# Using Bio-signals to Track the Effects of a Character-based Interface

**Junichiro Mori, Helmut Prendinger, Sonja Mayer, Hiroshi Dohi, and Mitsuru Ishizuka**
Department of Information and Communication Engineering
Graduate School of Information Science and Technology
University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
{jmori,dohi,ishizuka}@miv.t.u-tokyo.ac.jp
prendinger@acm.org SonjaCMayer@web.de

**Abstract**

The aim of the experimental study described in this paper is to show the impact of a character-based interface on users' emotional state which is derived from physiological signals of the user. Our results suggest that an embodied interface agent with emphatic behavior may significantly decrease user frustration and stress. Other findings are comparable to the 'Persona effect' of previously performed studies.

**Keywords.** Empirical methods to observe user behavior, embodied animated agents, empirical and other evaluation methods

## 1 Introduction

Interacting with computers is often responsible for negative emotional states of the user, such as frustration or stress. One way to alleviate the intensity of user frustration is to provide appropriate feedback. As people tend to respond to computers in an essentially natural way (Reeves & Nass, 1998), we suggest using an interface agent that gives affective feedback including the expression of empathy. In order to measure the effect of the agent's response on user emotions, we take physiological signals from the user.

Our work follows up to recent studies in the *Affective Computing* paradigm (Picard, 1997) that suggest employing bio-signals to detect user emotions (Schreirer et al., 2002) and affective feedback to reduce – deliberately induced – user frustration (Klein et al., 2002). By contrast, we employ an embodied interface agent rather than a text-based interface to communicate with the user. This design choice may also shed new light on the *Persona effect*, which relates (among others) to the credibility and motivation enhancing effects of character-based interfaces (Lester et al., 1997).

Unlike standard evaluation methods such as questionnaires, the use of physiological data may support a more accurate assessment of the affective state of users. In particular, the recorded history of users' bio-signals allows to precisely relate emotion occurrence with the (user-computer) interaction state. Furthermore, using both bio-signals and questionnaires enables to detect possible discrepancies between the interaction as perceived by the user and the user's physiological state.

## 2 Design of the Experiment

We implemented a simple mathematical quiz game where subjects are instructed to sum up five successively displayed numbers and are then asked to subtract the $i$-th number of the sequence ($i \leq 4$). Subjects compete for the best score in terms of correct answers and time. They received 1000 Yen for participation, and additionally 5000 Yen for the best score. Subjects were told that they would interact with a prototype interface that might still contain some bugs. Before game start, the "Shima" character shows some quiz examples that explain the game. This period also serves to collect physiological data of subjects that are needed in order to normalize data obtained during game play.

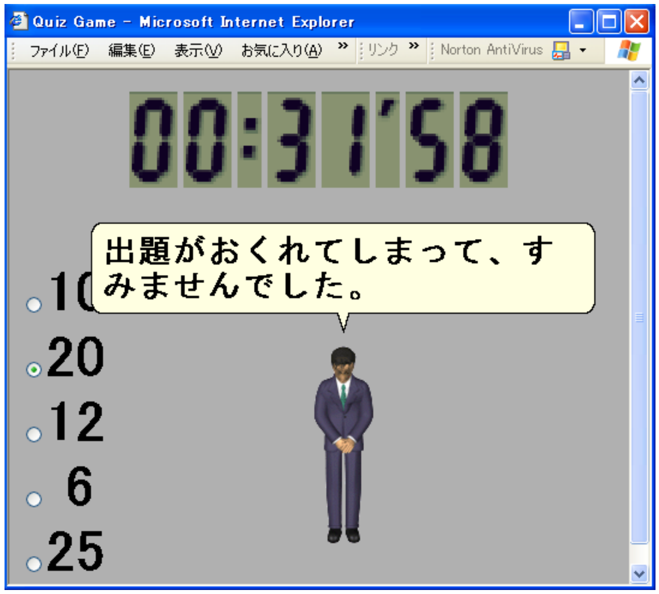


Figure 1: The "Shima" character apologizes.

In six out of a total of thirty quiz questions, a delay was inserted before showing the 5th number. The delay, between 6 and 14 sec. (9 sec. on average), is assumed to induce frustration as the subjects' goals of giving the correct answer and achieving a fast score are thwarted, called 'primary frustration' in behavioral psychology (Lawson, 1965).

In the experiment, subjects were twenty male students, all of them native speakers of Japanese. We randomly assigned subjects to one of two versions of the game (ten in each version).

- *Affective version.* Depending on whether the subject selects the correct or wrong answer from the menu displayed in the game window, the agent expresses 'happy for' (correct answer) and 'sorry for' (wrong answer) emotions both verbally and nonverbally (by "smiling", "hanging shoulders", "sad face", etc).

  If a delay in the game play happens, the agent expresses empathy for the user after the subject answers the question that was affected by the delay.

- *Non-affective version.* The agent does not give any affective feedback to the subjects. It simply replies "right" or "wrong" to the user's answer and does not comment on the occurrence of the delay.

Fig. 1 shows the agent expressing empathy to the user since a delay occurred. The agent displays a gesture that Japanese people perceive as a signal of the interlocutor's apology, and says: "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.

The "Shima" character is driven by the Microsoft Agent package (Microsoft, 1998) that allows to embed animated agents into a web page based JavaScript interface, and provides controls for 2D animation and synthetic speech. In order to make the character appear more life-like when no particular animation is programmed (e.g., while the subject is thinking about the correct answer), the animation engine randomly triggers 'idle behaviors' such as "blinking" or small movements of the body.

Subjects are attached to two types of sensors on the first three fingers of their non-dominant hand (see Fig. 2) that measure skin conductivity (SC) and heart rate (HR). Signals are recorded with the ProComp+ unit and visualized with BioGraph2.1 software (both available from (Thought Technology Ltd., 2003)).

- The galvanic skin response (GSR) signal is an indicator of SC. It has been shown that SC varies linearly with the overall level of arousal and increases with anxiety and stress (see the discussions in Picard (1997, p. 162) and Healey (2000, p. 25, 40)).

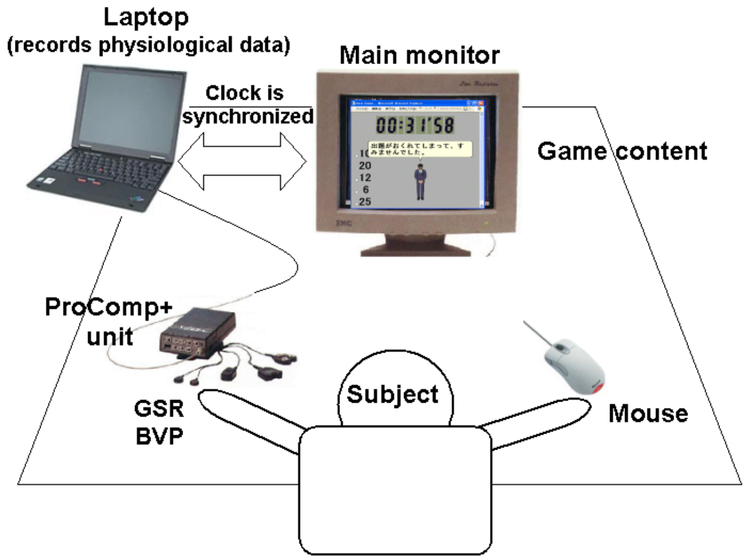


Figure 2: Experimental setup.

- The blood volume pressure (BVP) signal is an indicator of blood flow. (HR was automatically calculated from BVP with our software.) BVP increases with negatively valenced emotions such as fear and anxiety, and decreases with relaxation (Picard (1997, p. 162) and Healey (2000, p. 27)).

In order to obtain named emotions from signals, SC and HR can be mapped to the emotion model of (Lang, 1995) which shows that emotions can be located as coordinates of affective valence and arousal in a two-dimensional space. In our experiment, however, BVP data could be taken reliably in only six out of twenty cases. The BVP sensor is very sensitive to movement, which is the most likely cause for the noisy data. In effect, we could neither use Lang's model nor other, more sophisticated mappings from physiological signals to emotional states, as proposed, e.g., by Lisetti and coworkers (2003).

In order to show the effect of the agent's behavior, we have been interested in three specific segments of the game sequence (see Fig. 3):

- 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.
- 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 subject's answer.
- The RESPONSE segment refers to the agent's response to the subject's correct or wrong answer to the quiz question.

## 3 Results of the Experiment

The first observation relates to the use of delays in order to induce frustration and stress in subjects. All eighteen subjects showed a significant rise of SC in the DELAY segment, indicating an increased level of arousal.

Our general hypothesis about the positive effect of embodied agents with affective behavior on users can be divided into three specific hypotheses.

- Hypothesis 1 (*Empathy*): SC is lower when the agent shows empathy after a delay occurred, than when the agent does not show empathy.
- Hypothesis 2 (*Affective feedback*): When the agent tells whether the subject's answer is right or wrong, SC is lower in the affective version than in the non-affective version.
- Hypothesis 3 (*Score*): Subjects interacting with the affective version do not score significantly better (or worse) in the game than subjects interacting with the non-affective version.

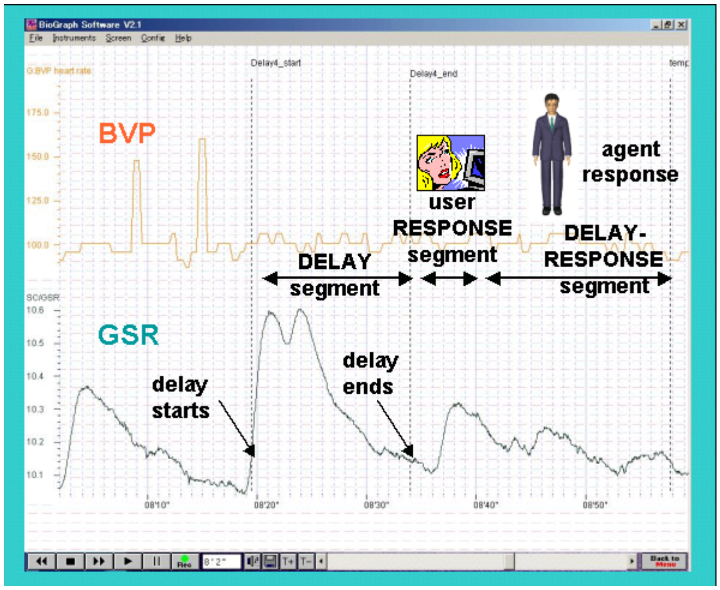


Figure 3: The three relevant segments of the game flow (DELAY, RESPONSE, DELAY-RESPONSE), exemplified by the bio-signals of one user interacting with the affective version of the game.

To test Hypothesis 1 (Empathy), we calculated the differences between the mean values of SC in the DELAY and DELAY-RESPONSE segments for each subject, by subtracting the means of the DELAY-RESPONSE segment from the DELAY segment.[1] In the non-affective version (no display of empathy), the difference is even negative (mean $= -0.08$). In the affective version (display of empathy), SC decreases when the character responds to the user (mean $= 0.14$). In the following, the $\alpha$ level is set to 0.05. The $t$-test (two-tailed, assuming unequal variances) showed a significant effect of the character's affective (emphatic) behavior as opposed to non-affective behavior ($t(16) = -2.47$; $p = 0.025$). This result suggests that an embodied agent expressing empathy may undo some of the frustration (or reduce stress) caused by a deficiency of the interface.

Hypothesis 2 (Affective feedback) compares the means of SC values of the RESPONSE segments for both versions of the game (the agent responses of all queries 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, the agent's affective behavior has seemingly no major impact on subjects' skin conductance.

Hypothesis 3 (Score) could be supported in the experiment. The average score in the affective version was 28.5 (from 30 answers), and 28.4 in the non-affective version. This result is in accordance with our expectation that (affective) life-like characters do not influence objective performance measures. In a comparable study, van Mulken and colleagues (1998) show that embodied interface agents have no significant effect on comprehension and recall. Considering the high scores achieved by subjects in both versions, another reason might be that the mathematical task was too simple, so that the character's behavior had no effect on game performance.

In addition to taking subjects' physiological data we asked subjects to fill out a short questionnaire after they completed the quiz. 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 ($t(17) = 1.74$; $p = 0.1$) somewhat related to the one observed in (van Mulken et al., 1998), namely, that a character may influence the subjects' *perception* of difficulty. This indicates that affective behavior influences the subjects' impression 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.

The scores for the second question indicate that subjects underestimate the extent to which they

[1]The data of two subjects of the non-affective version were discarded because of extremely deviant values.

Table 1: Mean scores for questions about interaction experience in affective ($A$) and non-affective ($NA$) 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 |

were frustrated in both versions of the game. Since the GSR signal significantly increased during the delay period, subjects were obviously frustrated during those periods which is not reflected in their answer to this question (non-extreme scores) in the questionnaire. This highlights the importance of using a more objective evaluation method, such as physiological user data assessment, which may detect user experiences that can hardly be revealed by using only questionnaires. Furthermore, bio-signal assessment is not affected by a well-known problem of the standard questionnaire method, namely that subjects answer the way they believe the experimenter expects them to answer.

Although the obtained results are still somewhat restricted, we believe that animated agents with affective behavior have the potential to alleviate user frustration similar to human interlocutors, and the assessment of user's physiological data is an adequate method to show the effects of agents.

# 4 Current and Future Work

We currently extend our work on emotion recognition and character-based interfaces along two main lines.

*First*, we want to process physiological data in real time and base the agent's behavior on the current emotional state of the user. We recently implemented a character-based job interview scenario featuring two embodied agents. One agent is in the role of an interviewer while a second agent reflects the emotional state of the user. The second agent may be used either as a 'mirror' of the user's current emotional state (see also (Lisetti et al., 2003)), or as a companion with a calming effect on the user. The great challenges of on-line emotion recognition include reliable bio-signal assessment of one-visit users and data synthesis. Besides the galvanic skin response signal, we employ the electrokardiogram (EKG) rhythm trace as an indicator of heart rate this time (Healey, 2000, pp. 109–111), rather than the blood volume pressure signal.

Our *second* goal is to develop an adaptive character-based interface that shows tailored agent reactions depending on more features of the interaction, such as user goals, personality, and interaction task (Conati, 2002; Hudlicka & McNeese, 2002).

# 5 Conclusions

Although life-like characters, or animated agents, are increasingly used in numerous computer-based applications (Prendinger & Ishizuka, 2003), their impact on human users is still largely unexplored. The aim of the experiment described in this paper was to investigate the impact of a character-based interface on the emotional state of users. The novel aspects of this investigation were (i) the use of an embodied character to address (intentionally induced) user frustration, and (ii) the utilization of physiological user data to track the effects of the interface on user emotions.

The main results of this empirical study can be summarized as follows:

- A life-like character verbally and nonverbally displaying empathy may significantly decrease user frustration and stress.

- Affective behavior of a life-like character has no impact on users' performance in playing a mathematical quiz game.

- A life-like character with affective behavior (including the display of empathy) has a positive effect on the users' perception of the task difficulty.

While the current paper focussed on bio-signals of users as an *evaluation method* for the impact of (character-based) interfaces, complementing the standard questionnaire evaluation method, we recently started to process users' physiological data in real-time, in order to *reflect* as well as *adapt to* the emotional state of computer users.

## Acknowledgements

This research was supported by the Research Grant (1999–2003) for the Future Program ("Mirai Kaitaku") from the Japanese Society for the Promotion of Science (JSPS). The third author was supported by a internship fellowship from the Carl Duisberg Society (Germany).

## References

Conati, C. (2002). Probabilistic assessment of user's emotions in educational games. *Applied Artificial Intelligence*, *16*, 555–575.

Healey, J. A. (2000). *Wearable and Automotive Systems for Affect Recognition from Physiology*. PhD thesis, Massachusetts Institute of Technology.

Hudlicka, E. & McNeese, M. D. (2002). Assessment of user affective and belief states for interface adaption: Application to an Air Force pilot task. *User Modeling and User-Adapted Interaction*, *12*, 1–47.

Klein, J., Moon, Y. & Picard, R. (2002). This computer responds to user frustration: Theory, design, and results. *Interacting with Computers*, *14*, 119–140.

Lang, P. J. (1995). The emotion probe: Studies of motivation and attention. *American Psychologist*, *50(5)*, 372–385.

Lawson, R. (1965). *Frustration: The Development of a Scientific Concept*. New York: MacMillan.

Lester, J. C., Converse, S. A., Kahler, S. E., Barlow, S. T., Stone, B. A. & Bhogal, R. S. (1997). The Persona effect: Affective impact of animated pedagogical agents. In *Proceedings of CHI-97* (pp. 359–366). ACM Press.

Lisetti, C., Nasoz, F., LeRouge, C., Ozyer, O. & Alvarez, K. (2003). Developing multimodal intelligent affective interfaces for tele-home health care. International Journal of Human-Computer Studies. To appear.

Microsoft (1998). *Developing for Microsoft Agent*. Microsoft Press.

van Mulken, S., André, E. & Müller, J. (1998). The Persona Effect: How substantial is it? In *Proceedings Human Computer Interaction (HCI-98)* (pp. 53–66). Berlin: Springer.

Picard, R. W. (1997). *Affective Computing*. The MIT Press.

Prendinger, H. & Ishizuka, M. (Eds.). (2003). *Life-like Characters. Tools, Affective Functions and Applications*. Cognitive Technologies. Springer Verlag. To appear.

Reeves, B. & Nass, C. (1998). *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.

Schreirer, J., Fernandez, R., Klein, J. & Picard, R. W. (2002). Frustrating the user on purpose: A step toward building an affective computer. *Interacting with Computers*, *14*, 93–118.

Thought Technology Ltd. (2003). URL: `http://www.thoughttechnology.com`.