16th IEEE International Conference on Robot & Human Interactive Communication
August 26 ~ 29, 2007 / Jeju, Korea

Effects of robotic agents' appearances on users' interpretations of the agents' attitudes: towards an expansion of "uncanny valley" assumption

Takanori Komatsu[1], Member, IEEE, and Seiji Yamada[2], Member, IEEE

[1]Department of Media Architecture, Future University-Hakodate, Hokkaido, Japan, e-mail : komatsu@fun.ac.jp
[2] National Institute of Informatics, Tokyo, Japan, e-mail:seiji@nii.ac.jp

***Abstract*— This paper describes experimental investigation how agents' appearances affect users' interpretations of agents' attitudes. Specifically, we conducted a psychological experiment that participants were presented artificial sounds as subtle expressions that can make human estimate specific agents' primitive attitudes from three kinds of different agents, e.g., Mindstorms robot, AIBO robot, and normal laptop PC, and they were asked to select the correct attitudes based on the expressed sounds from these three agents. As the result, the participants showed the higher interpretation rates when the sounds were presented from PC, while they did the lower rates from Mindstorms and AIBO robots, even though the artificial sounds expressed from these agents were the completely same sounds. Finally, the result was compared to the Mori's assumption that is about the relationship between an agents' appearance (likeness) and users' familiarity to the agents.**

I. Introduction

Recently, one of the hottest topics in Human-Computer Interaction or Human-Agent Interaction studies is "which kinds of appearances should the agents have in order to interact with users effectively." Because, it can be generally said that users establish the agents' behavior model in terms of their appearances so that it is expected that the agents' appearances significantly affect the interaction with these users [1]. For example, when users face the dog-like robot, they expect the dog-like behaviors from this robot and they would speak to this, like "sit, sit up, lie down" and so on. On the other hand, they did not do this way toward the cat-like robot.

Actually there are several studies that focus on the effects of the agents' appearances on the interactions with users [2,3,5]. Keisler [2] conducted a psychological experiment that the participants were asked to play the prisoner's dilemma game with virtual characters (human and dog) appearing on the computer display. The results showed that the participants who had some experiences in keeping dogs cooperated with the dog-like virtual agent significantly. Goetz et al [5] investigated the effects of the different design of head of humanoid robots on users' impressions. The results showed that the participants answered that the human-like head robots are good at achieving social tasks, while they did that the machine-like head ones are good at achieving industrial tasks.

Based on the findings of the above former studies, we are currently focusing on the following issue; "which kinds of appearance's agents should express which kinds of information to users?" As the first step to realize this issue completely, we investigated how human users interpret the presented sounds that can make users estimate specific attitudes from the agents having different appearances. We would then discuss about the relationship between our results and Mori's assumption that is about the one between the agents' appearance (likeness) and the familiarities human users would feel.

II. Agents' Attitudes, Appearance and Expressed Information

There would be several approaches to tackle with the above issue. In this study, we investigated the effects of the basic relationship between the agents' appearances and the expressed information on users' interpretation of agents' attitudes by means of a psychological experiment.

As the primitive attitudes that agent should express to users, we selected the positive and negative attitudes corresponded to the valence values. It is assumed that informing these two values to users are quite important for the agents required in interacting and cooperating with users. And as the agents having different appearances, we selected the following three artifacts; Mindstorms robot [7], AIBO robot [8], and a normal laptop PC (Let's note W2, product of Panasonic Inc,). These artifacts were corresponded to making users mechanical impressions, familial impressions, and no-agent-like impressions, respectively.

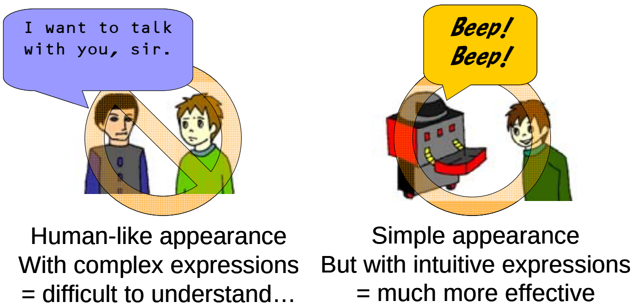

Fig.1 Conceptual diagram of our hypothesis

Concerning the effects of the relationship between agents' appearances and expressed information on users' interpretation of the agents' attitudes, we hypothesized that *the agents with rich appearance (quite similar with actual humans or pet animals) expressing the likely information (verbal information or animal like behaviors) are more confusing to users and are not really effective for interacting with users: Instead, the agents without rich appearance expressing the simple but intuitive information (e.g., subtle expressions* [6]*) are readily understood and are much more*

978-1-4244-1635-6/07/$25.00 ©2007 IEEE.

effective for their interaction (see Figure 1) [9]. If our hypothesis is verified, it would become possible to make users understand the agents' primitive attitudes without developing the dexterous and complex robotic or computer graphic systems consuming a bunch of costs. And then, eventually, our hypothesis would contribute to easily realizing the agents that can establish the intimate and natural interaction with users.

III. Preliminary Experiment

Before the actual experiment, we conducted the preliminary experiment to confirm which kinds of subtle expressions are effective to evoke certain attitudes, positive or negative, to humans. In this study, we focused on the artificial sounds as subtle expressions that were utilized in the former studies [4]. We then investigated which kinds of sounds were interpreted as positive or negative attitudes.

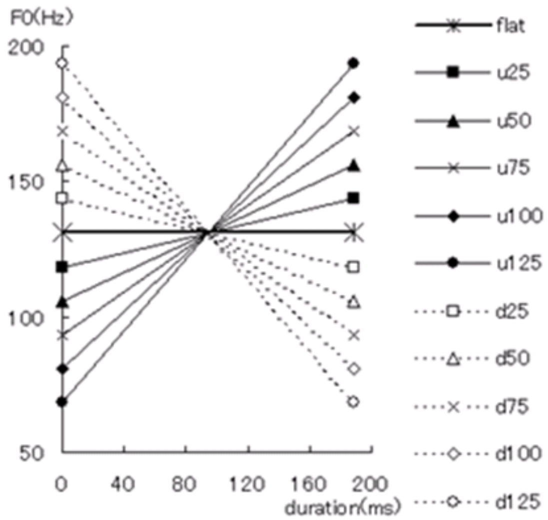

Fig.2 11 different F0 contours (duration: 189 ms). For example, "189u025" indicates that duration was 189 ms and F0 transition range was 25 Hz with upward slope (increasing intonation).

We prepared the 44 different types of triangle wave sounds with four different durations and 11 different F0 (fundamental frequency) contours. Specifically, the four durations were 189, 418, 639, and 819 ms. The 11 F0 contours were set so that the transition range of F0 values between the onset and endpoint in the sound stimuli were 0, ± 25, ±50, ±75, ±100, or ±125, and these were linearly downward or upward (Figure 2). All these 44 stimuli have the same F0 average 131 Hz. And these sounds' tone sounds like computer's beep sounds.

A. Participants

10 Japanese university students (6 men and 4 women; 19 – 23 years old) were participated.

B. Procedure

At first, an experimenter gave an instruction as "please estimate the attitudes of this laptop PC expressing some sounds." Then, they were asked to select one of the three attitudes "positive," "negative," or "undistinguishable" after presenting the one of the 44 prepared sounds. Specifically, these three attitudes were described to participants as follows:

- **Positive**: felt that the PC's internal state is good.
- **Negative**: felt that the PC's internal state is bad.
- **Undistinguishable**: could not select positive or negative.

As the concrete procedure, one randomly selected sound among 44 prepared sounds was presented to the participants. After presenting this sound, the participants were asked to select the one of the three attitudes described in the above. Each participant heard the all 44 prepared sounds. The order of the presenting sounds was counterbalanced among all 10 participants. Actually, these sounds were presented from a normal laptop PC (Let's note W2, product of Panasonic Inc,).

	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	*	−	−	−	−	−	*	−	−	−

Fig.3 Result of the preliminary experiment: This table indicates which kind of sounds were interpreted as positive (+), negative (-), or undistinguishable (*).

C. Results

The result of this preliminary experiment was summarized in Figure 3. This figure showed which kinds of sounds were interpreted as “positive,” “negative,” or “undistinguishable.” From this figure, there were four sounds that all 10 participants answered positive attitudes, that is, 189 ms with upward slope range 125 Hz (189u125), 418u125, 639u125, and 868u125. And there were four sounds that all answered negative, that is, 418 ms with downward slope range 25 Hz (418d25), 418d50, 418d75, and 639d50. Thus, the sounds with increasing intonation regardless of the duration were interpreted as positive attitudes, while the sounds that have around 500 ms duration with slightly decreasing were interpreted as negative ones.

These eight sounds were then selected to express to users from the agents having different appearances in the next actual experiment.

IV. EXPERIMENT

The purpose of this experiment is to investigate the effects of the agents’ appearance on participants’ interpretation of the agents’ attitudes. Specifically, the participants were presented the selected sounds in the above preliminary experiment from the agents having different appearances, e.g., Mindstorms robot, AIBO robot and normal laptop PC (see Figure 4). And they were asked to select the correct attitudes (positive, negative or undistinguishable) based on the expressed sounds from these three agents.

Fig.4 AIBO robot, Mindstorms robot and Laptop PC (from left to right).

A. Participants

9 Japanese university students (8 men and 1 woman; 21 – 24 years old) were participated. These participants were not familiar with robots or these toys, and did not participate in the preliminary experiment.

B. Procedure

At first, the participants were instructed that the concrete task of this experiment was to select one of the three attitudes (positive, negative or undistinguishable) based on the expressed sounds from the three agents.

As concrete experimental conditions that all participants experienced, the following four conditions were prepared.

1. Expressing the eight sounds from Mindstorms (**MS-sound condition**): expressing the eight sounds to participants from the FM radio tuner loaded on the Mindstorms. This radio tuner received the transmitted sounds from the Sound expressing PC (Figure 4).
2. Expressing the eight sounds from AIBO (**AIBO-sound condition**): AIBO expressing the sounds by using the AIBO operating software “AIBO entertainment player” installed in the AIBO operating PC.
3. Expressing the eight AIBO’s prepared behaviors (**AIBO-motion condition**): The AIBO entertainment player has about 80 prepared behaviors, such as “good morning” or “delightful.” Among these behaviors, we selected the following eight behaviors to express to the participants as primitive attitudes. As positive behaviors (four behaviors): “cheer1,” “cheer3,” “cheer4,” and “cheer5.” As negative behaviors (four behaviors): “angry1,” “angry2,” “sad1,” and “sad2.” These behaviors were selected based on the verbal labels of these behaviors.
4. Expressing the eight sounds from laptop PC (**PC-sound condition**): The same condition with the preliminary experiment. This laptop PC was remotely operated from the Sound expressing PC.

As the experiencing order of these conditions, at first, participants experienced MS-sounds, AIBO-sounds and AIBO-motion conditions in the random order. And then, they did the PC-sound condition. In all four conditions, the eight sounds or behaviors were randomly presented to the participants. The participants were asked to select the correct attitudes from the three attitudes (positive, negative, or undistinguishable) after the agent’s expressing certain information, so that this procedure was the nearly same with the preliminary experiment. The experiencing order of the experimental conditions and the presenting order of eight sounds or behaviors were counterbalanced among the participants.

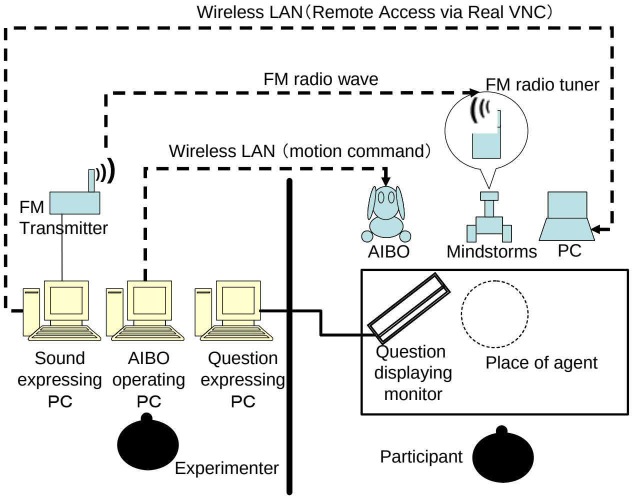

Fig.5 Experimental setting.

C. Results

We investigated the interpretation rates which indicated how many times the participants succeeded in estimating the agents' attitudes correctly in all four experimental conditions. As the results, the participants showed 3.33 times among eight experimental stimuli in MS-sound condition, 2.89 times in AIBO-sound condition, 3.33 times in AIBO-motion condition and 6.44 times in PC-sound condition (Figure 6). The results of ANOVA showed that there were significant differences among these four experimental conditions ($F(3,24)=8.26$, $p<.01$(**)), and multiple comparison by LSD test showed that there were significant differences between the PC-sound condition and the other three conditions (Mse=2.9421, 5% level).

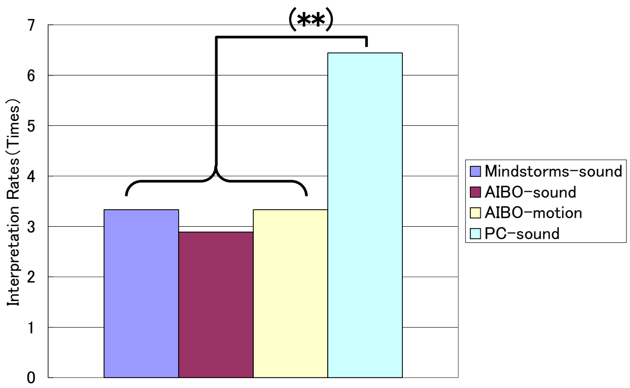

Fig.6 Participants' interpretation rates in four experimental conditions.

This result showed that the participants' interpretation rates of the same sound stimuli differed based on the each agent's appearance. Simply says, the participants showed the higher interpretation rates in PC-sound condition, while they showed the significant lower rates in MS-sounds and AIBO-sounds conditions, even though the same sounds were presented to participants in these three conditions. Moreover, the interpretation rates observed in AIBO-motion condition were also significantly lower than PC-sound condition, and these interpretation rates were the nearly same with MS-sound and AIBO-sound conditions, even thought these motions were prepared in commercial product.

D. Discussion

The eight artificial sounds selected in preliminary experiment were presented to the participants from Mindstorms, AIBO and PC. The participants' interpretation rates that indicate how many times the participants select the correct attitudes were then investigated. As the result, the interpretation rates when PC expressed these sounds were significantly higher than the rates when Mindstorms and AIBO expressed these. This result reveled that the agents' different appearances affected the users' interpretations of the agents' attitudes, even though these agents expressed the completely same information.

Let us consider the reasons why PC-sound condition showed the higher rates comparing with the other conditions. One reason is that these eight sounds were selected in the preliminary experiment when these sounds were presented from the laptop PC which was also used in the PC-sound condition. Thus, there were some possibilities that these sounds were only effective for informing the primitive attitudes when these were presented from the laptop PC. This phenomenon would be rooted on the fact that the PC expressing beep like sounds was very natural for all participants, in other words, the relationship between the appearances of PC and expressing beep like sounds would have higher affinities.

In any case, this result could reveal that agents' different appearances affected the users' interpretation of the agents' attitudes, however, this did not the actual reason why the interpretation rates in PC-sound condition were higher than in the other conditions. To tackle with this issue, it would require to carrying out the other experiments to consider the affinity between the appearance and its expressing information.

From the results of this experiment, there was an interesting phenomenon that the interpretation rates in AIBO-motion condition were lower than PC-sound condition like MS-sound and AIBO-sound conditions. From this result, it can be said that the AIBO's prepared behaviors were not really efficient to inform the participants the primitive attitudes, such as positive or negative. This result would support our hypothesis described in Introduction: *"the agents with rich appearance (quite similar with actual humans or pet animals) expressing the likely information (verbal information or animal like behaviors) are more confusing to users and are not really effective for interacting with users: Instead, the agents without rich appearance expressing the simple but intuitive information (e.g., subtle expressions) are readily understood and are much more effective for their interaction (see Figure 1).* Of course, the behaviors of AIBO were not designed for informing the primitive attitudes we estimated, however, this result would suggest the design policy to inform users certain attitudes effectively.

V. Discussion and Conclusions

A. Likeness and Familiarity

So far we discussed about the relationship between the agent's appearance and its expressed information for informing its attitudes for humans. On the other hand, about the effects of the agents' appearance (or human likeness) on the familiarities human users would feel, Mori [10] proposed the pioneering assumption "uncanny valley"; that is, the appearance of the agent is getting similar to ones of human beings, in some point, humans suddenly start feeling uncanny or losing the familiarities with this agent because the subtle differences on the appearances between actual human beings and the agents are emphasized. Mori described this relationship between the agents' likeness to human beings and familiarity that human users would feel as the following qualitative diagram shown in Figure 7.

Basically, we agree with this Mori's assumption because the relationship between the agents' appearance and the familiarities seems to succeed in explaining our hypothesis shown in Figure 1: *the agents with rich appearance (quite similar with actual humans or pet animals) expressing the likely information (verbal information or animal like behaviors) are more confusing to users and are not really effective for interacting with users: Instead, the agents without rich appearance expressing the simple but intuitive information such as subtle expressions are readily understood and are much more effective for their interaction.*

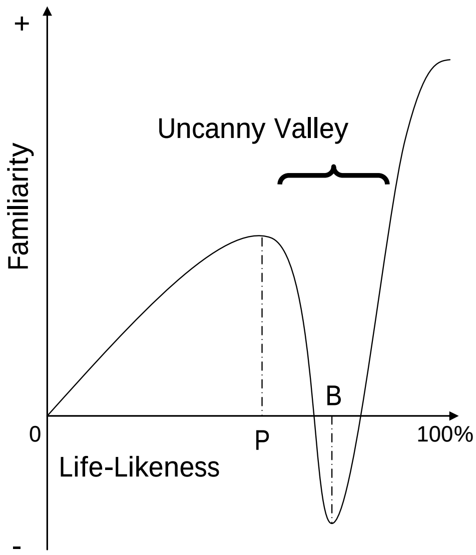

Fig.7 The concept diagram of "Uncanny Valley" [10].

We assumed that the former case in the above our hypothesis (the agents with rich appearance expressing the likely information) would correspond to the bottom (B) of the uncanny valley, so that it is expected that users would feel uncomfortable as shown in the left of Figure 1; On the other hand, we also assumed that the latter case (the agent without rich appearances expressing subtle expressions) would correspond to the peak (P) just before starting "uncanny valley," so that it is expected that users would feel comfortable as shown in the right of the Figure 1.

B. Likeness and Implementability

In this paper, we assumed that the agent without rich appearance expressing the simple but intuitive information, e.g., subtle expression, is much more effective for human users to understand the agents' internal states. One of the reasons to focus on utilizing "subtle expressions" is that it is technically and economically easy to implement these kinds of expressions into the agent. However, this does not imply that every kind of the agents should express the subtle expressions.

Corresponding to the results of this study, these subtle expressions should be designed according to the appearance of the agents, e.g., the beep sounds should be expressed from PC, not from robotic agents. Therefore, each agent would have each appropriate subtle expression, and then its "implementability" would be getting decreasing according with increasing the life-likeness shown in Figure 8. Simply saying, the agent without rich appearance (lower likeness) requires expressing rather simple expressions, while the agent with rich experience (higher likeness) does complex expressions. We assumed that the implementability is constantly decreasing according to increment of the likeness. And its value would become minus when the likeness value exceeds around (P).

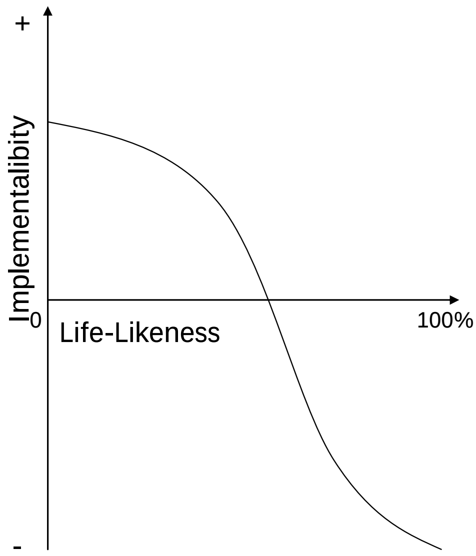

Fig.8 The hypothesized implementability according to its likeness.

C. Familiarity, Implementability and Likeness

So far we discussed about the relationship between the likeness of the agents and the familiarity human users would feel, and the one between the likeness and the implementability of expressed information. It can be said that the former relationship "familiarity" is about between the agent and the user, while the later one "implementability" is about between the agent and the designer.

$$Familiarity=q(agent, user) \quad (1)$$

$$Implementability=w(agent, designer) \quad (2)$$

Here, we assumed that the factor "feasibility" of the intimate interaction between users and the agents can be proposed by means of the two factors "familiarity" and "implementability" as the following equation (3) and (4):

$$Feasibility = Familiarity + Implementability \quad (3)$$

$$= q(agent, user) + w(agent, designer) \quad (4)$$

From the equation (4), this feasibility could include the three factors, "user," "agent" and "designer" that are the important factors to form the interaction design. Therefore, we expect that this feasibility factor could literally indicate whether the user and the agent could create the intimate interaction or not.

We hypothesized that it is possible to superpose the familiarity and the implementability according to the likeness. We could then acquire the concept diagram of the "feasibility" shown in Figure 9. From this figure, at first its feasibility values are getting increase around (P), however, this value suddenly decrease, as if the familiarity value falls into the uncanny valley. In addition, even though the likeness value is getting close to 100% likeness, the feasibility values could not recover to the same level of 50% likeness.

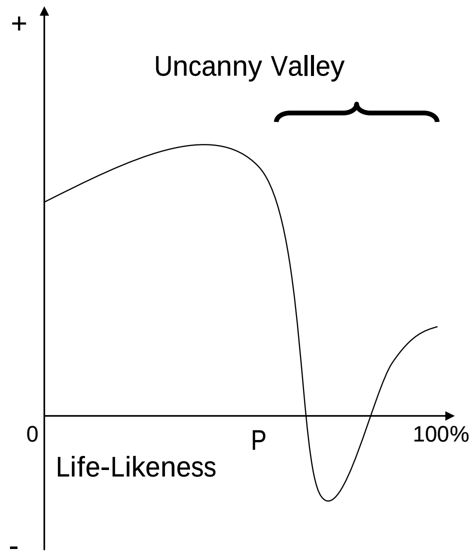

Fig.9 The hypothesized "feasibility" according to its likeness.

Thus, it can be said that this figure shows that the agent with higher likeness would have the lower feasibilities to create an intimate relationship between the users and the agents, while the agent with lower likeness would have the higher feasibilities. This feasibility concept would support our hypothesis shown in Figure 1, and accord the basic concept of the Mori's uncanny valley assumption.

This means that our feasibility concept mostly support Mori's uncanny valley assumption. However, there is one significant difference Mori's assumption: that is, once the feasibility value falls into the uncanny valley, this value never rises up even though its likeness value is getting close to 100% likeness. We assumed that this phenomenon about the feasibility diagram would support again our hypothesis shown in Figure 1.

As the consecutive studies based on the results of this study and our feasibility concept, we are planning to conduct the other experiments to reveal what is the most appropriate information according to the agents' appearance. For example, what is the appropriate information expressing from Mindstorms robot to inform its primitive attitudes? Are Starwars' R2D2 like behaviors appropriate? We expect that these consecutive studies would support our feasibility concept that can clarify the effective coupling between the appearance and its expressing information for realizing interactive agents readily and easily.

VI. References

[1] S. Yamada and T. Yamaguchi.: Training AIBO like a Dog, the 13th International Workshop on Robot and Human Interactive Communication, pp.431-436 (2004).

[2] S. Kiesler., L. Sproull., & K. Waters.: A Prisoner's Dilemma Experiment on Cooperation with People and Human-like Computers. Journal of Personality and Social Psychology, 79(1), 47-65 (1995).

[3] S. Kiesler., & L. Sproull.: Social Responses to "Social" Computers. In B. Friedman, Human values and the design of technology, CLSI, Publications (1997).

[4] T. Komatsu.: Audio Subtle Expressions Affecting User's Perceptions, the 9th International Conference on Intelligent User Interface, p.306-308 (2006).

[5] J. Goetz., S. Kiesler., & A. Powers.: Matching Robot Appearance and Behavior to Tasks to Improve Human-Robot Cooperation, 12th IEEE Workshop on Robot and Human Interactive Communication (2003).

[6] K. Liu., & W, R, Picard.: Subtle expressivity in a Robotic Computer, CHI2003 Workshop on Subtle Expressivity for Characters and Robots (2003).

[7] MindStorms: http://mindstorms.lego.com/

[8] AIBO: http://www.jp.aibo.com.

[9] S. Yamada, T. Komatsu: Designing Simple and Effective Expression of Robot's Primitive Minds to a Human, 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp.2614-2619 (2006).

[10] M. Mori.: Bukimi no tani (The Uncanny Valley). Energy, 7(4), 33–35 (1970). (Originally in Japanese)