Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive Communication, Technische Universität München, Munich, Germany, August 1-3, 2008

Smoothing Human-Robot Speech Interaction with Blinking-Light Expressions

Kazuki Kobayashi, Kotaro Funakoshi, Seiji Yamada, Mikio Nakano, Yasuhiko Kitamura, and Hiroshi Tsujino

Abstract— **We propose a method to enable smooth speech interactions between a user and a robot. Our method is based on subtle expression whereby a robot blinks a small LED attached to its chest. We performed experiments in which participants played a last-and-first games and counted the number of repetitions made by the participants and analyzed their impression of the game and the robot. The experimental results suggested that the blinking-light could prevent utterance collisions between a user and a robot and could create familiar and attentive impressions about the game on users.**

I. INTRODUCTION

Failures of turn-taking, especially speech overlaps between systems and users, harm smooth communication and degrade the usability of systems. When speech overlaps occur, users tend to interrupt their own speech. This prevents systems from responding adequately because interrupted speech is hard to recognize automatically [1].

In dialogues between a system and a user, undesired speech overlaps or utterance collisions arise from two situations: (a) the system misrecognizes a speech pause as the end of the user's turn and starts speaking, and (b) the user repeats her/his last utterance because of the lack of response from the system. One approach to avoid situation (a) is to make a long interval from the end of the user's speech signal. This approach, however, deteriorates system responsiveness and leads to situation (b).

To handle situation (a), one can use linguistic information to detect turn-ends [2], but this approach is affected by speech misrecognition. Although para-linguistic information such as F0 (fundamental frequency) can be used [3], it is not available for fragmented short utterances because a reliable F0 contour requires a substantial length of speech. Situation (b) might be handled by equipping the system to convey its intention of turn-taking to users by using body/eye movements as humans do. However, this approach is technically difficult and uneconomical. Another approach uses speech such as interjections (such as "well" and "uh"). Unfortunately, it is not easy to use interjections appropriately, either [4], [5]. Once the system fails to use them appropriately, users prefer systems not using them to the system using them [6].

This research verifies to what extent a method using blinking-lights can ameliorate this situation. In contrast to the above approaches that try to make machines resemble humans, there are other approaches that utilize expressions characteristic of artifacts [7]. Following these approaches, 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.

In this study, we devised a robot that expresses its internal state by using a blinking LED 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 user feedback. 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 repetitions decreases and participants' familiarity and attentiveness with the communication increase. The experimental results suggest that the blinking-light can prevent utterance collisions between a user and a robot.

II. BLINKING LED AS SUBTLE EXPRESSION

Although human communication is explicitly achieved through verbal utterances, non-verbal facial expressions, gaze, gestures, etc., also play an important role [8], [9]. Such non-verbal communication often influences the accuracy of utterance understanding [10].

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 human can intuitively understand subtle expression. Some studies have been done on applying subtle expression to human-agent interaction [11], [12], [13]. 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 [7]. Komatsu and Yamada reported that an agent's subtle expression of simple beeping sounds with decreasing/increasing frequency

Kazuki Kobayashi is with Research Center for Human Media, Kwansei Gakuin University, 2-1 Gakuen, Sanda, Hyogo 669-1337, Japan kby@kwansei.ac.jp

Kotaro Funakoshi is with Honda Research Institute Japan Co., Ltd., 8-1 Honcho, Wako, Saitama 351-0188, Japan funakoshi@jp.honda-ri.com

Seiji Yamada is with National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda, Tokyo 101-8430, Japan seiji@nii.ac.jp

Mikio Nakano is with Honda Research Institute Japan Co., Ltd., 8-1 Honcho, Wako, Saitama 351-0188, Japan nakano@jp.honda-ri.com

Yasuhiko Kitamura is with Department of Informatics, School of Science and Technology, Kwansei Gakuin University, 2-1 Gakuen, Sanda, Hyogo 669-1337, Japan ykitamura@kwansei.ac.jp

Hiroshi Tsujino is with Honda Research Institute Japan Co., Ltd., 8-1 Honcho, Wako, Saitama 351-0188, Japan tsujino@jp.honda-ri.com

978-1-4244-2213-5/08/$25.00 ©2008 IEEE

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 LED 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 engage in dialogue with a user and conducted preliminary experiments on participants to verify the effectiveness of the subtle expression.

III. 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.

A. 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.

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 20 seconds if the participant said nothing.

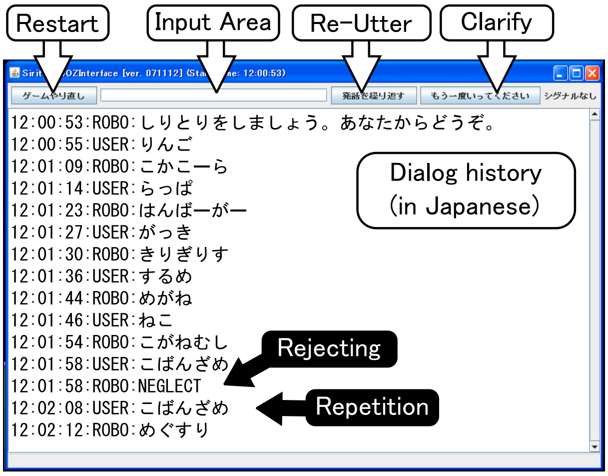

Fig. 1. Operational Interface

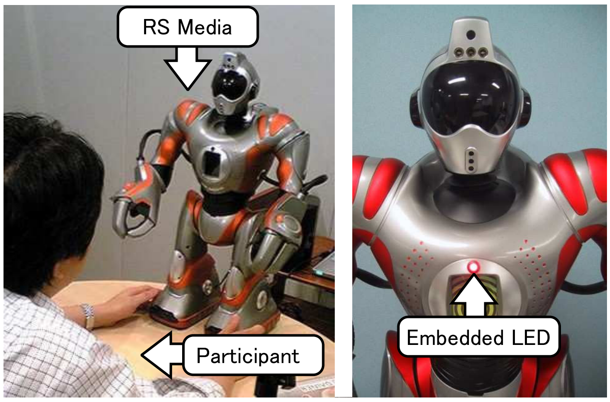

Fig. 2. RS Media and position of embedded LED

B. 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 blinking pattern was based on 1/f in order to express an emotional state [14]. The 1/f based generator provided random numbers in a range from -0.5 to 0.5. If a number was more than 0, the LED was turned on. If a number was less than 0, the LED was turned off. Figure 3 shows examples of 1/f random numbers and LED blinking pattern. A random number was generated every 1/30 seconds and then the LED status was updated.

C. Participants

Twelve participants were divided into two experimental groups: For six men (mean age: 22.8, S.D. = 1.2), LED blinked as explained above (the blinking-light condition), and

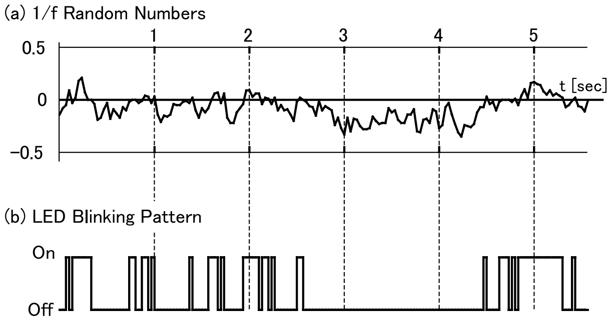

Fig. 3. LED Blinking Pattern

TABLE I
RULES OF THE JAPANESE LAST-AND-FIRST GAME (SHIRITORI)

1	Say a word which begins with the final kana (Japanese alphabet) of the previous word.
2	You lose if you say a work whose last character is "n(ん)."
3	Don't use same words.
4	You can remove a voiced/semivoiced sound mark. ex.) ga (が) → ka (か)
5	Remove a prolonged sound symbol. ex.) ko: (こー) → ko (こ)
6	Use the last character of a palatalized consonant symbol. ex.) kya (きゃ) → ya (や)

for five men and one woman (mean age: 23.8, S.D. =1.7), LED did not blink at all (the non blinking condition).

D. Experimental Method

The experiments were conducted in a small room (W: 256 D: 205 H: 215 cm) at Kwansei Gakuin University. 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 shown in Table I. They could answer as they liked if they forgot the rules. We 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 questionnaires about their impression of the game and the robot's behavior. To analyze the blinking-light effect, we counted the number of repetitions and assessed the impressions.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. Number of Participants' Repetitions

Table II shows the number of participants' repetitions for the case that the robot rejected their utterances. In total, the robot rejected 63 utterances in the blinking condition and 53 utterances in the non-blinking condition. The participants repeated their last utterances 37 times (58.7%) in the blinking condition and 20 times (37.7%) in the non-blinking condition. This represents a 21.0 point improvement (increase) of repetition during rejecting in the blinking condition. We performed Fisher's exact test and obtained a significant difference between conditions ($p < .05$).

TABLE II
NUMBER OF REPETITIONS IN REJECTING

condition	repetitions	no repetition	total
blinking	37	26	63
non-blinking	20	33	53
total	57	59	116

TABLE III
NUMBER OF REPETITIONS DURING PROCESSING AN UTTERANCE

condition	repetitions	no repetition	total
blinking	1	204	205
non-blinking	36	174	210
total	37	378	415

Table III shows the number of participants' repetitions for the case that the robot processed their utterances and prepared an answer. The robot processed 205 utterances in the blinking condition and 210 utterances in the non-blinking condition. The participants repeated their last utterance once (0.5 %) in the blinking condition and 37 times (17.1%) in the non-blinking condition. This represents a 16.6 point improvement (decrease) of repetition during processing in the blinking condition. We performed Fisher's exact test and obtained a significant difference between conditions ($p < .01$).

These results 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. This suggests that a blinking-light can prevent speech overlaps between a user and a robot.

B. Impression on the Last-and-First Game

Table IV shows the results of participants' ratings for the last and first game. The adjective pairs in the table are translated from Japanese words that we used in the questionnaire. The ratings are based on the 1-to-7 scales, where "1" equals strong agreement with a negative adjective and "7" equals strong agreement with a positive adjective. The questionnaire included nine reversal items.

Table V shows the factor loadings under the varimax rotation. We used 11 adjective pairs that had larger S.D. values than the omitted 7 pairs. The extraction method used unweighted least squares. We obtained three factors from the scree plot.

We interpreted the factors according to the adjective pairs that had loadings greater than 0.4 in Table V. The first factor (factor I) was named the *familiarity factor* and was based on the adjective pairs in the factor analysis of previous work [15]. The second factor (factor Ⅱ) was named the *looseness*

TABLE IV

RATED ADJECTIVE PAIRS FOR IMPRESSION OF THE GAME

adjective pairs		mean	S.D.
light	dark	3.33	1.23
exciting	dull	5.33	1.37
comfortable	uncomfortable	3.42	1.16
comprehensible	incomprehensible	3.92	1.56
easy	uneasy	3.33	1.07
informal	formal	3.08	1.08
spirited	spiritless	3.33	1.30
relaxed	tensional	2.67	1.30
casual	grave	3.83	1.27
pleasant	unpleasant	3.08	1.24
leisurely	hurried	2.92	1.08
likable	dislikable	3.33	1.07
peaceful	annoying	2.92	1.00
smooth	rough	3.42	0.90
warm	cold	2.83	0.83
decent	indecent	3.33	0.78
good	poor	3.50	0.67
interesting	boring	2.58	0.67

TABLE V

RESULTS OF FACTOR ANALYSIS OF IMPRESSION OF THE GAME (VARIMAX ROTATION, FACTOR LOADING MATRIX)

item	factor loadings			communality
	factor Ⅰ	factor Ⅱ	factor Ⅲ	
light	**0.90**	0.13	0.12	0.83
exciting	**0.77**	0.27	0.29	0.75
comfortable	**0.69**	**0.48**	−0.06	0.71
comprehensible	**0.44**	0.34	0.38	0.28
easy	0.02	**0.81**	**0.52**	0.67
informal	**0.64**	**0.69**	−0.10	0.90
spirited	**0.40**	**0.68**	0.21	0.60
relaxed	**0.44**	**0.48**	**0.42**	0.92
casual	0.25	**0.42**	−0.21	0.59
pleasant	−0.04	−0.09	**0.88**	0.46
leisurely	0.29	0.20	**0.68**	0.79
contribution(%)	27.06	22.69	18.41	

factor, since "easy" and "informal" had high loads. The third factor (factor Ⅲ) was named the *pleasantness factor*, since the "pleasant" had a high load.

We obtained the factor scores by using Bartlett's method and compared blinking condition with the non-blinking condition. Table VI shows the factor scores for the impression of the game. We performed the Mann Whitney test and found a significant difference between conditions in the factor I score ($U = 7.00, Z = -1.76, p = .05$) and 10% level difference in the factor Ⅱ score ($U = 6.00, Z = -1.92, p = .08$). There was no significant difference in the factor Ⅲ score ($U = 10.00, Z = -1.28, p = .20$). These results suggest that the blinking-light created familiar and attentive impression about the game on users.

C. Impression on the Robot

Table VII and Table VIII show the results of participants' ratings for the robot and the factor loadings under the varimax rotation, respectively. We used 11 adjective pairs that had larger S.D. values than the omitted 7 pairs. The extraction method used unweighted least squares. We obtained three factors from the scree plot. The questionnaire included ten reversal items.

TABLE VI

FACTOR SCORES FOR IMPRESSION OF THE GAME

factor	blinking(n=6)		non-blinking(n=6)	
	mean	S.D.	mean	S.D.
factor Ⅰ (familiarity)	0.49	0.77	−0.49	1.12
factor Ⅱ (looseness)	−0.60	0.53	0.60	1.13
factor Ⅲ (pleasantness)	0.42	0.67	−0.42	1.28

TABLE VII

RATED ADJECTIVE PAIRS FOR IMPRESSION OF THE ROBOT

adjective pairs		mean	S.D.
patient	irritable	4.17	1.53
friendly	unfriendly	3.42	1.24
kind	unkind	3.08	1.16
careful	careless	5.17	1.34
accessible	inaccessible	3.42	1.16
excited	cool	3.50	1.09
aggressive	defensive	4.33	1.44
active	inactive	3.67	1.56
impressive	unimpressive	4.92	1.08
sociable	unsociable	3.75	1.29
confident	unconfident	5.00	1.13
respectful	impudent	3.42	1.00
responsible	irresponsible	4.08	1.00
serious	frivolous	4.58	0.90
shy	shameless	3.50	0.80
decent	indecent	3.50	0.80
innocent	wicked	3.58	0.79
discreet	indiscreet	4.08	0.79
broad-minded	narrow-minded	3.50	0.67
pretty	provoking	3.58	0.67

We interpreted the factors according to the adjective pairs that had loadings greater than 0.4 in Table VIII. The first factor (factor I) was named the *friendliness factor* since the "patient" and "friendly" had high loads. The second factor (factor Ⅱ) was named the *optimism factor* since "careless" and "accessible" had high loads. The third factor (factor Ⅲ) was named the *activeness factor* since "active" and "impressive" had a high load.

We obtained the factor scores by using Bartlett's method and compared blinking condition with the non-blinking condition. Table IX shows the factor scores for the impression of the robot. We performed the Mann Whitney test and found no significant difference between conditions (factor I: $U = 16.00, Z = -0.32, p = .75$, factor Ⅱ : $U = 15.00, Z = -0.48, p = .63$, factor Ⅲ : $U = 18.00, Z = 0.00, p = 1.00$). These results suggest that the blinking-light caused no effect to users' impression about the robot which they talked to.

D. Waiting Time

In the questionnaire, we asked participants to answer the amount of time that the robot spent in replying. Table X shows their reported waiting time and actual waiting time. We performed Mann-Whitney test and found the significant difference between the conditions in reported time ($U = 3.00$, $Z = -2.42$, $p < .05$). There was no significant difference in actual time ($U = 15.00$, $Z = -.48$, $p = .63$). These results suggest that users felt their waiting time short by observing the blinking-light.

TABLE VIII

RESULTS OF FACTOR ANALYSIS OF IMPRESSION OF THE ROBOT (VARIMAX ROTATION, FACTOR LOADING MATRIX)

item	factor loadings			communality
	factor I	factor II	factor III	
patient	**0.95**	−0.32	−0.07	1.00
friendly	**0.79**	0.27	0.28	0.78
kind	**0.73**	−0.05	0.19	0.58
careful	0.28	**−0.92**	0.13	0.95
accessible	0.26	**0.74**	−0.10	0.63
excited	−0.38	**0.65**	0.08	0.58
aggressive	0.04	**0.44**	0.31	0.29
active	0.30	−0.20	**0.78**	0.74
impressive	0.22	−0.05	**0.78**	0.66
sociable	0.07	−0.13	**−0.64**	0.43
confident	**−0.58**	0.15	**0.62**	0.74
contribution (%)	25.86	20.77	20.46	

TABLE IX

FACTOR SCORES FOR IMPRESSION OF THE ROBOT

factor	blinking(n=6)		non-blinking(n=6)	
	mean	S.D.	mean	S.D.
factor I (friendliness)	−0.07	0.83	0.07	1.23
factor II (optimism)	0.14	1.36	−0.14	0.66
factor III (activeness)	0.00	1.29	0.00	0.91

E. Participants' Interpretation of Blinking-Light

Participants were asked to answer robot's behavior and its meaning that they noticed. Table XI and XII show the results of the questionnaire in the blinking and non-blinking condition, respectively. In the blinking condition, five of the six participants noticed the LED blinking and reported that its meaning was thinking or recognizing. In the non-blinking condition, four of the six participants noticed the behavior about last-and-first game and only one participant reported that changing answer time meant thinking. These results suggest that LED blinking was interpreted as thinking.

V. 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 played a last-and-first games. The number of repetitions made by the participants and their impression of the game and the robot were analyzed.

The analysis of the number of participants' repetitions showed that the participants adequately repeated their last utterance when the robot rejected them and they did not repeat when it processed them in the blinking condition. The analysis of participants' impression of the last-and-first game showed that the blinking-light created familiar and attentive impressions about the game on users. The blinking-light caused no effect to user's impression about the robot. Participants felt their waiting time short by observing the blinking-light. Most of participants interpreted the blinking as thinking. These results supported the effectiveness of the blinking-light expression. Although these experiments were small-scale and preliminary, we think they indicate that LED-based subtle expression is helpful for smoothing human-robot speech interaction.

TABLE X

REPORTED WAITING TIME AND ACTUAL WAITING TIME

waiting time	condition (n=6)	mean [sec]	S.D.
reported time	blinking	3.67	2.16
	non-blinking	8.83	3.76
actual time	blinking	7.21	1.04
	non-blinking	6.63	1.61

TABLE XI

REPORTED BEHAVIOR AND MEANING IN BLINKING CONDITION

No	Robot's behavior	Meaning
1	The LED turned on when the robot answered.	The robot recognized a participant's word when the LED turned on.
2	The LED turned on.	The robot said a next word.
3	(1) The LED turned on. (2) The LED stayed off.	(1) The robot was thinking. (2) The robot did not listen.
4	The red LED blinked.	The robot recognized a participant's word.
5	Nothing.	(no response)
6	The LED on its chest blinked.	The robot thought and understood a participant's word.

TABLE XII

REPORTED BEHAVIOR AND MEANING IN NON-BLINKING

No	Robot's behavior	Meaning
1	The robot recognized the speech and answered a word.	(no response)
2	The robot said the participant lost when it was not able to recognize a his/her word or he/she was quiet for a certain time.	(no response)
3	Nothing.	(no response)
4	(1) The robot changed answer time. (2) The robot uttered "Say it again."	(1) The robot was thinking. (2) The robot failed to recognize participant's speech
5	The robot asked again, repeated an word, and spoke in a monotone voice.	(no response)
6	Nothing.	(no response)

REFERENCES

[1] 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 Workshop on Discourse and Dialogue*, 2007.

[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] 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*, 2005, pp. 33–36.

[4] W. Tsukahara and N. Ward, "Evaluating responsiveness in spoken dialog systems," in *Proc. of International Conference on Spoken Language Processing*, 2000.

[5] N. Ward, “On the expressive competencies needed for responsive systems,” in *Proc. of CHI2003 workshop on Subtle Expressivity for Character s and Robots*, 2003.
[6] N. Kitaoka, M. Takeuchi, R. Nishimura, and S. N. a, “Response timing detection using prosodic and linguistic information for human- friendly spoken dialog systems,” *Journal of The Japanese Society for Artificial Intelligence*, vol. 20, no. 3, pp. 220–228, 2005.
[7] T. Komatsu and S. Yamada, “How do robotic agents' appearances affect people's interpretations of the agents' attitudes?” in *Proc. of CHI-2007*, 2007, pp. 2519–2524.
[8] A. Kendon, “Do gestures communicate?” *A Review. Research in Language and Social Interaction*, vol. 27, no. 3, pp. 175–200, 1994.
[9] ——, “Some functions of gaze direction in social interaction,” *Acta Psychologica*, vol. 26, pp. 1–47, 1967.
[10] W. Rogers, “The contribution of kinesic illustrators towards the comprehension of verbal behavior within utterances,” *Human Communication Research*, vol. 5, pp. 54–62, 1978.
[11] C. Bartneck and J. Reichenbach, “Subtle emotional expressions of synthetic characters,” *International Journal of Human-Computer Studies*, vol. 62, no. 2, pp. 179–192, 2005.
[12] S. Brave, C. Nass, and K. Hutchinson, “Computers that care: investigating the effects of orientation of emotion exhibited by an embodied computer agent,” *International Journal of Human-Computer Studies*, vol. 62, no. 2, pp. 161–178, 2005.
[13] H. Prendinger, J. Mori, and M. Ishizuka, “Using human physiology to evaluate subtle expressivity of a virtual quizmaster in a mathematical game,” *International Journal of Human-Computer Studies*, vol. 62, no. 2, pp. 231–245, 2005.
[14] S. Hosomi, M. Tsukamoto, and S. Nishio, “A system for controlling led blink in wearable fashion,” in *Proc. of International Conference on Wireless Communications and Mobile Computing*. New York, NY, USA: ACM, 2007, pp. 665–670.
[15] T. Kanda, H. Ishiguro, and T. Ishida, “Psychological analysis on human-robot interaction,” in *Proc. of IEEE International Conference on Robotics and Automation*, 2001, pp. 4166–4173.