INTL. JOURNAL OF HUMAN–COMPUTER INTERACTION, *27*(3), 260–279, 2011

ISSN: 1044-7318 print / 1532-7590 online
DOI: 10.1080/10447318.2011.537209

How Does the Agents' Appearance Affect Users' Interpretation of the Agents' Attitudes: Experimental Investigation on Expressing the Same Artificial Sounds From Agents With Different Appearances

Takanori Komatsu[1] and Seiji Yamada[2]
[1]Shinshu University, Nagano, Japan
[2]National Institute of Informatics/SOKENDAI, Tokyo, Japan

An experimental investigation into how the appearance of an agent such as a robot or PC affects people's interpretations of the agent's attitudes is presented. In general, people are said to create stereotypical agent behavioral models in their minds based on the agents' appearances, and these appearances significantly affect their way of interaction. Therefore, it is quite important to address with the following research question: How does an agent's appearance affect its interactions with people? Specifically, a preliminary experiment was conducted to select eight artificial sounds for which people can estimate two specific primitive attitudes (e.g., positive or negative). Then an experiment was conducted where the participants were presented with the selected artificial sounds through three kinds of agents: a MindStorms robot, AIBO robot, and laptop PC. In particular, the participants were asked to select the correct attitudes based on the sounds expressed by these three agents. The results showed that the participants had better interpretation rates when a PC presented the sounds and lower rates when the MindStorms and AIBO robots presented the sounds, even though the sounds expressed by these agents were the same. The results of this study contribute to the design policy of the interactive agents, such as, What types of appearances should agents have to effectively interact with people, and which kinds of information should these agents express to people?

1. INTRODUCTION

Various artificial agents, such as interactive robots (Imai, Hiraki, Miyasato, Nakatsu, & Anzai, 2003; Watanabe, Okubo, Nakashige, & Danbara, 2004) or virtual characters (Cassell et al., 2002; Prendinger & Ishizuka, 2004), have been developed to assist us with our daily tasks. In this article, "agent" means the artifacts that

Correspondence should be addressed to Takanori Komatsu, International Young Researcher Empowerment Center, Shinshu University, 3-15-1 Tokida, Ueda, Nagano 386-8567, Japan. E-mail: tkomat@shinshu-u.ac.jp

are especially developed for interacting with humans to assist in their daily lives. One of the hottest research topics in the fields of human–computer interaction and human–agent interaction is how an agent's appearance affects its interactions with people. People are said to create stereotypical agent behavioral models in their minds based on the agents' appearances, and these appearances significantly affect their way of interaction (Yamada & Yamaguchi, 2004). For example, when people encounter a doglike robot, they expect doglike behaviors from it and would naturally speak to it using verbal commands and other utterances for dogs, such as "sit," "lie down," or "fetch." However, they do not act this way toward catlike robots.

Several studies have already focused on the effects of an agent's appearance on its interactions with people (Kiesler & Sproull, 1997). For example, Kiesler, Sproull, and Waters (1995) conducted a psychological experiment where the participants were asked to play a prisoner's dilemma game with virtual characters that had a humanlike or doglike appearance. The results showed that the participants who had some experiences with dogs showed cooperative behaviors with the doglike virtual agents, whereas the participants without these experiences did not show such cooperative behaviors. Goetz, Kiesler, and Powers (2003) investigated the effects of the heads of humanoid robots on people's impressions in questionnaire-based approaches. The results showed that the participants thought that robots with humanlike heads are good at social tasks, whereas robots with machinelike heads are good at industrial tasks. These were pioneering studies concerning the effects of an agent's appearance on people's impressions.

On the basis of the findings of these studies, we focused on the issue of how people interpret the information expressed from agents with different appearances. Simply stated, we focused on the relationship between the agents' appearances and their expressed information. As just mentioned, when people encounter certain agents, they determine these agents' behavioral models and start interacting with them according to these predetermined models. Therefore, this study constitutes a significant step toward clarification of the recent hot topic in human–computer interaction and human–agent interaction, that is, how does an agent's appearance affect its interactions with people? Moreover, the results of this study contribute to the design policy of the interactive agents, such as, What types of appearances should agents have in order to effectively interact with people, and which kinds of information should these agents express to people?

In particular, we conducted a preliminary experiment to determine what kinds of information are effective in evoking certain interpretations. We then conducted an experiment where participants were presented with this information from agents with different appearances in order to observe how these participants interpreted the presented information and to investigate the effects of the agents' appearances on the people's interpretations.

2. RELATED STUDIES

Some approaches have focused on the effects of an agent's appearance in order to design the interactive agents required to create an intimate relationship with the users. For example, Matsumoto, Fujii, Goan, and Okada (2005) proposed a "minimal design policy" for designing interactive agents, that is, agents should have a minimalist appearance to avoid user overestimation or underestimation of the agents' behavioral models. In fact, they applied this minimal design policy for developing their interactive robot "Muu" (Okada, Sakamoto, & Suzuki, 2000) and the lifelike agent "Talking Eye" (Suzuki, Takeuchi, Ishii, & Okada, 2000). Although their study focused on the effects of agents' appearances on people's impressions, which is the same focus as Kiesler's studies previously mentioned, it did not mention the relationship between the agents' appearances and their expressed information. This study instead proposed a rather abstract design strategy for interactive agents.

Reeves and Nass (1996) showed in their "media equation" studies that anthropomorphized agents or computers might induce humans' natural behaviors, like their behavior toward actual humans. These studies focused on the characteristics by which people generally anthropomorphize artifacts, even though the artifacts do not have humanlike appearances; they did not focus on the relationship between the appearance and expressed information.

Kanda, Miyashita, Osada, Haikawa, and Ishiguro (2005) conducted psychological experiments to observe people's behaviors toward two different types of humanoid robots, ASIMO (Honda Motors Co., Ltd., Tokyo, Japan) and Robovie (ATR-ITC, Kyoto, Japan). The results of their study showed that the different appearances of these two robots actually affected the people's nonverbal expressions, such as gestures or other body movements. However, this study also did not mention the relationship between the appearance and the expressed information.

3. AGENTS' ATTITUDES, APPEARANCES, AND EXPRESSED INFORMATION

In this article, we focus on the agents' attitudes based on the most primitive information that the agents should express to the users in order to investigate the effects of the agents' appearances on the users' interpretations of the expressed information. In particular, we selected the "positive" and "negative" attitudes corresponding to valence values (Reeves & Nass, 1996) as the primitive attitudes that the agents should express. Informing people of these two primitive attitudes is quite important if the agents are to effectively interact with people (Kobayashi & Yamada, 2005).

Next, simple and intuitive information that did not include any verbal information was used to determine certain primitive agent attitudes that users may recognize. These simple but intuitive expressions are called subtle expressions (Hayashi, 1999; Liu & Picard, 2003; Suzuki, Kakehi, Takeuchi, & Okada, 2004; Takeuchi & Katagiri, 2001; Ward, 2003). Komatsu (2005, 2006) showed that people can estimate different primitive attitudes by means of simple beep-like sounds and simple animations with different durations and inflections or velocity. In particular, when the presented information, regardless of its type, drastically deviates from the natural energy flow (e.g., a natural damping phenomenon or momentum conservation law), this information is interpreted as "active or objection," whereas information accorded with the natural energy flow is interpreted as "passive or compliance." We also speculated that agents with lifelike appearances expressing true-to-life information are sometimes more confusing to users and are not effective enough for interacting with people; there are many cases where agents without such lifelike appearances expressing subtle expressions are readily understood and are much more effective (Komatsu & Yamada, 2007). This is the reason we selected subtle expressions for this experiment. There is also another advantage of using this type of information: the technical ease of implementation into agents.

Finally, we selected three artifacts to be used as the agents with different appearances: a MindStorms robot (The LEGO group), an AIBO robot (Sony Corporation; ERS-7, Tokyo, Japan), and a laptop PC (Panasonic Inc, Osaka, Japan; Let's note W2). These artifacts correspond to making mechanical, familial, and nonagent-like impressions on people, respectively.

As a concrete procedure for this study, we first investigated the specific artificial sounds to be used as subtle expressions that could convey positive or negative attitudes (preliminary experiment) to people. We then conducted an experiment where we presented the artificial sounds selected in the preliminary experiment through the three different agents to investigate how people interpret the presented sounds. Finally, we summed up the results and discussed the effects of the agents' appearances on people's interpretations of the agents' expressed information.

4. PRELIMINARY EXPERIMENT

4.1. Overview

We conducted a preliminary experiment to determine what kinds of artificial sounds for use as subtle expressions are effective at conveying certain attitudes, that is, positive or negative, to people. We then investigated what kinds of sounds were interpreted as being positive or negative attitudes.

We prepared 44 different types of triangular wave sounds with four different durations and 11 different fundamental frequency (F0) contours. The four durations were 189, 418, 639, and 868 ms. The 11 contours were set so that the transition ranges of the F0 values between the onset and endpoint in the sound stimuli were −125, −100, −75, −50, −25, 0, +25, +50, +75, +100, and +125 Hz; these were linearly downward or upward (Figure 1). In this figure, the duration of the sound stimuli was 189 ms. The "u25" sounds are represented by 189u25 in this article, which indicates that the duration was 189 ms and the F0 transition rage was 25 Hz with an upward slope (increasing intonation). These 44 stimuli had the same F0 average of 131 Hz and had the same sound pressure around the participants' heads: 60 dB (FAST, A). In addition, these sounds had a tone like a computer's beep. They were presented to the participants from experimental GUI software implemented in a laptop PC (Panasonic; Lets' note W2). These artificial sounds were also used as experimental stimuli in Komatsu's (2005) study.

4.2. Participants

Ten Japanese university students (six men, four women; 19–23 years old) participated in this study. A simple hearing test was conducted to investigate whether the

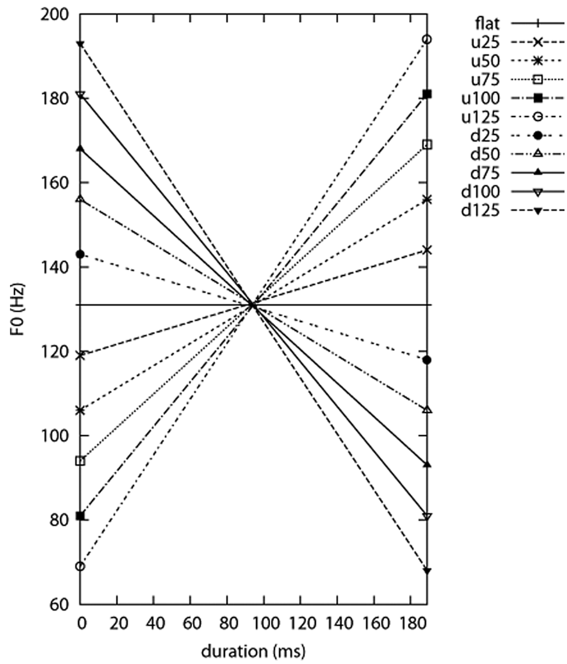

FIGURE 1 Eleven different F0 contours (duration: 189 ms; e.g., "u25" indicates that F0 transition range was 25 Hz with an upward slope [increasing intonation]).

participants had normal hearing. Specifically, the participants were asked whether they could hear three kinds of artificial sounds: flat 500-, 1000-, and 4000-Hz triangle wave sounds of 1.0-s duration (about 50 dB (FAST, A) at their head level). The test established that none of them had any hearing problems.

4.3. Procedure

First, an experimenter gave instructions to the participants, such as, "Please determine the attitudes of this PC based on the presented sounds." The participants were then asked to select one of three attitudes, "positive," "negative," and "undistinguishable," after hearing one of the 44 prepared sounds. These three attitudes were described to the participants as follows.

- Positive: The PC's internal state appears to be good. "Positive" includes attitudes like "active or objection."
- Negative: The PC's internal state appears to be bad. "Negative" includes attitudes like "passive or compliance."
- Undistinguishable: It is unclear whether the PC's internal state is positive or negative.

In the procedure, one randomly selected sound from the 44 prepared ones was presented by the PC to the participants. Afterward, the participants were asked

	Participants									
stimuli	A	B	C	D	E	F	G	H	I	J
189d025	−	−	*	−	−	*	−	−	−	−
189u025	+	*	−	*	−	*	−	+	−	+
189d050	−	+	*	−	−	−	−	−	*	−
189u050	*	*	+	−	+	+	+	+	−	+
189d075	−	*	−	−	−	+	−	−	−	−
189u075	+	+	+	+	−	+	+	+	−	+
189d100	+	*	−	−	−	−	−	*	*	−
189u100	+	+	+	+	−	−	+	+	*	+
189d125	−	*	+	−	−	−	−	−	−	−
189u125	+	+	+	+	+	+	+	+	+	+
189du0	*	−	−	−	−	*	−	−	−	−
418d025	−	−	−	−	−	−	−	−	−	−
418u025	*	*	*	+	−	*	*	+	−	−
418d050	−	−	−	−	−	−	−	−	−	−
418u050	+	+	+	+	−	+	*	*	+	+
418d075	−	−	−	−	−	−	−	−	−	−
418u075	+	+	+	−	−	+	+	+	−	*
418d100	+	−	−	−	−	−	−	−	−	−
418u100	+	+	+	+	+	+	−	+	+	+
418d125	−	−	−	−	−	−	−	−	−	−
418u125	+	+	+	+	+	+	+	+	+	+
418du0	*	−	−	−	−	*	−	−	−	−
639d025	−	*	−	−	−	−	−	−	−	−
639u025	*	*	+	+	+	*	+	+	+	−
639d050	−	−	−	−	−	−	−	−	−	−
639u050	+	*	+	−	−	+	+	+	+	+
639d075	−	−	−	−	−	−	−	−	−	+
639u075	+	+	+	+	+	+	+	+	+	*
639d100	−	+	−	−	−	−	−	−	+	−
639u100	+	+	+	+	+	+	+	+	+	+
639d125	−	*	−	−	−	−	−	−	+	−
639u125	+	+	+	+	+	+	+	+	+	+
639du0	*	−	−	−	−	*	*	−	−	−
868d025	−	*	−	−	−	*	−	−	*	−
868u025	−	*	−	*	+	*	*	*	+	−
868d050	+	*	−	−	−	*	*	−	−	−
868u050	*	+	*	+	+	+	+	+	+	+
868d075	*	−	−	−	−	−	*	−	*	−
868u075	+	*	+	*	+	+	+	+	+	+
868d100	−	*	*	−	−	−	*	−	+	−
868u100	+	+	*	+	−	+	+	+	+	−
868d125	−	−	−	*	−	−	−	−	−	−
868u125	+	+	+	+	+	+	+	+	+	+
868du0	*	−	−	−	−	−	*	−	−	−

FIGURE 2 Preliminary experiment results.

to select one of the three aforementioned attitudes that they thought appeared in the software. Each participant heard all 44 prepared sounds. The orders of sound presentation were randomly assigned for all 10 participants (see Figure 2).

4.4. Results

The results of this preliminary experiment are depicted in Figure 2; it shows which sounds were interpreted as "positive (+)," "negative (−)," and "undistinguishable (*)." The results showed that all 10 participants believed the PC had a positive attitude for five of the sounds: 189 ms with an upward slope range of 125 Hz (189u125), 418u125, 639u100, 639u125, and 868u125. In addition, they all believed the PC had negative attitudes for another five of the sounds: 418 ms with a downward slope range of 25 Hz (418d25), 418d50, 418d75, 418d125, and 639d50. Thus, the sounds with faster increasing intonation regardless of their durations were interpreted as conveying positive attitudes, whereas the sounds that were around 500 ms with slower decreasing intonation were interpreted as conveying negative attitudes. These results are in accordance with those observed in Komatsu's (2005) study, mentioned previously.

Here, of the five sounds interpreted as positive, we eliminated the sound labeled 639u100 because it had the slowest slope range. In addition, of the five sounds interpreted as negative, we eliminated the 418d125 sound because it had the fastest slope range. The remaining eight sounds were then selected for expression by the agents with different appearances in the next experiment.

5. EXPERIMENT

5.1. Overview

The purpose of this experiment was to investigate the effects of the agents' appearances on how the participants interpreted the agents' attitudes. In particular, the participants were presented the eight sounds selected in the preliminary experiment through agents with different appearances, that is, the MindStorms robot, AIBO robot, and a laptop PC (Figure 3). The participants were then asked to determine the attitude (i.e., positive, negative, or undistinguishable) based on the expressed sounds from these agents.

FIGURE 3 Photograph of AIBO robot, MindStorms robot, and laptop PC (from left to right).

5.2. Participants

Twenty Japanese university students (17 men, three women; 19–24 years old) participated in this experiment. They were not familiar with the robots or those kinds of toys and had not participated in the preliminary experiment. A simple hearing test established that none of them had hearing problems. In this hearing test, the participants were asked whether they could hear three kinds of artificial sounds: flat 500-, 1000-, and 4000-Hz triangle wave sounds of 1.0-s duration (about 50 dB (FAST, A)).

5.3. Procedure

First, an experimenter gave instructions to the participants, such as, "This is monitoring research to evaluate the three agents by means of a questionnaire." The participants were then told that the concrete task was to select one of the three attitudes after the agent expressed certain information. All the participants experienced the following three experimental conditions (Figure 4).

1. Eight sounds expressed by the MindStorms (MindStorms-condition) robot: The eight sounds came from an FM radio tuner mounted on the robot. This radio tuner received the transmitted sounds from a sound-expressing PC beside the experimenter.

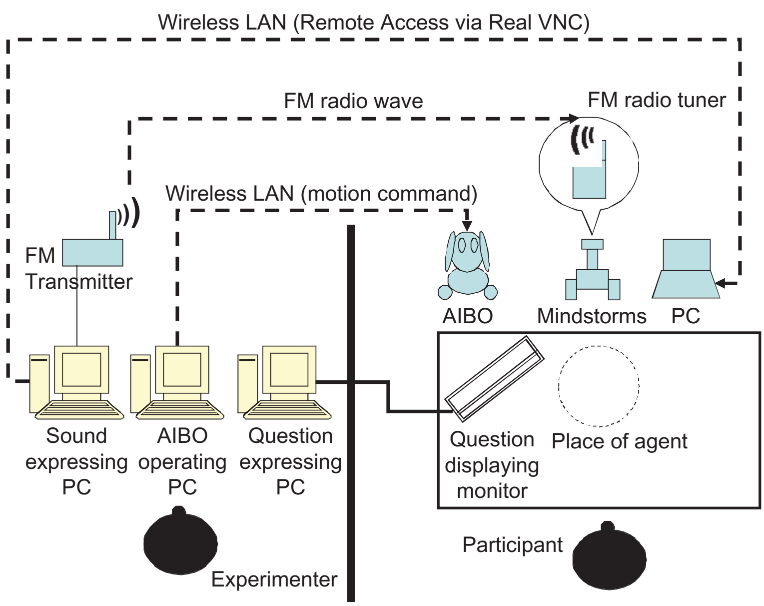

FIGURE 4 Experimental setup.

2. Eight sounds expressed by the AIBO (AIBO-condition) robot: The sounds were presented using AIBO operating software, "AIBO entertainment player," that was installed in an AIBO-operating PC. This software offers remote control of the AIBO.
3. Eight sounds expressed by the laptop PC (PC-condition): Nearly the same conditions as in the preliminary experiment were used. This laptop PC was remotely operated by a sound-expressing PC via Real VNC.

The same sound pressure was set around the participants' heads: 60 dB (FAST, A) for each condition. The eight sounds were expressed from different types of speakers implemented in each agent for each condition, so the sounds expressed from the agents differed slightly among these three conditions. However, we placed importance on the expression of the sounds from the agents themselves, and we assumed that implementing the same speaker on these agents would affect the agents' appearances. In addition, Komatsu (2005, 2006) reported that the effects of the durations and transitions of F0 were the most critical for user interpretation of the presented sounds, so we assumed that the differences in the speakers used would not affect the participants' interpretations of the agents' attitudes.

All participants experienced these three conditions in a counterbalanced order; that is, there are six different orders for experiencing the three conditions, so all participants experienced these six orders in turn. Under all three conditions, the eight selected sounds were randomly presented to the participants.

5.4. Results

Whether Participants Correctly Interpreted Agents' Attitudes

We calculated the interpretation rates of the participants, which indicate how many times the participants succeeded in correctly determining the agents' attitudes under all three experimental conditions. The results showed that the participants had interpretation rates of 4.45 for the eight experimental stimuli in the MindStorms-condition, 4.40 in the AIBO-condition, and 6.65 in the PC-condition (Figure 5).

The results of a Friedman test on the interpretation rates between the three experimental conditions (independent variable: experimental conditions; levels: MindStorms, AIBO, and PC) showed significant differences in these experimental conditions, $\chi^2(2) = 16.028$, $p = .00 < .01^{**}$, and a multiple comparison using pairwise comparison of the three conditions showed significant differences between the PC- and AIBO-conditions ($p = .017$) and between the PC- and MindStorms-conditions ($p = .001$). Thus, these results showed that the participants' interpretation rates for the same sound stimuli differed based on each agent's appearance. The participants showed higher interpretation rates for the PC-condition, whereas they had significantly lower rates for the MindStorms- and AIBO-conditions, even though the same sounds were presented in all three conditions.

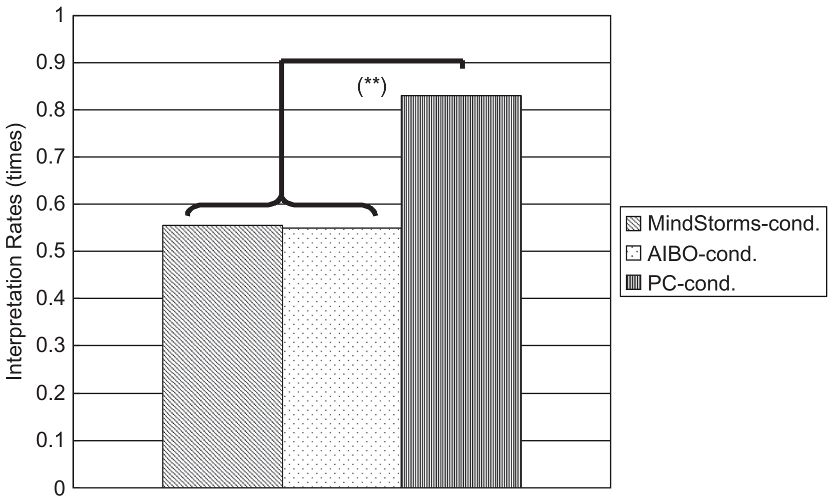

FIGURE 5 Participants' interpretation rates for three experimental conditions.

Relationship Between Types of Sounds and Interpretation Rates in Each Condition

Figures 6 to 8 depict the interpretation rates of the eight sound stimuli in the MindStorms-, AIBO-, and PC-conditions, respectively. We then investigated how

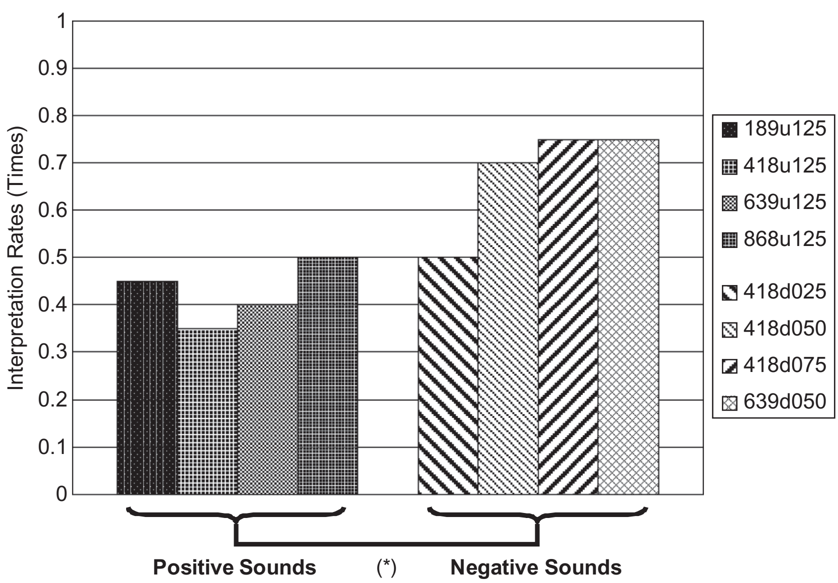

FIGURE 6 Interpretation rates of eight sounds for MindStorms-condition.

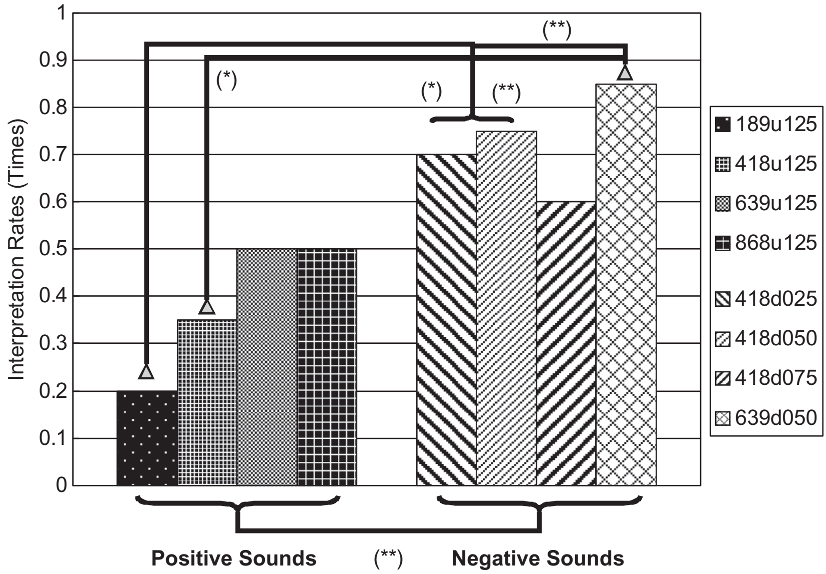

FIGURE 7 Interpretation rates of eight sounds for AIBO-condition.

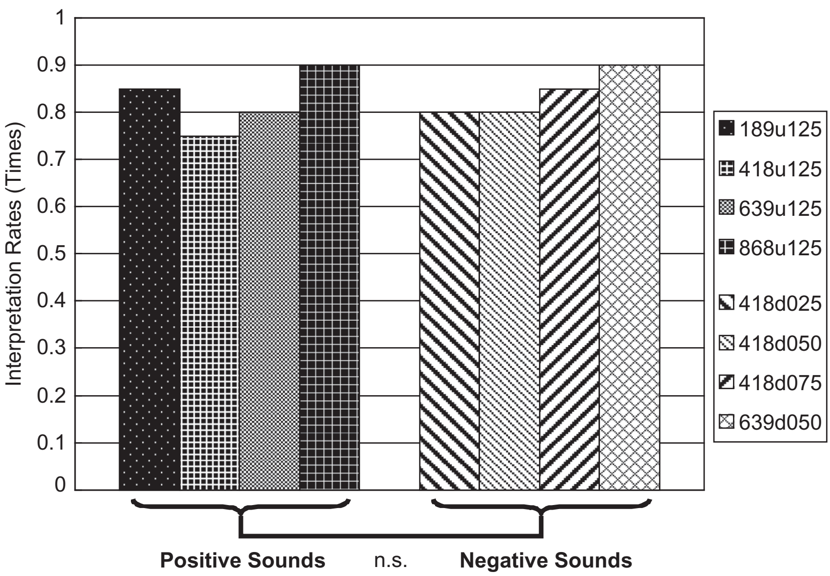

FIGURE 8 Interpretation rates of eight sounds for PC-condition.

the interpretation rates of the eight sounds were affected by the agents' appearances. Specifically, we first investigated the interpretation rates of the eight sounds within each experimental condition.

MindStorms-condition. The results of the Cochran's Q test on interpretation rates for the eight sounds in the MindStorms-condition (independent variable: sound stimuli; levels: eight different sounds) showed significant differences in the interpretation rate, $\chi^2(7) = 14.504$, $p = .043 < .05^*$. However, a multiple comparison using pairwise comparison of these sounds did not show significant differences.

AIBO-condition. The results of the Cochran's Q test for the AIBO-condition (independent variable: sound stimuli; levels: eight different sounds) showed significant differences in the interpretation rate, $\chi^2(7) = 28.526$, $p = .000 < .01^{**}$. A multiple comparison using pairwise comparison of these sounds showed significant differences in the following four sections: (a) between 189u125 and 418d025 ($p = .025$), (b) between 189u125 and 418d050 ($p = .007$), (c) between 189u125 and 639d050 ($p = .000$), and (d) between 418u125 and 639d050 ($p = .025$). That is, the negative sounds had higher interpretation rates than the positive ones in this condition.

PC-condition. The results of the Cochran's Q test did not show any significant differences in the interpretation rate, $\chi^2(7) = 4.282$, $p = .747$, *ns.*

Next, we investigated the interpretation rate for each of the eight sounds among the three experimental conditions.

189u125. The results of the Cochran's Q test (independent variable: experimental conditions; levels: MindStorms, AIBO, and PC) showed significant differences among the three conditions, $\chi^2(2) = 18.429$, $p = .001 < .01^{**}$. A multiple comparison using pairwise comparison of the three conditions showed significant differences between the PC- and MindStorms-conditions ($p = .026$), and between the PC- and AIBO-conditions ($p = .000$).

418u125. The results of the Cochran's Q test showed significant differences among the three conditions, $\chi^2(2) = 6.737$, $p = .034 < .05^*$. A multiple comparison using pairwise comparison of the three conditions showed marginally significant differences between the PC- and MindStorms-conditions ($p = .074$), and between the PC- and AIBO-conditions ($p = .074$).

639u125. The results of the Cochran's Q test showed significant differences among the three conditions, $\chi^2(2) = 6.933$, $p = .031 < .05^*$. A multiple comparison using pairwise comparison of the three conditions showed significant differences between the PC- and MindStorms-conditions ($p = .034$).

868u125. The results of the Cochran's Q test showed significant differences among the three conditions, $\chi^2(2) = 10.667$, $p = .005 < .01^{**}$. A multiple comparison using pairwise comparison of the three conditions showed significant differences between the PC- and MindStorms-conditions ($p = .014$), and between the PC- and AIBO-conditions ($p = .014$).

418d025. The results of the Cochran's Q test showed no significant differences among the three conditions, $\chi^2(2) = 4.308$, $p = .116$, *ns*.

418d050. The results of the Cochran's Q test showed no significant differences among the three conditions, $\chi^2(2) = 0.857$, $p = .651$, *ns*.

418d075. The results of the Cochran's Q test showed no significant differences among the three conditions, $\chi^2(2) = 4.570$, $p = .103$, *ns*.

639d050. The results of the Cochran's Q test showed no significant differences among the three conditions, $\chi^2(2) = 1.750$, $p = .417$, *ns*.

From the results of this analysis, we observed that the PC-condition had higher interpretation rates for all eight sound stimuli, whereas the MindStorms- and AIBO-conditions showed higher interpretation rates for only the four negative sounds. Then, we analyzed the average interpretation rates of the four positive and four negative sounds in each condition. The participants had average interpretation rates of 0.43 for the four positive sounds and 0.68 for the four negative sounds in the MindStorms-condition, 0.39 for the positive sounds and 0.73 for the negative sounds in the AIBO-condition, and 0.83 for the positive and 0.84 for the negative sounds in the PC-condition. The results of the Friedman test on the interpretation rates for the positive and negative sounds in each experimental condition (independent variable: attitudes of the sounds; levels: positive and negative) showed that there was no significant difference in the interpretation rate between the positive and negative sounds in the PC-condition, $\chi^2(1) = 0.111$, $p = .739$, *ns*, whereas there were significant differences in the MindStorms-condition, $\chi^2(1) = 4.571$, $p = .033 < .05^*$, and AIBO-condition, $\chi^2(1) = 13.00$, $p = .00 < .01^{**}$. This result reveals that the negative sounds showed significantly higher interpretation rates than the positive ones in the MindStorms- and AIBO-conditions.

Relationship Between Positive/Negative Sounds and Agents' Appearances

We next investigated the effects of the agents' appearances on the interpretation rates of the four positive and four negative sounds. The results of the Friedman test on the interpretation rates for the four positive sounds among the three experimental conditions (independent variables: experimental conditions; levels: MindStorms, AIBO, and PC) showed significant differences in the interpretation rates, $\chi^2(2) = 17.818$, $p = .00 < .01^{**}$. A multiple comparison using pairwise comparison of the three experimental conditions showed significant differences between the MindStorms- and PC-conditions ($p = .003$), and between the AIBO- and PC-conditions ($p = .003$). For the four negative sounds, the results of the Friedman test on the interpretation rates for these negative sounds among the three experimental conditions showed no significant differences among them, $\chi^2(2) = 4.481$, $p = .106$, *ns*.

From these results, the interpretation rates of the positive sounds were significantly higher in the PC-condition than in the MindStorms- and AIBO-conditions, and the interpretation rates of the negative sounds showed no differences among the three experimental conditions.

Summary

To summarize, we obtained the following results from our experiment.

- For the interpretation rates of all eight sounds, the PC-condition had significantly higher interpretation rates than the MindStorms- and AIBO-conditions.
- In the MindStorms- and AIBO-conditions, the interpretation rates for the four negative sounds were significantly higher than those for the four positive sounds.
- For the interpretation rates of the four positive sounds, the PC-condition had significantly higher interpretation rates than the MindStorms- and AIBO-conditions. In addition, for the interpretation rates of the four negative sounds, there were no significant differences among these three experimental conditions.

6. DISCUSSION

The results of our experiment in which the eight artificial sounds selected in the preliminary experiment were presented through agents having different appearances can be summarized in terms of each experimental condition.

- MindStorms-condition: The average interpretation rate of the eight sounds selected in the preliminary experiment was 0.56, and this was significantly lower than that in the PC-condition. Among these eight sounds, the interpretation rates of the four positive sounds were significantly lower than the ones for the four negative ones.
- AIBO-condition: The average interpretation rate of the eight sounds was 0.55, and this was significantly lower than that in the PC-condition. Among these eight sounds, the interpretation rates of the four positive sounds were significantly lower than the ones for the four negative ones.
- PC-condition: The average interpretation rate of the eight sounds was 0.83, and this was significantly higher than those for the other conditions. In addition, there was no significant difference between the interpretation rates of the four positive and four negative sounds. The interpretation rates of the

positive sounds were significantly higher than those for the other conditions, and the ones for the negative sounds showed no differences between the three conditions.

These results clearly showed that the agents' appearances affected the participants' interpretations of the agents' primitive attitudes, even though these agents expressed sounds that were the same. In particular, in terms of the interpretation of the positive sounds, the PC-condition had significantly higher interpretation rates than the other two conditions, whereas in terms of the interpretation of the negative sounds, there were no significant differences between the three experimental conditions. Here, let us consider the results of the AIBO- and MindStorms-conditions and a comparison between the conditions as follows.

6.1. Results in the AIBO-Condition

When the eight artificial sounds were presented from the AIBO robot, the four positive sounds had lower interpretation rates than the four negative ones. In addition, the interpretation rates of the negative sounds did not have significant differences from the ones in the PC-condition. This means that these positive sounds were interpreted as negative, so all eight sounds were eventually interpreted as negative.

On the basis of the idea mentioned in the introduction that people are said to create the behavioral models of agents based on the agents' appearances, the participants facing the AIBO robot would expect doglike behaviors from it, for example, wagging the tail cheerfully means a positive attitude, whereas barking actively is a negative attitude. In this case, it was assumed that beep-like sounds expressed from the AIBO robot were completely unexpected, and these sounds would evoke the thought that "some errors have happened" with the robot, because beep-like sounds are generally utilized to signify some kind of warning from a computer to users. We assumed that this was the reason the participants interpreted all eight sounds as negative attitudes. These participants had higher interpretation rates for the negative sounds that were nearly the same level as the ones in the PC-condition.

6.2. Results in the MindStorms-Condition

When the eight artificial sounds were presented from the MindStorms robot, the four positive sounds had lower interpretation rates than the four negative ones. This means that most of the positive sounds were also interpreted as negative. Thus, at a glance, the results of the MindStorms-condition were similar to those of the AIBO-condition. However, the differences in interpretation rates between the positive and negative sounds were not very clear compared to those in the AIBO-condition (Figures 6 and 7).

When taking into account the participants' expected expressions from the MindStorms robot, it is assumed that most participants would expect machinelike behaviors from this robot, unlike those from the AIBO robot. In addition, the eight artificial sounds used in this study were similar to a computer's beep, so these sounds were not entirely unexpected for the participants; this was the reason that all eight sounds were not interpreted as negative attitudes, as in the AIBO-condition.

It is still unclear why the participants had lower interpretation rates in the MindStorms-condition than in the PC-condition, even though most of the participants expected machinelike expressions from the MindStorms robot and actual machinelike sounds were expressed to them. An interesting questionnaire-based investigation to consider explained that robots with machinelike appearances could evoke an impression of higher functions than ones with a familiar appearance (Komatsu & Nambu, 2008). Thus, it was assumed that most of the participants thought the MindStorms robot was an agent with higher functions compared to a laptop PC. In this case, the beep-like sounds did not really fit this higher functioned agent, and this lead to lower interpretation rates in this condition than in the PC-condition.

6.3. Differences Between MindStorms- and AIBO-conditions

From the results of the experiment, it would be expected that the participants showed higher interpretation rates in the PC-condition. However, they also showed lower interpretation rates in the MindStorms- and AIBO-conditions for the positive sounds in particular. The reason for this phenomenon could be based on the aforementioned argument that the MindStorms and AIBO's expression of artificial sounds was unexpected for the participants. Thus, it would be expected that these expressed sounds deviated from the participants' mental behavior models about MindStorms and AIBO and that these sounds were interpreted by the participants as error sounds, because these sounds were very similar to a computer's beeping.

There are obvious differences in the interpretation rates of the eight artificial sounds between the MindStorms- and AIBO-conditions (see Figures 6 and 7), that is, no significant differences were observed in pairs of the eight sounds within the MindStorms-condition (Figure 6), whereas several significant differences were observed within the AIBO-condition (e.g., between 189u125 and 418d025, 418d050, and 639d050, and between 418u125 and 639d050; see Figure 7). Therefore, it can be said that the participants clearly interpreted the eight sounds from AIBO as negative sounds, whereas they did not clearly interpret the sounds from MindStorms. We then assumed that the participants' different interpretations for the sounds from MindStorms and AIBO depended on how accurately the participants formed mental behavior models about MindStorms and AIBO. For example, because of the AIBO's doglike appearance and publicity about it, the participants would form very concrete mental models about AIBO in which sophisticated doglike sounds or behaviors are expected, so they could immediately and clearly judge that the AIBO's expressed artificial sounds deviated completely from such mental models. In contrast, because of the MindStorms's characterless mechanical appearance, the participants would form ambiguous mental models about MindStorms in which specific sounds or behaviors were not really expected, so they could not clearly judge whether the MindStorms's sounds deviated from such a model. From this assumption acquired from the comparison of the results of the MindStorms- and AIBO-conditions, we could determine that the agents' expressions should be designed based on the users' mental models.

We were able to determine the requirements for expressing primitive attitudes to users using the previous discussions about the three experimental conditions. In particular, we assumed that these agents should satisfy the following two requirements.

1. Selecting the appropriate expression information that agrees with the agents' appearances.
2. Assigning certain attributes that could satisfy Komatsu's (2005, 2006) arguments, that is, when the presented information drastically deviates from the natural energy flow, this information is interpreted as positive, whereas the information that agrees with the natural energy flow is interpreted as negative.

On the basis of the aforementioned assumptions, we are now planning a consecutive study to find which kinds of expressed information are appropriate for the AIBO and MindStorms robots to express their primitive attitudes to users. For the AIBO, it is expected that this robot should express doglike behaviors to users; the AIBO operating software, "AIBO entertainment player," in fact offers about 80 basic preset motions, like "cheer up" and "good morning." In our former study (Yamada & Komatsu, 2006), we chose six motions (e.g., the motion named "Happy 1" as a positive attitude and "Unhappy 1" as a negative attitude) from among these 80 motions. We then presented them to participants who were asked to select one of the three primitive attitudes as in this study. However, the participants showed much lower interpretation rates for these motions. We assumed that these preset motions were confusing for users trying to understand the AIBO's primitive attitudes, because the AIBO's motions were not originally designed to inform people of its primitive attitudes. This means that these preset AIBO motions actually satisfied the first requirement previously mentioned but not the second one. Therefore, we expect that the kinds of expressed information appropriate for the AIBO would be the doglike behaviors assigned with the attributes, whether or not these behaviors agree with the natural energy flow. For example, the barking motions of AIBO with changing durations and inflections would be candidates. For the MindStorms robot, the preparation of machinelike expressions is required to assist users in imagining much more sophisticated functions. For example, the blinking of several LEDs at changing intervals and velocities, like the front light of the KITT car in the television program *Knight Rider*, would be a candidate for such expressions.

In any case, this requires an objective methodology to define the appropriate relationship between the agents' appearances and their expressed information. In particular, we are now planning to conduct the following investigations.

1. In our preliminary experiment for this study, we selected eight sounds from among 44 expressed from a laptop PC. We are now planning other preliminary experiments to select more appropriate sounds from among the 44 sounds expressed from the MindStorms or AIBO robots to investigate which kinds of artificial sounds are selected as positive or negative attitudes according to the agents' appearance.
2. We are planning to conduct a questionnaire-based investigation to better comprehend the users' opinions on "which kinds of expressions should certain agents use to express their positive or negative attitudes." The results of this questionnaire will be analyzed in terms of user gender, age, and preferences.

To conduct these two experiments, we should clarify the appropriate relationship between the agents' appearances and their expressed information and should verify our assumptions to satisfy the two requirements just mentioned. This leads us to propose that an objective methodology is required for designing an interactive agent that can effectively present its primitive attitudes to create an intimate relationship with users.

7. CONCLUSION

We conducted an experiment to clarify the relationship between agents' appearances and their expressed information by presenting participants with artificial sounds that can make people estimate specific agents' attitudes using three kinds of agents. The participants' interpretation rates, indicating how many times they selected the correct attitudes, were then investigated. The results revealed that the agents' different appearances affected people's interpretations of the agents' attitudes, even though these agents expressed the same information. In particular, the interpretation rates when a PC expressed these sounds were significantly higher than those when a MindStorms or AIBO robot expressed them. Moreover, when these sounds were expressed from the MindStorms and AIBO robots, most were interpreted as negative attitudes. We assumed that the reason for these lower interpretation rates was that the expressed information was unexpected by the participants. In other words, the agents should express imaginable or predictable expressions to inform people of their attitudes (e.g., real doglike expressions from the AIBO robot, and expressions such as those by KITT's moving light in *Knight Rider* from the MindStorms robot). We are planning to investigate the most appropriate information based on the agents' appearances. We expect this follow-up study will contribute to establishing a design policy that can clarify the effective coupling between agents' appearances and their expressed information.

REFERENCES

Cassell, J., Stocky, T., Bickmore, T., Gao, Y., Nakano, Y., Ryokai, K., Tversky, D., Vaucelle, C., & Vilhjálmsson, H. (2002). MACK: Media lab Autonomous Conversational Kiosk. *Proceedings of Imagina02*. February 12–15, Monte Carlo.

Goetz, J., Kiesler, S.,& Powers, A. 2003. Matching robot appearance and behavior to tasks to improve human-robot cooperation. *Proceedings of the 12th IEEE International Workshops on Robot and Human Interactive Communication*, 55–60.

Hayashi, Y. (1999). Recognition of vocal expression of emotions in Japanese: Using the interjection "eh." *Proceedings of the 14th International Congress of Phonetic Sciences*, 2355–2358.

Imai, M., Hiraki, K., Miyasato, T., Nakatsu, R., & Anzai, Y. (2003). Interaction with robots: Physical constraints on the interpretation of demonstrative pronouns. *International Journal of Human–Computer Interaction, 16*, 367–384.

Kanda, T., Miyashita, T., Osada, T., Haikawa, Y., & Ishiguro, H. (2005). Analysis of humanoid appearances in human–robot interaction. *Proceedings of the 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS2005)*, 62–69.

Kiesler, S., & Sproull, L. (1997). Social responses to "social" computers. In B. Friedman (Ed.), *Human values and the design of computer technology*. (pp. 191–199). Stanford, CA: CLSI.

Kiesler, S., Sproull, L., & Waters, K. (1995). A prisoner's dilemma experiment on cooperation with people and human-like computers. *Journal of Personality and Social Psychology, 79*, 47–65.

Kobayashi, K., & Yamada, S. (2005). Informing a user of robot's mind by motion. *Proceedings of the 3rd International Conference on Computational Intelligence, Robotics and Autonomous Systems (CIRAS2005)* (CD-ROM).

Komatsu, T. (2005). Toward making humans empathize with artificial agents by means of subtle expressions. *Proceedings of the 1st International Conference on Affective Computing and Intelligent Interaction (ACII2005)*, 458–465.

Komatsu, T. (2006). Audio subtle expressions affecting users' perceptions. *Proceedings of 2006 International Conference on Intelligent User Interface (IUI2006)*, 306–308.

Komatsu, T., & Nambu, M. (2008). Effects of the agents' appearance on people's estimations about the agents' abilities: Questionnaire investigations for liberal arts and informatics students. *Proceedings of the 17th IEEE International Workshops on Robot and Human Interactive Communication*, 142–147.

Komatsu, T., & Yamada, S. (2007). How do robotic agents' appearances affect people's interpretation of the agents' attitudes? *ACM-CHI2007* [Extended abstract], 2519–2525.

Liu, K., & Picard., W. R. (2003). Subtle expressivity in a robotic computer. *Proceedings of the CHI2003 Workshop on Subtle Expressivity for Characters and Robot*, 5–9.

Matsumoto, N., Fujii, H., Goan, M., & Okada, M. (2005). Minimal design strategy for embodied communication agents. *Proceedings of the 14th IEEE International Workshops on Robot and Human Interactive Communication*, 335–340.

Okada, M., Sakamoto, S., & Suzuki, N. (2000). Muu: Artificial creatures as an embodied interface. *Proceedings of the 27th International Conference on Computer Graphics and Interactive Techniques (SIGGRAPH 2000)*, 91.

Prendinger, H., & Ishizuka, M. (Eds.). (2004). *Life-like characters*. New York, NY: Springer.

Reeves, B., & Nass, C. (1996). *The media equation*. Stanford, CA: CSLI.

Suzuki, N., Kakehi, K., Takeuchi, Y., & Okada, M. (2004). Social effects of the speed of hummed sounds on human-computer interaction. *International Journal of Human-Computer Studies, 60*, 455–468.

Suzuki, N., Takeuchi, Y., Ishii, K., & Okada, M. (2000). Talking eye: Autonomous creatures for augmented chatting. *Robotics and Autonomous Systems, 31*, 171–184.

Takeuchi, Y., & Katagiri, Y. (2001). Emotional feedback by synthesized voice and facial expression. *Proceedings of the 3rd International Conference on Cognitive Science*, 697–698.

Ward, N. (2003). On the expressive competencies needed for responsive systems. *Proceedings of the CHI2003 Workshop on Subtle Expressivity for Characters and Robot*, 33–34.

Watanabe, T., Okubo, M., Nakashige, M., & Danbara, R. (2004). InterActor: Speech-driven embodied interactive actor. *International Journal of Human-Computer Interaction, 17*, 43–60.
Yamada, S., & Komatsu, T. (2006). Designing simple and effective expressions of robot's primitive minds to a human. *Proceedings of 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2614–2619.
Yamada, S., & Yamaguchi T. (2004). Training AIBO like a dog. *Proceedings of the 13th IEEE International Workshops on Robot and Human Interactive Communication*, 431–436.