Smoothing Human-robot Speech Interactions by Using a Blinking-Light as Subtle Expression

Kotaro Funakoshi
Honda Research Institute
Japan Co., Ltd.
8-1 Honcho, Wako-shi
Saitama, Japan
funakoshi@jp.honda-ri.com

Kazuki Kobayashi
Shinshu University
4-17-1 Wakasato, Nagano
Nagano, Japan
kkobayashi@cs.shinshu-u.ac.jp

Mikio Nakano
Honda Research Institute
Japan Co., Ltd.
8-1 Honcho, Wako-shi
Saitama, Japan
nakano@jp.honda-ri.com

Seiji Yamada
National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda
Tokyo, Japan
seiji@nii.ac.jp

Yasuhiko Kitamura
Kwansei Gakuin University
2-1 Gakuen, Sanda
Hyogo, Japan
ykitamura@kwansei.ac.jp

Hiroshi Tsujino
Honda Research Institute
Japan Co., Ltd.
8-1 Honcho, Wako-shi
Saitama, Japan
tsujino@jp.honda-ri.com

ABSTRACT

Speech overlaps, undesired collisions of utterances between systems and users, harm smooth communication and degrade the usability of systems. We propose a method to enable smooth speech interactions between a user and a robot, which enables subtle expressions by the robot in the form of a blinking LED attached to its chest. In concrete terms, we show that, by blinking an LED from the end of the user's speech until the robot's speech, the number of undesirable repetitions, which are responsible for speech overlaps, decreases, while that of desirable repetitions increases. In experiments, participants played a last-and-first game with the robot. The experimental results suggest that the blinking-light can prevent speech overlaps between a user and a robot, speed up dialogues, and improve user's impressions.

Categories and Subject Descriptors

H.5.m [**Information interfaces and presentation (e.g., HCI)**]: Miscellaneous

General Terms

Experimentation, Verification

Keywords

Turn-taking, speech overlap, subtle expression, human-robot interaction

ICMI'08, October 20–22, 2008, Chania, Crete, Greece.
Copyright 2008 ACM 978-1-60558-198-9/08/10 ...$5.00.

1. INTRODUCTION

Speech overlaps, undesired collisions of utterances between systems and users, harm smooth communication and degrade system usability. When speech overlaps occur, users tend to stop speaking. This prevents systems from responding adequately because interrupted speech is hard to recognize automatically [7].

Often a speech overlap arises when the system misrecognizes a speech pause as a turn-end and starts speaking. Detection accuracy of turn-ends can be improved by using linguistic information [2] and para-linguistic information such as F0 (fundamental frequency) [8]. However, these methods are not sufficient. Linguistic information is affected by speech misrecognition. F0 is not available for fragmented short utterances because a reliable F0 contour requires a substantial length of speech.

Another approach to avoid turn-end misrecognitions is to make a long interval after the user's speech signal ends and before the system responds. This approach is simple and steady, but deteriorates system responsiveness and leads to another speech overlapping situation, that is, the user repeats her/his last utterance because of the lack of response from the system.

Equipping the system to express its turn-taking intention by using body/eye movements as humans do may resolve this situation. However, such an approach is technically difficult and uneconomical. Using spoken back-channel feedback (such as "well" and "uh") is another option. It is, however, not an easy matter [10] and a user study found that users preferred systems using spoken back-channel feedback to systems not using them only in cases systems could use them appropriately [4].

In contrast to the above approaches trying to make machines resemble humans, there is another approach that utilizes expressions characteristic of artifacts [6]. Following such an approach, we believe there is a way that we can avoid the difficulties in using human-like expressions but still gain the desired effect from a simple and economical implementation.

This paper verifies to what extent a method using a blinking-light can ameliorate this situation. We devised a robot that expresses its internal state by using a blinking-light and conducted experiments in which users and the robot engaged in last-and-first games (shiritori) through speech. The smoothness of the resulting dialogues was analyzed from the number of user repetitions of last utterances and users' answers to questionnaires. In concrete terms, we show that, by blinking an light from the end of the user's speech until the robot's speech, the number of undesirable repetitions decreases while that of desirable repetitions increases, and participants' impressions improve.

2. BLINKING LIGHT AS SUBTLE EXPRESSION

Although human communication is explicitly achieved through verbal utterances, non-verbal facial expressions, gaze, gestures, etc., also play an important role [3]. Such non-verbal communication often influences the accuracy of utterance understanding [9].

Furthermore, researchers have reported that very small changes (called *subtle expression*) in facial expressions and gestures might influence human communication. We believe that we can utilize such subtle expression to make humans easily understand a robot's internal state because humans can intuitively understand subtle expression. Some studies have been done on applying subtle expression to human-agent interaction (e.g., [1]). However, since they tried to enable subtle expression on real faces and with real arms, their implementations were considerably expensive.

In contrast with such approaches, subtle expression has been studied for artifacts like a robot or PC. Komatsu and Yamada [6] reported that an agent's subtle expression of simple beeping sounds with decreasing/increasing frequency enabled humans to interpret the agent's positive/negative states. Their work indicated the effectiveness of subtle expressions such as varying beeping sounds for a robot or agent.

In this work, we propose the use of a blinking light as a means of subtle expression to intuitively notify a user about robot's internal states (such as processing or busy). We implemented the subtle expression on a robot that engaged in dialogue with a user and conducted experiments on participants to verify the effectiveness of the subtle expression.

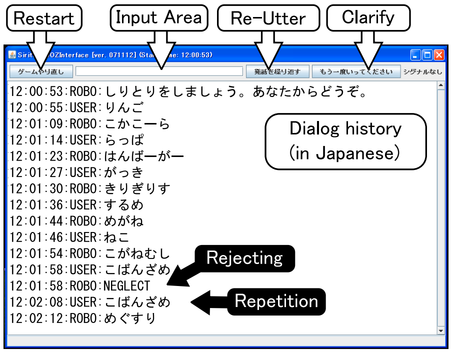
3. EXPERIMENTS

We conducted experiments in which participants played a last-and-first game with a robot. A last-and-first game was an appropriate task to investigate speech overlaps because it involved a lot of turn-taking. We counted the number of times the participants repeated their last utterances and estimated their impression of the dialogue. The dialogue system, robot, light-blinking expression, and experimental method are explained as follows.

3.1 Last-and-First Dialogue System

We chose a Wizard-of-Oz method to avoid speech recognition errors. The operator listened to the participant and operated the robot using the interface shown in Figure 1. The robot's utterances were voiced by a commercial speech synthesizer (NTT-IT FineVoice). The operator performed the following operations when needed.

1. Input participant's answer: The operator inputs the participant's answer immediately after the participant utters it. If the answer violates rules, the robot claims a foul and restarts another game. Otherwise, it chooses its next answer from a predefined lexicon.
2. Order to re-utter: The robot's speech is sometimes hard to listen to. Upon a participant's request, the operator directs the robot to re-utter its last utterance.
3. Order to clarify: Sometimes the operator cannot catch the participant's answer. The operator directs the robot to utter "Say it again."
4. Order to restart: Dialogue can break down for a variety of reasons. In such cases, the operator directs the robot to restart another game by uttering "Let's start again."
5. Record the participant's repetition: If the participant repeats her/his last utterance, the operator records it by pressing the ESC key.



```
12:00:53:ROBO:しりとりをしましょう。あなたからどうぞ。
12:00:55:USER:りんご
12:01:09:ROBO:こかこーら
12:01:14:USER:らっぱ
12:01:23:ROBO:はんばーがー
12:01:27:USER:がっき
12:01:30:ROBO:きりぎりす
12:01:36:USER:するめ
12:01:44:ROBO:めがね
12:01:46:USER:ねこ
12:01:54:ROBO:こがねむし
12:01:58:USER:こばんざめ
12:01:58:ROBO:NEGLECT
12:02:08:USER:こばんざめ
12:02:12:ROBO:めぐすり
```

Figure 1: Operational Interface

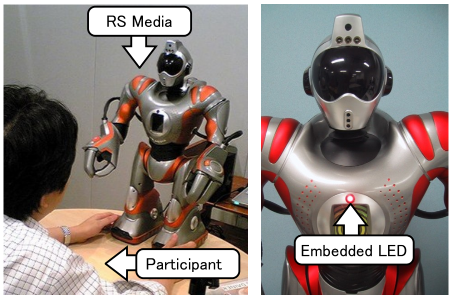


Figure 2: RS Media and position of embedded LED

We controlled the experiments as follows so that the target phenomena could be clearly observed. First, the intervals between the users' answers and the robot's answers were randomly varied between 0 and 15 seconds by inserting a waiting time. In actual spoken dialogue systems, these intervals correspond to delays due to internal processing. Second, the robot neglected the participants' answers with a probability of 1/4. In actual dialogue systems, this corresponds to rejecting speech as noise by mistake. Third, to resolve a standoff, the robot automatically re-uttered its last utterances after 30 seconds if the participant said nothing.

3.2 Robot and Blinking-Light Expression

The participants talked to the small human-like robot "RS Media (WowWee)" shown in Figure 2. In our experiments, the robot did not drive any actuators. We embedded a red LED (diameter: 4 mm) in its chest. The LED started blinking when an operator began to input a participant's utterance and stopped blinking when the robot began to utter. In other words, the LED blinked while the robot processed the participant's utterance to prepare an answer. When the robot rejected utterances, the LED stayed off. The LED blinked at an even interval of 1/30 seconds.

3.3 Participants

Forty seven participants of 19 to 62 years old were divided into two experimental groups:

(1) the blinking condition
(11 men, 12 women, mean age 32.0, S.D. = 11.0), and

(2) the non-blinking condition
(11 men, 13 women, mean age 31.0, S.D. = 10.3).

3.4 Experimental Method

The experiments were conducted in a small room. Participants entered the room, sat on a chair in front of a desk. They answered a questionnaire after they had been given instructions about the experiments.

After answering the questionnaire, they were asked to play a last-and-first game with the robot for 10 minutes and explained the rules of the game (see [5] for the rules). They could answer as they liked if they forgot the rules. The experimenter told them that the robot sometimes replied quickly, sometimes leisurely, and sometimes rejected their utterances. Moreover, they were requested to continue the game as long as they could. The meaning of the light-blinking expression was not explained to them. After giving the instructions, the experimenter left the room, and the participants began to play a game when the robot started to talk to them. The game finished after ten minutes, when the robot said that it was over.

After finishing the game, the participants answered another questionnaire about their impression of the game and the robot's behavior, and other particular questions.

4. RESULTS AND DISCUSSION

4.1 Participants' Repetitions

We classified participants' repetitions into two cases:

(1) the in-processing case (uttered while the robot was processing their utterances), and

(2) the rejection case (uttered when the robot rejected their utterances, i.e., the robot failed to catch the utterances).

Table 1 shows the mean repetition rates of the in-processing and rejection cases.

The mean repetition rate of the in-processing case is the average of the rates in which each participant repeated his/her last word while the robot was recognizing the word and preparing its next answer. In the in-processing case, a significant tendency between the mean repetition rates in the blinking condition and the non-blinking condition was found by using a t-test ($t = 1.81$, d.f. $=$ 45, $p = .079$). This supports that a blinking-light can suppress users' undesirable repetitions, and thus can prevent speech overlaps between a user and a robot.

The mean repetition rate of the rejection case is the average of the rates in which each participant repeated his/her last word when the robot failed to catch what the participant said. In the rejection case, a significant difference between the mean repetition rates in the two conditions was found ($t = 2.67$, d.f. $= 45$, $p = .010$). This suggests that a blinking-light can speed up dialogues because dialogues are suspended when the robot rejected the participants' utterances unless they repeat or the robot prompts.

Indeed, the average number of the robot's answering was significantly greater in the blinking condition than in the non-blinking condition (avg. 25.8 vs 21.1, $t = 3.06$, d.f. $= 43.29$, $p < .005$). This suggests that the dialogues in the blinking condition were faster.

Table 1: Mean repetition rate

condition	in-processing	rejection
blinking	1.3% (S.D. = 3.6)	55.3% (S.D. = 50.0)
non-blinking	5.4% (S.D. = 10.2)	24.3% (S.D. = 34.0)

Table 2: Rated adjective pairs for impression of the game

adjective pairs		blinking		non-blinking	
positive	negative	mean	S.D.	mean	S.D.
casual	grave	4.61	1.62	4.04	1.46
smooth	rough	2.83	1.23	2.33	1.09
comfortable	uncomfortable	3.96	1.30	4.25	1.26
exciting	dull	4.91	1.59	4.04	1.65
relaxed	tensional	3.78	1.41	3.71	1.30
easy	uneasy	3.83	1.23	3.42	1.35
warm	cold	3.96	1.11	3.67	1.05
pleasant	unpleasant	4.48	0.99	4.25	1.42
leisurely	hurried	5.26	0.96	5.17	1.20
informal	formal	4.40	1.40	3.71	1.23
light	dark	4.04	1.07	4.25	0.99
comprehensible	incomprehensible	3.74	1.57	3.21	1.10
likable	dislikable	4.35	1.03	4.21	1.41
good	poor	4.17	0.98	4.21	1.06
peaceful	annoying	3.83	1.19	3.67	1.52
interesting	boring	4.61	1.56	4.46	1.74
encouraging	discouraging	3.17	0.72	3.50	1.44
settled	unsettled	4.13	1.14	3.54	1.25

4.2 Impression of the Game and the Robot

The participants evaluated their impressions of the game and the robot by rating adjective pairs based on a scale from 1 to 7, where "1" equals strong agreement with a negative adjective and "7" equals strong agreement with a positive adjective. Table 2 and Table 3 show the rating results of the game and the robot respectively. Original adjective pairs in the questionnaire were in Japanese. We performed factor analysis (principal factor method) in the same way as [5], obtained factor scores with a regression method, and compared the two conditions by using a t-test.

No factor with a significant difference between the conditions was extracted with regard to the impression of the game. On the other hand, with regard to the impression of the robot, a factor with a significant difference ($t = 2.68$, d.f. $= 45$, $p = .010$) was extracted, to which the two adjective pairs of "responsible/irresponsible" and "broad-minded/narrow-minded" contribute positively for the blinking condition. This result suggests that the blinking light gives users the sincere impression of the robot.

4.3 Interpretation of the Blinking-Light

As free format questions in the questionnaire, we asked the participants about distinctive behaviors of the robot, situations when participants repeated their utterances and reasons they repeated, etc. From the answers to these questions, we extracted in what manner the participants in the blinking condition interpreted the blinking-light. Table 4 shows the classification of their interpretations. This shows that at least 21 participants out of 23 in the blinking condition were aware of the blinking-light, and most of them interpreted the meaning of the blinking light as we intended.

We can expect that the rate of repetitions will change over time if it requires a certain amount of interactions for participants to interpret the meaning of the blinking-light. Therefore we conducted t-tests between the mean repetition rates in the first five minutes and the last five minutes of dialogues. In the blinking condition, no significant difference was found. This suggests that the participants grasped the meaning of the blinking-light in a short time.

Table 3: Rated adjective pairs for impression of the robot

adjective pairs		blinking		non-blinking	
positive	negative	mean	S.D.	mean	S.D.
aggressive	defensive	4.22	1.31	4.21	1.02
innocent	wicked	4.17	1.27	4.29	1.46
respectful	impudent	4.00	1.48	4.50	1.56
accessible	inaccessible	3.96	1.19	4.17	1.31
pretty	provoking	4.26	1.14	4.08	1.41
broad-minded	narrow-minded	4.22	1.09	3.67	1.27
sociable	unsociable	4.13	1.46	3.83	1.17
responsible	irresponsible	4.52	1.08	4.50	1.44
careful	careless	4.78	1.00	4.50	1.44
shy	shameless	4.04	0.56	3.79	0.72
serious	frivolous	4.35	1.15	4.21	1.35
excited	gloom	3.65	1.19	3.79	0.78
regal	servile	4.96	1.15	4.83	1.13
decent	indecent	4.22	1.13	3.92	1.35
discreet	indiscreet	4.52	1.04	4.46	1.18
friendly	unfriendly	4.09	1.53	3.96	1.57
active	inactive	4.22	1.17	4.21	0.88
confident	unconfident	5.00	1.21	4.46	1.22
patient	irritable	4.65	1.23	4.83	1.13
kind	unkind	3.52	1.34	3.71	1.33

Table 4: Interpretation of the blinking-light

interpretation	# of participants	(percentage)
recognizing the user's answer	14	(60.9%)
preparing the next answer	4	(17.4%)
other	3	(13.0%)
(no answer)	2	(8.7%)

On the other hand, in the non-blinking condition, significant differences were found both in the in-processing case and in the rejection case. In the case of in-processing, the mean repetition rate of the first half was 5.9% and that of the last half was 3.6% ($t = 2.32$, d.f. $= 23$, $p < .05$). In the case of rejection, the mean repetition rate of the first half was 29.7% and that of the last half was 12.9% ($t = 2.11$, d.f. $= 23$, $p < .05$). This might be because the participants found that their repetitions were ineffective. The participants' repetitions were effective only in the case of rejection. However, in the non-blinking condition, the participants could not distinguish rejection cases from in-processing cases while in-processing cases had a higher proportion.

These results suggest that the adequate interpretation of the blinking-light can be obtained easily and immediately.

4.4 Reasons for Repetitions

Table 5 shows the classification of reasons for repetitions. Although about 30% of participants in the non-blinking condition made repetitions because of the robot's late response, no participants in the blinking condition reported such a reason. This suggests that the blinking-light dilutes the negative impression of late responses.

5. CONCLUSION

We proposed a method to enable smooth speech interactions between a user and a robot. Our method was based on subtle expression whereby a robot blinks a small LED attached to its chest. We performed experiments in which participants were divided into two groups: the blinking condition and the non-blinking condition, and played last-and-first games. The number of repetitions made by the participants and their answers to the post-experiment questionnaire were analyzed.

The analysis of the number of participants' repetitions showed that, in the blinking condition, the participants adequately repeated their last utterance when the robot rejected them and they did not repeat when it processed them. The analysis of the questionnaire suggested that (1) the blinking-light created sincere impressions about the robot on users, (2) most of the participants interpreted the blinking adequately in a short time, and (3) the blinking suppressed the negative impression of late responses. These results supported the effectiveness of the blinking-light expression. We think they indicate that LED-based subtle expression is helpful for smoothing human-robot speech interaction by preventing speech overlaps between a user and a robot, by speeding up dialogues, and by improving user's impressions.

Table 5: Reasons for repetitions

condition	reason	#	(percentage)
blinking	the LED did not blink	16	(69.6%)
	the robot repeated its last word	5	(21.7%)
	(no repetition)	2	(8.7%)
non-blinking	the robot repeated its last word	10	(41.7%)
	the robot's response was late	7	(29.2%)
	other	5	(20.8%)
	(no repetition)	2	(8.3%)

In future work, we would like to bring the proposed method into task-oriented dialogues, and demonstrate the method can improve task achievement ratio in addition to smoothness. Comparisons with other approaches are also to be addressed.

6. REFERENCES

[1] C. Bartneck and J. Reichenbach. Subtle emotional expressions of synthetic characters. *International Journal of Human-Computer Studies*, 62(2):179–192, 2005.

[2] L. Bell, J. Boye, and J. Gustafson. Real-time handling of fragmented utterances. In *Proc. of NAACL-2001 workshop on Adaptation in Dialogue Systems*, 2001.

[3] A. Kendon. Do gestures communicate? *A Review. Research in Language and Social Interaction*, 27(3):175–200, 1994.

[4] N. Kitaoka, M. Takeuchi, R. Nishimura, and S. Nakagawa. Response timing detection using prosodic and linguistic information for human-friendly spoken dialog systems. *Journal of The Japanese Society for Artificial Intellignece*, 20(3):220–228, 2005.

[5] K. Kobayashi, K. Funakoshi, S. Yamada, M. Nakano, Y. Kitamura, and H. Tsujino. Smoothing human-robot speech interaction with blinking-light expressions. In *Proc. of RO-MAN 2008*, 2008.

[6] T. Komatsu and S. Yamada. How do robotic agents' appearances affect people's interpretations of the agents' attitudes? In *Proc. of CHI-2007*, pages 2519–2524, 2007.

[7] M. Nakano, Y. Nagano, K. Funakoshi, T. Ito, K. Araki, Y. Hasegawa, and H. Tsujino. Analysis of user reactions to turn-taking failures in spoken dialogue systems. In *Proc. of SIGdial-2007*, 2007.

[8] T. Ohsuga, M. Nishida, Y. Horiuchi, and A. Ichikawa. Investigation of the relationship between turn-taking and prosodic features in spontaneous dialogue. In *Proc. of European Conference on Speech Communication and Technology*, pages 33–36, 2005.

[9] W. Rogers. The contribution of kinesic illustrators towards the comprehension of verbal behavior within utterances. *Human Communication Research*, 5:54–62, 1978.

[10] N. Ward. On the expressive competencies needed for responsive systems. In *Proc. of the CHI2003 workshop on Subtle Expressivity for Characters and Robots*, 2003.