Differences in Effect of Robot and Screen Agent Recommendations on Human Decision-Making

Kazuhiko Shinozawa [a,*] Futoshi Naya [a] Junji Yamato [b]
Kiyoshi Kogure [a]

[a] *ATR Intelligent Robotics and Communication Laboratories, 2-2-2 Hikaridai, Keihanna Science City, Kyoto, Japan, 619-0288*

[b] *NTT Communication Science Laboratories, NTT, 3-1 Morinosato Wakamiya, Atsugi, Kanagawa Japan, 243-0198*

Abstract

This paper compares the effect of a robot's and on-screen agent's recommendations on human decision making using a quantitative evaluation method. We are interested in whether a robot's physical body produces some differences in the effect or not. Previous research investigated the advantage of a physical body; however, the advantage was not clarified quantitatively and there was not enough evidence to give the results credibility. Our method based on quantitative evaluation clarifies the effect of a robot's and on-screen agent's behavior on user decision making. Comparing a robot's behavior with an on-screen agent's, we show that the degree of the effect firmly depends on the interaction environment and that geometrical consistency between the interaction environment and embodied social agents (ESAs), which include robots and on-screen agents, is important in the recommendation situation.

Key words: Human-robot interaction, interactive social agents, social responses to technology, persuasion.

* Corresponding Author, Tel: +81-774-95-1488, Fax: +81-774-95-1408
Email addresses: shino@atr.jp (Kazuhiko Shinozawa), naya@atr.jp (Futoshi Naya), yamato@eye.brl.ntt.co.jp (Junji Yamato), kogure@atr.jp (Kiyoshi Kogure).

1 Introduction

The recent rapid growth of computer power has enabled the development of on-screen agents that interact with their users in a natural and friendly manner through speech recognition, synthesized voice, and action display. In addition, robots or robotic companions that have a physical body are attracting attention as communication partners.

Such embodied social agents (ESAs) make interaction more meaningful than it is when interfaces do not appropriately display actions or speak (Beskow and McGlashan, 1997). It is known that people's attitudes towards computerized media are similar to the attitudes they have towards other people (Reeves and Nass, 1996). Even if people only read text or hear a voice from computers, they tend to assign some social existence to them. More social richness, defined as more complete human-like presentations, promises to make computers more attractive, productive, and easy to use. Some research has provided fruitful results and suggestions for presentation, i.e., graphical appearance (Massaro, 1998), non-verbal behavior (Cassell and Thórisson, 1999), and speech characteristics (Nass and Gong, 1999), and for personality (Nass and Isbister, 1998), emotion (Ball and Breese, 1998; Becheiraz and Thalmann, 1998), ethnicity (Takeuchi et al., 1998), and interpersonal communication strategy (Shinozawa et al., 2001) as well. Much of such research suggests that an ESA should be an effective interface for interactions with media.

The above research mainly focused on graphical on-screen agents and computers. Recently, robots having a physical body have attracted some attention as useful physical agents, and the above research results may apply to interaction with such robots. However, when we consider robots as ESAs, a new research topic, "dimensionality", appears. A robot has a three-dimensional (3D) physical body while an on-screen agent has a two-dimensional (2D) one. This leads to several questions: Does increasing dimension make a big difference or not? Does the physical 3D appearance affect us in a significant way during the interaction? When both a 2D agent like an on-screen agent and a physical 3D agent like a robot have a similar shape and use the same voice, what is the significance of the difference in dimensionality? Little research has focused on dimensionality, and we still have no solid answers.

We live and work in three-dimensional space. Everything has three dimensions and is located in 3D space. With a 3D body, pointing to some location makes it easy to understand what is being pointed at. When a robot navigates a person, the combination of the robot's gestures and its body's direction has a strong relationship with high ratios of successful task completion (Ono et al., 2001). This suggests that the dimensionality produces a difference in the effect of interaction and that a 3D body makes interaction more meaningful. However,

to confirm this hypothesis, we have to directly compare the effect of 2D and 3D agents in various tasks, where the locations to which agents should point are changed.

2 Task

The color-name selection task was introduced to quantitatively measure the influence of ESAs on human decision-making (Shinozawa et al., 2001). The ESA's role is either recommendation, such as in an advertisement, or assistance in a navigation task. When the ESA is an assistant, it can easily influence users' decisions because users want to know appropriate information. Generally speaking, however, changing a user's mind is difficult in the advertisement situation. Advertising is an important application of ESAs, and we can also say that the recommendation includes helpful interaction like assistance, because the initial recommendation does not always depend on what the user wants. So, we compared the effect of a robot with that of an on-screen agent by measuring how ESA recommendations influence user decision-making in color-name selection tasks.

In a color-name selection task, a subject selects the color name from two options and the selection ratio of the recommended color names is measured. The ratio is treated as showing the degree of a recommendation's influence. We compared the ratios between robot and on-screen agent's recommendations to investigate the degree of influence.

2.1 Condition

Two situations are considered. In one, an ESA points to an object located in 3D space and in the other, it points to an object in 2D space during interaction. Accordingly, we prepared two scenes for interacting with an ESA.

One scene is equivalent to the original one in the color-name selection task (Shinozawa et al., 2001). An ESA recommends a color name while it points to or looks at the color region and two color-name options on a CRT display when a subject should select one of them. In this case, objects used in the selection task are mainly located in 2D space. We therefore call this, the 2D world condition.

The other scene is a new one that we call the 3D world condition. The color region that an ESA points to is in 3D space. Actually, we developed two new machines for displaying color regions in 3D space. One displays printed

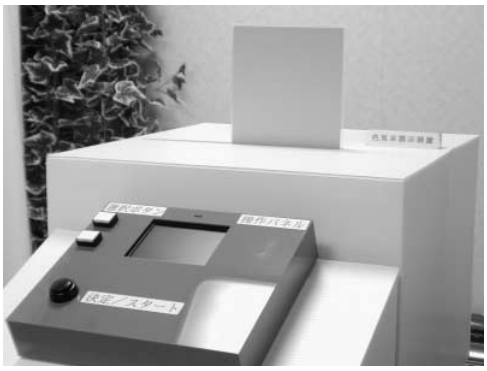


Fig. 1. Display machines

color plates according to external PC control. The other is a button box for displaying and selecting a color name. In the 3D world condition, a subject looks at printed color plates and selects color names using the button box (Fig. 1)

The ESA recommended one of the two options under these two conditions, and we investigated the dimensionality by comparing the ESA's effect on user decision-making.

2.2 Robot and On-screen agent

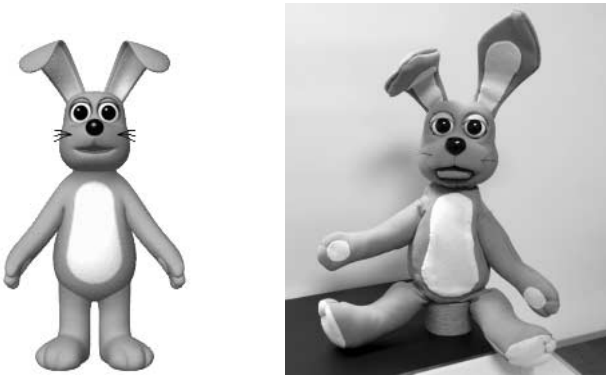
Appearance is important for robots as well as for on-screen agents. Humans tend to recognize social roles, gender, or character by analogy with appearance. Prior knowledge according to appearance has much influence on subjective evaluation (Shibata and Tanie, 2001).

In this research, to avoid such influence, the appearance of robots and on-screen agents are made as equivalent to each other as possible (Fig. 2(a) and 2(b)). The robot's height is 300 mm and on-screen agent's height is similar to the robot's.

Similarly, voice plays an important role in molding the robot's or on-screen agent's character. Both the robot and on-screen agent use the same voice, which was made by "Fluet", the Japanese speech synthesizer developed by NTT (Mizuno and Nakajima, 1998). The robot was also developed by NTT.

2.3 Gestures

We prepared 27 gestures for both the robot and on-screen agent, which included pointing to a color region, nodding, blinking, and so on. We made the



(a) On-screen agent (b) Robot

Fig. 2. Appearance of ESAs

robot's motions similar to the on-screen agent's motions.

2.4 Color names

Before starting an experiment, subjects were told that this was a color name selection task and that they should make a selection based on their own feeling and that there were no correct answers. Most of the color regions and options for color names in the experiment, such as vermilion or carmine, are unfamiliar to ordinary people.

2.5 Speaking words

Instead of having definite knowledge about the displayed color, the robot and agent offered their personal opinions, for example, "I think it is vermilion", "This shade is vermilion, isn't it?", instead of making statements that would indicate it had definite knowledge about the displayed color. This was to avoid the effect of the subject's attributing any authority to the robot and agent.

2.6 Recommended color names

We carried out pretests without an ESA's recommendation and determined what color names the ESA should recommend and the orders of color name options. In both the 2D and 3D world condition, the same order of color names and the same recommended color name options were used.

2.7 Displayed Color region

In both world cases, the size of the displayed color region was about 270 mm × 160 mm, and the average distance from subjects to color regions was about 600 mm.

In the 2D world case, the colors displayed on the CRT display were measured by the CRT color analyzer three times a day. The changes in these values were small (less than 10%) for the whole experiment. In the 3D world case, the colors displayed on printed plates were measured by a tristimulus colorimeter once a day. The changes in these values were also small (less than 7%) for the entire experiment.

Therefore, all of the subjects saw the same color in each of the world cases.

2.8 Subjects

Six experiments were conducted to manage all combinations described above.

(1) In the 2D world case
   (a) No recommendation (Group **No2**)
   (b) Agent recommendation (Group **Ag2**)
   (c) Robot recommendation (Group **Ro2**)
(2) In the 3D world case
   (a) No recommendation (Group **No3**)
   (b) Agent recommendation (Group **Ag3**)
   (c) Robot recommendation (Group **Ro3**)

None of the subjects were experts on color names and all were recruited from the general public. Each subject participated in only one experiment; never more than one. Table 1 shows information about the subjects in each group.

2.9 No recommendation case

To investigate the influence of an ESA's recommendation on user decisions, we must know the mean of the selection ratios without recommendation. We therefore conducted no-recommendation experiments for the 2D and 3D world condition. In these experiments, subjects did not see any on-screen agents or robots and chose a color name with no recommendation. In all recommendation conditions, the recommended color name options were fixed due to the pretest described above. The difference in selection ratios between the

Table 1
Subjects in each group

	Group					
	No2	**Ag2**	**Ro2**	**No3**	**Ag3**	**Ro3**
Number	30	30	30	31	27	30
Mean age	23.87	27.60	23.30	25.40	25.00	26.29
Max. age	49	39	36	36	39	45
Min. age	19	19	18	18	20	18
Ratio of males	0.43	0.50	0.50	0.50	0.44	0.48

no-recommendation case and recommendation case shows the degree of the recommendation's influence on user decision-making. When selection ratios in the recommendation case are greater than in the no-recommendation case, the influence is considered to be positive.

2.10 Procedure

Upon arriving at the lab, subjects were told that the purpose of this experiment was to mainly investigate the relationship between color region and color name, and that they should make a selection based on their own feeling because there were no correct answers. After this explanation, they signed an informed consent statement.

2.10.1 No-recommendation case

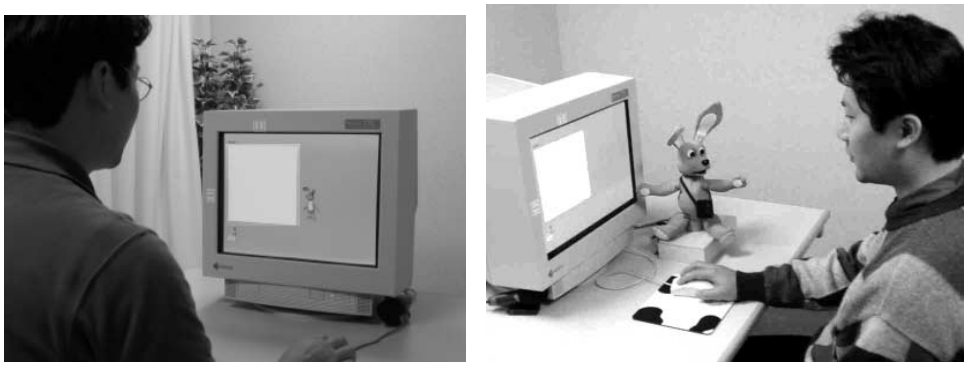
In the no-recommendation condition, subjects did the task without the ESA's recommendation and there was no ESA near them.

2.10.2 Recommendation case

They encountered the robots or on-screen agents for the first time when they entered the experimental room.

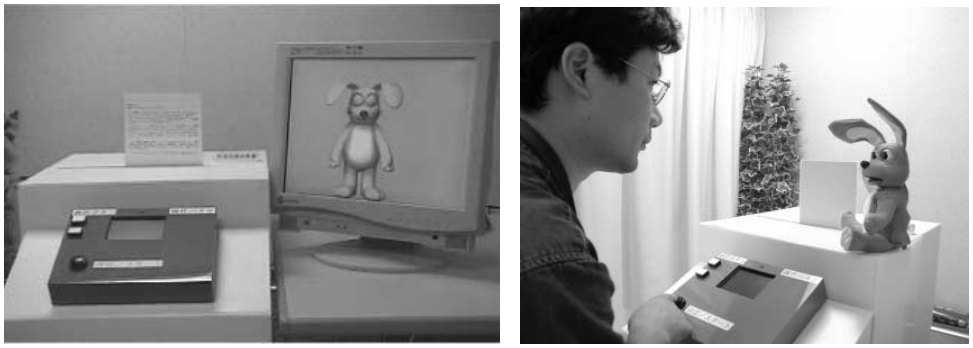
The ESA behaved like it was asleep until the subject pushed a button. Once the button had been pushed, the ESA behaved like it had been awakened and introduced itself, and the experiment started.

The experiment consisted of thirty questions, and each question had two possible responses. When presenting each question, the ESA made a statement endorsing one of the two possible responses.



(a) Agent recommendation (b) Robot recommendation

Fig. 3. Scene in 2D world case



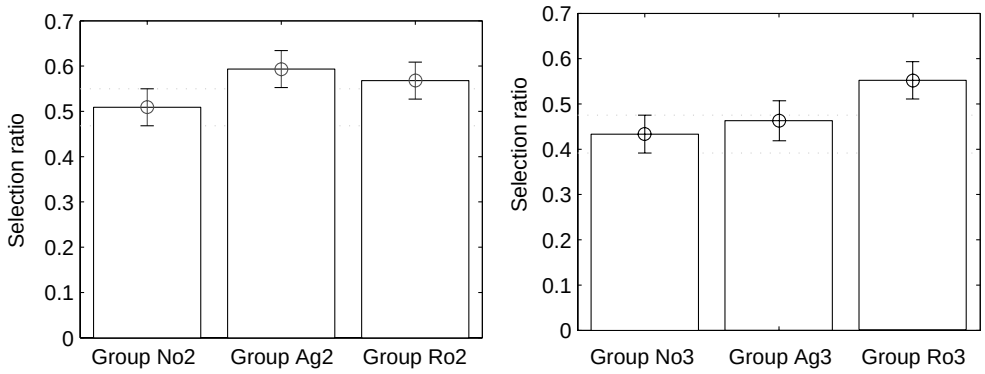
(a) Agent recommendation (b) Robot recommendation

Fig. 4. Scene in 3D world case

While the ESA was asking the question and presenting the two choices, these options appeared on the computer display in the 2D world condition and on the button box display in the 3D world condition. The subject in both conditions indicated his/her choice by clicking a radio button on the computer display and by pressing the corresponding button. The subject then pressed the "OK" button to send the selection to the computer.

If the choice matched the ESA's suggestion, the ESA nodded with approval while expressing a positive statement. If it did not match the suggestion, the ESA bowed and shook its head slowly while responding with a negative statement. This continued until all questions were answered. When the interaction finished, the experimenter gave the subject a questionnaire.

Figures 3(a) and (b) show scenes of the on-screen agent and robot experiment in the 2D world condition, respectively. Figures 4(a) and (b) show scenes of the on-screen agent and robot experiment in the 3D world condition, respectively.



(a) 2D world case (b) 3D world case

Fig. 5. Selection ratios

Almost all subjects finished one experiment in less than 20 minutes. The options that subjects selected were automatically recorded in a computer when subjects pushed the OK button. The scenes in one experiment were videotaped.

3 Results

We calculated the mean selection ratios of the color names that the agent or robot successfully recommended to each subject. The mean ratios in the groups were also calculated. In the no-recommendation case, subjects did not get any recommendation, but the same color name options as in the recommendation case were presented. In the recommendation case, one of the color names was recommended. In estimating the mean selection ratios in the no-recommendation case, the mean selection ratios of the color names that were recommended in the recommendation case were calculated.

3.1 2D world case

Figure 5(a) shows the mean selection ratios for each group in the 2D world case. Factorial analyses of variance (ANOVA) were conducted for each mean of selection ratios. We compared the mean between no-recommendation, on-screen agent's recommendation, and robot's recommendation. There was a significant difference only between Group **No2** and Group **Ag2** (ANOVA, $F = 3.457, p = 0.036$, Scheffé, $p = 0.043$). The difference between Group **No2** and Group **Ro2** was not statistically significant.

Table 2
Effect of recommendation

	2D world case	3D world case
Agent	◯	·
Robot	·	◯

3.2 3D world case

Figure 5(b) shows the mean selection ratios for each group in the 3D world case. Again, factorial analyses of variance (ANOVA) were conducted for each mean of selection ratios. We again compared the mean between no-recommendation, on-screen agent's recommendation and robot's recommendation. There were significant differences between Group **No3** and Group **Ro3** and between Group **Ag3** and Group **Ro3** (ANOVA, $F = 6.725, p = 0.002$, Scheffé, $p = 0.003, p = 0.042$).

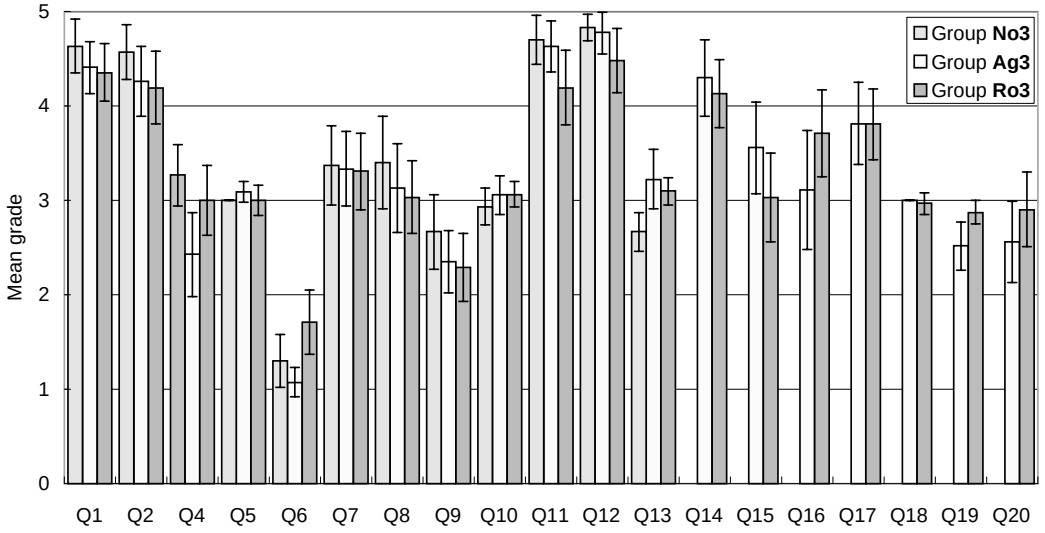
Table 2 summarizes the experiment results. The circles mean that the difference from the no-recommendation case is statistically significant.

3.3 Questionnaire

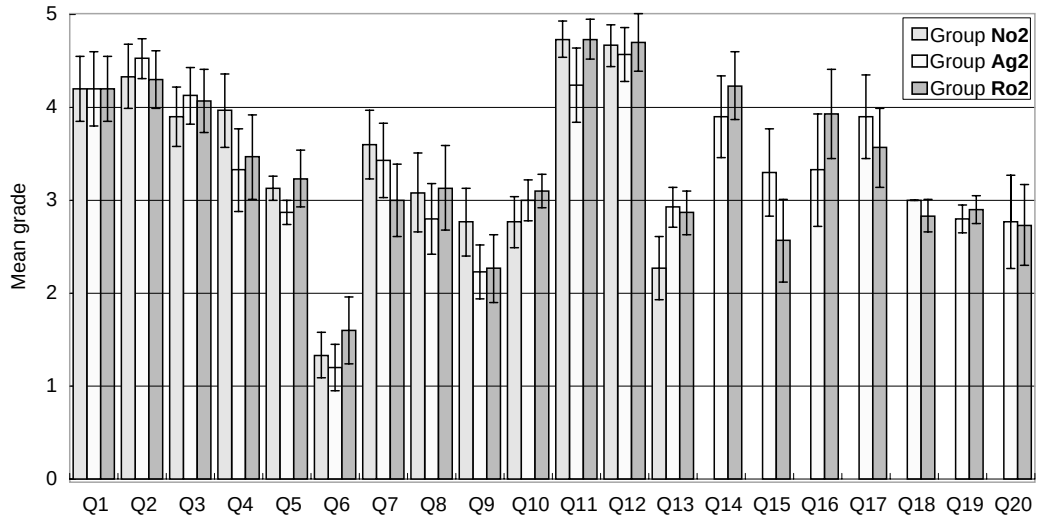
Lastly, we confirmed that the conditions were equal for all subjects by analyzing the questionnaire results. All questionnaire items are listed in appendix A. Each item was assigned a value from 5 to 1 according to the answers.

Figure 6(a) shows the mean values of the questionnaire in the 2D world case. Q13's value for Group **No2** is less than that for both Group **Ag2** and Group **Ro2** (ANOVA, $F = 7.045, p = 0.000$, Scheffé, $p = 0.002, p = 0.004$). The robot's and the on-screen agent's behavior made the experiment a little long. However, mean values are close to 3, so almost none of the subjects thought the experiment was long. Q11's values also have significant difference (ANOVA, $F = 4.149, p = 0.019$). And **Ag2**'s value is less than both **No2**'s and **Ro2**'s (Scheffé, $p = 0.049, p = 0.049$). However, all mean values are also much higher than 3, so none of subjects had difficulty understanding the method of questioning.

Figure 6(b) shows the mean values of the questionnaire in the 3D world case. Q4's values are significantly different (ANOVA, $F = 5.183, p = 0.008$). **Ag3**'s value is less than **No3**'s (Scheffé, $p = 0.009$). Q4's values mean that the subjects in **Ag3** thought that the system's response was a little slow. Even when comparing the mean selection ratios for subjects who gave a score higher than 2, the significant differences of the mean selection ratios between **Ro3**



(a) 2D world case



(b) 3D world case

Fig. 6. Mean questionnaire values

and the other groups still remain (ANOVA, $F = 10.932, p = 0.000$, [Scheffé, $p = 0.000$ (**No3** and **Ro3**), $p = 0.019$ (**Ag3** and **Ro3**)]). Q6's values are significantly different (ANOVA, $F = 5.558, p = 0.005$). **Ro3**'s value is greater than **Ag3**'s (Scheffé, $p = 0.007$). Q6's values mean that **Ro3** subjects did not feel that the experiment room was quieter than the other groups did. However, the mean values are much less than 3, indicating that the subjects did not feel that it was noisy either. Q13's values are significantly different (ANOVA, $F = 6.996, p = 0.002$). **No3**'s value is less than the other groups' (Scheffé, $p = 0.003, p = 0.022$). The reason is the same as that for the 2D world case

and seems to be no problem. Q19's values are significantly different (t-test, $t = 2.66, p = 0.01$). So, **Ag3**'s value is smaller than **Ro3**'s value. Q19's values mean that the subjects felt that the speed of the on-screen agent's speech was a little slow. However, even when comparing the mean selection ratios in the subjects who gave a score higher than 1, the significant differences of the mean selection ratios between **Ro3** and **No3** still remain (ANOVA $F = 6.834, p = 0.002$, Scheffé, $p = 0.002$ (**No3** and **Ro3**)).

In both world cases, there are positive correlations between Q20's values (credibility) and the selection ratios ($0.265, p = 0.013$), while the familiarity value (Q16) has no statistically significant correlation to the selection ratios ($p = 0.348$). This means that the subjects' choice mainly changed because of the ESA's credibility rather than its familiarity.

4 Discussion

The results show that the dimensionality of the ESA causes differences in the recommendation's effect on user decision-making. The 3D body was not always superior to the 2D body for recommendation, and on-screen agents seem to have weak points, too.

Those differences cannot be explained only by the advantages or disadvantage of pointing. In the 2D world condition, the color region was presented on the computer display. The display was in 3D space, so the color region was presented in the 3D space. From this point of view, there should be no difference between the 2D and 3D world conditions and therefore no difference in the effect. There must be some other reasons.

The results changed according to the combination of the location pointed to and the ESA's dimensionality. This seems to be evidence supporting the importance of consistency in the dimensionality between communication partners and the environment consisting of the pointing location and manipulated objects. In the 2D world condition, the color region was displayed on the CRT and color name options that should be selected were also on the CRT. In the 2D world and on-screen agent case, all were contained in the frame of the computer display. The frame might have emphasized the appropriate consistency and the on-screen agent might have had a strong influence through its recommendation. In the 3D world condition, the communication environment was in the 3D physical space. In the 3D world and robot case, the color region and color name options were contained in the physical 3D space, although they were physically separated and there was no frame.

In addition, the effect of robot's recommendation was much greater than that

of the on-screen agent's. The robot's body might have had a strong influence for emphasizing appropriate consistency to 3D space without a visual frame. In addition, behavior in communication contains ambiguous meanings and depends on the situation and communication environment (Sperber and Wilson, 1993). So, humans tend to quickly recognize an environment where the communication partner exists for interpretation.

The results of our experiment seem to provide evidence that humans tend to quickly recognize communication environments even in interaction with an ESA, and also suggest that we must not forget the communication environment in designing ESA behavior.

Another issue is the effect of the ESA's tracking function. We have found that geometrical consistency between an ESA and communication environment plays an important role in the recommendation situation and also confirmed that pointing with robot's 3D body worked well in 3D space. Humans have a 3D shape. This suggests that a robot's pointing to humans works well in 3D space and has much influence on user decision-making. We will investigate this further using the color name selection task and a robot.

5 Conclusion

We experimentally confirmed through quantitative evaluation that the degree of recommendation effect firmly depends on the interaction environment. The results show that a three-dimensional body has some advantage when the interaction environment is a three-dimensional space, but has less advantage in two-dimensional space than a two-dimensional body does. This suggest that geometrical consistency between an ESA and the interaction environment plays an important role in communication.

Acknowledgements

The authors would like to thank Kevin Wise and Byron Reeves for helpful comments on this paper.

Appendix

A Questionnaire

(1) In the 2D world case.
  Q1 The computer display was easy (5) or difficult (1) to see.
  Q2 The letters of the display were easy (5) or difficult (1) to see.
  Q3 The brightness of the display was high (5) or dull (1).
(2) In the 3D world case.
  Q1 The color plate was easy (5) or difficult (1) to see.
  Q2 The text on the button box was easy (5) or difficult (1) to read.
(3) In both cases.
  Q4 The system's response was fast (5) or slow (1).
  Q5 The experiment room was hot (5) or cold (1).
  Q6 The experiment room was noisy (5) or quiet (1).
  Q7 You were sure (5) or unsure (1) of your answers.
  Q8 There were many familiar (5) or unfamiliar (1) colors.
  Q9 There were many familiar (5) or unfamiliar (1) names.
 Q10 There were too many (5) or too few (1) questions.
 Q11 The method of questioning was easy (5) or difficult (1) to understand.
 Q12 The explanation about this experiment was easy (5) or difficult (1) to understand.
 Q13 The time for this experiment was too long (5) or too short (1).
(4) Concerning the agent.
 Q14 The agent's phrases were easy (5) or difficult (1) to understand.
 Q15 The agent's motions were natural (5) or unnatural (1).
 Q16 You felt familiarity (5) or unfamiliarity (1) with the agent.
 Q17 The agent's voice was clear (5) or unclear (1).
 Q18 The agent's voice was too loud (5) or too quiet (1).
 Q19 The speed of the agent's voice was too fast (5) or too slow (1).
 Q20 The agent was believable (5) or unbelievable (1).

In the no-recommendation case, subjects answered items 1 to 13. In the recommendation cases, they answered items 1 to 20. In the robot recommendation case, the word "agent" in item 13 to 20 was changed to "robot". The subjects evaluated these items using five ranks. Each number in the items shows the corresponding rank.

References

Ball, G., Breese, J., 1998. Emotion and personality in a conversation character. In: Proceedings of the Workshop on Embodied Conversation Characters.

Becheiraz, P., Thalmann, D., 1998. A behavioral animation system for autonomous actors personified by emotions. In: Proceedings of the 1998 Workshop on Embodied Conversational Characters.

Beskow, J., McGlashan, S., 1997. Olga - conversational agent with gestures. In: Proceedings of the IJCAI-97 Workshop on Animated Interface Agents: Making them Intelligent.

Cassell, J., Thórisson, K. R., 1999. The power of a nod and a glance: Envelope vs. emotional feedback in animated conversational agents. In: Applied Artificial Intelligence. Vol. 13. pp. 519–538.

Massaro, D., 1998. Perceiving talking faces: From speech perception to a behavioral principle. MIT Press.

Mizuno, O., Nakajima, S., 1998. Synthetic speech/sound control language: Mscl. In: 3rd ESCA/COCOSDA Proceedings of International Workshop on Speech Synthesis. pp. 21–26.

Nass, C., Gong, L., 1999. Maximized modality or constrained consistency? In: Proceedings of the AVSP '99 Conference.

Nass, C., Isbister, K., 1998. 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.

Ono, T., Imai, M., Ishiguro, H., 2001. A model of embodied communications with gestures between humans and robots. In: Proceedings of Twenty-third Annual Meeting of the Cognitive Science Society. pp. 732–737.

Reeves, B., Nass, C., 1996. The Media Equation. Cambridge University Press.

Shibata, T., Tanie, K., 2001. Physical and affective interaction between human and mental commit robot. In: Proceedings of IEEE International Conference on Robotics and Automation. pp. 2572–2577.

Shinozawa, K., Yamato, J., Naya, F., Kogure, K., 2001. Quantitative evaluation of effect of embodied conversational agents on user decision. In: Proceedings of HCI International 2001. pp. 998–1002.

Sperber, D., Wilson, D., 1993. RELEVANCE: Communication and cognition. Oxford: blackwell.

Takeuchi, Y., Katagiri, Y., Nass, C., Fogg, B., 1998. Social response and cultural dependency in human-computer interaction. In: Proceedings of PRICAI. pp. 114–123.