Induction of Intentional Stance in Human-Agent Interaction by Presenting Goal-Oriented Behavior using Multimodal Information

Yoshimasa Ohmoto*, Jun Furutani* and Toyoaki Nishida*
*Department of Intelligence Science
and Technology
Graduate School of Informatics
Kyoto University
Kyoto, Japan
Email: ohmoto@i.kyoto-u.ac.jp, jfurutani@ii.ist.i.kyoto-u.ac.jp, nishida@i.kyoto-u.ac.jp

***Abstract*—We have made noticeable progress in developing robots and virtual agents; human-like robots and agents are closer than ever to becoming a reality. We want to develop an embodied conversational agent that is regarded as a social partner, not just multimodal interface. However, the mental stance of people when they interact with agents is usually different from when they interact with humans. Therefore, in some cases, it is difficult for people to speculate on an agent's emotion and it is also difficult for an agent to persuade people. To solve this problem, we focused on "intentional stance". Intentional stance is a mental state in which we think that an interaction partner has intention. We hypothesized that agents could induce the intentional stance by performing goal-oriented actions in human-agent interaction. To investigate the effect of induction of intentional stance, we made two agents: a "trial-and-error agent" that performed goal-oriented actions using multimodal behavior and a "text display agent" that displayed its behavioral intention via text. We conducted an experiment in which two participants played customized tag in virtual reality with one of the agents. The results showed that participants continuously tried to communicate with the trial-and-error agent, which did not respond to the participant's actions except when necessary for performing the task. We found that the participants felt that the agent using multimodal nonverbal behavior was more goal-oriented, more intelligent and understood their intentions more than the agent that displayed text above its head. Thus, we were able to induce the intentional stance by presenting a trial-and-error process using multimodal behavior.**

Keywords–Multi-modal interaction, human-agent interaction, intentional stance.

I. INTRODUCTION

In recent years, noticeable progress has been made in developing Embodied Conversational Agents (ECAs), such as robots and virtual agents. Human-like ECAs are closer than ever to becoming a reality. We want to develop an ECA that is regarded as a social partner, rather than just a multimodal interface. Many issues must be dealt with in the production of a social partner agent, such as flexible conversation ability, learning ability in novel situations and so on. We focus here on the issues that relate to the construction of human-agent relationships.

Roubroeks et al. [1] reported the occurrence of psychological reactance when artificial social agents are used to persuade people. In that study, participants read advice on how to conserve energy when using a washing machine. The advice was either provided as text-only, as text accompanied by a still picture of a robotic agent, or as text accompanied by a short film clip of the same robotic agent. The results of the experiment indicated that the text-only advice was more accepted than either advice with the still picture of the robotic agent or the advice with the short film clip of the robotic agent. Social agency theory proposes that more social cues lead to more social interaction, but the result was the exact opposite. This is caused by differences in people's mental state with respect to humans or agents. These differences provide a critical barrier for an ECA to cross before it can be accepted as a social partner. It is thus important that the mental state of people when they interact with the agents is the same as that when they interact with humans.

The mental states that humans can be in with respect to an agent can be defined as physical stance, design stance and intentional stance [2]. When we take the physical stance, we pay attention to physical features, such as the power of the motor, the spec of the display and so on. When we take the design stance, we expect that the agent works mechanically according to predefined rules. When we take the intentional stance, we consider that the agent has subjective thoughts and intentions. When a human interacts with another human, they usually take the intentional stance. In this case, they and their communication partner respect each other. When a human interacts with a machine, they usually take the design stance. In this case, they usually interact with the machine from a self-centered perspective because they do not consider that the machine has its own intentions. To establish social relationships between a human and an artificial agent, the agent has to induce the intentional stance.

The purpose of this study is to investigate how to induce the intentional stance in human-agent interaction. The final goal is to establish social partner relationships between humans and agents. For this purpose, we propose a method to induce the intentional stance, implement the method in an agent and experimentally investigate the effect of inducing the intentional stance.

The paper is organized as follows. Section 2 briefly introduces previous work on the intentional stance. Section 3 explains the outline of the proposed method to induce the intentional stance. Section 4 describes an experiment for comparing two types of methods and then presents the results. Section 5 discusses the achievements and limitations. Section 6 concludes and discusses future work.

 ISBN: 978-1-61208-390-2

II. Related work

Heider and Simmel [3] demonstrated that observers attribute elaborate motivations, intentions and goals to even simple geometric shapes based solely on the purposeful pattern of their movements. Dittrich and Lea [4] discussed that the perception of object's motion as animate depended not only on the interaction between the objects, but also on goal-oriented behavior conveyed by it. From these studies, it can be concluded that goal-oriented behavior is important in the induction of the intentional stance.

If an agent resembles a human or an animal in appearance, people tend to spontaneously think that the agent has intentions. Friedman et al. [5] reported that 42% members of discussion forums about a animal robot named AIBO, a robotic pet, spoke of AIBO having intentions or that AIBO engaged in intentional behavior. On the other hand, some people think that AIBO is just programmed robot. They usually get bored with interacting with AIBO in a short time. These show that the mental stance can dynamically change throughout interaction.

In this study, we attempt to induce the intentional stance by presenting goal-oriented behavior. In short interactions, people take the intentional stance when the agent has similar appearance to a human. However, we aimed at long-term interaction because our final goal is to establish social partner relationships between humans and agents. Therefore, we evaluated how successful the induction of the intentional stance was after a certain length of interaction.

Chen et al. [6] reported that the perceived intent of the robot significantly influenced people's responses when a robotic nurse autonomously touched and wiped each participant's forearm. The participants responded less favorably when they believed the robot touched them to comfort them versus when they believed the robot touched them to clean their arms. In this study, they used the robot's speech, the actions of its arm, and the nursing scenario to convey intent of the robot. These explicit cues could quickly induce intentional stance. We expect, however, that the affective relationship between participants and robot may be short-lived because they can easily estimate the mechanisms of the robot behavior and they feel that the robot mechanically interacts with them.

III. A method for presenting goal-oriented behavior

Recognizing the goal-oriented actions of the artificial agent is important for taking the intentional stance. There are many ways to present goal-oriented actions, but, we think, they are not always useful in inducing the intentional stance. For example, optimized actions for a particular goal are goal-oriented actions, but we do not tend to think that an optimized agent has human-like intentions. In this study, we propose two methods for presenting goal-oriented actions: showing a trial-and-error progression towards a goal using multimodal behavior, and displaying the agent's behavioral intention using text. We compared the differences between these methods and investigate how effective they are at inducing the intentional stance.

A. Task description

In this study, we used a ”customized tag” game in virtual space as an interaction task. Some rules were added to the rules of normal tag, such as ”a tagger cannot tag to players who stand on higher place than the tagger's place,” ”a tagger can tag to players on higher place after the tagger stops in front of one of the players on higher place and counts to five,” ”players can only move limited area separated by the virtual water.” The virtual game environment did not automatically controlled. This means that the players (two humans and one agent) themselves had to judge and communicate about whether the tagger had changed or not, whether the ”five count” was finished or not and whether the players moved the valid region. The game settings encouraged players to consider different objectives to enjoy the game, such as chasing a fastest runner as often as possible, forcing all of the players to be a tagger at least once and so on.

When people play a playground game like a tag game, each player has a different objective in their enjoyment of the game. To ensure that all players enjoy the game, it is important to understand each other's different objectives or goals. Therefore, when playing the customized tag game, participants can take the intentional stance depending on the behavior of the playing partners. In addition, using this game for an experiment allows us to obtain good data for analysis because participants become quite involved in the game [7].

In the customized tag game, the players actively communicate with each other to make all players enjoy the game, such as discussing about the way to control the game (e.g., how to judge the tagger change), advising other players (e.g., ”wait! wait!” and ”please chase other player!”) and seeking to approval (e.g., ”he is cheat! I cannot go to the place!”). The communication behavior is not expressed to non-player characters in general video game. So, we focused on the communication behavior to evaluate whether the intentional stance was induced or not.

B. Outline of architecture

The agents in this study decide their behavior through three layers: a goal layer, a behavior category layer and a concrete behavior layer. The elements of each layer are predefined by a designer of the agent.

The goal layer is the most abstract layer. The elements of the layer show the task goal. In our task, this layer has three goals: chasing other players for an extended time, making the time during which a player is a tagger equal among all the players, and making the number of times that a player becomes a tagger equal among all the players. The first goal is predefined, but when other players do not accept the goal, the goal is changed depending on the other players' behavior.

The elements of the behavior category layer show the category of possible behavior. In this study, the categories of behavior include ”chasing”, ”provocation”, ”dissatisfaction”, ”escape” and ”hiding”. Each category has a parameter named ”effect level”, which indicates how effective it is at achieving a selected goal in the goal layer. Each category is a subgoal of a concrete behavior.

The elements of the concrete behavior layer show the concrete behavior produced by an agent. Each concrete behavior has a parameter named ”expression strength”, which indicates how clearly the behavior expresses the subgoal of the behavior.

The outline of the system architecture is shown in Figure 1 and was developed based on a Belief-Desire-Intention (BDI) model. The overview of each component is briefly explained below.

 ISBN: 978-1-61208-390-2

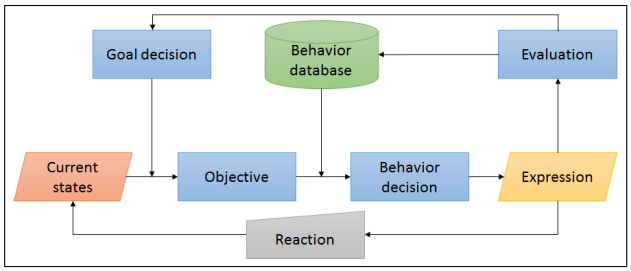

Figure 1. The outline of the system architecture.

Current states:
These are the inputs to the system. The inputs include who is the tagger, each player's position, the time taken to chase other players, how long a player has been a tagger and so on.

Goal decision:
This component determines which goal to achieve. The goal is selected in the goal layer based on the predefined rules.

Objective:
This component determines a category of the behavior. The category is selected in the behavior category layer. This component also determines the values of the "effect level" and the "expression strength" that are needed to achieve the goal.

Behavior database:
The database contains all of the possible behaviors and the structure of the behaviors that are defined in the behavior category layer. The database also contains the current "effect level" and the predefined "expression strength".

Behavior decision:
This component decides a concrete behavior based on the received values of the "effect level" and the "expression strength".

Expression:
This component produces the selected behavior.

Evaluation:
This component evaluates the effect of the concrete behavior on achieving the selected goal. The current values of the "effect level" in the behavior database depend on the evaluation.

C. Method 1: presenting a trial-and-error process using multimodal behavior

In this method, we present a trial-and-error process of achieving a goal using multimodal behavior, such as hand gestures, body orientation, moving speed and iconic motions. Here, we hypothesize that the mental stance, such as the design stance and the intentional stance, changes depending on the agent behavior estimation model. People construct a behavior estimation model but imperfect through interacting with the agent with this method, because it is difficult to precisely interpret multimodal behavior in terms of estimating the behavioral goal. Since we expect consistent goal-directed behavior from the agent, people regard any uncertainties as being caused by the "intentions" of the agent.

The agent uses both the "effect level" and "expression strength" parameters. The expression strength parameter is rated from one to four for each concrete behavior by the designer of the agent behavior, but its value does not change during the task. The value of the effect level changes depending on the output of the evaluation component.

The agent selects its behavior category depending on the value of the effect level. When the same behavior category is selected and achieves the subgoal of the concrete behavior, the agent produces a concrete behavior with higher expression strength than before. When the agent is less able to achieve the selected goal, the value of the effect level decreases. When the effect level of the particular category is less than that of other categories, the behavior category is changed. For example, when the agent wants to provoke a tagged player, first action is "standing near the tagged player." After the provocation is contributed a selected goal (for example, provocation often encourages extending the time to chase other players), the action which has greater value of "effect level" is selected in next phase of the game, such as "waving hand near the tagged player" or "jumping and waving hand near the tagged player." Of course, if the action does not contribute the goal, the agent changes the behavioral category and tries to encourage the goal.

Changes in the expression and the behavior category are thus made in a trial-and-error fashion to achieve the goal. Therefore, when they observe this kind of trial-and-error process using multimodal behavior, people construct a behavior estimation model containing some uncertainties.

D. Method 2: displaying the agent's behavioral intention using text

This method encourages the construction of a behavior estimation model of an agent by displaying intentional agent behavior via text. In this study, the intention of the agent's behavior is a category of the behavior, because each category is a subgoal of a concrete behavior. People construct a behavior estimation model with no black boxes through interacting with the agent with this method, because they can precisely understand the intention of the agent. If only presenting goal-oriented behavior is important in taking the intentional stance, then people interacting with the agent with this method should take the intentional stance.

The agent produces patterns of text corresponding to the behavior category. The diversity of the representations of goal-oriented behavior is the same as the method of presenting a trial-and-error process using multimodal behavior. For example, when the agent wants to provoke a tagged player, the agent displays one of the text expressions, such as he is waiting your chase, he is relaxing and a little bored, "you can't catch me" or "are you tired?" The expression strength is not rated in each text and the text in the same category is randomly selected when the concrete behavior is determined. The value of the effect level changes depending on the output of the evaluation component. The category of the behavior changes in the same way as in the trial-and-error agent.

IV. EXPERIMENT

To investigate the effect of inducing intentional stance, we conducted an experiment using two agents: one was a "trial-and-error agent" that performed goal-oriented actions using multimodal behavior and the other was a "text display agent" that displayed its behavioral intention using text. These agents

 ISBN: 978-1-61208-390-2

were controlled manually (Wizard of Oz) but the behavior planning of the game and the expressions of the multimodal behavior and the text were automatically controlled. We used a virtual reality ”customized tag task”.

To evaluate the effect, we asked the participants to answer questionnaires after the experiment. In addition, we analyzed the number of communicative actions towards the agent throughout the experiment. Since both agents do not respond to the participant's actions except when they need to respond to allow them to perform the task (for example, when the participant is chasing the agent and when the participant argues that the agent is tagged), the number of communicative actions of participants towards the agent decreases. However, we consider, when the participants have intentional stance towards the agent, they unconsciously try to communicate with the agent in the same ways with humans (e.g., calling the agent's name when they excited in the game, waving hand towards the agent who was chased by other player, asking the way to go to the place near the agent and offering a particular action towards the agent). We focused on such communicative actions. The communicative actions were annotated by two annotators and we adopted communicative actions which was annotated by both annotators. We compared the experimental results between a group in which participants interacted with the ”trial-and-error agent” and a group in which participants interacted with the ”text display agent.”

A. Task

Two humans and an agent (which randomly selected the trial-and-error or text display behavior) participated in the customized tag task. We used the virtual space showed in Figure 2, and added a rule to limit the movement range using a region of virtual water.

The game was not controlled automatically. The players (the humans and the agent) judged whether the chaser had changed, whether the count to five had finished and whether the players moved to a valid region on their own.

The human players were allowed free communication using verbal and nonverbal information. Both agents only communicated when it was necessary to perform the task. They did not respond to the utterances of human players in other situations.

B. Experimental setup

In this study, we used Immersive Collaborative Interaction Environment (we call this ICIE) [8] and Unity[9] to construct the virtual environment and the two agents. ICIE uses a 360-degree immersive display that is composed of eight portrait orientation monitors with a 65-inch screen size in an octagonal shape. In this environment, participants could easily look around in the virtual space with low cognitive load like in the real world. The player's virtual avatar could be controlled by their body motions using a motion capture system embedded in the ICIE. The participants could thus easily interact using body motions with low physical constraints. To move in the virtual space, the players used the Wii controller to move the virtual environment. The controller did not interfere with the player's body motions because it was lightweight and had a wireless connection.

Two video cameras recorded the participants' behavior; one was placed on the screen facing the participants, and another was placed behind them. The participants' voices were recorded using microphones. All of the input and output in the virtual space were also recorded.

C. Participants

Sixteen students (14 males and 2 females) participated in the experiment. They were undergraduate students from 18 to 25 years old (an average of 21.5 years old). All of them interacted with one of the agents for 40 minutes. Eight participants (7 males and 1 females) interacted with the trial-and-error agent and the rest interacted with the text display agent. The experimenter gave the following instructions about the agent: ”the agent can recognize your speech. The agent has a lot of knowledge about the customized tag task.” We expected that the participants thought that the agent had conversation ability at least at the beginning of the human-agent interaction.

D. Results

To investigate the degree of induction of the intentional stance, we analyzed the number of communicative actions towards the agent and the participants' subjective impressions of the agent using questionnaires.

1) Analysis of the number of communicative actions towards the agent: The purpose of this analysis is to investigate whether performing goal-oriented behavior influenced the actual communication behavior related to the intentional stance. For this purpose, we counted the number of communicative actions towards the agent. We expected that the number of the actions would decrease when a participant took the design stance because he/she think that the agent never react to his/her communicative actions. On the other hand, we would consistently observe actions because the participant unconsciously produced the communicative actions when they took the intentional stance.

To analyze the changes in the number of the communicative actions throughout the experiment, we divided the time series of the experiment evenly into four periods and counted the number of communicative actions in each period. After that, we compared the number in the second period with that in the fourth period. We did not use the number in the first period because during this period the participants were still learning how to control their avatars and how to play the game. In other words, the first period was the ”ice breaking” period.

T-tests were used on the data from the trial-and-error agent and the text display agent for comparing the numbers in the second and fourth periods. The results are shown in Figure 3: the number in the fourth period was significantly less than that in the second period only for the text display agent ($p = 0.0003$). The participants could clearly estimate the behavior model of the text display agent and in the end they took the design stance towards this agent. The number in the second period for the text display agent was more than that for the trial-and-error agent (but there was no significant difference). We assume that the participant could easily estimate the goal of the text display agent's behavior because its intention was clear. From these results, we suggest that clearly presenting the goal of the behavior can quickly induce the intentional stance, but that the stance quickly changes to the design stance because humans can construct a precise behavior model.

On the other hand, three participants out of eight increased the communicative actions in the trial-and-error agent group

 ISBN: 978-1-61208-390-2

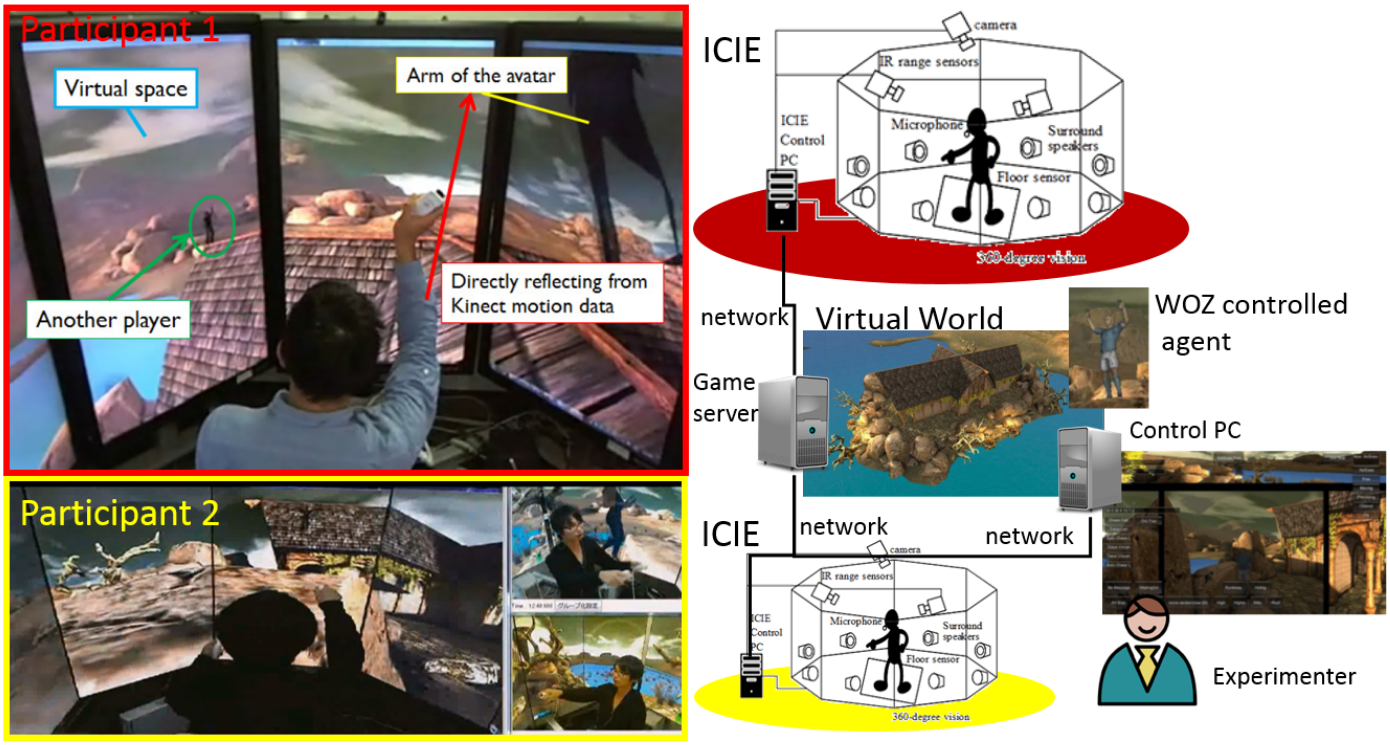

Figure 2. Images during the experiment and the experimental environment.

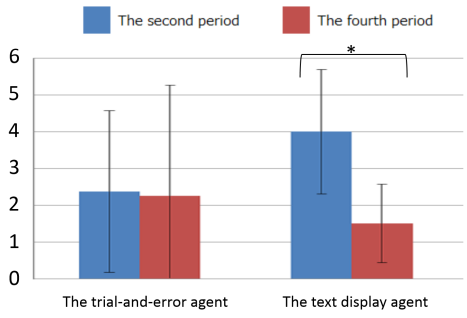

Figure 3. Means of the number of communicative actions towards the agent.

Figure 4. Means of the scores of questionnaires.

though no participant increased in the text display agent group. This means that the trial-and-error agent cannot quickly induce but relatively maintain participant's intentional stance.

2) Questionnaire analysis: The purpose of this analysis is to investigate how the presentation method influenced participants' subjective impressions. The participants answered five rating questions on the ECA's behavior using a seven-point scale. The scale was presented as seven ticks on a black line without numbers, which we scored from 1 to 7. The results are shown in Figure 4. We performed a Mann-Whitney U test on the data in the questionnaire. This analysis shows the final impressions of the agent throughout the experiment.

How human-like do you feel that the agent is?
The average score of the trial-and-error agent was higher than that of the text display agent but there was no significant difference. One reason is that the communication ability of both agents was the same and was poorer than a human's. Since, however, the participants' behavior was changed, we suggest that they unconsciously took the intentional stance.

How strongly do you think the agent has a definite goal?
The participants felt that the trial-and-error agent had significantly more definite goals than the text display agent ($p = 0.039$). This suggests that the agent can effectively provide its goal by performing goal-oriented actions using multimodal behavior. One reason why the text display agent could not do that is that the participants took the design stance because of the artificial "text" and precisely estimated the behavior model.

How smart is the agent?

 ISBN: 978-1-61208-390-2

The participants felt that the trial-and-error agent was significantly smarter than the text display agent ($p = 0.015$). This result also shows that obviously presenting the goal or the intentions is not an effective way to induce the intentional stance.

How well does the agent understand your intentions?

The participants felt that the trial-and-error agent understood their intentions better than the text display agent ($p = 0.054$, marginally significant difference) but the average scores were not high. This result shows the same tendency as the question "how human-like do you feel that the agent is?" The main reason is that the communication ability of both agents was poor. The reason why the text display agent had a lower rating is that the text above the agent's head did not change when the participants communicated.

How much did you enjoy the game?

The participants enjoyed the game significantly more with the trial-and-error agent than with the text display agent ($p = 0.025$), and the scores for both agent were fairly high. This means that the participants were involved in the experiment. We assume that the significant difference was caused by inducing the intentional stance.

V. Discussion

To sum up the experimental results: the trial-and-error agent could not quickly induce participant's intentional stance but, when induced once, the agent maintained intentional stance more than the text display agent. We suggest that the process of constructing a behavior estimation model influences the mental stance participants take to interact with the agent. In addition, an obvious presentation of the inner state of the agent is not effective because the way that an agent presents that is different from the way that a human does.

Clark [10] said that a conversation is a form of joint action. Joint action involves individuals performing individual actions that are intended to carry out a jointly intended shared action. We have also previously proposed that, in some cases of decision-making, the decision or intention is extemporarily shaped, based on the underlying and ambiguous wish (which is one of the sources of the decision and intention) through the interaction [11]. If a person could completely predict and understand a communication partner's behavior and intentions in communication, the communication is the same as conducting a monolog. Therefore, for the text display agent in this study, the participants did not take the intentional stance because they could easily understand the agent's goal. On the other hand, it is difficult to directly understand the goal of the trial-and-error agent from its multimodal behavior. Since the trial-and-error process presented the goal indirectly, the participants had to estimate the goal of the agent through interacting with it.

In future work, we intend to induce the intentional stance and clearly present the goal and the intentions at the same time. We think that a method of implicitly presenting the inner state of agents will be useful for this research.

VI. Conclusions

In this study, we investigated how to induce the intentional stance in human-agent interaction. For this purpose, we tried to induce the intentional stance by presenting goal-oriented behavior in long-term interaction. We proposed two methods of presenting goal-oriented actions and implemented two agents: one was a "trial-and-error agent" that performed goal-oriented actions using multimodal behavior and the other was a "text display agent" that displayed its behavioral intentions via text. We conducted an experiment to evaluate the effect of inducing the intentional stance using these agents. The results showed that participants continuously tried to communicate with the trial-and-error agent, which did not respond to the participant's actions except when necessary for performing the task, and we found that the participants felt that this agent was more goal-oriented, more smart and understood the participants' intentions more than the text-display agent.

References

[1] M. Roubroeks, J. Ham, and C. Midden, "When artificial social agents try to persuade people: The role of social agency on the occurrence of psychological reactance," International Journal of Social Robotics, vol. 3, no. 2, 2011, pp. 155–165.

[2] D. C. Dennett, The intentional stance. MIT press, 1989.

[3] F. Heider and M. Simmel, "An experimental study of apparent behavior," The American Journal of Psychology, 1944, pp. 243–259.

[4] W. H. Dittrich and S. E. Lea, "Visual perception of intentional motion," PERCEPTION-LONDON-, vol. 23, 1994, pp. 253–253.

[5] B. Friedman, P. H. Kahn Jr, and J. Hagman, "Hardware companions?: What online aibo discussion forums reveal about the human-robotic relationship," in Proceedings of the SIGCHI conference on Human factors in computing systems. ACM, 2003, pp. 273–280.

[6] T. L. Chen, C.-H. A. King, A. L. Thomaz, and C. C. Kemp, "An investigation of responses to robot-initiated touch in a nursing context," International Journal of Social Robotics, vol. 6, no. 1, 2014, pp. 141–161.

[7] K. Collins, K. Kanev, and B. Kapralos, "Using games as a method of evaluation of usability and user experience in human-computer interaction design," in Proceedings of the 13th International Conference on Humans and Computers. University of Aizu Press, 2010, pp. 5–10.

[8] Y. Ohmoto, D. Lala, H. Saiga, H. Ohashi, S. Mori, K. Sakamoto, K. Kinoshita, and T. Nishida, "Design of immersive environment for social interaction based on socio-spatial information and the applications." J. Inf. Sci. Eng., vol. 29, no. 4, 2013, pp. 663–679.

[9] "Unity," http://unity3d.com/ (2014/12/01).

[10] H. H. Clark, Using language. Cambridge University Press, 1996.

[11] Y. Ohmoto, M. Kataoka, and T. Nishida, "The effect of convergent interaction using subjective opinions in the decision-making process," in Proc. the 36th Annual Conference of the Cognitive Science Society, 2014, pp. 2711–2716.

 ISBN: 978-1-61208-390-2