exploited to achieve dialogs of the former and latter sections, respectively. The system conducts two dialog sections in the fixed order to achieve the mutual self-disclosure.

In the experiments, we selected "travel" as a topic of the dialog. The users talk about three subtopics in the dialog: 1) a place a speaker traveled recently, 2) the best place the speaker has traveled to, and 3) a place the speaker wants to travel next. A dialog example of the proposed system is shown in Table I.

B. System's self-disclosure section

The example-based approach is used for the system's self-disclosure section. The example-based approach is a popular one to realize the non-task-oriented system. The system first calculates the similarities between the user utterance and example sentences in the database; then selects a response

TABLE I
EXAMPLE OF DIALOG WITH THE SYSTEM FOR MUTUAL SELF-DISCLOSURE

User's self-disclosure section	
Sys	Let's talk about travel. Which is the best place you have visited?
Usr	Spain.
Sys	With whom did you go to Spain?
Usr	College friends.
Sys	Do you get along with the colleague friends?
Usr	Yes.
System's self-disclosure section	
Sys	Good. By the way, I have been to Thailand. It was a really good place.
Usr	Where did you go?
Sys	I went to Thailand at February last year.
Usr	How was the climate of Thailand?
Sys	It was very hot and humid. I felt uncomfortable.
Usr	What did you eat in Thailand?
Sys	I recommend a traditional fried noodle called *pad thai*.

TABLE II
EXAMPLE OF FOCUS ANNOTATED BY IOB2 FORMAT

Sentence	*Hokkaido*	*no*	*Otaru*	*ni*	*iki*	*mashi*	*ta*
	(I visited Otaru in Hokkaido.)						
Focus labels	B-F	I-F	I-F	O	O	O	O

A sentence of italic type is original Japanese sentence, and a sentence in brackets is translation. Underlines represent a focus phrase.

TABLE III
EXAMPLES OF THE TEMPLATE FOR FOLLOW-UP QUESTION

Question category	Template	Focus category
Place	Where did you go in [FW]?	Place
Place	What did you eat in [FW]?	Food
Person	With whom did you go to [FW]?	Person
Food	Did [FW] taste good?	Other
Person	Do you often travel with [FW] ?	Other

[FW] denotes placeholders.

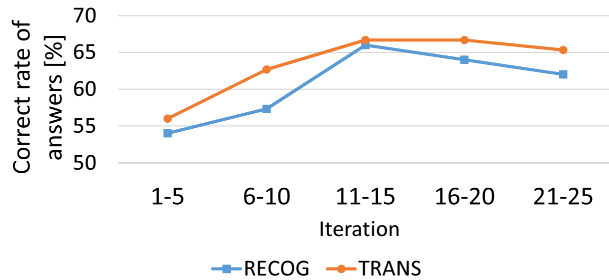

Fig. 1. Correct rate of answers to the number of the dialog. The value is the average of subtopics per 5 interactions.

corresponding to the most similar example. The cosine similarity is typically used for the similarity calculation. Let $\boldsymbol{q}$ be the one-hot vector of the user utterance, $\boldsymbol{d}$ be the one-hot vector of an example sentence, and q_i, d_i be the i-th element of $\boldsymbol{q}$, $\boldsymbol{d}$. The cosine similarity is calculated as follows:

$$\cos(\boldsymbol{q}, \boldsymbol{d}) = \frac{\boldsymbol{q} \cdot \boldsymbol{d}}{|\boldsymbol{q}||\boldsymbol{d}|} = \frac{\sum_i q_i d_i}{\sqrt{\sum_i q_i^2}\sqrt{\sum_i d_i^2}} \quad (1)$$

In the experiments, we prepared the example-response database subtopic by subtopic.

C. User's self-disclosure section

The user's self-disclosure section exploits the template-based dialog approach. The system asks not only predefined questions, but also follow-up questions related to a focus detected from the user utterance. Here, the focus is defined as nouns and noun phrases that the system can ask the follow-up questions [15], [16], and detected based on the sequential labeling. Table II shows an example of the focus represented by the IOB2 format. B-F and I-F denote the beginning of the focus and the inside of the focus, respectively.

In the user's self-disclosure section, the system poses the predefined question at first. When detecting the focus in the user utterance, the system makes the follow-up question by substituting the detected focus for a placeholder of a template. Here, each template has a question category as shown in Table III, and the system selects the template which has the question category same to the focus category. The focus category is determined by the template of the system question. We prepared about 17 templates for each subtopic, and five focus categories: Place, Food, Person, Vehicle, and Other.

III. ENHANCEMENT OF MODULES CONSTRUCT EXPERIMENTAL SYSTEM

A. Collection of example-response pairs

The example-response databases were constructed according to the procedure conducted in [17]. In this method, the example-response pairs are created based on the dialog with the user. It is reported that the system can collect appropriate examples for the actual conversation effectively.

We prepared an initial database that contained 300 example-response pairs for each subtopic. 25 users participated in the example collection. Fig. 1 shows the change of the rate of the answers evaluated as appropriate by the users. RECOG denotes the values when using the speech recognition results for the user input, and TRANS denotes the values when using transcriptions. The graph shows that the rate of the appropriate answers improved as increasing the number of the dialog. However, the correct rate plateaued at 11–15th dialog, and the correct rate of answers was 64% for RECOG and 67% for TRANS. The results were lower than the previous study [17] because we created the responses so that the variation of the user utterance became large. We collected 681 pairs for the subtopic 1), 681 pairs for the subtopic 2), and 636 pairs for the subtopic 3).

B. Training of focus detection model

Yoshino *et al*. [15] proposed a method using the conditional random field (CRF) and phrase-level features for the focus detection. In particular, they employed the features related to the predicate argument (P-A) structure of the sentence. However, the P-A structure analysis is not easy for the utterances in the chat-like talk. Therefore, we used word-level features to detect the focus. The feature set is summarized in Table IV.

TABLE IV
PROPOSED FEATURE SET FOR FOCUS DETECTION

Feature type	Feature
WORD	uni-gram and bi-gram of the previous, current, and next word
POS	uni-gram and bi-gram of POS tags of the previous, current, and next word
Sub-POS	uni-gram and bi-gram of sub-categories of the POS tags of the previous, current, next word

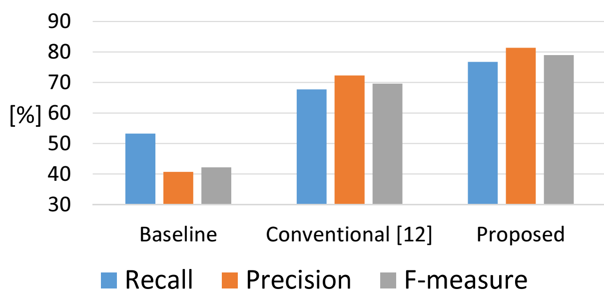

Fig. 2. Results of focus detection

We compared the performance of the focus detection between the proposed feature set and the conventional method. We used the response sentences collected in Section III-A as the training data. The number of samples was 548, excluding overlap. We used user utterances collected on the wizard of Oz basis as the test data. The number of samples of the test data was 90. Three evaluators annotated the labels of the focus. The evaluators were instructed to append the labels according to the definition of the focus mentioned in Section II-C. Here, Fleiss' κ of the labels are 0.539 that denotes medium agreement. The union of the labels was employed as the definitive labels because some sentences had multiple focuses. The CRF Suite[1] was used to train the model, and the hyper parameters were decided by the grid-searching.

Fig. 2 shows the detection results. In addition to the conventional method, we added a simple baseline result, where all of the nouns we regarded as focuses. As shown in the graph, the proposed feature set gave the best result. In particular, we observed improvements in the utterances that include multiple focuses such as "I ate a <u>*Ningyo yaki*</u>, a <u>cheesecake</u>, and a <u>fried chicken</u>" (the underlines denote focuses) by using our feature set.

IV. Dialog Experiments

A. Prepared systems

In addition to the proposed system, we prepared two systems for comparison. The three systems are summarized as below:

SYSTEM:
A system which only conducts the system's self-disclosure section. The user asks the questions and the system discloses the information about itself.

USER:
A system which only conducts the user's self-disclosure section. The system asks the questions and the user discloses the information about herself/himself.

MUTUAL:
A system which conducts two sections alternately. Both interlocutors disclose the information about themselves.

The experimental systems employed Google Speech API [2] for the speech recognition, and projected a female agent during dialog.

B. Experimental procedure

The dialog was conducted on the assumption that the participants talk with the system about their experience of travel. Each participant talked with one of the systems mentioned above. We controlled the experiments so that the participants interchange the same number of the utterances. The total number of user utterances was fixed to 12 with any systems.

Fig. 3 shows an overview of the dialog with each system. SYSTEM and USER conduct the dialog about all of three subtopics. On the other hand, MUTUAL only conducts the dialog about two of the three subtopics because it has to conduct two sections in each subtopic. Here, MUTUAL system conducted the user's self-disclosure section first, then conducted the system's self-disclosure section. Two subtopics were selected randomly. At the end of each section, the systems encouraged the user to change the subtopic or section.

C. Subjective evaluation

At the end of the experiments, the participants answered the following five questions using the five-grade Likert scale, one (not at all) to five (very much).

Satisfaction:
How the participant was satisfied with the dialog.

Friendliness:
How friendly the participant felt the dialog system is.

User's intention of talk:
How strongly the participant wants to talk.

System's intention of talk:
How strongly the participant felt the system wants to talk.

Naturalness:
How natural the participant felt the dialog is.

The participants also evaluated the responses of the system between "appropriate," "acceptable" and "not acceptable." Ten participants conducted the dialog with each system, and the total number of the participants was 30 (Female: 6, Male: 24).

D. Experimental results

Fig. 4 shows the results of the subjective evaluation. As a result, MUTUAL obtained the best score in terms of Satisfaction

[1] http://www.chokkan.org/software/crfsuite/

[2] https://cloud.google.com/speech/?hl=ja

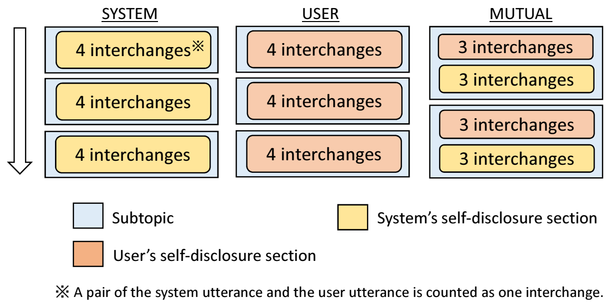

Fig. 3. Overview of dialog experiments. USER and SYSTEM conduct dialog about three subtopics, and MUTUAL only conducts two of the three subtopics.

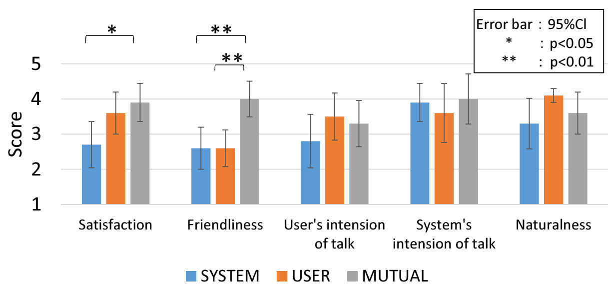

Fig. 4. Average scores of subjective evaluation

and Friendliness. We observed 5% significant difference at Satisfaction and 1% significant difference in Friendliness by the one-way layout ANOVA factoring the system condition. Then, the Bonferroni's multiple comparison tests were conducted to investigate the difference between the systems. We obtained 5% significant difference between MUTUAL and SYSTEM in Satisfaction, and 1% significant difference between MUTUAL and other two systems in Friendliness. In the experiments, many of the participants reported that they felt much closer to the system by exchanging the self-disclosures although they involved in the short interaction. These results suggest that the mutual self-disclosure is effective to improve the user impression also in the human-machine dialog. On the other hand, the participants that talked with USER reported that they felt easy to talk to because the system asks the questions. Few participants felt that one-sided conversation was uncomfortable, while some participants reported that the dialog with USER is like an interview. These opinions seem to be the reason why the score of USER was lower than that of MUTUAL in terms of Friendliness.

Fig. 5 shows the result of subjective evaluation of the utterances given by three systems. In the graph, "Correct" shows the percentage of utterances evaluated as "appropriate," and "Acceptable" shows that evaluated as either "appropriate" or "acceptable." As shown in the figure, the scores of the USER was about ten points higher than that of the other systems. Considering that a system made twelve utterances in one dialog, the point improvement of the appropriateness of answers was worth one utterance.

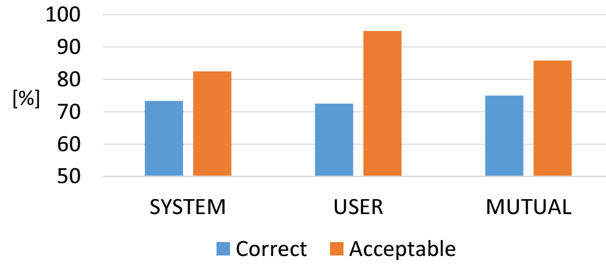

Fig. 5. Rate of appropriate/acceptable answers of each system

V. Discussion

The experiment in Section IV indicated that the mutual self-disclosure improves the user impression regarding the satisfaction and the friendliness. Therefore, it is expected that the spoken dialog system can build a better relationship with the user by implementing mutual self-disclosure. However, there are several remaining issues regarding introducing the mutual self-disclosure to the actual dialog system.

The current our system switches sections and subtopics at a fixed time. However, such interaction is unnatural in the actual dialog, and the system has to be able to judge the timing of the topic switching in accordance with the flow of the conversation. Various studies examined the topic switching in the dialog with the system (e.g., [18], [19]), and the methods proposed in such studies are considered to be effective to achieve natural section switching of our system. The analysis of the human-human dialog is also effective to build the natural topic switching model. Another problem is a generation of the self-disclosure sentence. The proposed system employed the example- and the template-based approaches for the response generation. The results showed most of the system responses generated in these approaches were appropriate or acceptable. However, more large-scale example-response database is required to generate self-disclosure utterances of the arbitrary topic. For the non-task-oriented dialog system, Sugiyama et al. examined the approach that manually creates the pairs to construct the large-scale example-response database [20]. In addition, Kageyama et al. also examined the approach that collects the pairs from the actual conversation [17]. We have to consider such studies and create the database to achieve natural and flexible response generation in the system self-disclosure section.

VI. Conclusions

This study investigated the effectiveness of the mutual self-disclosure in the human-machine dialog. We developed a spoken dialog system and conducted dialog experiments where the user and the system exchange the information about themselves alternately. Example- and template-based approaches were employed in the sections of the system's self-disclosure and the user's self-disclosure respectively. The experimental results revealed that the mutual self-disclosure can improve the user impression regarding the satisfaction and the friendliness compared to the dialog that either of them discloses the information.

In future work, we will examine methods to switch the subtopics and the sections naturally at an appropriate timing by analyzing the human-human dialog.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Numbers JP15H02720, JP16K13253, JP17H00823, JP18K18136.

References

[1] T. Bickmore and R. Picard, "Establishing and maintaining long-term human-computer relationships," *ACM Transactions on Computer-Human Interaction*, vol. 12, no. 2, pp. 297–327, 2005.

[2] T. Meguro, R. Higashikata, Y. Minami, and K. Dohsaka, "Controlling listening-oriented dialogue using partially observable Markov decision processes", In *Proc. the 23rd International Conference on Computational Linguistics*, 2010, pp. 761–769.

[3] Z. Yu, L. Nicolich-Henkin, A. W. Black, and A.Rudnicky, "A wizard of Oz study on a non-task-oriented dialog systems that reacts to user engagement.", In *Proc. the 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, 2016, pp. 55–63.

[4] P. Milhorat, D. Lala, K. Inoue, T. Zhao, M. Ishida, K. Takanashi, S. Nakamura, and T. Kawahara, "A conversational dialogue manager for the humanoid robot ERICA", In *Proc. the Ninth International Workshop on Spoken Dialogue Systems Technology*, 2017, pp. 1–12.

[5] T. Kobori, M. Nakano, and T. Nakamura, "Small talk improves user impressions of interview dialogue systems", In *Proc. the 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, 2016, pp. 370–380.

[6] Z. Miyashita, T. Kanda, M. Shiomi, H. Ishiguro, and N. Hagita, "A robot in a shopping mall that affectively guide customers", *Journal of Robotics Society of Japan*, vol. 26, no. 7, pp. 821–832, 2008.

[7] Y. Kim, S. Kwak, and M. Kim, "Am I acceptable to you? Effect of a robot's verbal language forms on people's social distance from robots," *Computers in Human Behavior*, no. 29, pp. 1091–1101, 2012.

[8] Y. Kageyama, Y. Chiba, T. Nose, and A. Ito, "Improving User Impression in Spoken Dialog System with Gradual Speech Form Control," *Proc. of the 19th Annual SIGdial Meeting on Discourse and Dialogue*, pp. 235–240, 2018.

[9] S.M. Jourard, "Self-disclosure: An experimental analysis of the transparent self," New York: Wiley-Interscience, 1971.

[10] I. Altman and D. Taylor, "Social penetration: The development of interpersonal relationships," New York: Holt, Rinehart & Winston, 1973.

[11] M. Worthy, A.L. Gary, and G.M. Kahn, "Self-disclosure as an exchange process," *Journal of Personality and Social Psychology*, vol. 13, no. 1, pp. 59–63, 1969.

[12] J.A. Kohen, "The development of reciprocal self-disclosure in opposite sex interaction," *Journal of Counseling Psychology*, vol. 22, no. 5, pp. 404–410, 1975.

[13] A. Ravichander and A.W. Black, "An Empirical Study of Self-Disclosure in Spoken Dialogue Systems," In *Proc. the 19th Annual SIGdial Meeting on Discourse and Dialogue*, pp. 253–263, 2018.

[14] S.Y. Lee and J. Choi, "Enhancing user experience with conversational agent for movie recommendation: Effects of self-disclosure and reciprocity," *International Journal of Human-Computer Studies*, vol. 103, pp. 95–105, 2017.

[15] K. Yoshino and T. Kawahara, "Information navigation system based on POMDP that tracks user focus," In *Proc. the 15th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, 2014, pp. 32–40.

[16] S. Tada, Y. Chiba, T. Nose, and A. Ito, "Response classification of interview-based dialog system using user focus and semantic orientation," in *Proc. the International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, 2017, pp. 84–90.

[17] Y. Kageyama, Y. Chiba, T. Nose, and A. Ito, "Analyses of example sentences collected by conversation for example-based non-task-oriented dialog system," *IAENG International Journal of Computer Science*, vol. 45, no. 1, pp. 285–293, 2018.

[18] X. Song, K. Maeda, H. Kunimasa, H. Toyota, and D. Han, "Topic control in a free conversation system," *Proc. International Conference on Natural Language Processing and Knowledge Engineering*, 2009, pp. 1–6

[19] N. Glas, K. Prepin, and C. Pelachaud, "Engagement driven topic selection for an information-giving agent," *Proc. Workshop on the Semantics and Pragmatics of Dialogue*, 2015, pp. 1–11

[20] H. Sugiyama, T. Meguro, R. Higashinaka, and Y. Minami, "Large-scale collection and analysis of personal question-answer pairs for conversational agents," *Proc. International Conference on Intelligent Virtual Agents*, 2014, pp. 420–433