Agents with Faces: The Effect of Personification

Tomoko Koda / Pattie Maes

MIT Media Laboratory
20 Ames Street
Cambridge, MA 02139 USA
tomoko/pattie@media.mit.edu

Abstract

It is still an open question whether software agents should be personified in the interface. In order to study the effects of faces and facial expressions in the interface, a series of experiments was conducted to compare subjects' responses to and evaluation of different faces and facial expressions.

The experimental results obtained demonstrate that: 1) personified interfaces help users engage in a task, and are well suited for an entertainment domain; 2) people's impressions of a face in a task are different from ones of the face in isolation. Perceived intelligence of a face is determined not by the agent's appearance but by its competence; 3) there is a dichotomy between user groups which have opposite opinions about personification. Thus, agent-based interfaces should be flexible to support the diversity of users' preferences and the nature of tasks.

1. Introduction

There is a growing interest in interface agents which assist a user in performing daily computer tasks. Software agents will make software more active and work autonomously without waiting for a user's command. This "agent-based" interface should be social and active to support cooperative tasks between a user and an agent [5]. Maes addresses that the key HCI issue in building a successful interface agent is to help the user understands what the agent's capacities are, its limitations, and its way of operating, etc. [4].

Many software agents are personified in terms of a human or caricature face. "Phil" in Apple's Knowledge Navigator video [3], Apple's "Guides" project [6], the MIT Media Lab's Maxims system [2] used faces to help users understand the agent's characteristics or behaviors. The above research hypothesizes that users interact with agents more easily by being able to predict the agent's characteristics and behaviors via their external traits. Though the above research reports positive results for a personified interface, the results are obtained by observation or interviews, not by quantitative analysis. The primary focus of this research is on implementation of a prototype rather than user testing.

However, other research with quantitative analysis indicates that adding a face to an interface does not necessarily result in better human-computer interactions. Walker et al. [9] report that having a face is engaging and takes more effort and attention from the user. A face with more expression leads to greater engagement, but is not always preferable. Takeuchi et al. report that users respond differently to systems having a face than to those without. They suggest that a face in an interface takes more effort from the user because people try to interpret the human images [8].

The research reported in this paper includes both implementation and quantitative analysis of subjects' impression about a personified interface. This paper argues that employing a face as the representation of an agent is engaging and makes a user pay attention. It is possible to make advantages of personification outweigh the drawbacks by using faces in appropriate application domains, where conveying the agent's behavior to the user is important for making a user comfortable with the system, or in an entertainment domain where engagement is crucial and users don't try to minimize efforts. This paper reports the results of an investigation on people's impressions of faces in an interactive entertainment environment.

2. Experiment

We built an application to test agent-based interfaces, namely a poker game, and experiments were performed on the World Wide Web so as to have as many versatile subjects as possible. The results were analyzed quantitatively.

2.1 Purpose

The purpose of the experiment is to understand 1) the effect of having a face and facial expressions in an interface, such as required attention, engagement, and distraction, 2) what kind of facial features (gender, humanity, realism) make the agent look intelligent, likable, and comfortable to work with, 3) whether people's impression of an agent is

IEEE International Workshop on
Robot and Human Communication
0-7803-3253-9/96 $5.00 ©1996 IEEE

determined by its representation or by its performance or both, and 4) whether people's impression of the faces differs by their gender or opinion about personification.

2.2 System Design

We designed and built a web-based poker game in which computer poker playing agents play against each other and a user. The task for the user is to play this poker game against the poker playing agents which have personified representations. Figure 1 shows the interface of the game. The poker playing agents used in the experiments have seven different graphical representations: Realistic Human faces, Caricature Human faces, a Caricature Dog's face, a line-drawn Smiley face, and the Invisible Man, which has no face (see Figure 2). They vary in terms of humanity and realism. Each agent has ten facial expressions (neutral, pleased, displeased, satisfied, surprised, excited, anxious, and very excited). It changes its expressions when it deals, bets, and wins/loses. For example, a dog's face might show an anxious expression while betting, or show an excited face when bluffing. Figure 3 shows the ten expressions of the female caricature face. The emotions were derived from the Ortony, Clore, and Collins (OCC) model of emotion types [7].

2.3 Procedure

A typical scenario of an experiment is as follows:

A subject accesses the poker web site. The subject reads the consent form and instructions and agrees to participate in the experiment. The subject's name, email address, and login time are sent to the server and stored in a disk. The poker java applet is downloaded to the subject's computer. The subject plays a poker game with one out of five conditions for 15 rounds, which takes from 10 to 15 minutes to complete. The subject's and agents' hands, actions, and the ending money state are logged into the server. A questionnaire appears after playing 15 rounds. The subject answers and submits the questionnaire, and the answers are stored in the server.

2.4 Conditions of the Experiments

There are five web-based experiments which are randomly assigned to a user. The FACE experiment is designed to study the effect of having a face. The representations are the Caricature Female face and a white square with the same size which is called "Invisible Man". The GENDER experiment studies the difference in subjects' impressions of characters' gender. The representations used in this experiment are the Realistic Male and the Realistic Female face. The HUMANITY experiment is designed to compare the difference in subjects' impressions between a human face and a non-human face. The representations used are the Caricature Male face and the Dog's face. The REALISM experiment uses three faces to compare three different levels of realism. The representations used are the Realistic Male

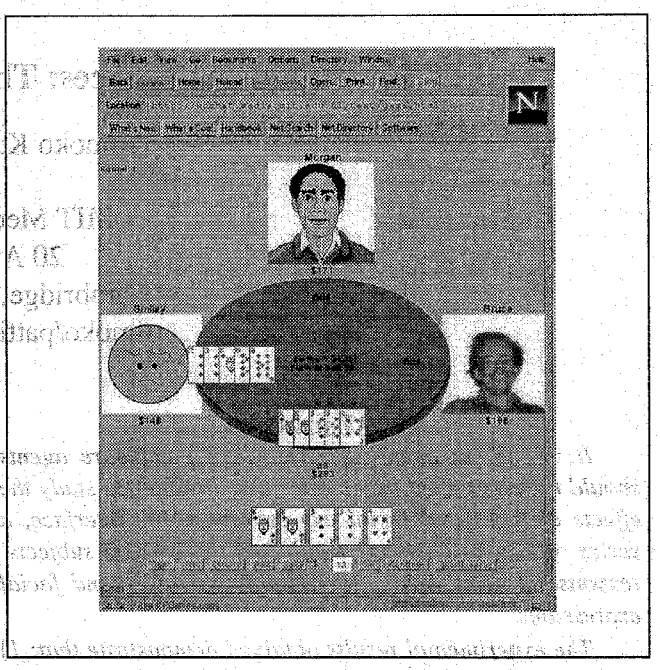

Figure 1: Web interface of the poker game (REALISM experiment)

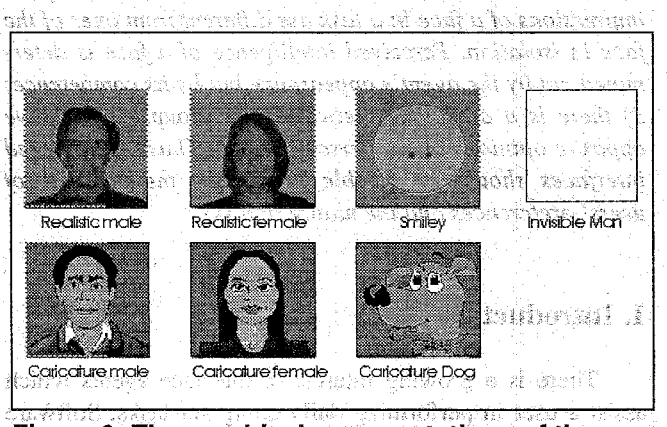

Figure 2: The graphical representations of the poker playing agents

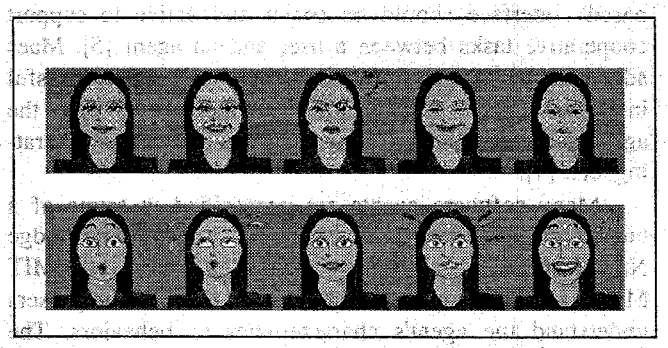

Figure 3: Ten facial expressions of the Caricature Female face

face, the Caricature Male face, and the Smiley face. The EXPRESSIVENESS experiment has three players with the same facial representation. This experiment was designed to evaluate the difference in subjects' impression between three modes of expressiveness. One agent has Honest expressions, the other has Deceiving expressions, and the last one has Stoic expressions. Table 1 shows the facial representations used in each experiment. All facial images used in the experiments have the same size, same resolution, and the same background color. Except for the EXPRESSIVENESS experiment, all poker playing agents are in Honest mode. The agents play poker game using exactly the same poker playing strategy, which the subjects were not made aware of.

Table 1: The facial representations used in each experiment

	Player 1	Player 2	Player 3
FACE	Caricature Female	Invisible man	-
GENDER	Realistic Male	Realistic Female	-
HUMANITY	Caricature Male	Caricature Dog	-
REALISM	Realistic Male	Caricature Male	Smiley
EXPRES-SIVENESS	Honest Caricature Female	Deceiving Caricature Female	Stoic Caricature Female

Three pilot experiments were conducted prior to the five web-based experiments. The purpose of the pilot experiments was to collect subjects' impressions of each face based solely upon its visual appearance. Subjects were shown one of the sets of faces used in the GENDER, HUMANITY, and REALISM experiments with a neutral expression. The subjects answered questions about each face's intelligence, likability, and engagingness as a opponent poker player, just by looking at the facial images.

2.5 Questionnaire

Questions varied according to the experiment. There are 4 background questions (subjects's gender, age, computer expertise, and their opinion about personification), and 4 to 8 questions about the subjects' impressions of the faces against which they played poker. The questions about their impressions are answered on a seven point scale (1: strongly disagree to 7: strongly agree).

The common questions across the experiments are perceived intelligence of playing poker game, likability, engagingness, and level of comfort for each face. Specific questions are attention required and level of distraction (for FACE and EXPRESSIVENESS experiments), level of correspondence to the actual intelligence of the player (for HUMANITY and REALISM experiments), believability of the facial expressions, and the usefulness of the facial expressions (for EXPRESSIVENSS experiment).

2.6 Subjects

Experiments continued for one and a half months. Subjects participated in the experiments from all over the world using the World Wide Web. More than 1,000 people accessed the poker game site and 157 of them answered the questionnaire, for a response rate of 15%. Of these subjects, 78% were male. The age range of the subjects was from 10 to 50 years old. Fifty-seven percent of them were in their 20's, 26% were teens, and 14% were in their 30's. Fifty-two percent of them were advanced computer users, 40% were intermediate users. When asked about personifying an interface (VOTE), 51% of them supported having a face on the screen ("AGREE" group), the rest were against having a face ("DISAGREE" group).

3. Results

3.1 Impressions Based on Appearance

The pilot experiments evaluate subjects' impressions based solely on agents' appearance, in particular, gender, humanity, and realism. The result of the comparison for gender shows that there is no difference between people's impression of the Male face and Female face in terms of its perceived intelligence, likability, and engagingness, as shown in Table 2. The result of the comparison for humanity shows that the Human face is perceived as more intelligent than the Dog's face based on visual appearances, but less likable and engaging as a representation for a poker player, as shown in Table 3. In the experiment to compare different levels of realism, subjects rated the Realistic face to be more intelligent, likable, and engaging than the Caricature face and Smiley, as shown in Table 4.

Table 2: The mean value of variables for a Male and Female face based on appearance

Variables	Male (n=15)	Female (n=15)	t (paired)
INT (Intelligence)	4.73	5.13	t(14) = -.88
LIKE (Likability)	4.80	4.73	t(14) =.19
ENG (Engagingness)	5.13	4.80	t(14) =.77

** p<.01, ** p<.05, *** p<.10, no mark = not significant*

Table 3:The mean value of variables for a Human and Dog's face based on appearance

Variables	Human (n=16)	Dog (n=16)	t (paired)
INT (Intelligence)	3.88	2.63	t(15) = 3.10*
LIKE (Likability)	3.81	5.37	t(15) = -3.13*
ENG (Engagingness)	3.50	5.56	t(15) = -5.57*

** p<.01, ** p<.05, *** p<.10, no mark = not significant*

Table 4:The mean value of variables for a Smiley, Caricature, and Realistic face based on appearance

Variables	Smiley (n=16)	Caricature (n=16)	Realistic (n=16)	F(value, df)
INT (Intelligence)	3.53	3.59	4.94	F (6.12, 2)*
LIKE (Likability)	3.76	4.06	4.88	F (5.06, 2)**
ENG (Engagingness)	3.65	3.82	5.12	F (6.25, 2)*

** p<.01, ** p<.05, *** p<.10, no mark = not significant*

3.2 Difference for Face and Noface

As shown in Table 5, subjects paid attention to the face (ATT = 5.28, where 4.0 is neutral), but were not distracted by the existence of the face (DIS = 2.84, where 4.0 is neutral). Subjects rated having a face to be significantly more likable (t(24)=2.76, p<.05), more engaging (t(24)=2.85, p<.01), and more comfortable to play against (t(24)=3.36, p<.01), than not having a face. However, both poker playing agents are rated to be equally intelligent regardless of the existence of a face.

Table 5:The mean value of variables for a Caricature face(Face) and the Invisible Man (NoFace) in the game

Variables	Face (n=25)	NoFace(n =25)	t (paired)
INT (Intelligence)	3.64	3.80	t(24) = -.27
LIKE (Likability)	5.08	3.72	t(24) = 2.76**
ENG (Engagingness)	4.60	3.32	t(24) = 2.85*
COM (Comfortableness)	4.88	3.36	t(24) = 3.06*
ATT (Attention to the face)	5.28	-	-
DIS (Distracted by the face)	2.84	-	-

** p<.01, ** p<.05, *** p<.10, no mark = not significant*

3.3 Difference for Characters's Gender

As shown in Table 6, there is no main effect of the agent's gender in any variables. However, there are significant interactions between VOTE (people's opinion about personification) and perceived intelligence (F(4.73, 1), p <.05), and slight interactions between VOTE and likability(F(3.62, 1), p <.10). The AGREE group (those who are for personification) rated intelligence and likability for the Male face higher and the Female face lower. While the DISAGREE group (those who are against personification) rated the Male and the Female face in the opposite way. No other 2-way or 3-way interactions were found.

3.4 Difference for Degree of Humanity

As shown in Table 7, there is no main effect of HUMANITY (Human or Dog's face) in any variables. However, there are significant interactions between VOTE and likability (F(5.40, 1), p<.05), comfortableness (F(11.42, 1), p <.01), and level of correspondence to the actual poker playing skill (F(6.97, 1), p <.05). The AGREE group rated likability, comfortableness, and correspondence of the Human face significantly lower while they rated the Dog's face significantly higher. The DISAGREE group rated the Human and the Dog's face in the opposite way. There are also interactions between subjects' gender and likability (F(9.24, 1), p <.01), engagingness (F(5.94, 1), p <.05), and comfortableness (F(7.26, 1), p <.05).

Table 6: The mean value of variables for a Male and Female face in the game.

Variables	Male (n=37)	Female (n=37)	t (paired) F(value, df)
INT (Intelligence)	3.35	3.83	t(36) = -1.33
Vote Agree (n=18)	3.94	3.55	F(4.73, 1)**
Vote Disagree (n=19)	2.78	4.10	
LIKE (Likability)	4.49	4.62	t(36) = -.55
Vote Agree (n = 18)	5.11	4.83	F(3.62,1)***
Vote Disagree (n = 19)	3.89	4.42	
ENG (Engagingness)	4.32	4.43	t(36) = -.55
COM (Comfortableness)	4.00	3.92	t(36) =.30

Table 7:The mean value of variables for Human and Dog's face in the game

Variables	Human (n=23)	Dog (n=23)	t (paired) F(value, df)
INT (Intelligence)	3.73	3.82	t(22) = -.18
LIKE (Likability)	4.56	4.39	t(22) =.53
Vote Agree (n = 11)	4.72	4.73	F(5.40, 1)**
Vote Disagree (n = 12)	4.45	4.09	
Male subject (n = 17)	4.47	4.65	F(9.24, 1)*
Female subject (n = 6)	4.83	3.67	
ENG (Engagingness)	3.95	4.13	t(22) = -1.00
Male subject (n = 17)	3.88	4.29	F(5.94, 1)**
Female subject (n = 6)	4.17	3.67	
COM (Comfortableness)	4.08	4.04	t(22) =.13
Vote Agree (n = 11)	3.91	4.55	F(11.42, 1)*
Vote Disagree (n = 12)	4.27	3.55	
Male subject (n = 17)	4.29	4.41	F(7.26, 1)**
Female subject (n = 6)	3.50	3.00	
COR (Correspondence)	3.91	3.50	t(22) = 1.16
Vote Agree (n = 11)	3.82	4.18	F(6.97, 1)**
Vote Disagree (n = 12)	4.00	2.82	

** p<.01, ** p<.05, *** p<.10, no mark = not significant*

3.5 Difference for Degree of Realism

As shown in Table 8, there is no main effect of REALISM in perceived intelligence, likability, and engagingness.

The results show that the Realistic face is slightly more intelligent, likable, engaging than other faces, though not significantly so. There is a main effect of REALISM in comfortableness (F(.05, 2), p <.10) and a significant main effect in correspondence to the skill level (F(.18, 2), p <.05). The Realistic face is rated as more comfortable to play against than the Smiley face. In terms of correspondence to the actual poker playing skill, subjects rated the Smiley face as the most appropriate representation, and the caricature and Realistic faces are less appropriate. No 2-way or 3-way interactions were found.

Table 8: The mean value of variables for Smiley. Caricature, and Realistic face in the game

Variables	Smiley (n=30)	Caricature (n=30)	Realistic (n=30)	F(value, df) t (paired)
INT (Intelligence)	3.37	3.53	3.70	F (.30, 2)
LIKE (Likability)	4.20	4.10	4.70	F (.59, 2)
ENG (Engagingness)	4.17	4.13	4.76	F (.58, 2)
COM (Comfortableness)	4.07	4.00	4.63	F (.05, 2)***
Smiley vs. Caricature				t(29) =.14
Caricature vs. Realistic				t(29) = -1.60
Smiley vs. Realistic				t(29) = -1.71***
COR (Correspondence)	4.67	3.90	4.03	F (.18, 2)**
Smiley vs. Caricature				t(29) = 2.25**
Caricature vs. Realistic				t(29)= -.36
Smiley vs. Realistic				t(29) = 2.00**

* p<.01, ** p<.05, *** p<.10, no mark = not significant

3.6 Difference between Levels of Expressiveness

As shown in Table 9, there is no main effect of EXPRESSIVENESS in any variables. However, there are significant interactions between VOTE and perceived intelligence (F(8.91, 2), p <.01) and comfortableness(F(4.47, 2), p <.05). The AGREE group rated the Honest face as most intelligent, while the DISAGREE group rated the Stoic face as the most intelligent and the Honest face as the least. The AGREE group rated the Stoic face as least comfortable, while the DISAGREE rated the same face as most comfortable.

4 Discussion

Subjects rated a poker playing agent with and without a face as equally intelligent. This suggests that attaching a face did not add any perceived intelligence to the poker player. However, having a face is considered more likable, engaging, and comfortable to play against regardless of subjects' opinion about personification. As described earlier, Takeuchi's [8] and Walker's [9] studies show a similar result, that having a face is engaging. It is encouraging that the face did not create a negative effect on the subjects' likability of the experience, even to those who have a negative

Table 9:The mean value of variables for Honest, Deceiving, and Stoic face in the game

Variables	Honest (n=20)	Deceiving (n=20)	Stoic (n=20)	F(value, df)
INT (Intelligence)	3.84	3.37	3.89	F(1.98, 2)
Vote Agree (n = 10)	4.20	3.10	3.30	F(8.91, 2)*
Vote Disagree (n = 10)	3.44	3.66	4.55	
LIKE (Likability)	4.37	4.37	4.68	F(1.28, 2)
ENG (Engagingness)	3.95	4.21	4.21	F(.56, 2)
COM (Comfortableness)	3.95	3.74	3.95	F(.14, 2)
Vote Agree (n = 10)	3.90	4.10	3.60	F(4.47, 2)**
Vote Disagree (n = 10)	4.00	3.33	4.33	
BEL (Correspondence)	4.79	4.79	4.63	F(6.44, 2)
ATT (Required attention)	4.32	4.47	4.37	F(1.46, 2)
DIS (distraction)	3.26	3.21	3.37	F(5.26, 2)
USE (usefulness)	4.21	4.42	4.11	F(1.81, 2)

* p<.01, ** p<.05, *** p<.10, no mark = not significant

opinion about personification, contrary to what Walker's study found [9]. This difference may be caused by the nature of the tasks and faces used in these studies. Walker used a woman's synthetic talking face in a questionnaire survey, while this study used a female caricature face in a poker game environment. Subjects' responses to the Human and Dog's face when seen in isolation show reasonable expectations from the appearance of each face regardless of their opinion about personification. King's study [1] shows a similar result in perceived intelligence, that human forms are perceived to be more intelligent than other forms. However, subjects rated the Human and Dog's faces to be equally intelligent (or unintelligent) in the poker game. This means that they did not rate a player's intelligence based on its appearance. Their impressions are based on the player's actual competence in playing poker. However, their opinions about personification affect their attitude toward each face. The AGREE group are more "strict" about personification, since they prefer the Dog's face to the Human face, felt more comfortable with the Dog's face, and considered the Dog's face more appropriate for representing the actual competence of the player in playing poker. On the other hand, the DISAGREE group are more "generous" about personification. The subject's gender works in the same way as the subjects' opinion about personification. Males are more "strict", while female are more "generous" about personification. However, due to the small number of female subjects, further study is needed to conclude that users' gender affects their impressions of agent's personification.

In terms of realism, the Realistic face was rated as more intelligent, engaging, and likable than the less realistic faces when seen in isolation. Subjects categorized the faces in two ways--realistic face and not-realistic (Caricature and Smiley) when they evaluate the faces based on appearance. The

level of realism does not affect the overall impression of faces, unless they are not realistic. This result is similar to King's study [1] which indicates that fully articulated human forms are rated to have higher intelligence and agency than either the caricatures or the Chernoff faces (corresponding to Smiley face in this experiment), and caricatures and Chernoff faces are rated to have similar intelligence and agency. However, subjects rated the three faces differently after playing poker in this study. They judged the level of perceived intelligence of a face based not on its appearance but instead on its competence. Subjects rated the three faces to be equally intelligent, and they thought that the Smiley face (less realistic) face represented the true level of competence more appropriately.

It is interesting that people's opinion about personification affects the way they feel about expressiveness. The AGREE group rated expressive faces (Honest and Deceiving) to be more comfortable, and the Honest face to be most intelligent. The DISAGREE group rated the stoic face as most intelligent and comfortable. This means, the DISAGREE group is more strict about simulating a real poker situation, where players are not expected to show honest expressions.

5 Conclusion

The first finding is that having a face is considered more likable, engaging, and comfortable to see in a poker game environment. It is encouraging that people are favorable to having a face in an interface regardless of their opinion about personification. People were not distracted by the presence of a face or facial expressions. Moreover, people tried to interpret faces and facial expressions, which makes the users pay attention to the face and engage in the task. It is clear that faces are useful for entertainment purposes, since engagement is essential for games and people don't care about taking more effort for entertainment. Hence it may also be useful for applications which require an engaged user for success, such as education and training.

The second finding is that people's impressions of a face are different when they see a face in isolation versus when they interact with a face within a task. People evaluate a face not based on appearance but its competence or performance. For example, people rated perceived intelligence, likability, and engagingness of the Human face and the Dog's face, or Smiley, the Caricature, and the Realistic face differently when they rated the faces based on their appearance. While there were no differences in evaluations when they rated the same faces in the poker game. Most psychological HCI studies use static facial images separately from applications. Considering what we understand from this study, we have to evaluate effects of personification within a context,--i.e. the context of software agent applications. Of course not all software agents require personified interfaces. The goal of HCI work should be to understand when a personified interface is appropriate.

The third finding is that there is a dichotomy between user groups which have opposite opinions about personification. Differences in facial features--such as character's gender, humanity, expressiveness--cause opposite evaluations by these two subject groups. For example, the Dog's face is preferred by those who are in favor of personification, while the Human face is preferred by those who are against. Another example is that those who are in favor of personification attributed more intelligence to the expressive face, while those who are against thought the stoic face had higher level of intelligence. Some of the experiments performed indicate that there is also a potential difference in evaluation of a human face and a non-human face between the subjects' gender. Subjects' computer expertise, age range were not taken into consideration in this study. We need to consider the target users when designing a personified interface. The future personified interface should be flexible so that it can provide options to choose a preferred face for each user.

Acknowledgments

This research was supported by Apple Computer and British Telecom.

References

1. King, J. *et al.* "The Representation of Agents: Anthropomorphism, Agency, and Intelligence." in Proc. of CHI '96. ACM Press, 1996. 289-290.

2. Lashkari, Y. *et al.* "Collaborative Interface Agents." In Proc. of the National Conference on Artificial Intelligence, Cambridge, MA.: MIT Press, 1994

3. Laurel, B. "Interface Agents: Metaphors with Character." In The Art of Human-computer Interface Design, edited by Brenda Laurel. Reading, MA: Addison-Wesley, 1990. 355-365.

4. Maes, P. "Agents That Reduce Work and Information Overload". In Communications of the ACM. Vol. 37, No. 7 (July, 1994). 31-40,146.

5. Negroponte, N. Being Digital. 101-102. The Alfred A. Knorpf, Inc. 1995.

6. Oren, T. *et al.* "Guides: Characterizing the Interface." In The Art of Human-computer Interface Design, edited by Brenda Laurel. Reading, MA: Addison-Wesley, 1990. 367-381.

7. Ortony, A. *et al.* "The Cognitive Structure of Emotions." Cambridge, MA: Cambridge University Press. 1988

8. Takeuchi, A. *et al.* "Situated Facial Displays: Towards Social Interaction," in Proc. of CHI '94, ACM Press, 1994. 450-454.

9. Walker, J. *et al.* "Using a Human Face in an Interface," in Proc. of CHI '94, ACM Press, 1994. 85-91.