Persona Effect Revisited

Using Bio-signals to Measure and Reflect the Impact of Character-based Interfaces

Helmut Prendinger, Sonja Mayer, Junichiro Mori, and Mitsuru Ishizuka

Department of Information and Communication Engineering
Graduate School of Information Science and Technology
University of Tokyo
prendinger@acm.org, SonjaCMayer@web.de,
{jmori,ishizuka}@miv.t.u-tokyo.ac.jp

Abstract. The so-called 'persona effect' describes the phenomenon that a life-like interface agent can have a positive effect on the user's perception of a computer-based interaction task. Whereas previous empirical studies rely on questionnaires to evaluate the persona effect, we utilize bio-signals of users in order to precisely associate the occurrence of interface events with users' autonomic nervous system (ANS) activity. In this paper, we first report on the results of an experiment with an agent-guided mathematical game suggesting that an interface character with affective behavior may significantly decrease user stress. Then, we describe a character-based job interview scenario where a user's affective state derived from physiological data is projected back (or 'mirrored') to the user in real-time. Rather than measuring the effect of an interface agent, the focus here is on employing a character as a medium to reflect the user's emotional state, a concept with some potential for emotional intelligence training and the medical domain, especially e-Healthcare.

1 Introduction

While animated agents, or life-like characters, start populating the interfaces of numerous computer-based applications [10], their impact on human users is still largely unexplored, or at least very general in its formulation. In the context of educational software, Lester et al. [5] identified the *persona effect* which refers to (i) the credibility and motivation enhancing effects of character-based interfaces, as well as to (ii) the positive effect of animated agents on the users' perception of the learning experience. Van Mulken et al. [14] conducted a follow-up study to [5] where a life-like character acts as a presenter of technical and non-technical information. In their experiment, the positive effect of an animated interface agent on the 'subjective measures' entertainment and perceived difficulty is supported (for technical information), whereas no significant effect on 'objective' measures of the interaction, such as comprehension and recall, could be shown. Both of the mentioned studies rely on questionnaires as an evaluation method that does not allow for a precise temporal assessment of which particular behavior of agents is responsible for their good overall perception.

In this paper, we propose to take physiological data of users during the interaction with a character-based interface as an evaluation method. As bio-signals have recently been shown to be indicative of the affective state of users [12], we may gain new insights concerning subjective measures of the persona effect.[1] The recorded history of users' bio-signals will enable to precisely relate ANS activity with the (user-computer) interaction state, and hence track the impact of agent behavior. Furthermore, by using both bio-signals and questionnaires as evaluation methods, we may detect possible discrepancies between the interaction as *perceived* by the user and the *factual* physiological state of the user.

The promising results of the experiment sparked our interest to reflect the effect of an interface persona to the user in a more direct way, thus allowing the user to inspect his or her arousal in real-time. The *Emotion Mirror* is a web-based application depicting a job interview scenario. The arousal level of users—again inferred from their ANS activity—during the interview is 'mirrored' to them by employing a life-like character as an embodied mediator of their experienced stress or relaxation states. Although the current demonstrator system is fairly simple, it allows to gather valuable experiences for the next generation of emotional intelligence [1] training systems and e-Healthcare applications [7].

The following section reports on the experiment, and Section 3 briefly explains the Emotion Mirror application. Section 4 concludes the paper.

2 Character-based Quiz Game – An Empirical Study

The experiment about a simple quiz game described in this section investigates the effect of a life-like character with affective behavior on users' affective state which is derived from physiological data. The primary hypothesis of this study can be formulated as follows: *If a life-like interface agent provides affective feedback to the user, it can effectively reduce user stress.* To our knowledge, this is the first investigation that explores the possibility of employing an animated agent to respond to presumed negative feelings on the part of the user. Other research used an embodied character without addressing the issue of user frustration (Mulken et al. [14]) or provided only text-based response as a feedback medium to the (deliberately frustrated) user (Klein et al. [3]).

2.1 Theory and Game Design

We implemented a simple mathematical quiz game where subjects are instructed to sum up five consecutively displayed numbers and are then asked to subtract the i-th number of the sequence ($i \leq 4$). The instruction is given by the"Shima" character, an animated cartoon-style 2D agent, using synthetic speech and appropriate gestures. The numbers are also displayed in a balloon adjacent to the character. Subjects compete for the best score in terms of correct answers and

[1] It is important to note that our work differs from other research on the persona effect [5, 14] in that we compare "affective persona" vs. "non-affective persona" conditions rather than "persona" vs. "no persona" conditions.

time. Subjects were told that they would interact with a prototype interface that may still contain some bugs. This warning was essential since in some quiz questions, a delay was inserted before showing the 5th number. The delay was assumed to induce frustration as the subjects' goals of giving the correct answer and achieving a fast score are thwarted.

In order to measure user frustration (or stress), we took users' galvanic skin response (GSR) signal which is an indicator of skin conductance.[2] It has been shown that skin conductance varies linearly with the overall level of arousal and increases with anxiety and stress (see Picard [9], Healey [2]).

2.2 Method

Subjects and Design. Participants of the experiment were twenty male students of the School of Engineering at the University of Tokyo, on average 24 years of age, and all of them native speakers of Japanese. According to the independent variables, *affective* vs. *non-affective* feedback of a life-like character, two versions of the quiz game have been prepared:

- *Affective version.* Depending on whether the subject selects the correct or wrong answer from the menu displayed in the game window (see the numbers in Fig. 1), the character expresses 'happy for' and 'sorry for' emotions both verbally and nonverbally, e.g., by "smiling" (for happiness) and "hanging shoulders" (for sorriness). When a delay in the game flow happens, the character expresses empathy for the subject after the subject answers the question that was affected by the delay (see Fig. 1).
- *Non-affective version.* The character does not give any affective feedback to the subjects. It simply replies "right" or "wrong" to the answer of the subjects. If a delay happens, the agent does not comment on the occurrence of the delay, and simply remains silent for a short period of time.

If a delay occurs (in the affective version), the character expresses empathy to the subjects by displaying a gesture that Japanese people will easily understand as a signal of the interlocutor's apology (see Fig. 1), and uttering: "I apologize that there was a delay in posing the question" (English translation). Note that the apology is given *after* the occurrence of the delay, immediately after the subject's answer (and not during the delay period).

In order to show the effect of the character's behavior on the physiological state of subjects, we consider specific segments. (i) The DELAY segment refers to the period after which the agent suddenly stops activity while the question is not completed until the moment when the agent continues with the question; (ii) the DELAY-RESPONSE segment refers to the period when the agent expresses empathy concerning the delay, or ignores the occurrence of the delay—which follows the agent's response (regarding the correctness of the answer) to the

[2] We also recorded subjects' blood volume pulse (BVP) signal from which the heart rate of subjects can be calculated. Unfortunately, the low reliability of our method used to gather the BVP signal precluded its consideration in the analysis.

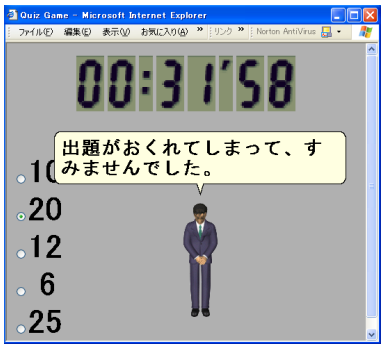

Fig. 1. Shima character: "I apologize that there was a delay in posing the question."

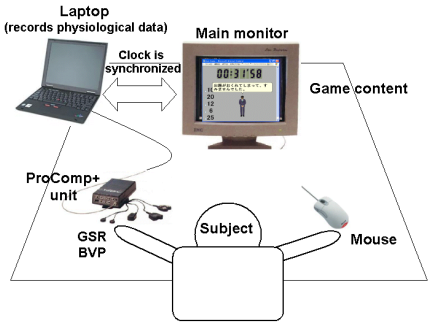

Fig. 2. Schematic of the experimental setup.

subject's answer; (iii) the RESPONSE segment refers to the agent's response to the subject's correct or wrong answer to the quiz question.

Procedure and Apparatus. The subjects were recruited directly by the experimenter and offered 1000 Yen for participation, and additionally 5000 Yen for the best score. Subjects have been randomly assigned to one of the two versions of the game. The experiment was conducted in Japanese, and lasted for about 25 minutes (15 minutes for game play, and 10 minutes for experimenter instructions, attaching the sensors, etc). Subjects came to the testing room individually and were seated in front of a computer display, keyboard, and mouse. After briefing the subjects about the experiment and asking them to sign the consent form, they were attached to galvanic skin response and blood volume pulse sensors on the first three fingers of their non-dominant hand (see Fig. 2).

Before subjects actually started to play the game, the character shows some quiz examples that explain the game. This period also serves to collect physiological data of subjects that are needed as a baseline to normalize data obtained during game play. In six out of a total of thirty quiz questions, a delay was inserted before showing the 5th number. The duration of delays was 6–14 secs. (9 secs. on average). While subjects played the game the experimenter remained in the room and monitored their physiological activity on a laptop computer. The experimenter and laptop were hidden from the view of the subjects. After the subjects completed the quiz, the sensors have been removed from their hand, and they were asked to fill out a short questionnaire, which contained questions about the difficulty and their impression of playing the game. Finally, subjects were told to keep checking a web page that will announce the best score.

The game was displayed on a 20 inch color monitor, running Internet Explorer with browsing buttons deactivated. The Microsoft Agent package [8] was used to control character animations and synthetic speech. Two flat speakers produced the sound. Physiological signals have been recorded with the ProComp+ unit and visualized with BioGraph2.1 software (both from Thought Technology Ltd. [13]).

2.3 Results

The first observation relates to the use of delays in order to induce stress in subjects. All eighteen subjects showed a significant rise of skin conductance in the DELAY segment, indicating an increased level of arousal. The data of two subjects of the non-affective version were discarded because of extremely deviant values. In the following, the confidence level α is set to 0.05.

The general hypothesis about the positive effect of life-like characters with affective behavior on a subjective measure, here the users' stress level, can be divided into two specific hypotheses (*Empathy* and *Affective Feedback*).

- *Empathy* Hypothesis: Skin conductance (stress) is lower when the character shows empathy after a delay occurred, than when the character does not show empathy.
- *Affective Feedback* Hypothesis: When the character tells whether the subject's answer is right or wrong, skin conductance is lower in the affective version than in the non-affective version.

To support the Empathy Hypothesis, the differences between the mean values of the GSR signal (in micro-Siemens) in the DELAY and DELAY-RESPONSE segments have been calculated for each subject. In the non-affective version (no display of empathy), the difference is even negative (mean $= -0.08$). In the affective version (display of empathy), GSR decreases when the character responds to the user (mean $= 0.14$). The t-test (two-tailed, assuming unequal variances) showed a significant effect of the character's emphatic behavior as opposed to non-affective behavior ($t(16) = -2.47$; $p = 0.025$). This result suggests that an animated agent expressing empathy may undo some of the frustration (or reduce stress) caused by a deficiency of the interface.

The Affective Feedback Hypothesis compares the means of GSR values of the RESPONSE segments for both versions of the game. Note that the character responses of all queries, not only the queries affected by a delay, are considered here. However, the t-test showed no significant effect ($t(16) = 1.75$; $p = 0.099$). When responding to the subject's answer, affective behavior of the character has seemingly no major impact on subjects' skin conductance.

In line with the study of van Mulken et al. [14] who show that embodied interface agents have no significant effect on comprehension and recall, we expected that affective life-like characters do not influence objective performance measures. Our expectation was confirmed as the average score in the affective version was 28.5 (from 30 answers), and 28.4 in the non-affective version.

In addition to taking physiological data of subjects, they were asked to fill out a short questionnaire. Table 1 shows the mean scores for some questions. None of the differences in rating reached the level of significance. Only the scores for the first question suggest a tendency about the subjects' impression of the difficulty of the game ($t(17) = 1.74$; $p = 0.1$). This result can be compared to the findings of Mulken et al. [14], which show that a character may influence the subjects' perception of difficulty. In their experiment though, van Mulken and coworkers compare "persona" vs. "no persona" conditions rather than "affective persona" vs. "non-affective persona" conditions.

Table 1. Mean scores for questions about interaction experience in non-affective (*NA*) and affective (*A*) game version. Ratings range from 1 (disagreement) to 10 (agreement).

Question	*NA*	*A*
I experienced the quiz as difficult.	7.5	5.4
I was frustrated with the delays.	5.2	4.2
I enjoyed playing the quiz game.	6.6	7.2

3 The Emotion Mirror – Future Work

This section briefly describes the Emotion Mirror, a character-based application aimed at training interpersonal communication skills known as *emotional intelligence* [1], specifically the abilities to be aware of and to regulate one's emotions. A job interview situation is one example where emotional intelligence is beneficial, as the interviewee has to manage his or her emotions when confronted with unpleasant and probing questions of the interviewer.[3] Since physiological manifestations of stress may reflect negatively on the interviewer's impression of the interviewee, a virtual job interview alerting the user (as interviewee) about his or her arousal level might serve as a valuable preparatory training environment. The Emotion Mirror application assumes that users are biased to conceive life-like characters as veritable social actors (the 'Media Equation' [11]), and hence *actually* get aroused when interviewed by a virtual agent.[4]

The job interview scenario features two life-like characters, the interviewer to the left and the 'Mirror Agent' to the right (see Fig. 3). Users in the role of interviewees are attached to the sensors of the ProComp+ device. As in the study described above, we currently take the galvanic skin response (GSR) signal only. Unlike the implementation of the experiment, however, the Emotion Mirror application requires to process physiological data in real-time. This was achieved by using Visual C++ and the ProComp+ data capture library, with the Active Template Library (ATL) as an interface to the JavaScript code and the Microsoft Agent controls [8] that drive the agents' animation and speech engines.

The baseline for subsequent bio-signal changes is obtained during an initial relaxation period of 40 secs. where the user listens to music from Cafe del Mar (Vol. 9), as the average of GSR values. An interview episode consists of four segments: (i) The interviewer character asks a question; (ii) the user selects an answer from a set of given options (the lower part in Fig. 3); (iii) the interviewer responds to the user's answer; (iv) the Mirror Agent displays the user's arousal level calculated from the data gathered during segments (i)–(iii). More precisely, we take values every 50 msec., for a period of 5 secs. The psychophysiological literature, e.g., Levenson [6, p. 30], suggests 0.5–4 secs. as an approximation for the duration of an emotion.

[3] A job interview scenario featuring an 'affective mirror' has been suggested by Picard [9, p. 86], but to our knowledge, it was never implemented.

[4] It is certainly true that an online interview cannot induce the stress level of a face-to-face or phone interview.

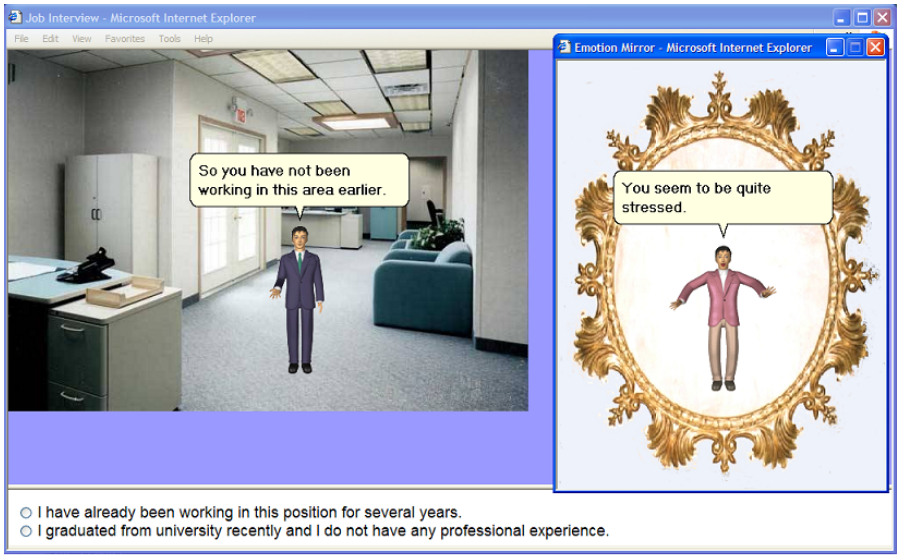

Fig. 3. Job Interview Scenario with Emotion Mirror.

When the (average) GSR signal is 15–30% above the baseline, the user's arousal level is assumed as 'high'. If the signal is on average higher than 30%, the user is assumed to be very aroused, and the Mirror Agent would display a gesture expressing anxiety and utter, e.g.,"You seem to be quite stressed". The Mirror Agent reflects the arousal state of the user in a rather exaggerated way, in order to alert the user of his or her presumed impression on the interviewer.

While initial experiences with the Emotion Mirror are promising—people *do* get aroused for some questions—the Mirror Agent's reaction is still too unspecific about the user's emotion. We currently implement the electrocardiogram (ECG) rhythm trace as an indicator of heart rate, which together with skin conductance will allow to identify named emotions in Lang's [4] two-dimensional model. We also plan to use a Bayesian network to combine different sensor data and to account for the uncertainty of the domain.

4 Conclusions

This paper proposes a new approach to the persona effect, a phenomenon that describes the positive effect of life-like interface agents on human users. In addition to questionnaires, we suggest to utilize bio-signals in order to show the persona effect in terms of users' arousal level. While user stress could not be directly deduced from the sensor data, the design of the experiment suggests this interpretation. The main results of the empirical study discussed in the paper are: (i) a character displaying empathy may significantly decrease user stress; (ii) the character's affective behavior has no impact on users' performance in a

simple mathematical game; but (iii) it has a (almost significant) positive effect on the users' perception of task difficulty.

The Emotion Mirror application offers a direct interpretation of the persona effect, by reflecting the user's arousal level in a job interview scenario to the user in real-time. This concept is already under consideration for e-Healthcare systems [7]. We also hope that character-based interfaces with emotion recognition capability will prove useful for social skills training and software testing.

Acknowledgments. This research is supported by the JSPS Research Grant (1999-2003) for the Future Program. The authors would like to thank Naoaki Okazaki for his generous help with implementing the ATL interface.

References

1. D. Goleman. *Emotional Intelligence.* Bantam Books, New York, 1995.
2. J. A. Healey. *Wearable and Automotive Systems for Affect Recognition from Physiology.* PhD thesis, Massachusetts Institute of Technology, 2000.
3. J. Klein, Y. Moon, and R. Picard. This computer responds to user frustration: Theory, design, and results. *Interacting with Computers*, 14:119–140, 2002.
4. P. J. Lang. The emotion probe: Studies of motivation and attention. *American Psychologist*, 50(5):372–385, 1995.
5. J. C. Lester, S. A. Converse, S. E. Kahler, S. T. Barlow, B. A. Stone, and R. S. Bhogal. The Persona effect: Affective impact of animated pedagogical agents. In *Proceedings of CHI-97*, pages 359–366. ACM Press, 1997.
6. R. W. Levenson. Emotion and the autonomic nervous system: A prospectus for research on autonomic specificity. In H. L. Wagner, editor, *Social Psychophysiology and Emotion: Theory and Clinical Applications*, pages 17–42. John Wiley & Sons, Hoboken, NJ, 1988.
7. C. Lisetti, F. Nasoz, C. LeRouge, O. Ozyer, and K. Alvarez. Developing multimodal intelligent affective interfaces for tele-home health care, 2003. International Journal of Human-Computer Studies. To appear.
8. Microsoft. *Developing for Microsoft Agent.* Microsoft Press, 1998.
9. R. W. Picard. *Affective Computing.* The MIT Press, 1997.
10. H. Prendinger and M. Ishizuka, editors. *Life-like Characters. Tools, Affective Functions and Applications.* Cognitive Technologies. Springer Verlag, 2003. To appear.
11. B. Reeves and C. Nass. *The Media Equation. How People Treat Computers, Television and New Media Like Real People and Places.* CSLI Publications, Center for the Study of Language and Information. Cambridge University Press, 1998.
12. J. Schreirer, R. Fernandez, J. Klein, and R. W. Picard. Frustrating the user on purpose: A step toward building an affective computer. *Interacting with Computers*, 14:93–118, 2002.
13. Thought Technology Ltd. URL: `http://www.thoughttechnology.com`.
14. S. van Mulken, E. André, and J. Müller. The Persona Effect: How substantial is it? In *Proceedings Human Computer Interaction (HCI-98)*, pages 53–66, Berlin, 1998. Springer.