An Interactive Agent Supporting First Meetings Based on Adaptive Entrainment Control

Tatsuya Hayamizu
Graduate School of Information Science and Technology
Osaka Institute of Technology
1-79-1 Kitayama, Hirakata City, Osaka, Japan
e-mail: hayamizu@is.oit.ac.jp

Kenzaburo Miyawaki
Facility of Information Science and Technology
Osaka Institute of Technology
1-79-1 Kitayama, Hirakata City, Osaka, Japan
e-mail: miyawaki@is.oit.ac.jp

Mutsuo Sano
Facility of Information Science and Technology
Osaka Institute of Technology
1-79-1 Kitayama, Hirakata City, Osaka, Japan
e-mail: sano@is.oit.ac.jp

Kentarou Mukai
Facility of Information Science and Technology
Osaka Institute of Technology
1-79-1 Kitayama, Hirakata City, Osaka, Japan
e-mail: m1m09a21@info.oit.ac.jp

***Abstract*—This paper describes an agent that can facilitate first meeting communications. In this situation, a communication mediator is important because people can feel stress and an inability to talk comfortably. Our agent reduces this stress using embodied entrainment and promoting communication. In previous research into embodied entrainment, appropriate back-channel feedback has been discussed but communication studies have been limited. We propose an embodied entrainment control system that recognizes a state of communication and is adaptive to each situation with effective nonverbal communication. In this way, our agent mediates a balanced, two-way conversation. Our experiments with the agent confirmed its effectiveness across various social skills levels. We demonstrate that the embodied entrainment of our agent in first meetings benefits people who have low social skills, thereby verifying the efficacy of our agent.**

Keywords-Embodied Entrainment; Nonverbal Communication; Introducer Agent; Group Communication Introduction; Social Skills

I. INTRODUCTION

Wider communication capabilities are required for robots to support humans' daily life. In today's world, it is not enough to have a simple relationship between a human and a robot; if robots are to support daily life, we need to develop true socialized robots.

We propose a robot design that can mediate a smooth group conversation and encourage human communication. Humans communicate with each other via many channels that can be categorized as verbal and nonverbal. Nonverbal communication channels such as intonation, accent, eye-contact, gestures such as nodding or other emotional expressions all encourage synchronization, embodied entrainment, and friendship [1]–[7]. Watanabe et al. [8] have investigated such a robotic agent, which includes embodied entrainment, and they developed a speech-driven interactive actor, the Inter Actor, which can predict appropriate timing of nods and gestures with voice information from a communication partner.

However, the Inter Actor focused on synchronization with a listener without regard to group communication that involves both the talker and listener in a balanced, two-way conversation. Kanda et al. [9], [10] adopted a joint gaze function for their robot and have investigated a more adaptive robot through a design-based approach. However, the research focused on rule-based natural behavior of robots rather than a robot design aiming to facilitate communication. In addition, the investigation did not mention group communication.

This paper focuses on group communication and presents a control method of embodied entrainment that is adaptive to the situation.

We begin with an introducer robot for communication between persons meeting for the first time (Fig. 1). Next, we explain our method for adaptive control of embodied entrainment and show the experimental results that verify the basic effectiveness of our agent. Finally, we analyze our agent's effectiveness from the viewpoint of social skills.

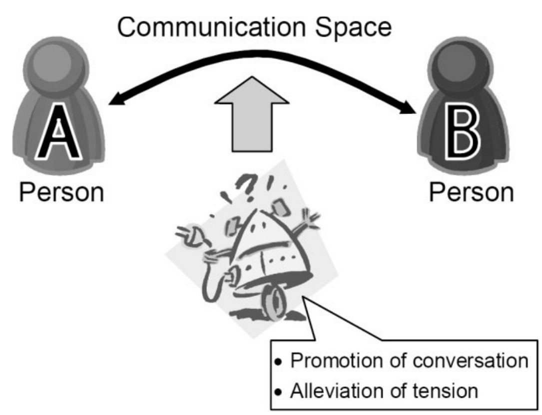

Figure 1. Introducer agent of a first meeting pair.

II. BASIC STRATEGIES TO PROMOTE CONVERSATIONS

This section describes the five mediating behaviors of our introducer agent to promote conversation.

 ISBN: 978-1-61208-250-9

A. Utterances inducement

The agent asks questions related to the participants in order to establish a mutual relationship between itself and each participant.

B. Gaze Leading

For transmission of information from participant A to participant B, the agent moves its gaze to A and asks him/her about the information. Then, the agent moves its gaze from A to B in order to achieve an equal eye-contact.

C. Gaze Distribution and Synchronizing Nod

When the participants talk to each other, the agent makes equal eye-contact with the participants and adjusts its nod timing to those of the participants'.

D. Dynamically Synchronizing

The synchronizing method is dynamically changed according to the state of the two participants as follows.

1) Natural synchronizing to a listener.

When a listener synchronizes with a talker by nodding and other back-channel feedback, the agent echoes these according to the listener's timing to encourage further communication.

2) Synchronizing to a talker to invite the other participant to the conversation.

When a listener does not listen to the talker, the agent tunes its back-channel feedback rhythm to that of the talker and makes eye-contact with the listener. By this behavior, the agent influences the listener and encourages flow of the conversation.

Fig. 2 shows the strategy details. Note that the symbols in the figure represent communication members as follows.

H1 Participant 1.
H2 Participant 2.
R Introducer agent.

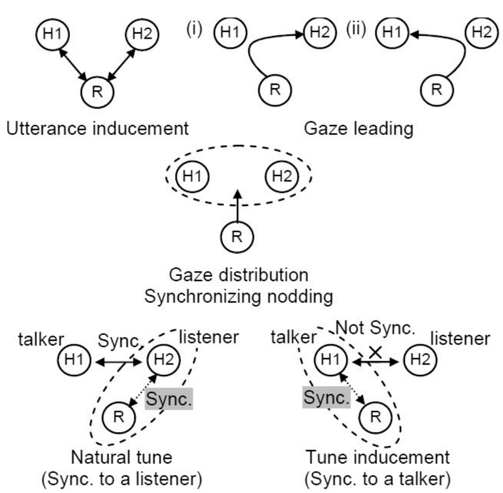

Figure 2. Introducer agent of a basic strategy.

III. ADAPTIVE CONTROL OF EMBODIED ENTRAINMENT

A. Macro Strategy to Promote Group Conversation

These behaviors must be executed in a conversational situation; therefore, we modeled the group communication on a first-time meeting and designed a macro strategy for the promotion of conversation. Fig. 3 describes the strategy and state transition model. The strategy consists of the grounding and enhancement processes on the top layer of the figure. The grounding process establishes a rapport and enhancement process promotes communications within the group. The bottom layer of Fig. 3 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 according to the circumstances of the participants' communication as follows:

1) Greeting

The agent introduces itself and offers a brief explanation of the situation. Then, it introduces the participants' names.

2) Grounding

In this state, the agent tries to establish a relationship between itself and each participant. For this, the agent cites the participants' profiles and asks them simple questions. It aims to encourage them into the rhythm of the agent using utterance inducement. It is assumed that the agent has data such as name, hobbies, or other details of the participants who are being introduced.

3) Topic Offering

This state encourages conversation between the participants. The agent offers information and profiles of one participant to the other or asks a simple question in order to start a conversation between them. By such behavior, the agent manipulates the participants' gaze to make them communicate face to face.

4) Focusing on a Specified Topic

In this state, the agent tries to join in a conversation between the participants. The agent focuses on a topic that was offered at the previous state, Topic Offering.

5) Hearing Conversations

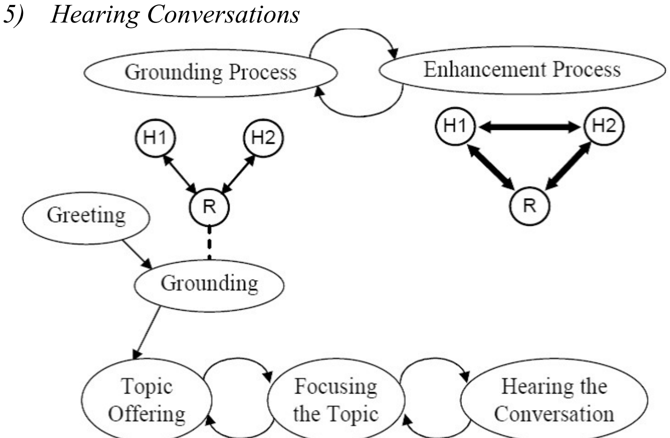

Figure 3. Macro strategy and state transition of an introduction scene.

After successfully making a close and friendly relationship between the participants, the agent keeps hearing their conversations quietly. In this state, the agent nods and looks at the talker using basic embodied entrainment strategies with correct timing.

B. Communication Activity

The agent monitors the participants' communication and estimates the current state as shown in Fig. 3. For the estimation, we define communication activity measurement. The agent calculates the activity in every time slice, defined by the average of speech level and mutual eye-contact of the participants. When this measurement satisfies the condition of a state transition, the agent moves to the next state. To explain this, we define the essential actions of communication by the following symbols that represent functions which return 1 when each action is detected by sensors. Otherwise, they return 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 rancewhen t	X talks to Y
Terminate Utterance (X, t)	X terminates his/her speech.
Gaze (X e/her sp	X directs his/her gaze to Y
Face (X irects h	X looks toward Y
Gaze (X ⇔ Y, t)	X's gaze accords to Y's
Face (X ⇔ Y, t)	X and Y look toward each other
Turn Gaze (R, X azeook	The agent guides X's gaze toward Y
Turn Utterance (R, X tter-ancee	R encourages X to talk to Y

Note that, in this paper:
Gaze (X → Y, t) = Face (X → Y, t) and
Gaze (X ⇔ Y, t) = Face (X ⇔ Y, t)

Our method calculates the communication activity using these symbols and is based on the following two equations:

1) *Ratio of gaze sharing*

$$R_{gazeShare}(x \leftrightarrow y, \Delta t) = \frac{\sum_{t=t0}^{t0+\Delta t} Gaze(x \leftrightarrow y, t)}{\Delta t}$$

2) *Average of utterance power*

$$Average_{Pu}(x, \Delta t) = \frac{\sum_{t=t0}^{t0+\Delta t} UtterancePower(x, t)}{\Delta t} + P_n$$

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

$$Average_{Pu}(x, y, \Delta t) = \frac{\sum_{t=t0}^{t0+\Delta t} average\{P_u(x,t), P_u(y,t)\}}{\Delta t} + P_n$$

Note that Δt is a time span for calculation of the measurement.

Conditions of state transitions depend on the communication activity and are defined as follows.

1) *From Greeting, Grounding to Topic Offering*

In these states, the agent behaves according to the prescribed action scenario that is composed of general protocols of greeting and introduction.

2) *From Topic Offering to Focusing the Topic*

$$R_{gazeShare}(x \leftrightarrow y, \Delta t) \geq Threshold_{gaze}$$

3) *From Focusing the Topic to Hearing Conversation*

$$R_{gazeShare}(x \leftrightarrow y, \Delta t) \geq Threshold_{gaze}$$
and
$$Average_{Pu}(x, y, \Delta t) \geq Threshold_{utter}$$

In the equations, both the $threshold_{gaze}$ and $threshold_{utter}$ are acquired through learning.

C. Learning for Adaptive Control of Embodied Entrainment

The agent retrieves participants' information by sensors and determines its course of action to encourage entrainment. We built a decision tree (Fig. 4) of inductive learning for this interaction. The following three rules are extracted from the tree.

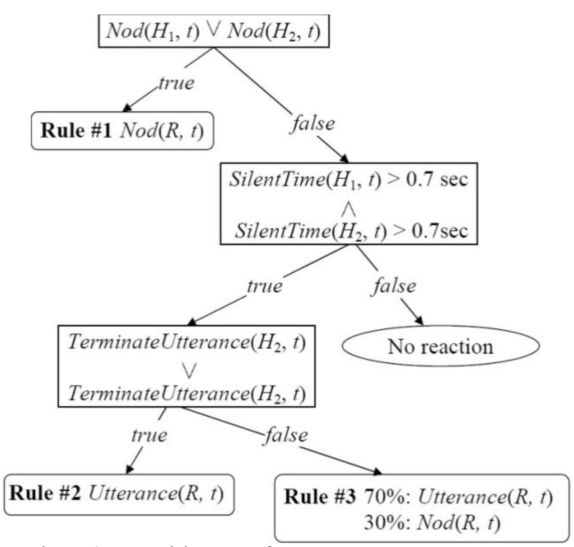

Figure 4. Decision tree of an agent.

1) Rule #1-Reactive nodding

This rule represents reactive nodding to a participant's nod.

2) Rule #2-An utterance

When one participant has finished speaking and the other participant stays quiet, it can be assumed that the conversation is over. The agent recognizes this situation and offers a topic to avoid any silence.

3) Rule #3-An utterance and nod to a participant's speech

This rule is very similar to Rule #2 but represents the situation where a participant stops speaking before finishing. In this situation, the agent is 70% likely to interject and 30% likely to nod. The criteria to determine whether a participant has finished speaking depends on the final part of the utterance. We extract the final part by Japanese morphological analysis.

IV. BASIC EFFECTIVENESS OF A FIRST MEETING INTRODUCER AGENT

Our aim is to promote exchange of information in introductory meetings and we conducted our experiment using an agent to mediate between strangers.

To validate the communication model in first-time meetings, one experiment explored communication between an agent and two strangers (condition with an agent), and the second experiment explored communication between two strangers (condition without an agent). To eliminate potential problems such as hearing difficulties, we excluded elderly participants from these experiments.

Fig. 5 shows our sensing environment allowing real-time gaze tracking of the strangers using multiple cameras (Figs. 5 (e)–(f)). A virtual agent is shown in the display used for this experiment ((Fig. 5 (g)). Fig. 6 shows our virtual agent. An acceleration sensor that was attached to participants detects gestures, whereas a microphone detects their speech (Fig. 7).

Experimental participants included 10 pairs of strangers, who were all university students.

From results of previous experiments, the time of this experiment was set to 6 min. In this experiment, topics offered by the agent included information about each stranger such as hobbies and special skills.

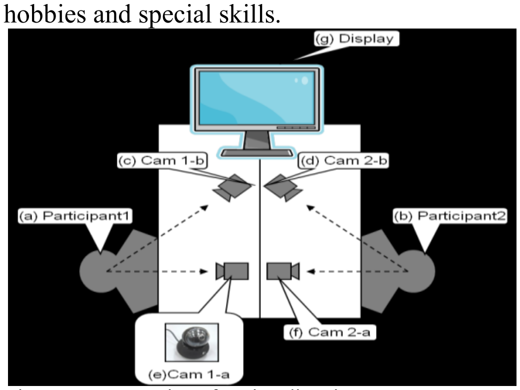

Figure 5. Detection of gazing direction.

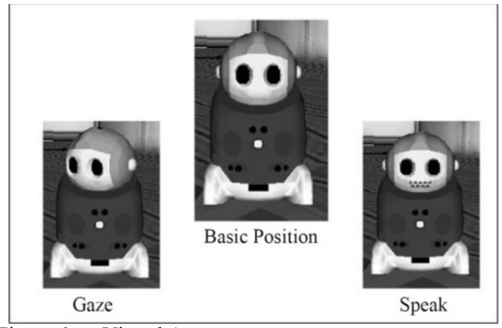

Figure 6. Virtual Agent.

Figure 7. Acceleration censor and microphone.

Fig. 8 shows an example of a communication field transition in this experiment.

Table 1 shows a statistical summary of the questionnaire that included questions such as "Was the conversation stimulating?", "Did you feel comfortable with your partner?", and "Did you come to know your partner well?". Answers were evaluated between 1 (the worst) and 5 (the best).

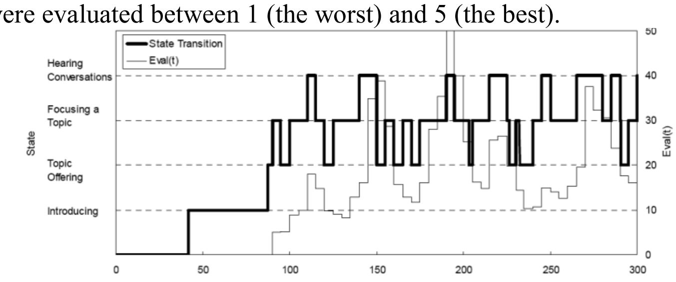

Figure 8. Communication field of state transition and activity.

TABLE 1. STATISTICAL SUMMARY OF QUESTIONNAIRE

Condition	Statistics	Is conversation excit	Did you feel your partner favor	Did you come to know your partner well ?
With Agent	Average	3.9	3.5	4.05
	Median	4	4	4.5
	Covariance	1.165287	1.192079	1.316894
Without Agent	Average	4.1	4.1	4.6
	Median	4	4	5
	Covariance	0.788069	0.788069	0.502625

We compared the questionnaire results of two experiment conditions, i.e., with and without an agent, using a t-test. The results of the t-test showed a higher score with an

 ISBN: 978-1-61208-250-9

agent than without an agent particularly in regard to questions 2 and 3.

This validates the efficacy of our model for mediating first meetings.

V. Analysis of Agent from a Viewpoint of Social Skill

We examined the role of the agent's mediating behaviors with regard to participants' social skills, having already verified the basic effectiveness of the five mediating behaviors previously mentioned.

In this section, we hypothesize that embodied entrainment in mediating behaviors does not affect people with high social skills, but is effective for people with low social skills. Our experiments verify this.

People with low social skills often mistime speech, while people with high social skills risk monopolizing the conversation. We considered that our five mediating behaviors of utterance inducement, gaze leading, gaze distribution, synchronizing to a listener, and synchronizing to a talker are effective communication methods to defuse these situations for people with low social skills.

This experiment was conducted with pairs of strangers and an introducer agent and in conditions of with and without gaze sharing and nodding. In our previous experiment, the subjects were university students. In this test, we used elderly people in order to verify applicability of communication support in geriatric care as communication support for people with low social skills. The number of participants was 38 that included 18 elderly men and 20 elderly women.

The experimental environment was the same as the previous experiment, but this time, we used a real robotic agent for a more effective embodied entrainment than a virtual agent (Fig. 9).

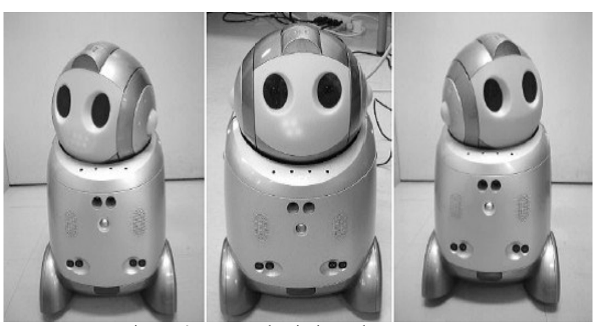

Figure 9. Robotic introducer agent.

We conducted a questionnaire before the experiment and inputted the information obtained to an introducer agent. We explain robot's mediating behaviors to two experimental participants and we shows questionnaire item in monitor to participants.

Each participant completed KiSS-18 for classification into groups of high and low social skills. Table 2 shows statistical summary of participants' social skills.

TABLE 2. AVERAGE OF PARTICIPANTS' SOCIAL SKILL

	Low	High
Sample	19	19
Average	57.2632	72.4737
Standard Deviatio	9.949	3.5492

Table 3 shows the national average of social skill [11].

TABLE 3. NATIONAL AVERAGE OF SOCIAL SKILL

	National Average	
Man	61.82	(n=45,SD=9.41)
Woman	60.1	(n=121,SD=10.5)

After the experiment, the participants answered a questionnaire about the experiment with gestures and that without gestures. We conducted a t-test on the questionnaire with the results shown in Table 4.

TABLE 4. RESULTS OF THE T-TEST

Statistics	Low	High
t-Value	1.9509	1.2865
Degrees of Freedom	16	18
p-Value	0.0344(*)	0.1073

As shown by the t-test results, groups of high social skills cannot verify the efficacy of the agent either with or without gestures. However, groups with low social skills can verify the efficacy of an agent with and without gestures.

This result means that participants who have low social skills gain more from the five mediating behaviors.

VI. Conclusion

Our agent could change a state of communication such as Topic Offering and Focusing and Hearing on the basis of the communication activity measurement, which is calculated from mutual eye-contact and utterance power. On the basis of this state transition, we could stimulate communication through body entrainment behaviors such as nodding, eye-contact, and other back-channel feedback pertinent to each communication state. We conducted communication experiments between strangers and confirmed the effectiveness of adaptive state transitions and smooth interaction control. Our questionnaire confirmed that our agent could promote understanding between the communication pairs. From the evaluation, it is supposed that our system could deal with a stressful situation between strangers.

We aim to adapt our agent to promote information exchange between elderly strangers. For this, we conducted experiments in conditions of with and without gaze sharing and nodding.

 ISBN: 978-1-61208-250-9

From our results, we confirm the efficacy of our agent using gestures in groups with low social skills to encourage conversation between participants and demonstrate that the five mediating behaviors work effectively for participants with low social skills.

ACKNOWLEDGMENT

This research was supported by Mr. Ryouhei Sasama, Mr. Tomoharu Yamaguchi, and Mr. Keiji Yamada of NEC Corp. The authors would like to thank Enago (www.enago.jp) for the English language review.

REFERENCES

[1] J. D. Matarazzo, M. Weitman, G. Saslow, and A. N. Wiens, "Interviewer influence on durations of interviewee speech," J. Verb. Learn. Verb. Beh., vol. 1, pp. 451–489, May 1963.

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

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

[4] W. S. Condon and L. W. Sander, "Neonate movement is synchronized with adult speech: Interactional participation and language acquisition," Science, vol. 183, pp. 99–101, January 1974.

[5] F. J. Bernieri, J. S. Reznick, and R. Rosenthal, "Synchrony, pseudosyn chrony, and dissynchrony: Measuring the entrainment process in mother-infant interactions," Journal of Personality and Social Psychology, vol. 54, pp. 243–253, 1988.

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

[7] H. Kozima, C. Nakagawa, and H. Yano, "Can a robot empathize with people?," International Journal of Artificial Life and Robotics, vol. 8, pp. 83–88, September 2004.

[8] T. Watanabe, "E-cosmic: embodied communication system for mind connection," In Proc. of the 13th IEEE International Workshop on Robot and Human Interactive Communication (ROMAN), pp. 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] T. Kanda, H. Ishiguro, M. Imai, and T. Ono, "Development and evaluation of interactive humanoid robots," Proceedings of the IEEE (Special issue on Human Interactive Robot for Psychological Enrichment), vol. 92, pp. 1839–1850, 2004.

[11] H. Hori and F. Yoshida, "Psychometric scale collection II," Science, 2001.

 ISBN: 978-1-61208-250-9