Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics
San Antonio, TX, USA - October 2009

Adaptive Embodied Entrainment Control and Interaction Design of the First Meeting Introducer Robot

Kenzaburo Miyawaki, Mutsuo Sano
Fuculty of Information Science and Technology,
Osaka Institute of Technology,
1-79-1 Kitayama, Hirakata City, Osaka, Japan
{miyawaki, sano}@is.oit.ac.jp

Ryohei Sasama, Tomoharu Yamaguchi, Keiji Yamada
NEC Corporation, C & C Innovation Research Laboratories
8916-47 Takayama-Cho, Ikoma, Nara, Japan
{r-sasama@cp, yamaguchi@az, kg-yamada@cp}.jp.nec.com

***Abstract*—This paper describes a robotic agent that can promote communication when people first meet. When two people meet for the first time, a communication mediator can be important because people often feel stressed and cannot talk comfortably. Our agent reduces their stress by using embodied entrainment and thus promotes communication. In the research field of embodied entrainment, suitable timing of a nod or back-channel feedback has been discussed, but situations to communicate in are limited. We have developed an embodied entrainment control system that recognizes communication states and adapts to them accordingly. For this, we focus on effective non-verbal information for communication. Using this information, our agent helps a talker and a listener to alternate their roles appropriately. We conducted communication experiments with the agent and confirmed its effectiveness. In the experiments, we used and compared different representation of the agent: an embodied robot agent, a computer graphics agent, and no agent. We report the comparison results and discuss representations for communication agents.**

***Index Terms*—Embodied Entrainment, Nonverbal Communication, Robotic Introducer Agent, Group Communication**

I. Introduction

Recently, wider communication capability has become required for robots to better support peoples' daily life. In these situations, investigating a simple relationship between one person and one robot is not enough to develop true socialized robots. We must design a robot that can mediate a smooth group conversation and expand the circle of human communication. People communicate with each other via many verbal and nonverbal channels. It has been confirmed that the nonverbal communication channels (such as intonation, accent, gazing, gestures, nodding, or emotional expressions) encourage synchronization, embodied entrainment, and friendship [1]–[7]. Watanabe et al. [8] have investigated such a robotic agent, which includes embodied entrainment. Their research has developed a speech-driven interactive actor, the Inter Actor, which can predict the appropriate timing of a nod and a gesture with voice information from a communication partner. The Inter Actor focused on synchronization with a listener, but application to group communication which involves a talker, a listener, appropriately alternating between them, and smoothly leading utterances was not mentioned. Kanda et al. [9], [10] applied a joint gaze function to a robot and has investigated a more adaptive robot through a design-based approach. However, the research mainly discussed rule-based natural behavior of robots, but did not sufficiently discuss a robot design that aims to stimulate communication. Moreover, the investigation did not mention group communication. This paper focuses on group communication and presents a control method for embodied entrainment that is adaptive to various situations. On the basis of this method, we design a new agent to promote communication. First of all, our investigation starts from research into an introducer robot for communication between people meeting for the first time (Fig. 1). Next, we explain our method for adaptive control of embodied entrainment and show experimental results of our method. In the experiments, we used and compared different representations of the agent (an embodied agent and a computer graphic agent) to investigate what influence different forms generate. Finally, the effectiveness is verified by results of a questionnaire, and we discuss representations of the communication agents.

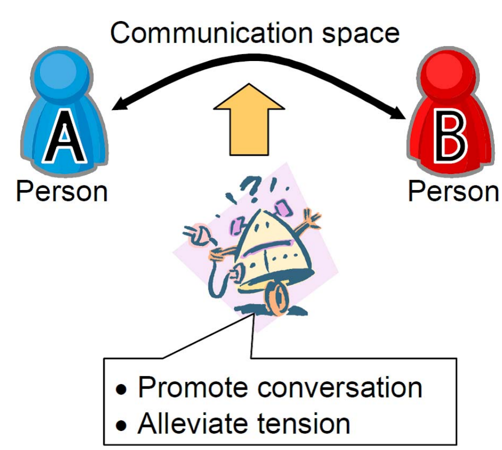

Fig. 1. Introducer robot for two people at their first meeting

978-1-4244-2794-9/09/$25.00 ©2009 IEEE

II. Basic Strategies to Promote Group Conversation

Figure 2 shows our robotic agent. The purpose of this introducer agent is to support conversations between two strangers. For this purpose, we designed various basic behavior strategies for the agent, the details of which are as follows.

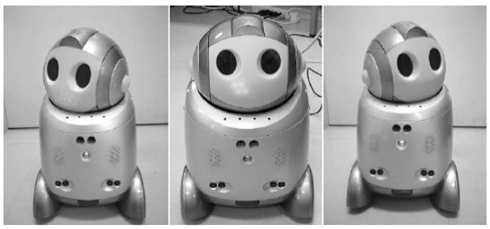

Fig. 2. Robotic Introducer Agent

1) **Utterances inducement**
   The agent asks a question to participants to establish a binary trusting relationship between itself and each participant.
2) **Gaze leading**
   To transmit information about person A to person B, the agent moves its gaze to A and asks him/her a question. After that, the agent leads A's gaze to B to accord A's gaze and B's gaze.
3) **Gaze distribution and nod Synchronization**
   When both participants talk to each other, the agent moves its gaze to them equally and synchronizes its nod timing with their nods.
4) **Dynamic Synchronization**
   The synchronization method is dynamically changed in accordance with the state of the two people as follows.
   a) **Natural synchronization with a listener.**
      When a listener listens to talker's speech and synchronizes their nods and back-channel feedback with those of the talker, our robot tunes its rhythm to listener's timing to widen the communication further.
   b) **Synchronization with a talker to invite the other participant in to the conversation.**
      When a listener does not listen to a talker's speech, the robot tunes its rhythm to talker's and accords its gaze to the listener. By doing this, the robot leads the listener and creates a rhythm for communication.

Figure 3 shows the strategies' details. Note that, the symbols in the figure represent communication members as follows.

H_1 Person A.
H_2 Person B.
R The introducer agent.

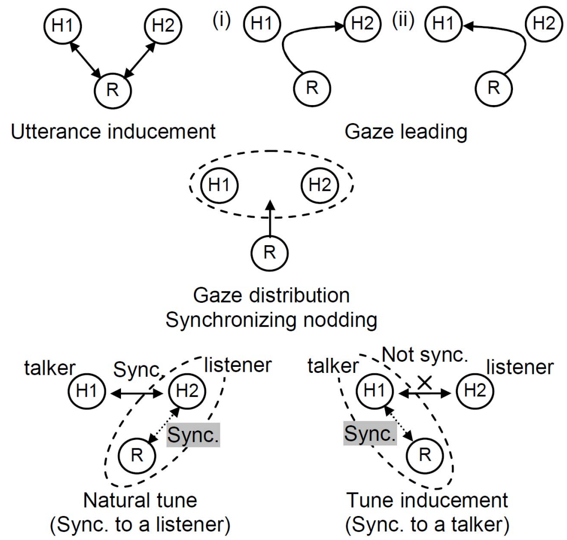

Fig. 3. Basic strategies to promote pair conversation

III. Adaptive Control of Embodied Entrainment

A. Macro Strategy to Promote Pair Conversation

The behaviors in basic strategies must suit the communication situation. To create appropriate behavior, we modeled pair communication on a first meeting between two people and designed a macro strategy to promote conversation. Figure 4 describes the strategy and the state transition model. The strategy consists of the grounding process and the enhancement process shown at the top of the figure. The grounding process establishes a rapport, and the enhancement process promotes communications between the pair. The bottom of Fig. 4 shows the state transition model. We segmented an introduction scene into five states on the basis of preliminary observations. The agent moves among these states in accordance with the participants' communication situation. The details of the states and agent behaviors in each state are as follows.

1) **Greeting**
   The agent introduces itself and briefly explains the situation. Then it simply tells each participant the other's name.
2) **Grounding**
   In this state, the agent tries to create a trusting relationship between itself and each participant. For this, the agent cites the participants' profiles and asks simple questions to them. It aims to entrain them into its rhythm by using the utterance inducement strategy. Here, it is assumed that the agent knows the data, such as name, interests, or other information, on the persons who are introduced.
3) **Offering Topics**
   This state encourages conversations between the participants. The agent offers the information and profiles on one participant to the other, or asks simple questions to

the participants in order to make a favorable situation for the participants to converse. By such behavior, the agent guides the participants' gaze to make them communicate face to face with each other.

4) **Focusing on a Specified Topic**
In this state, the agent tries to join in the conversation. The agent focuses on the previously brought up topic.

5) **Listening to Conversations**
After succeeding in making an open and friendly relationship between the participants, the agent keeps listening to their conversations quietly. In this state, the agent nods and looks at whoever is talking using basic embodied entrainment strategies with proper timing.

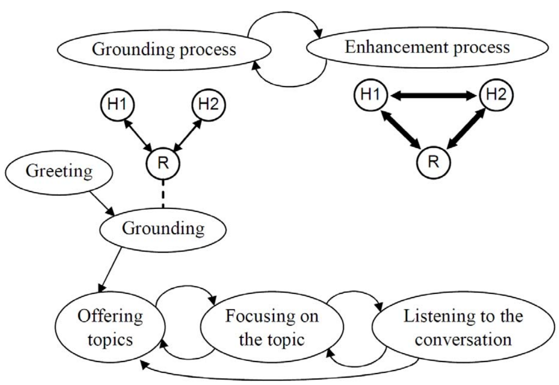

Fig. 4. Macro strategy and state transition of an introduction scene

B. Communication Activity Measurement

The robotic agent monitors the participants' communication and estimates the current state, as show in Fig. 4. For the estimation, we define communication activity measurement. The agent calculates the activity in every time slice. It is defined by the average speech volume, gaze direction, and gaze accordance of participants. When this measurement satisfies state transition conditions, the agent moves to the next state as directed by the transition. To explain the measurement, we define essential actions of communication members using the following symbols. These symbols represent functions that return to 1 when that action is detected by sensors. In other cases, they return to 0. Henceforth, the parameter t always denotes the time when the actions occur.

$Nod(X,t)$	X nods.
$Utterance(X,t)$	X talks.
$Utterance(X \rightarrow Y,t)$	X talks to Y.
$TerminateUtterance(X,t)$	X terminates his/her speech.
$Gaze(X \rightarrow Y,t)$	X directs his/her gaze to Y.
$Face(X \rightarrow Y,t)$	X looks toward Y.
$Gaze(X \Leftrightarrow Y,t)$	X's gaze accords to Y's.
$Face(X \Leftrightarrow Y,t)$	X and Y look toward each other.
$TurnGaze(R, X \rightarrow Y,t)$	The agent guides the X's gaze toward Y.
$TurnUtterance(R, X \rightarrow Y,t)$	R encourages X to talk to Y.

Note that in this paper
$Gaze(X \rightarrow Y,t) = Face(X \rightarrow Y,t)$, and
$Gaze(X \Leftrightarrow Y,t) = Face(X \Leftrightarrow Y,t)$
Our method calculates the communication activity by using these symbols. This is based on the following two equations.

1) **Ratio of gaze sharing**

$$R_{gazeShare}(X \Leftrightarrow Y, \Delta T) = \frac{\sum_{t=t_0}^{t_0+\Delta T} Gaze(X \Leftrightarrow Y,t)}{\Delta T} \quad (1)$$

2) **Average utterance power**

$$Average_{P_u}(X, \Delta T) = \frac{\sum_{t=t_0}^{t_0+\Delta T} P_u(X,t)}{\Delta T} + P_n \quad (2)$$

Note that $P_u(x,t)$ represents the power of x's utterance at the time t. P_n is a term of background noise. The following equation is an extended formula to calculate $Average_{P_u}$ of multiple participants.

$$Average_{P_u}(X, Y, \Delta T) = \frac{\sum_{t=t_0}^{t_0+\Delta T} average\{P_u(X,t), P_u(Y,t)\}}{\Delta T} + P_n \quad (3)$$

Note that ΔT is a time taken to calculate the measurement. State transition conditions depend on the communication activity measurement and are defined as follows.

1) **from Greeting to Grounding and then to Topic Offering**
In these states, the agent behaves as prescribed in the action scenario. The scenario is composed of general etiquette for greetings and introductions.

2) **from Topic Offering to Focusing on the Topic**

$$R_{gazeShare}(X \Leftrightarrow Y, \Delta T) \geq Th_{gaze} \quad (4)$$

3) **from Focusing the Topic to Listening to the Conversation**

$$R_{gazeShare}(X \Leftrightarrow Y, \Delta T) \geq Th_{gaze} \;\; and \;\; Average_{P_u}(X, Y, \Delta T) \geq Th_{utter} \quad (5)$$

In the equations, the Th_{gaze} and the Th_{utter} are acquired through learning.

C. Global Evaluation Value of Communication Activity

This value represents a communication level that the agent raises and is calculated from averages of participants' utterance power and gaze accordance ratio. The value is used to evaluate our algorithm, and is described in the following equations.

$$Eval(t) = \sqrt{\gamma} \tag{6}$$

$$\gamma = \alpha \, \{ \; Average_{P_u}(X,T)^2 + Average_{P_u}(Y,T)^2 \; \} + \beta \, \{ \; R_{gazeShare}(X \Leftrightarrow Y,T)^2 \; \} \tag{7}$$

In the equations, T means time elapsed from the start to the end of communication. α and β are weight factors to normalize a range.

D. Learning for Adaptive Control of Embodied Entrainment

Our agent performs some actions to encourage embodied entrainment. The agent retrieves the information on the participants by sensors and determines its own action. We built a decision-tree, shown in Fig. 5, by inductive learning for this interaction. The three rules below are extracted from the tree.

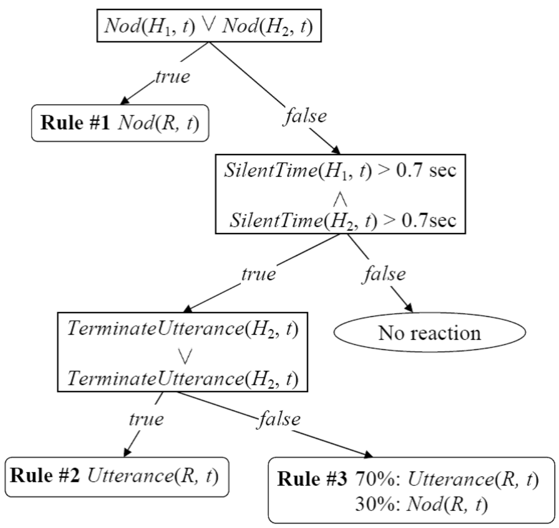

Fig. 5. Decision tree for embodied entrainment

1) **Rule #1 Reactive nodding**
   This rule represents reactive nodding to a participant's nod. The agent nods when one of the participants nods.
2) **Rule #2 An utterance**
   This rule explains the following situation. When a participant keeps quiet and the other participant stops speaking, it can be supposed that they are taking a break from the conversation. Therefore, the agent recognizes this situation and offers a new topic to avoid silence.
3) **Rule #3 An utterance and a nod to a participant's speech**
   This rule is very similar to **Rule #2** but represents the situation in which the participant does not completely finish his/her utterance. In this situation, the agent offers a topic for 70% and nods for 30%. The criterion to determine whether a participant finished his/her utterance or not depends on whether the final part of the utterance is comprehensible. We extract the final part by Japanese morphological analysis.

IV. Experiments

A. Experimental System

Information on an ongoing situation is recognized by a situation cognitive system composed of voice recognition, voice power detection, gaze detection, and nodding detection. After doing all this, an interaction generator retrieves the information and determines the agent behavior with the participants' profiles. Detection modules in the cognitive system retrieve raw sensor data and extract the participants' state and action. The details of the modules and sensors are listed below.

1) **Gaze Detection**
   Multiple cameras are used for this module. We point one camera at each participant. Each camera detects the participant's face, and the gaze detector estimates whether the participant is turning his/her gaze to the camera, which can detect the participant's face.
2) **Nodding Detection**
   We put acceleration sensors on the participants' heads and measured their head movement. On the basis of this movement, this module detects when he/she is nodding.
3) **Voice Recognition, Voice Power Detection**
   A pin microphone was placed on each participant.

B. Experimental Result

Figure 6 shows an experiment with our introducer agent. We asked two participants who had never met to talk to each other for six minutes. In this situation, we checked whether our method could control the agent's behavior appropriately. Figure 7 plots the global evaluation value of communication activity as $Eval(t)$ and state transitions of the communication. In Fig. 7, we can see favorable state transition of communication through sequential changes between greetings, offering topics, focusing, and listening. Additionally, the global evaluation of communication activity ($Eval(t)$) seems to change along with the state transition. Figure 8 shows the frequency

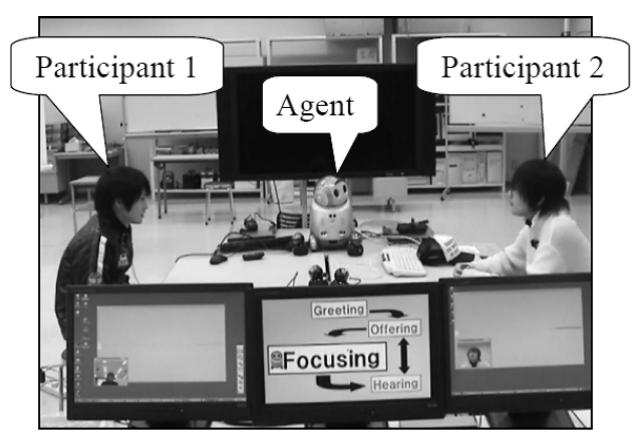

Fig. 6. Experiments with the agent

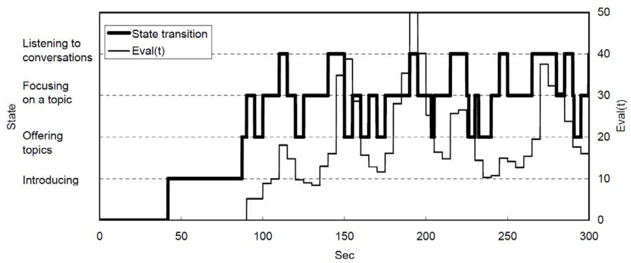

Fig. 7. $Eval(t)$ and state transition

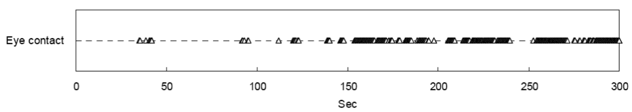

Fig. 8. Participants' eye contact

with which the participants made eye contact. From the figure, we can see increases in events exceeding 60% gaze accordance ratio. Note that $\Delta T = 5(sec)$ in the experiments. This value was based on preliminary observations. In the experiments, we confirmed many cases in which the agent prompted what the participants said. The following dialog is an example.

An example dialog

Greeting / Grounding

1: R (To H_1) I came here from *Hirakata*[1]. Where did you come from?

2: H_1 I came from *Neyagawa*[2], which is where is was born.

3: R (To H_2) I caught a train here today. How did you come here?

4: H_2 I also took a train.

Topic Offering

5: R (To H_1) Is there any food you don't like to eat?

6: H_1 I don't like mushrooms, such as *Shiitake*. And you (to H_2) ?

7: H_2 I hate beans!

8: H_1 Me too. Particularly sweet boiled beans; I don't like sweet food.

9: R (To H_2) You're into computer games, aren't you?

10:H_2 Yes, I am.

11:H_1 So am I.

12:H_2 What kind of games do you play?

13:H_1 I like action games and FPSs[3]

14:H_2 I don't like FPSs because they make me feel really sick.

15:H_1 I can understand that.

Focusing the Topic

16:H_2 You have you move your head quickly to see everything when playing FPSs, so I end up feeling sick.

17:H_1 Yeah, that happens to quite a few people.

17:R Please tell me more about this topic.

18:$H_{1,2}$ Alright, alright (laughing).

19:H_1 We can't play FPSs too well.

20:H_2 Yeah. Maybe Japanese people aren't made for them.

C. Comparison of Different Types of Communication

To investigate the influence of embodying the agent, we conducted a comparison experiment involving the following situations.

1) Communication with an embodied robotic agent.
2) Communication with a computer graphics (CG) agent. (Fig. 9)
3) Communication without an agent.

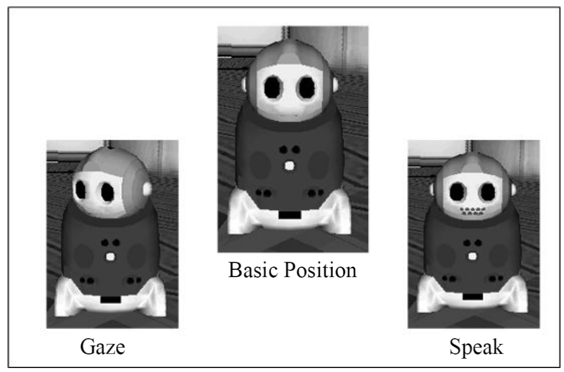

Fig. 9. The computer graphics agent

The subjects were 10 pairs who did not know each other. Each of above cases was carried out 10 times with 20 (2×10) subjects. The robotic and computer graphics agents behaved the same way. In the experiments, we measured the communication activity and then conducted a survey. Note that all questions were answered with grades from one to five, "one" being the worst and "five" being the best. The results were compared with a one-sided t-test. The results of this survey are as follows

- In answer to "Did you feel get on well with your partner?", we could see a statistically significant change of 5%. (Table.I)
- In answer to "Could you get to know your partner well?", we could also see a statistically significant 5% change. (Table.II)

TABLE I
AN ANALYSIS OF "DID YOU GET ON WELL?"

	Without Agent	With Agent
Mean	4.05	4.60
Variance	1.73	0.25
Result	$p = 0.045 (< .05)$	

The results describe how our agent helped the participants to better communicate with their partners.

[1,2]Cities in Osaka Prefecture, Japan.
[3]First person shooters.

TABLE II
AN ANALYSIS OF "COULD YOU GET TO KNOW HIM/HER WELL ?"

	Without Agent	With Agent
Mean	3.50	4.10
Variance	1.42	0.62
Result	$p = 0.034 (< .05)$	

Table III compares the embodied and the CG agents. The details are as follows.

- In answer to "Was the agent useful for you?", 70% of subjects gave the agent a positive evaluation. (Table.III)
- In answer to "Did you feel with the agent was friendly?", 90% positively evaluated the embodied robot and 65% the CG agent.
- In answer to "Did the conversation become more lively?", negative evaluations were 0 % for the embodied agent, 0% for computer graphics agent and 15% when there was no-agent.

TABLE III
"WAS THE AGENT USEFUL FOR YOU?"

	CG Agent	Embodied Robot
Very useful	45%	40%
Useful	20%	30%

Finally, we show interesting comments of participants. When without the agent, one subject said that,

- "I couldn't get the conversation going."

Figure 10 explains why they communicated poorly. This figure shows the $Eval(t)$ and state transition without an agent. The $Eval(t)$ does not move up, and the state remains almost totally in the introducing state.

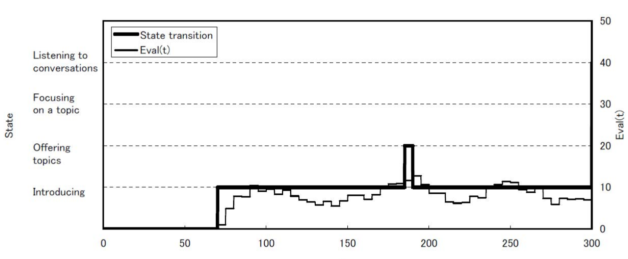

Fig. 10. $Eval(t)$ and state transition in the case without agent

On the other hand, we received the following favorable comments when an agent was used.

- "Because the agent mediates our communication, I didn't worry about what to say."
- "I appreciated having the agent because it asked questions that then prompted us to converse."

From these comments, we can confirm the effectiveness of our agent system. The agent supported communication between strangers well.

V. CONCLUSION

Our agent helped people change their communication states, such as offering topics, focusing, and listening, appropriately on the basis of communication activity measurements, which were estimated from gaze accordance and utterance power. On the basis of this state transition, the agent could activate communication through entraining behavior, (such as nods, gazes, and back-channel feedback) suitably in each communication state. We conducted experiments with pairs communicating in their first meeting, and we confirmed the agent adapted to changing states and smoothly controlled interaction. In addition, we compared three cases of communication: one with an embodied robotic agent, one with a computer graphics agent and with no agent. We then conducted a questionnaire, the results of which showed that our agent could promote conversation between the pairs of subjects. From the evaluation, it is supposed that our system can deal with a stressful situation between a pair meeting for the first time. Additionally, it was confirmed that the embodied robotic agent was better than the computer graphics agent. Now, we are trying to enhance this method to N people's group communications schema, and we will apply it for use in a retirement center.

REFERENCES

[1] J. D. Matarazzo, M. Weitman, and A. N. Saslow, G.and Wiens. Interviewer influence on durations of interviewee speech. *Journal of Verbal Learning and Verbal Behavior*, 1:451–489, 1963.

[2] J.T. Webb. Interview synchrony: An investigation of two speech rate measures in an automated standardized interview. In *B. Pope and A.W. Siegman (Eds.), Studies in Dyadic Communication*, pages 115–133, New York, Pergamon, 1972.

[3] A. Kendon. Movement coordination in social interaction: Some examples described. *Acta Psychologica*, 32:101–125, 1970.

[4] WS. Condon and Sander LW. Neonate movement is synchronized with adult speech: interactional participation and language acquisition. *Science*, 183:99–101, 1974.

[5] F.J. Bernieri, J.S. Reznick, and R Rosenthal. Synchrony, pseudosynchrony, and dissynchrony: Measuring the entrainment process in motherinfant interactions. *Journal of Personality and Social Psychology*, 54:243–253, 1988.

[6] Takeuchi.Y and Katagiri.Y. Emotional feedback by synthesized voice and facial expressions. In *Proccedings of the 3rd intenational conference of cognitive science*, pages 697–698, 2001.

[7] Kojima H., Nakagawa.C., and Yano.H. Can a robot empathize with people ? *International Journal of Artificial Life and Robotics*, 8:83–88, 2004.

[8] Tomio Watanabe. E-cosmic: Embodied communication system for mind connection. In *Proc. of the 13th IEEE International Workshop on Robot-Human Interactive Communication (RO-MAN 2004)*, pages 1–6, 2004.

[9] T. Kanda, H. Ishiguro, T. Ono, M. Imai, T. Maeda, and R. Nakatsu. Development of frobovief as platform of everyday-robot research. *The transactions of the Institute of Electronics, Information and Communication Engineers*, J85-D-I(4):380–389, 2002.

[10] Takayuki Kanda, Hiroshi Ishiguro, Michita Imai, and Tetsuo Ono. Development and evaluation of interactive humanoid robots. *Proceedings of the IEEE (Special issue on Human Interactive Robot for Psychological Enrichment)*, 92(11):1839–1850, 2004.