Improving User Impression in Spoken Dialog System with Gradual Speech Form Control

Yukiko Kageyama, Yuya Chiba, Takashi Nose, and Akinori Ito
Graduate School of Engineering, Tohoku University
6-6-05, Aoba, Aramaki-Aza, Aoba-ku, Sendai, Miyagi, Japan
{kageyama@spcom, yuya@spcom, tnose@m, aito@spcom}.ecei.tohoku.ac.jp

Abstract

This paper examines a method to improve the user impression of a spoken dialog system by introducing a mechanism that gradually changes form of utterances every time the user uses the system. In some languages, including Japanese, the form of utterances changes corresponding to social relationship between the talker and the listener. Thus, this mechanism can be effective to express the system's intention to make social distance to the user closer; however, an actual effect of this method is not investigated enough when introduced to the dialog system. In this paper, we conduct dialog experiments and show that controlling the form of system utterances can improve the users' impression.

1 Introduction

Demand for a spoken dialog system has raised, including AI speakers or personal assistant systems (Bellegarda, 2014). Not only the conventional task-oriented dialog systems (Aust et al., 1995; Zue et al., 2000), but also non-task-oriented systems (Bickmore and Picard, 2005; Meguro et al., 2010; Yu et al., 2016; Akasaki and Kaji, 2017) have attracted the attention in recent years. In order for such dialog systems to become ubiquitous in the society, it is important to improve the user impression to the dialog with the system.

Miyashita et al. (2008) conducted a research that increases the user's intention to talk with the system by gradually increasing the behavior of a robot that expresses intimacy. Their study showed that the user felt the robot more friendly and increased desire to use the robot continuously by the robot's behavior. This research showed that, expressing intimacy with the user is effective to promote the user's desire to use the system.

In this research, we focused on a linguistic form of system utterances to improve the user impression. Several languages, including Japanese, have a mechanism called "honorifics" by which the speech form changes according to the relative social position or closeness of the social distance to the dialog partner (Brown and Ford, 1961). The honorific is often treated as one of the categories of politeness (Brown and Levinson, 1978, 1987) although several arguments have been raised (Ide, 1989; Agha, 1994). Brown and Levinson (1987) claimed that the speaker can choose strategy according to the politeness level depending on the social distance or relative power between the speakers. In Japanese, the speakers try to close the social distance by gradually decreasing the use of honorific form.

This paper examines effectiveness of introducing such mechanism to the dialog system. Kim et al. (2012) conducted experiments of human-robot interaction in Korean language, and indicated that the robot is perceived more friendly when calling the user in the familiar form, but the effect of the speech form itself was limited. In contrast, we investigate the effect of changing speech form on the user impression including the friendliness.

2 Changing Form of System Utterances Considering Social Distance

2.1 Expressions of Japanese for social distance, politeness and familiarity

This study exploits the expressions of Japanese that express politeness and social distance between the talker and the listener. Thus, we first explain such mechanism of Japanese briefly. The Japanese language has a system of speaking form

Proceedings of the SIGDIAL 2018 Conference, pages 235–240,
Melbourne, Australia, 12-14 July 2018.

called "the honorifics (*keigo*)", that indicates social relationships between the speaker and the listener or the speaker and the persons referred in the utterance using the linguistic form. For example, the verb *tsukuru* (to make) can be used as either *tsukuru* (normal form) or *tsukuri-masu* (polite form). Another way of expressing closeness is to use the ending particles, such as *tsukuri-masu* (polite, far) or *tsukuri-masu-yo* (polite, closer). In addition to the honorifics, it is possible to express closeness using different wording, such as *hai* (a positive answer or a backchannel, polite) and *un* (casual). When the interlocutors are familiar with each other, the form of utterances become less polite, closer and more casual. In this experiment, we defined "honorific form" as polite, less close and formal expressions, and "normal form" as less polite, closer and casual expressions.

2.2 Gradual control of system speech form based on speech level shift

The changes of the speech form are caused by several factors, such as the social entrainment (Hirschberg, 2008). One of the main factors is the changes of the social distance. When two persons make conversations several times, it was mentioned that the proportion of honorific form decreases, and that of normal form increases as they make more conversations (Ikuta, 1983). This phenomenon is called "speech level shift" or "speech style shift" (Ikuta, 1983; Hasegawa, 2004). The "speech level" or "speech style" means the expressions in the utterances that express closeness of the interlocutors. Thus, the "speech level shift" means the switching of speech level that occurs in conversations between the same persons.

To make the dialog system express that the system and the user gradually become more friendly, we propose a method to use the speech level shift. In the experiment, the subjects talked with the system for three consecutive days and evaluated the impression on the system and the dialog with the system. We changed the speech level step by step within the three-day experiment, as shown in Table 1. In Japanese, it is natural to use the honorific form when persons meet for the first time; thus, all of the system utterances were in the honorific form in the first conversation.

	Proposed system	
	Honorific	Normal
Day 1	100%	0%
Day 2	50%	50%
Day 3	0%	100%

Table 1: The ratio of utterance form corresponding to day of experiment for proposed system

3 Experimental Dialog System

3.1 System architecture

An experimental system is based on an example-based dialog system (Takeuchi et al., 2007; Lee et al., 2009) commonly used for the non-task-oriented system. A computer-based female agent was employed. In the example-based dialog system, the system calculates the similarities between the user's utterance and example sentences in the database, and then selects a response corresponding to the most similar example. This study employed the cosine similarity for the similarity calculation.

3.2 Topic-dependent example-response database for non-task-oriented dialog

The example-response databases for the experiments were constructed through the actual dialogs with the system and users (Kageyama et al., 2017). We focused on chatting between friends, which is one of the non-task-oriented dialog, and prepared four databases corresponding to the different dialog topic. To collect the dialog data, the users asked the agent what she had done yesterday on the assumption that she had led a human-like life in the dialog collection. The topics of the database were cooking, movies, and meal. A dialog example is appended at **Appendix A**. The number of pairs included in the constructed database was ranged from 1,000 to 1,125. The responses of the system were composed in the honorific form.

3.3 Preparation of the system utterances in normal form

The databases of the normal form were constructed by rewriting the form of the response sentences of the collected databases. 26 persons rewrote the sentences into the normal form. In the rewriting, the rewriting rules shown at **Appendix B** were provided to the rewriters for the consis-

tency.

4 Dialog Experiments by Gradually Changing Expression

4.1 Experimental condition

The experiments were conducted in a sound-proof chamber for 3 consecutive days. The participants interacted with the system once a day, where a participant made 10 utterances to control the number of interchanges. The topic of the conversation was different from day to day, where the order of the topics was randomly determined from participant to participant. The rate of the system utterances in the honorific and normal form was changed according to Table 1. After the conversation, they evaluated the impression on the spoken dialog system using a questionnaire. For comparison, we prepared the dialog systems speaking in only the honorific form and the normal form in all three days. These two systems are denoted as "Honorific" and "Normal" hereafter. In the experiments, 14 participants talked with one of the three systems, and thus the total number of the participants was 42 (3 systems $\times$ 14 participants). Each group contained 7 male and 7 female participants.

We first presented the participants all the topics the dialog system could handle, and the participants were instructed to ask what the agent did yesterday for the specific topic. We also presented a dialog example to the participants. Then the participants made conversation with the system on the presented topic. The participants were allowed to make self-disclosure utterances.

We expected the system and the participant made conversations within the given topic, but the conversation broke down when the participant made an unanticipated utterance. The participants were instructed to talk with the system until making the specified number of utterances even when the conversation broke down.

4.2 Procedure of dialog experiments

The experimental procedure is as below:

Step 1: The topic is announced to the participant.

Step 2: The participant asks the system what the agent did yesterday.

Step 3: The participant made 10 interchanges with the system.

Step 4: The participant answered a questionnaire on the impression of the dialog.

	Day 1	Day 2	Day 3	Total
Proposed	67.1	72.1	70.7	70.0
Honorific	65.0	71.4	73.6	70.0
Normal	69.3	67.9	66.4	67.9

Table 2: Rate of correct answer [%]

Step 5: The steps 1 to 3 were repeated for 3 consecutive days changing the topic every day

4.3 Evaluation method

At the end of the every conversation, the participants answered the following four 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

Impression of speech form: How adequate the participant felt of the system's speech form

Intention of talk: How strongly the participant wants to use the system again

In addition, we asked the participants who talked with the proposed system, whether they noticed the changes of the speech form or not after the last experiment.

5 Analysis of Experimental Results

5.1 Analysis of response rates

Table 2 shows the rates of the correct answers made by the system in the experiments. The correctness was judged by the participant based on the naturalness of the response to the question.

As shown in the table, the rate of correct answer of each system through three days experiments is about 70%, and this is almost equal to the previous results (Kageyama et al., 2017). From the one-way layout ANOVA factoring the condition of speech form, the significant difference was not observed. Therefore, the effect of response error in the subjective evaluation is considered to be almost equal between systems.

5.2 Experimental results of subjective evaluation

Figure 1 shows the average scores of the subjective evaluation per day. The graph shows that the subjective scores of the proposed system tend to

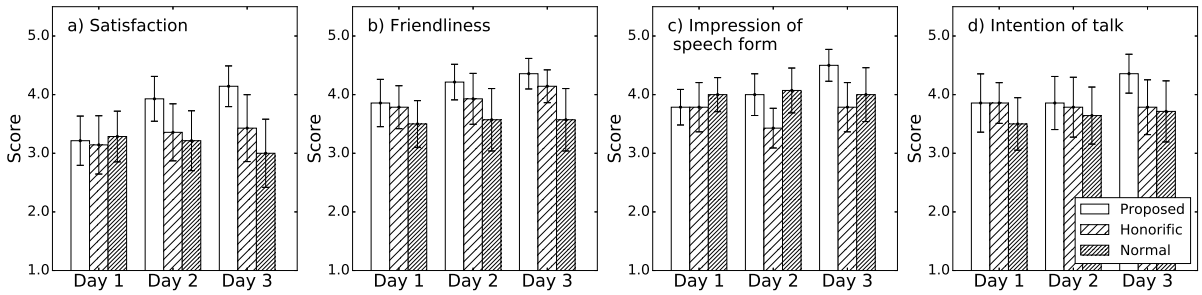

Figure 1: Average scores of subjective evaluation per day (error bar: 95% confidential interval)

	Satisfaction		Friendliness		Impression of speech form	
	Mean diff. (95%CI)	p-value	Mean diff. (95%CI)	p-value	Mean diff. (95%CI)	p-value
Proposed - Honorific	0.45 (-0.04, 0.93)	0.07	0.19(-0.20, 0.59)	0.49	0.43 (0.06, 0.80)	0.02*
Proposed - Normal	0.60 (0.12, 1.07)	0.01*	0.60 (0.20, 0.99)	<0.01**	0.07 (-0.30, 0.44)	0.89
Normal - Honorific	-0.14 (-0.62, 0.33)	0.76	-0.40 (-0.800, -0.01)	0.04*	0.36 (-0.01, 0.72)	0.06

Table 3: Results of Tukey-Kramer multiple-comparison test (Mean diff.: difference of average score, CI: confidence interval, *$p < 0.05$, **$p < 0.01$)

increase day by day, whereas those of the "Honorific" and the "Normal" systems tend to be flat. The scores of "Proposed" and "Honorific" are almost same at the first day because the all of utterances conducted in the honorific form. Interestingly, we can observe the difference between the scores of "Proposed" and "Normal" at Day 3 even both systems spoke in the same form. This result reflects that the effect of the changing form of the utterance by number of interactions.

Here, we conducted the two-way layout ANOVA to compare the condition of the speech form and the number of the interaction, and obtained the significant difference at the speech form factor in Satisfaction ($p \leq 0.01, F = 3.07$), Impression of speech form ($p = 0.01, F = 3.07$), and Friendliness ($p \leq 0.01, F = 3.07$). Then, we conducted the Tukey-Kramer tests to investigate the difference between the conditions. The results are summarized in Table 3.

As shown in the table, "Proposed" surpassed "Honorific" in terms of Impression of speech form, and surpassed "Normal" in terms of Satisfaction and Friendliness. These results suggest that the proposed system tends to obtain the better subjective score comparing to the simple systems without changing the form of utterance.

5.3 Perception of changes of speech form

In the experiments, 5 out of 14 participants that used the proposed system did not perceive the changes of the speech form. Here, we compared

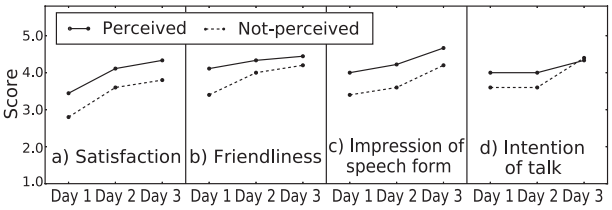

Figure 2: Score depending on perception

the scores between the groups of participants who perceived (denoted as "Perceived") and did not perceive (denoted as "Not-perceived") the changes of the form. Figure 2 shows the variation of the average scores of each group. From the figure, we can observe that all of the subjective scores of "Not-perceived" tend to increase as same with the scores of "Perceived." This result suggests that it is possible that the proposed method is able to improve the user impression unconsciously.

6 Conclusion

In this paper, we examined a method to improve the user impression by changing the form of system utterance according to number of uses. The dialog experiments showed that the proposed method can improve the subjective scores, such as the satisfaction compared to the simple systems unchanging the speech form, even the user could not perceive the changes of the expression.

In a future work, we will examine a method to change the form of the sentences considering the relationship between the speakers (Li et al., 2016).

Acknowledgments

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

References

Asif Agha. 1994. Honorification. *Annual Review of Anthropology*, 23:277–302.

Satoshi Akasaki and Nobuhiro Kaji. 2017. Chat detection in an intelligent assistant: Combining task-oriented and non-task-oriented spoken dialogue system. *arXiv preprint arXiv:1705.00746*.

Harald Aust, Martin Oerder, Frank Seide, and Volker Steinbiss. 1995. The Philips automatic train timetable information system. *Speech Communication*, 17(3–4):249–262.

Jerome R Bellegarda. 2014. Spoken language understanding for natural interaction: The Siri experience. In *Natural Interaction with Robots, Knowbots and Smartphones*, pages 3–14. Springer.

Timothy Bickmore and Rosalind Picard. 2005. Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction*.

Penelope Brown and Stephen C Levinson. 1978. *Questions and politeness: strategies in social interaction*, chapter Universals in language usage: Politeness phenomena. Cambridge University Press.

Penelope Brown and Stephen C Levinson. 1987. *Politeness: Some universals in language use*, volume 4. Cambridge University Press.

Roger Brown and Marguerite Ford. 1961. Address in American English. *Journal of Abnormal and Social Psychology*, 62(2):375–385.

Yoko Hasegawa. 2004. Speech-style shifts and intimate exaltation in Japanese. In *Proc. of the 38th Annual Meeting of the Chicago Linguistic Society*, pages 269–284.

Julia Hirschberg. 2008. Speaking more like you: lexical, acoustic/prosodic, and discourse entrainment in spoken dialogue systems. In *Proc. of the 9th SIGdial Workshop on Discourse and Dialogue*, pages 128–128.

Sachiko Ide. 1989. Formal forms and discernment: Two neglected aspects of universals of linguistic politeness. *Multilingua*, 8(2–3):223–248.

Shoko Ikuta. 1983. Speech level shift and conversational strategy in Japanese discourse. *Language Sciences*, 5(1):37–53.

Yukiko Kageyama, Yuya Chiba, Takashi Nose, and Akinori Ito. 2017. Collection of example sentences for non-task-oriented dialog using a spoken dialog system and comparison with hanf-crafted DB. In *Proc. HCI International*, pages 458–563.

Yunkyung Kim, Sonya S. Kwak, and Myung-suk Kim. 2012. Am I acceptable to you? Effect of a robot's verbal language forms on people's social distance from robots. *Computers in Human Behavior*, 29:1091–1101.

Cheongjae Lee, Sangkeun Jung, Seokhwan Kim, and Gary Geunbae Lee. 2009. Example-based dialog modeling for practical multi-domain dialog system. *Speech Communication*, 51(5):466–484.

Jiwei Li, Michel Galley, Chris Brockett, Georgios Spithourakis, Jianfeng Gao, and Bill Dolan. 2016. A persona-based neural conversation model. In *Proc. of the 54th Annual Meeting of the Association for Computational Linguistics*, pages 994–1003.

Toyomi Meguro, Ryuichiro Higashinaka, Yasuhiro Minami, and Kohji Dohsaka. 2010. Controlling listening-oriented dialogue using partially observable Markov decision processes. In *Proc. of the 23rd International Conference on Computational Linguistics*, pages 761–769.

Zenta Miyashita, Takayuki Kanda, Masahiro Shiomi, Hiroshi Ishiguro, and Norihiro Hagita. 2008. A robot in a shopping mall that affectively guide customers. *Journal of Robotics Society of Japan*, 26(7):821–832.

Shota Takeuchi, Tobias Cincarek, Hiromichi Kawanami, Hiroshi Saruwatari, and Kiyohiro Shikano. 2007. Construction and optimization of a question and answer database for a real-environment speech-oriented guidance system. In *Proc. Oriental COCOSDA*, pages 149–154.

Zhou Yu, Leah Nicolich-Henkin, Alan W Black, and Alexander Rudnicky. 2016. A wizard-of-Oz study on a non-task-oriented dialog systems that reacts to user engagement. In *Proc. of the 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 55–63.

Victor Zue, Stephanie Seneff, James R Glass, Joseph Polifroni, Christine Pao, Timothy J Hazen, and Lee Hetherington. 2000. JUPlTER: a telephone-based conversational interface for weather information. *IEEE Transactions on speech and audio processing*, 8(1):85–96.

A Dialog example

Speaker	Utterance
User	What did you cook?
System	I cooked *nikujaga*.
User	Are you good at cooking *nikujaga*?
System	I can't say, but I often cook it.
User	Do you make anything else?
System	I often cook *stir fry vegetables*.
User	Did you invite your friend?
System	I used *Cookpad*. Have you ever used it?
User	Did you invite your friend?
System	I went for a light run after cooking.
User	Do you often go jogging?
System	Yes. I'm trying to always go jogging for my health.
User	How long did you run?
System	I ran about 5 km.
User	Where did you run?
System	The jogging path by the *Hirose river*.
User	Did you run alone?
System	Yes, I ran on the jogging path by the *Hirose river*.
User	Have you ever participated in a marathon?
System	Yes, I have participated in a half-marathon several times.

Table 4: Dialog example of cooking (translation from Japanese)

B Example of rewriting sentence

	Example rewriting sentences
Honorific	*Hai, ie de yoku tsukutte masu yo* Yes, I often make it at home.
Normal	*Un, ie de yoku tsukutte iru yo* Yeah, I often make it home

Table 5: Example of rewriting presented to the rewriters (italic: original, upright: translated, under line: different parts depending on form)