An Utterance Attitude Model in Human-Agent Communication - From Good Turn-taking to Better Human-Agent Understanding -

Masahide Yuasa
Tokyo Denki University
2-1200 Muzai Gakuendai
Inzai, Chiba, 270-1382, Japan
yuasa @sie.dendai.ac.jp

Naoki Mukawa
Tokyo Denki University
2-1200 Muzai Gakuendai
Inzai, Chiba, 270-1382, Japan

Koji Kimura
Tokyo Denki University
2-1200 Muzai Gakuendai
Inzai, Chiba, 270-1382, Japan

Hiroko Tokunaga
Tokyo Denki University
2-1200 Muzai Gakuendai
Inzai, Chiba, 270-1382, Japan

Hitoshi Terai
Tokyo Denki University
2-1200 Muzai Gakuendai
Inzai, Chiba, 270-1382, Japan

CHI 2010, April 10–15, 2010, Atlanta, Georgia, USA.
ACM 978-1-60558-930-5/10/04.

Abstract
In this study, we discuss a novel expression and comprehension model of the utterance attitude of speaking/hearing during conversations. Humans who participate in conversation display these implicit and explicit attitudes, and use them to understand the other participants in advance of turn-taking. We design abstract animated agents that mimic human turn-taking in conversations to confirm the validity of our model. The subjective evaluation tests show that the expressions of the agents are understandable. The model may facilitate turn-taking in human-agent interaction.

Keywords
Turn-taking, attitude, nonverbal behavior, agent

ACM Classification Keywords
H5.1. Information interfaces and presentation (e.g., HCI): Multimedia Information Systems.

General Terms
Design, Human Factors

Introduction

In human conversation, when a participant wants to speak, he/she often expresses his/her attitude through utterances, such as "I want to speak/I do not want to speak." Humans tend to equally share the opportunity to speak among conversation partners, which may produce a sense of community and pleasantness. When a human wants to speak or to let some other participant speak, he/she expresses the utterance attitudes through nonverbal behaviors, namely, facial expressions, gazes, head orientation, body postures, etc. By observing these expressions, participants understand the partner's attitude and can infer who the next speaker should be. Thus, the participants can avoid overlapping speech or undesirable silence in conversations and enjoy their conversations.

Assuming that humans display expressions of the "utterance attitudes'' through facial expressions, gazes, etc., we propose an attitude expression model for human conversation. With regard to the application of this model to human-agent interactions, this model contributes to designing conversational robots and agents that not only convey information but also share feelings or thoughts with humans. In order to read these feelings or thoughts, these robots and agents need to have the ability to deal with not only explicit behaviors but also implicit behaviors, such as subtle feelings or thoughts, that are expected to be noticed by others. In this work, we propose a model that explains both explicit and implicit behaviors.

Related Work

In pragmatics research and conversation analysis, researchers have been investigating "discourse markers" [1, 5], such as "oh", "well", "then", "so", "because", and "but." Discourse markers function as topic changers, discourse planners, back channelers, etc.; however, unlike nonverbal behaviors, they do not provide information that will help infer the speaker's mental states.

Poggi et al. [7,8] have termed the taxonomy of a speaker's behavior as called "Mind Markers," which appear through various modalities, including facial expressions and gazes that represent the speaker's beliefs, goals and emotions [7]. They clarify the relationship between nonverbal behavior and translated intentions. On the basis of the taxonomy, they have designed embodied multimodal conversational agents that communicate with users [8]. However, they did not propose a comprehensive model consisting of a limited number of basic behaviors and the relevant fundamental implicated meanings.

Therefore, we need a model that explains implicated meanings from the participants' behaviors and that is applicable to the development of robots and agents that have the ability to express subtle or direct attitude and understand those of humans. Especially, it is very important to express and to understand implicit expressive behaviors that were not found in previous researches, in order to read subtle human feelings or thoughts and to produce heart-to-heart enjoyable conversations with humans. As the first step in our research, we primarily focus on the expression model of utterance attitude that is understandable by humans.

The Utterance Attitude Model

In this section, we propose an utterance attitude model. In order to model utterance attitudes that are displayed in advance of turn-takings in conversation, we

observed and analyzed a 20-minute conversational scenes of three female university students recorded on video. We analyzed the chatting on the basis of ethnographic conversation analysis. We observed and carefully transcribed every action of these participants, such as the words (what the participants were saying), and nonverbal actions, such as the gazes, the head orientations, the facial expressions, etc. After the analysis, we proposed an utterance attitude expression model (Figure 1). As shown in the figure, we found that utterance attitudes can be categorized into nine classes on a two dimensional plane: the horizontal axis represents the expressions of a person who "wants to speak/not to speak," and the vertical axis that represents a person who "wants someone to speak/not to speak." The plane also shows that expressions can be classified into two types: subtle implicit expressive behaviors that are expected to be noticed by others and direct explicit behaviors that purposively control the utterance behaviors of others. The inner ring of the model shows implicit attitudes displayed by participants, for example, "(1) I want to speak." The outer ring indicates explicit attitudes that can control the other participants, for example, the "(7) I want him/her to speak" attitude.

The Design of Abstract Animated Agent

We design abstract animated agents that can only express limited fundamental elements of the utterance attitudes. We do not use human-like characters because they have too many variations of expressions for utterance attitudes and also they make the model complicated.

We adopt a sphere with a black circle in the front of the sphere as the shape of agent. The circle directs frontal direction of communication like face and implies the pupils of the eye. Figure 2 and 3 shows the appearance of our agent. The agent has a hand that connected to the body.

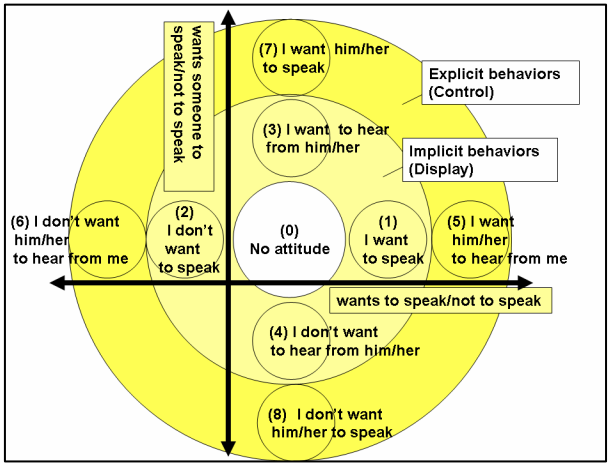

Figure 1. Utterance Attitude Model

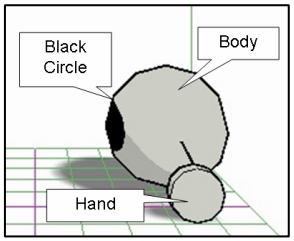

Figure 2. Abstract Animated Agent

Figure 3 shows behaviors of animated agents. The agent's body expanding in height shown in (1) of Fig. 3 represents positive utterance attitude and contracting in height in (2) of Fig. 3 represents the negative utterance attitude. The effects of expanding/contracting

behaviors are well known in the literature of animation movies as in [3, 9]. The frontal black circle of the agent enlarges and reduces in size to represent the agent's interests and activeness to the communications as in (3) and (4) in the Figure 3. The motion imitates that of a pupil that enlarges when a person has interests to talks of someone [2, 10]. Note that the movements of the body and the frontal black circle display the agent's implicit attitudes. The agent has a hand for purposively controlling partners' utterances as shown in Figure 3, (5) inviting, (6) forbidding, (7) promoting and (8) blocking [4, 6] whose movements shows explicit attitudes to control the other's turn.

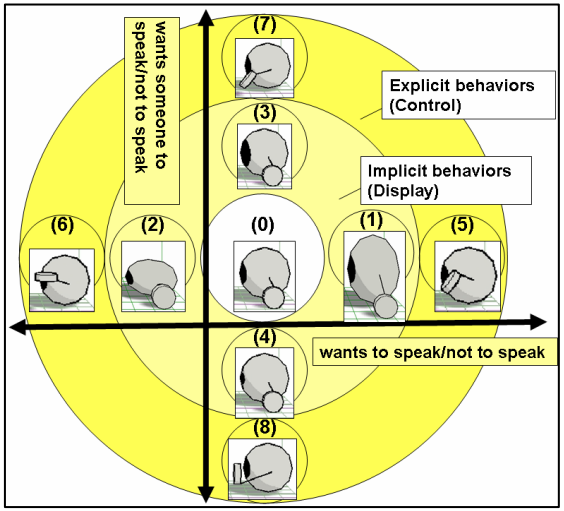

Figure 3. Behaviors of Animated Agent

Experiment

In the experiment, we use nine short video clips for stimulus designed based on the expression model (Figure 3 (0)-(8)). Two agents communicate with each other in the video clips (Figure 4); the agent in left acts as a speaker and the right one is a hearer. As shown in Figure 4, the video starts from a blackout frames (a), then initial scene of (b) of communication.

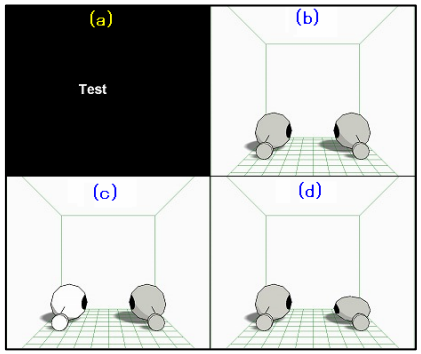

Figure 4. Interaction Scene of Two Agents in the Experiments (Contraction).

Next the left begins blinking that indicates a message or speaking behavior conveyance to the right one as shown in (c). Finally, the right one expresses its utterance attitude. Figure 4 (d) shows "contraction." The duration of each stimulus was approximately six seconds. After watching the right agent's behaviors of each stimulus, the participants in the experiments were requested to rate (Q1-Q4) the utterance attitude of the agent on the right hand side on a scale of 1–5.

Q1: I think the right agent wants to start speaking. **Q2:** I think the right agent wants to hear from the left one. **Q3:** I think the right agent wants the left one to hear from the right one. **Q4:** I think the right agent wants the left one to start speaking.

Results

The experiments involved 22 students from a department of computer science. Figure 5 shows experimental results. We conducted a one-way ANOVA and there was a significant main effect ($p < .01$) for each question (Q1-Q4). As shown in the figure, the expansion behavior (1) can be interpreted as "the right agent express 'wants to speak' attitude" ($p < .01$, Fisher's LSD test) and "contraction (2)" can be as "the right agent expresses 'wants to hear' attitude" ($p < .05$), more precisely than "no attitude expression (0)". Therefore, the expansion and contraction behaviors are operative expressions of attitudes within our model. The black circle enlarging behavior (3) can be interpreted as "the agent expresses 'want to hear' attitude," rather than "reducing circle (4)" ($p < .01$) and "no attitude expression (0)" ($p < .01$). Almost all implicit bodily behaviors of the agents were understandable by participants. In the case of explicit behaviors, the "Inviting behavior (5)" was interpreted as "the agent expresses `wants the other one to hear' attitude" rather than the "Forbidding (6)" ($p < .01$) and "no attitude expression (0)" ($p < .01$). And "Promoting (7)" and "Blocking (8)" were interpreted as "the agent expresses 'wants the other to hear' attitude" more clearly than "no attitude (0)" ($p < .01$, $p < .05$). "Promoting (7)" can be also interpreted rather than "Blocking (8)" ($p < .01$). These results showed that almost all explicit hand behaviors were also understandable by the participants.

Discussion

We cannot show detailed experimental results because of limited pages, but experiments also showed that the hand movements can express the agent's attitude more precisely than the movements of the body and the black-circle. However, some of the behaviors were not clearly interpreted contrary to our initial expectations. It was found that the motion of "expansion" can be interpreted as "I want to hear" as well as "I want to speak." The "Inviting" motion by hand can be interpreted as "I want to hear from it" as well as "I want him/her to speak."

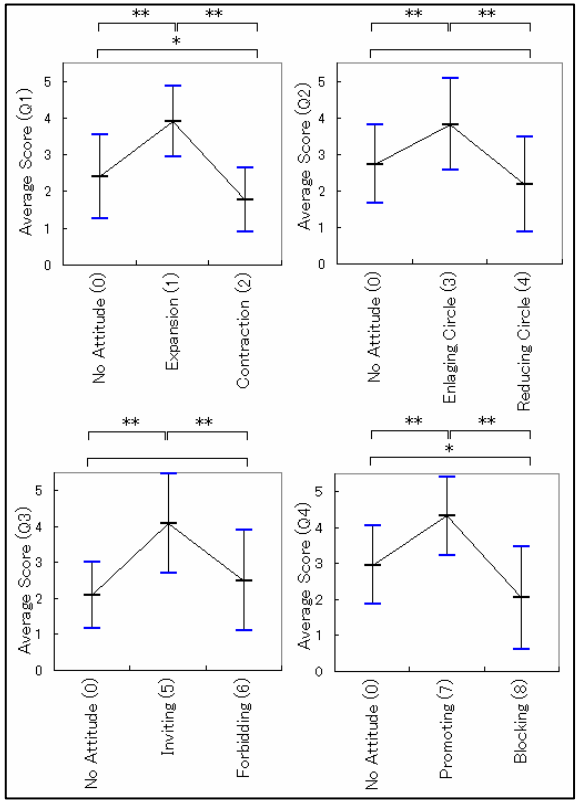

Figure 5. Average Scores for Q1-Q4. (N = 22, p-value: **<.01, *<.05, Fisher's LSD test. The bars indicate the SD.)

We need additional experiments to design more understandable expressions, by referring to psychological findings on nonverbal behaviors. In this work, we did not take into account the cultural aspects of nonverbal behaviors. In the future, we need cross-cultural investigation to establish a universal model. We must admit that we did not take into account combinations of attitudes; "expansion" (the wants-to-speak attitude) and "blocking" (the does-not-want-a-partner-to-speak attitude), etc. Further experiments are required on the combinatorial effects of attitude.

We must emphasize that the role of conversation among humans is not only to exchange the information but also to get a sense of community among friends and family. This work revealed that utterance attitudes of abstract agents which are designed so as to be recognized in the course of a speaker's utterances in advance of the turns, understood by humans. Note that they are not expressed at the very point of the turn-takings. Consequently, they will convey cooperative concerns to human and will work to share desires for communications and sense of unity among humans and agents in the future.

Conclusion

We proposed an utterance attitude model based on observations of human behaviors. It was shown that utterance behaviors are categorized into four subtle implicit expressive behaviors and four direct explicit behaviors. The implicit expressive behaviors that may be noticed by others were not found in previous researches. It is significant to take into account implicit behaviors to design behaviors of conversational agents and robots, in order to produce heart-to-heart enjoyable conversations.

Acknowledgement

This research was partially supported by the Research Center for Advanced Technologies at Tokyo Denki University, and the Research Institute for Science and Technology at Q07J-14, and by a Grant-in-Aid for Scientific Research on Innovative Areas, "Face perception and recognition", by the Ministry of Education, Culture, Sports, Science and Technology, Japan.

References

[1] Fraser, B. An Approach to Discourse Markers, *Pragmatics, 14,* (1990).

[2] Hess, E. H. Attitude and Pupil Size. *Scientific American, 212,* (1965), 46-54.

[3] Lasseter, J. Principles of Traditional Animation Applied to 3D Computer Animation, Computer Graphics, In *Proc. SIGGRAPH 87,* (1987), 35-44.

[4] McNeil, D. *Hand and Mind: What Gestures Reveal About Thought,* University of Chicago Press, (1992).

[5] Schiffrin, D. *Discourse Markers,* Cambridge University Press (1987).

[6] Pease, P. and Pease, B. *The Definitive Book of Body Language,* Bantam, (2006).

[7] Poggi, I. Mind Markers. In *Proc. 5th International Pragmatics Conference*, (1996).

[8] Poggi, I., Pelachaud, C., and Caldognetto, E. M. Gestural Mind Markers in ECAs, In *Proc. AAMAS 2003,* (2003), 1098-1099.

[9] Thomas, F., Johnston, C., and Johnson, O. *The Illusion of Life: Disney Animation,* Disney Editions, (1995).

[10] Vargas, M. F. *Louder Than Words: An Introduction to Nonverbal Communication,* Iowa State Press, (1986).