PROT – An Embodied Agent for Intelligible and User-Friendly Human-Robot Interaction

Ryota Fujimura, Kazuhiro Nakadai, Michita Imai and Ren Ohmura

Abstract— A system has been developed that can project an embodied agent's image and sound anywhere in a room. It can thus overcome the problems inherent to other embodied agents in dealing with 2D on-screen information and 3D physical information simultaneously. An experiment demonstrated that this 'PROT' agent can effectively present both on-screen information and real-world physical information. Because the PROT agent combines the advantages of a robot agent with those of an on-screen agent, it should improve human-robot interaction.

I. INTRODUCTION

Embodied agents have drawn attention recently for use as an interface to facilitate human-computer interaction. Since most embodied agents are controlled by a networked computer, they can generally handle a lot of information accurately. They can gather information from the Internet and can access physical world information through a sensor network. They can present information in an easily understandable from to a use using such modalities as 'speech,' 'gesture,' and 'movement' [1], [2], [3], [4], [5], [6], [7], [8].

Such embodied agents are generally classified as either "on-screen" or "robot." On-screen agents, such as Peedy [9] in Microsoft® Agent, basically have two-dimensional bodies and appear only on a device display. Robot agents such as HONDA's ASIMO® [10] and ATR's Robovie® [11], have three-dimensional physical bodies. Previous studies on information presentation by embodied agents [12], [13], [14], [15] have shown that, for on-screen domain tasks, on-screen agents present information more effectively and more understandably. This is because they can move quickly across the screen and clearly point to on-screen objects [12], [15]. Studies have also shown that robot agents interact more naturally with the user. This is because they have 3D physical bodies similar to that of a person [13], [14]. That is, the presentation of information by a robot agent' tends to make a better impression on the user.

Since on-screen and robot agents have different areas of specialty advantages, combining an on-screen agent with a robot agent should result in better overall performance. For example, an on-screen agents can point to on-screen objects more clearly than robot agents, which are less capable of moving and pointing. Conversely, a robot agent

Ryota Fujimura is on the Faculty of Science and Technology, Keio University (fujimura@ayu.ics.keio.ac.jp).
Kazuhiro Nakadai is at the Honda Research Institute Japan Co., Ltd. (nakadai@jp.honda-ri.com)
Michita Imai and Ren Ohmura are on the Faculty of Science and Technology, Keio University (michita@ayu.ics.keio.ac.jp, ren@ayu.ics.keio.ac.jp)

can interact with the user more naturally. Nishimura *et al.* suggested combining an on-screen agent with a robot agent for more effective performance of on-screen domain presentation tasks [12].

However, an embodied combination agent has trouble dealing with 2D on-screen information and 3D physical information at the same time because both types of agent have difficulty pointing at physical world objects. For instance, an on-screen agent can easily point to a physical object near the display [16] but not to one far from the display. In contrast, although a robot agent can naturally indicate a physical world object by using deictic words and pointing gestures [17], the pointing is not always accurate. One way to make the pointing more accurate is to move the robot closer to the target object. This is problematic, however, because the motion speed and moving speed of a robot are usually limited by hardware capabilities.

We have developed an embodied agent that can effectively deal with 2D on-screen information and 3D physical information at the same time. Since overcoming the hardware limitations of a robot agent will be a lengthy undertaking, we gave an on-screen agent the ability to handle physical world objects. Our "PROT agent" can quickly and intelligibly point at physical world objects distant from the device location.

We experimentally compared the performance of an on-screen agent, a robot agent, and our PROT agent for 2D and 3D information presentation tasks. Our identification of the differences between the three types of agents clarified the type of agent best suited for different types of tasks. Our findings for the PROT agent showed that an on-screen agent can be applied to physical world tasks.

The rest of the paper is organized as follows. In Section II, we look at previous related work on the use of agents to present information. In Section III we describe the operation of the PROT agent. The experiment used to compare the three agent types is explained in Section IV. In Sections V and VI, we present the results of the experiment and discuss our findings. We conclude in Section VII with a summary of the key points and a look at future work.

II. PREVIOUS WORK

The three main axis for evaluating information presentation by an embodied agent are 1) clarity of presentation, 2) ease of interaction, and 3) user's impression of presentation.

Previous studies have shown that robot agents are more effective than on-screen agents in terms of the ease of interaction and the user's impression of the presentation [12], [13], [14], [15]. This suggests that the strong consistency

between a 3D physical-body robot agent and the 3D physical world is an important factor in information presentation [13], [15].

Several studies have focused on the clarity of information presentation [12], [15], [16], [17], [18]. Nishimura *et al.* showed that, for a presentation using PowerPoint-like slides, an on-screen agent is better in terms of the clarity of the presentation [12]. Shinozawa *et al.* showed that it is also better at recommending 2D on-screen objects than a robot agent is at recommending 3D real-world objects [15]. Nakano *et al.* showed that an on-screen agent can point at physical world objects that are close to the display. Osawa *et al.* compared the humanoid Robovie robot with an agent using object direct anthropomorphization for an information presentation task and found that the an agent using object direct anthropomorphization is better at presenting information about itself [18].

These findings suggest that the clarity of the information presentation is strongly affected by the physical distance between the presenter agent and the target information. That is, an embodied agent should move close to the target information and point at it for the information presentation to be intelligible. Since an on-screen agent can move around freely on the display, it can quickly and intelligibly point at objects on the screen. It cannot intelligibly point at physical environment objects because it cannot move around in the physical environment. On the other hand, a robot agent can move around in the physical environment and point at 2D and 3D objects using a pointing motion; however, the moving and pointing take a long time at present.

III. OPERATION OF PROT AGENT

As mentioned in the Introduction, our PROT agent can quickly and intelligibly point at physical world objects distant from the device location. It does this by projecting an image and sound by the PROT system to a specific point in the physical environment.

The PROT system consists of hardware and software parts. The following sections describe each part.

A. Hardware configuration

As shown in Fig. 1, the PROT system consists of a rotating mirror, a projector, a Web camera, and an ultrasonic directional speaker [19]. These four devices are installed in a pole. The mirror and the projector have a shared servo motor for yaw-axis control, and the mirror has another servo motor for pitch-axis control. There ultrasonic directional speaker and the web camera have two servo motors for controlling yaw-axis and pitch-axis independently. Thus, the PROT system has six degrees-of-freedom. PROT projects an embodied agent on the wall, on the floor, on the ceiling or on an object, and moves the agent freely by controlling the mirror and the projector. In addition, it can produce a speech signal as if the agent speaks by using the ultrasonic directional speaker thanks to its interesting characteristics [19]. An audible-sound-modulated ultrasonic wave is transmitted from the ultrasonic directional speaker. The audible sound appears at
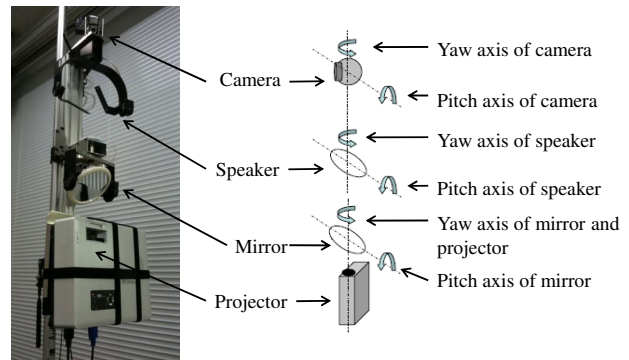

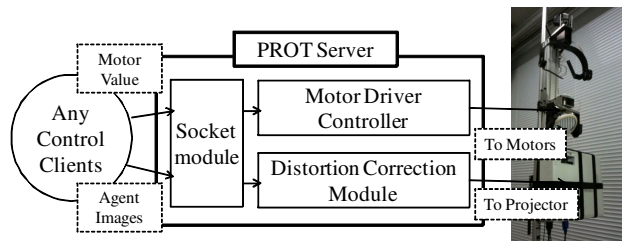Fig. 1.   Hardware configuration of PROT system

Fig. 2.   Software configuration of PROT server system

the point where the ultrasonic wave is reected by objects like the walls and the floor. Thus, by using the ultrasonic speaker, a user can hear voice from the projected image. The web camera is prepared to be used as a telecommunication tool in the future.

B. Software configuration

The PROT system adopt a client-server arrangement. Fig. 2 shows the configuration of PROT server system. There are two main modules: Distortion correction module, and Motor driver controler.

Distortion correction module receives image data from any service clients using TCP/IP socket communication, and projects the received image in a room. Such flexible projection generates distortion caused by the Keystone effect. Thus, the Distortion correction module has a function of real-time Keystone distortion correction. We implemented it by using projective transformation.

The Motor Control Module receives a motor rotation value via socket communication. The angle range of each servo motor is 0 - 300 degrees, and the resolution of control is 0.3 degrees.

These mechanism allows PROT to project a variety of contents because any service client can be connected to the PROT system.

IV. EXPERIMENT

To evaluate the operation of our PROT agent, we carried out a psychological experiment using a task that needs both 3D physical world information and 2D on-screen information. We compared the performances of the three agents in terms of the three axes described in Section II (clarity of
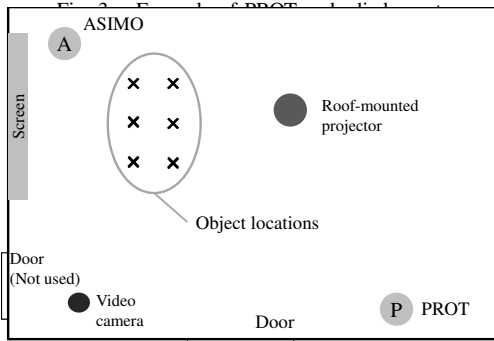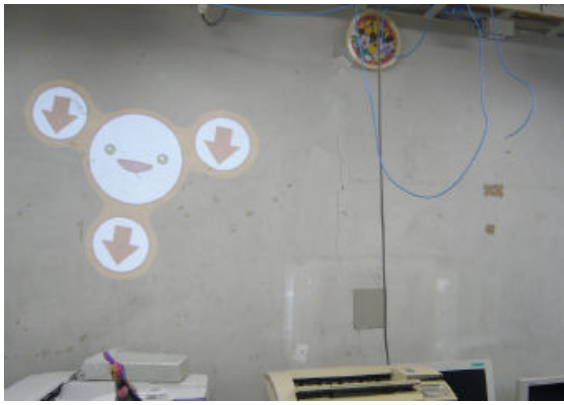
Fig. 4. Experimental setup

presentation, ease of interaction, and user's impression of presentation).

We used the semantic differential (SD) method to investigate the impressions made by the agents and their presentations and used questionnaires to evaluate the differences between the three agents. We also analyzed the verbal interactions between the users and the agents and the time it took to complete the task.

Prior to the experiment, we developed two hypotheses.

H1   A PROT agent is the best in terms of clarity for presentations that associate real-world physical information with on-screen information.

H2   A robot agent is better in terms of presentation impression and interaction ease.

A. Overview and setup

The participants (16 men and 2 women; aged 22 to 46) were asked to select the 3D object that he or she felt best matched a specified 2D character.

Six objects were placed on the floor, and six characters were displayed on a screen. The agent being tested presented them to the participant by pointing at them, answered questions posed by the participant, and recommended an object to select.

The experimental setup is shown in Fig. 4. It was in a 4×7 m room that contained the six objects on the floor, a 100×180 cm display screen, the PROT system, a Honda ASHIMO robot, a projector, and a video camera. The six objects were located 50 cm apart. Three types of object where used: plastic

bottles, cellular phones, and books. For each run, only one types of object was used. The six characters associated with the six objects were projected on the screen.

B. Conditions

Three conditions were tested: on-screen agent, robot agent, and PROT agent. The agent recommended which object to select by using reference terms in Japanese and gesturing. For example, it might say "I think this object matches that character." Three reference terms were used to indicate an object's location: *kore*, *sore*, and *are*. *Kore* refers to an object close to the speaker. *Sore* refers to an object close to the listener or in the middle between the speaker and listener. *Are* refers to an object that is neither close to the speaker nor to the listener [20]. The words used for each presentation were the same except for the reference term. The speech was synthesized in the same manner for each condition. The participant's actions were captured accurately by using the Wizard of Oz method. All the utterances and gestures by the agent were in accordance with a specified scenario. The instructions were given to the participants verbally by the person conducting the experiment.

- On-screen agent condition
  The on-screen agent was projected on the screen (as shown in Figure 5 a)), and it moved around on the screen. It moved close to the on-screen character that was to be matched and pointed at it. To recommend the matching object on the floor, it pointed at it and used the reference term *sore*.
- Robot agent condition
  The robot agent did not move from the position marked A in Figure 4 to avoid impairing the tempo of the presentation. It pointed at the on-screen character and then the recommended object with its arm (as shown in Fig. 5 b)). For the each, it used *kore*, *sore*, or *are* depending on the relative location of the character and the object.
- PROT agent condition
  The PROT agent was first projected on the screen to indicate the character and then on the floor to indicate a matching object. It was also projected on the wall when it was moved from the screen to the floor. It was positioned next to the character and then next to the recommended object (as shown in Figure 5 c)). It used a pointing gesture and the reference term *kore* to indicate which one.

We used a cartoon-like image of ASIMO for the on-screen and PROT agents, as shown in Fig. 6 to avoid the effects of different appearances as much as possible. Of course, the appearance of the robot agent was much different than that of the other two agents, but the projection of a robot's real picture sometimes creates negative effects [21], [22]. In addition, SD analysis (as shown in Section V-A) showed that factor I, 'character,' was about the same for all three agents. This indicates that a cartoon-like image is better for both the on-screen and PROT agent in terms of a fair evaluation.
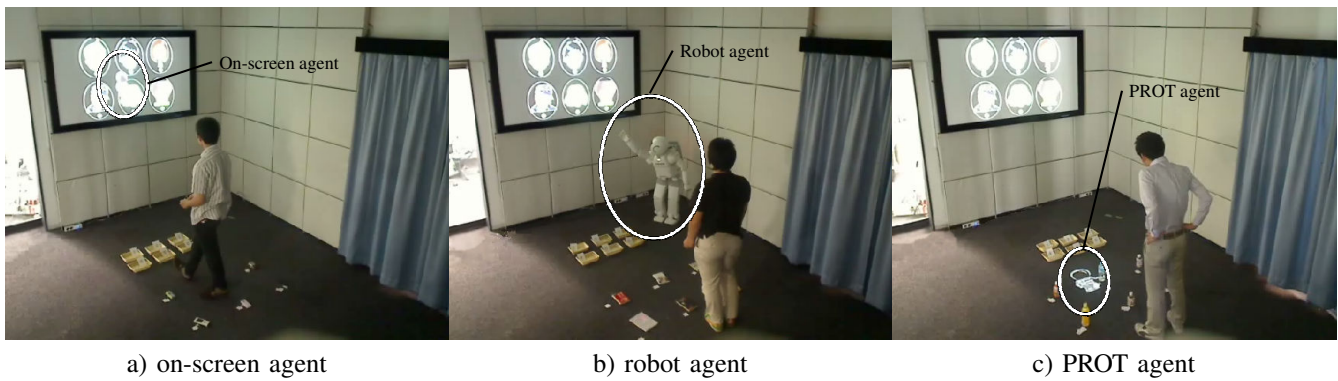
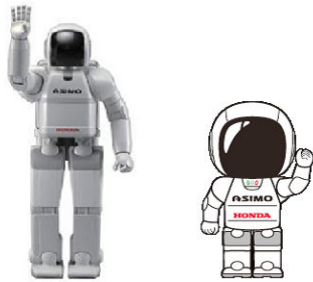|a) on-screen agent | b) robot agent | c) PROT agent |

Fig. 5.   Three experimental conditions

Fig. 6.   Robot agent (left); on-screen and PROT agent (right)

C. Procedure

The experiment had a within-subject design, and the order of all experimental trials was counterbalanced.

1) Each participant was told that the purpose of the experiment was mainly to investigate the relationship between a character and an object.

 • The was told select the object from the six on the floor that was most appropriate for the indicated two-dimensional character.
 • The agent assisted him/her in making the selection.
 • The agent made only a recommendation. The participant made the decision.
 • The agent answered questions asked by the participant.

2) When the participant entered the room, he/she encountered one of the three agents.
3) The agent made a recommendation for each object selection.
4) Immediately after making the six selections, the participant was asked to fill in a questionnaire giving his/her impressions about the agent.
5) The participant completed steps 2) to 4) three times, once for each agent condition. The runs repeated after the participant filled in a questionnaire.

The experiment sessions were videotaped.

D. Prediction

We made predictions of the results of experiment from the hypotheses and previous studies.

1) Predictions from H1: We made predictions from H1 and previous studies as follows:

 • The PROT agent condition get highest scores on the questionnaire item and the adjective pairs regarding clarity of the agent's assistance.
 • The number of confirmation utterances is the lowest for the PROT agent condition.
 • The task completion time for the PROT agent is short as well as the on-screen agent because the PROT agent has about the same range of motion speed as the on-screen agent.

2) Predictions from H2: We made predictions from H2 and previous studies as follows:

 • The robot agent condition get higher scores on the questionnaire item and the adjective pairs regarding ease of interaction.
 • The robot agent condition get higher scores on the questionnaire item and the adjective pairs regarding presentation impression.
 • The number of positive utterances is the highest for the robot agent condition.
 • The task completion time for the robot agent is the longest because the motion speed of the robot agent were limited by hardware capabilities.

V. RESULTS

A. SD method

Twenty-six pairs of adjectives were used for the SD analysis. The judgments were expressed as values ranging from 1 to 7, with 1 being the most negative evaluation (e.g., highly bad) and being the most positive evaluation (e.g., highly good). We carried out a factor analysis for the adjective pairs, and, on the basis of the difference in eigenvalues, we adopted chose a solution consisting of four factors. Table I shows the generated factor matrix, which was rotated using the varimax method, and the scores for each adjective pair. Since the question were written in Japanese, the original meaning of the Japanese adjectives may differ somewhat from those of the English adjectives shown in the table.

TABLE I

SD Factor Analysis

Adjective Pairs		Varimax Rotated Factor Matrix					Mean of Scores			
		FactorI	FactorII	FactorIII	FactorIV	Communality	O_n	R_o	P_R	Significance
Warm	Cold	**0.882**	0.047	0.071	0.029	0.800	3.59	3.82	3.88	N.D
Cheerful	Lonely	**0.813**	0.149	0.139	0.079	0.773	4.24	4.47	4.53	N.D
Light	Dark	**0.769**	0.316	0.165	-0.011	0.728	4.18	4.12	4.35	N.D
Humane	Inhumane	**0.747**	0.177	0.202	-0.181	0.667	4.24	4.35	4.53	N.D
Living	Dead	**0.730**	0.220	0.080	0.036	0.612	3.76	4.35	3.76	N.D
Friendly	Hostile	**0.656**	0.295	-0.019	-0.235	0.607	4.24	4.59	4.47	N.D
Humanlike	Mechanical	**0.574**	-0.117	0.249	-0.001	0.451	3.24	2.94	3.06	N.D
Flexible	Inflexible	**0.501**	-0.084	0.123	-0.348	0.414	3.53	3.41	3.65	N.D
Cooperative	Uncooperative	**0.434**	-0.005	0.366	-0.438	0.659	3.53	3.41	3.76	N.D
Satisfactory	Unsatisfactory	0.077	**0.907**	0.140	0.020	0.872	4.06	4.65	4.47	N.D
Likable	Dislikable	0.201	**0.858**	-0.013	0.044	0.849	4.18	4.76	4.47	N.D
Beautiful	Ugly	0.241	**0.772**	0.156	0.085	0.750	4.53	5.00	5.12	N.D
Good	Bad	0.147	**0.747**	0.443	0.252	0.883	4.12	4.41	4.88	$P>O(p = 7.5e^{-2})$
Exciting	Dull	0.123	**0.669**	0.194	0.143	0.727	3.71	4.65	4.65	$R>O(p = 6.3e^{-2})$, $P>O(p = 6.3e^{-2})$
Pleasant	Unpleasant	0.249	**0.649**	0.159	0.108	0.552	4.24	4.41	4.88	N.D
Distinct	Vague	-0.065	**0.576**	0.042	0.263	0.451	3.94	4.41	4.29	N.D
Dynamic	Static	-0.017	**0.223**	-0.074	0.119	0.238	4.59	4.88	4.24	N.D
Reasonable	Unreasonable	0.428	0.060	**0.667**	-0.075	0.753	3.82	4.00	4.06	N.D
Intelligible	Unintelligible	-0.102	0.293	**0.631**	0.086	0.503	3.24	3.06	5.41	$P>O(p = 5.7e^{-5})$, $P>R(p = 1.6e^{-5})$
Successful	Unsuccessful	0.366	0.289	**0.618**	-0.155	0.650	3.82	4.18	4.12	N.D
Quick	Slow	0.482	-0.046	**0.567**	-0.061	0.648	4.06	2.76	4.12	$O>R(p = 1.4e^{-2})$, $P>R(p = 9.9e^{-3})$
Superior	Inferior	0.256	0.120	**0.560**	0.355	0.595	4.06	4.29	4.12	N.D
Full	Empty	0.392	0.345	**0.510**	0.133	0.660	3.59	4.18	4.00	N.D
Complicated	Simple	0.272	0.085	**0.394**	-0.240	0.999	2.94	3.65	3.29	N.D
Aggressive	Unaggressive	-0.099	0.345	0.013	**0.790**	0.769	4.82	5.00	5.00	N.D
Active	Passive	-0.041	0.116	0.091	**0.665**	0.469	4.47	5.00	4.53	N.D

Factor I represents 'Character' because such adjectives as 'warm,' 'cheerful,' 'light,' and 'humane' have high absolute factor loadings. Factor II represent 'Goodness' because such adjectives as 'satisfactory,' 'likable,' and 'beautiful' have high absolute factor loadings. Factor III represents 'Intelligibility' because such adjectives as 'reasonable,' and 'intelligible' bear heavily on performance. Factor IV represents 'Activeness' because such adjectives as 'aggressive,' and 'active' have high absolute factor loadings.

Figure 7 shows the standardized factor scores. Tukey's test for the factor scores and adjective pairs showed that the *Intelligibility* factor differed significantly between the robot agent and PROT agent conditions ($p = 2.47e^{-2}$). For the 'intelligibility' adjective pairs with a significant difference, the highest score was for the PROT agent condition. Similarly, the 'quick' scores were significantly different, and the scores for the on-screen agent and PROT agent conditions were higher than those for the robot agent condition. These result show that the presentations by the PROT agent were generally intelligible. Tukey's test on the other factors did not yield significant differences.

B. Questionnaire

Figure 8 shows the averaged results for the questionnaire completed by the 18 participants for each agent condition. There were three questions.

Q1 How easy was it to understand the agent's purpose? (1:extremely difficult; 7:very easy)

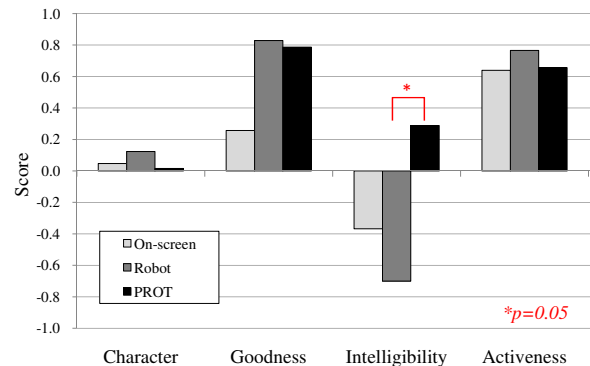Q2 How easy was it to talk to the agent? (1:extremely difficult; 7:very easy)

Fig. 7.   Standardized scores for each factor

Q3 What kind of an impression did the agent make?(1:very weak; 7:very strong)

Tukey's test was performed for each question. There was a significant difference for all three questions: Q1 (PROT>on-screen($p = 9.75e^{-6}$), PROT>robot($p = 3.86e^{-4}$)), Q2 (robot>PROT($p = 8.58e^{-2}$), robot>on-screen($p = 5.54e^{-2}$)), and Q3 (robot>on-screen($p = 1.78e^{-2}$))).

C. Utterance analysis

We transcribed the utterances made by the participants from the recorded videos and classified them into six categories.

C1 Greetings (e.g., "Hello.")

C2 Confirmations about 3D physical world objects (e.g., "That one?")

C3 Confirmations about 2D on-screen characters (e.g., "To which character did you point?")
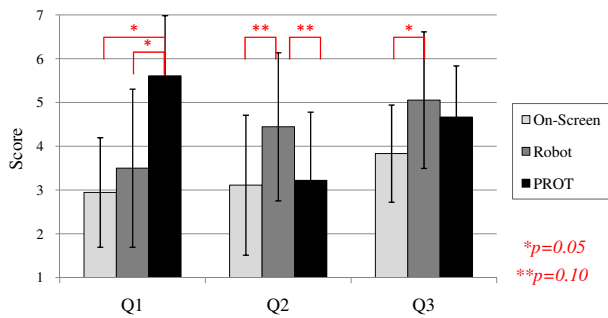
Fig. 8. Averaged results for the questionnaire

C4 Requests for a recommendation (e.g., "What do you think about this object?")
C5 Questions about a recommendation (e.g., "Why do you recommend that one?")
C6 Others (e.g., "I choose this one.")

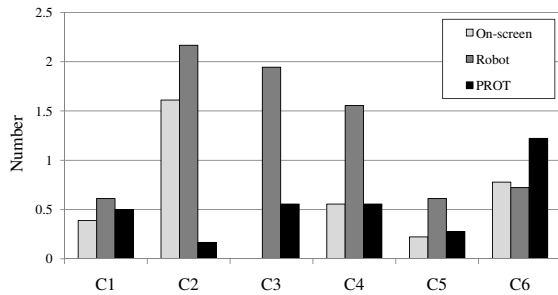Figure. 9 shows the averaged number of utterances for each category.

Fig. 9. Averaged number of utterances by category

The number of confirmation utterances about 3D physical world objects and 2D on-screen characters was the lowest for the PROT-agent condition. For the robot agent condition, there were many confirmation utterances about the 2D/3D information; however, there were many positive utterances such as requests for a recommendation. For the on-screen agent condition, there were many confirmation utterances about the 3D physical world objects but none about the 2D on-screen characters.

D. Task completion time

We measured the time it tool each participant to make the six selections, i.e., the "task completion time." Figure 10 shows the averaged time for each agent condition. Tukey's test for each condition revealed significant differences between robot and on-screen and between robot and PROT. These results show that the speed of motion and clarity of presentation strongly affect task completion time.

VI. DISCUSSION

A. Clarity of information presentation

We first discuss the clarity of the information presentation, which is affected by two factors: "accuracy of pointing" and "speed of motion."
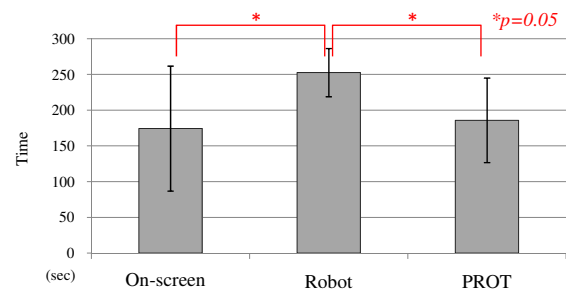

Fig. 10. Averaged task completion time

1) Accuracy of pointing:

The scores for SD Factor III, 'Intelligibility' were significantly different (PROT > robot). In addition, analysis of the answers to Q1 revealed a significant difference in the ease of understanding the agent's purpose (PROT > robot & on-screen). These findings indicate that the accuracy of pointing greatly affected the impression of intelligibility and the understanding of agent's purpose. The pointing of the PROT agent was clear for both the on-screen and physical real-world environments because it could quickly move close to the target. The pointing of the robot agent was ambiguous for both the on-screen and real-world environments, and the pointing of the on-screen agent was clearest for the on-screen characters and the most unclear for the real-world objects.

Analysis of the C2 and C3 type utterances, which reflect the clarity of the presentation, support the results of the SD analysis(Fig.9). The number of confirmation utterances for physical real-world objects was robot > on-screen > PROT. The number of confirmation utterances for the PROT agent condition was about one-tenth that for the other conditions. This shows that the presentations by the PROT agent were clear and easy to understand in the physical real-world environment. The number of confirmation utterances for the on-screen characters was robot > PROT > on-screen. This shows that the presentations by the on-screen agent were the clearest for on-screen characters. Those by the PROT agent were the second clearest, and the number of confirmation utterances was about one-third that for the robot agent condition.

2) Speed of motion:

Figure. 10 shows that the robot agent took more time than the other two types of agent. This finding supports the insight of a previous study [12]. The PROT agent had about the same range of motion speed as the on-screen agent.

3) Summary:

Our findings related to clarity of information presentation support H1; i.e., a PROT agent is better in terms of clarity for presentations that associate real-world physical information with on-screen information.

- When interaction takes place in an on-screen environment, the presentation is clearest under the **on-screen agent** condition.
- When interaction takes place in a physical real-world environment, the presentation is clearest under the **PROT agent** condition.

- When interaction takes place in both an on-screen and physical real-world environment, the presentation is clearest under the **PROT agent** condition.
- The pointing by a robot agent is less clear than that of the other agent types and takes longer because the robot motion is slower.

B. Ease of interaction

Analysis of the answers to Q5 showed that the ease of talking to the agent had marginally significant differences (robot > PROT & on-screen). Furthermore, the participants tended to talk more in the robot-agent condition than under the other two conditions (Fig.9).

These results show that the robot agent was the most effective agent in terms of the ease of interaction. In addition, there was little difference between the PROT agent and on-screen agent conditions. This might be related to the robot's embodiment, since the robot's 3D physical body has a strong presence and a strong connection with the 3D environment.

In short, there were two key findings.

- The **robot agent** had the best performance in terms of the ease of interaction.
- There was little difference between the **PROT agent** and **on-screen agent** in terms of the ease of interaction.

These findings are consistent with those of previous studies [12], [13], [14], [15]. In addition, there was not much difference between the PROT agent and on-screen agent in terms of the ease of interaction.

C. User's impression of information presentation

We think the impression made by the agent affects the impression mad by the information presentation.

The scores for SD Factor II, 'Goodness,' were not significantly different between the three conditions. That is, there was only a small difference between the on-screen agent and obot agent conditions in terms of user impression. For the 'Exciting' adjective pair, there were significant differences between the on-screen agent, robot agent, and PROT agent.

Analysis of Q3 supports these findings since the results for Q3 show that the strength of the impression made by the agent differed significantly (robot>on-screen) and that the scores of for the robot agent and PROT agent were about the same.

These results show that the robot agent was more effective than the on-screen agent in terms of the impression made by information presentation. This finding is similar to that of a previous study [12]. In addition, the PROT agent was more effective than the on-screen agent and received about the same scores as the robot agent.

D. Limitations

Since this study is one of the first systematic studies of the PROT agent, we focused on the primary aspects, i.e., the differences from other types of embodied agents for the information presentation task. Given that many factors remain unexplored, the generality of our findings is limited.

1) Object alignment:

In our experiment, we evaluated the clarity of pointing, which is affected by the characteristics and placement of the objects (e.g., size of object, distance from agent, distance from other objects). Evaluation of the effects of the characteristics and placement on the clarity of pointing remain for future work. Nevertheless, the PROT agent at least has the advantage of being able to quickly move close to the target object.

2) PROT hardware:

The PROT agent is projected onto a plain, flat surface, such as a floor or wall, by a projector. Accordingly, there are several locations at which the PROT agent cannot be projected, for example, on a steel shelf, or on a glassy desk. In addition, if the PROT agent were to be used in a well-lighted area (e.g., out doors), its viewability would be degraded because of the limitation on the brightness of the projector light. In such an environment, the clarity of the pointing by the PROT agent would likely be lower.

E. Summary

Our findings related to impression made by information presentation support H2; i.e., a robot agent is better in terms of presentation impression and interaction ease.

- **robot agent** had the best performance in terms of ease of interaction and impression made by information presentation.
- **PROT agent** had the best performance in terms of clarity of information presentation. It also had better performance in terms of impression made by information presentation than the on-screen agent.

VII. CONCLUSION

Our PROT agent system can project an agent's image and sound anywhere in a room. It can thus overcome the problems inherent to other embodied agents in dealing with 2D on-screen information and 3D physical information simultaneously. Our experiment demonstrated that a PROT agent can effectively present both on-screen information and real-world physical information. Because our PROT agent combines the advantages of a robot agent with those of an on-screen agent, it should improve human-robot interaction. We are currently working in such a hybrid system and will report the results in a future paper.

REFERENCES

[1] G. Ball and J. Breese, "Emotion and personality in a conversational agent," in *Embodied conversational agents*. The MIT Press, 2001, pp. 189–219.

[2] P. Becheiraz and D. Thalmann, "A behavioral animation system for autonomous actors personified by emotions," in *Proceedings of the 1998 Workshop on Embodied Conversational Characters*, 1998.

[3] J. Beskow and S. Mcglashan, "Olga - a conversational agent with gestures," in *In Proceedings of the IJCAI'97 workshop on Animated Interface Agents - Making them Intelligent*. MorganKaufmann Publishers, 1997.

[4] J. Cassell and K. R. Thorisson, "The power of a nod and a glance: Envelope vs. emotional feedback in animated conversational agents," *International Journal of Applied Artificial Intelligence*, vol. 13, no. 4-5, pp. 519–538, 1999.

[5] D. W. Massaro, *Perceiving talking faces: From speech perception to a behavioral principle*. The MIT Press, 1997.

[6] C. Nass and L. Gong, "Maximized modality or constrained consistency?" in *Proceedings of the AVSP'99 Conference*, 1999.

[7] C. Nass and K. Isbister, "Personality in conversational characters: Building better digital interaction partners using knowledge about human personality preferences and perceptions," in *Proceedings of the 1998 Workshop on Embodied Conversational Characters*, 1998, pp. 105–114.

[8] B. Reeves and C. Nass, *The Media Equation*. Cambridge University Press, 1996.

[9] "Microsoft agent home page." http://www.microsoft.com/msagent.

[10] "Honda asimo official site," http://asimo.honda.com.

[11] M. Imai, T. Ono, and H. Ishiguro, "Robovie: Communication technologies for a social robot," *International Journal of Artificial Life and Robotics*, vol. 6, pp. 73–77, 2003.

[12] Y. Nishimura, K. Kushida, H. Dohi, M. Ishizuka, J. Takeuchi, M. Nakanoa, and H. Tsujino, "Development of multimodal presentation markup language mpml-hr for humanoid robots and its psychological evaluation," *International Journal of Humanoid Robotics*, vol. 4, no. 1, pp. 1–20, 2007.

[13] J. Wainer, D. J. Feil-Seifer, D. A. Sell, and J. Mataric, "Embodiment and human-robot interaction: A task-based perspective," in *Proceedings of the 16th IEEE International Conference on Robot & Human Interactive Communication*, 2006, pp. 872–877.

[14] A. Powers, S. Kiesler, S. Fussel, and C. Torrey, "Comparing a computer agent with a humanoid robot," in *Proceedings of the 2nd ACM/IEEE International Conference on Human-robot Interactions*, 2007, pp. 145–152.

[15] K. Shinozawa, F. Naya, J. Yamato, and K. Kogure, "Differences in effects of robot and screen-agent recommendations on human decision-making," *International Journal of Human-Computer Studies*, vol. 62, pp. 267–279, 2004.

[16] Y. I. Nakano, G. Reinstein, T. Stocky, and J. Cassell, "Towards a model of face-to-face grounding," in *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics (ACL03)*, July 2003, pp. 553–561.

[17] O. Sugiyama, T. Kanda, M. Imai, H. Ishiguro, N. Hagita, and Y. Anzai, "Humanlike conversation with gesutures and verbal cues based on a three-layer attention-drawing model," *Connection Science*, pp. 379–402, 12 2006.

[18] H. Osawa, R. Ohmura, and M. Imai, "Using attachable humanoid parts for realizing imaginary intention and body image," *International Journal of Social Robotics*, vol. 1, no. 1, p. DOI 10.1007/s, 2009.

[19] K. Nakadai and H. Tsujino, "Towards new human-humanoid communication: Listening during speaker by using ultrasonic directional speaker," in *Proc. ICRA*, 2005, pp. 1495–1500.

[20] O. Sugiyama, T. Kanda, M. Imai, H. Ishiguro, and N. Hagita, "Three-layered draw-attention model for humanoid robots with gestures and verbal cues," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS2005)*, 2005, pp. 2140–2145.

[21] N. Yee, J. N. Bailenson, and K. Rickertsen, "A meta-analysis of the impact of the inclusion and realism of human-like faces on user experiences in interfaces," in *Proceedings of Computer Human Interaction*, 2007, pp. 1–10.

[22] C. Bartneck, "How convincing is mr. data's smile: Affective expressions of machines," *User Modeling and User-Adapted Interaction*, vol. 11, no. 4, pp. 279–295, 2001.