Available online at www.sciencedirect.com
SCIENCE DIRECT
Int. J. Human-Computer Studies 62 (2005) 231–245
International Journal of Human-Computer Studies
www.elsevier.com/locate/ijhcs

Using human physiology to evaluate subtle expressivity of a virtual quizmaster in a mathematical game ☆

Helmut Prendinger[a,*], Junichiro Mori[b], Mitsuru Ishizuka[b]

[a]*National Institute of Informatics (NII), 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan*
[b]*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*

Available online 31 December 2004

Abstract

The aim of the experimental study described in this article is to investigate the effect of a life-like character with subtle expressivity on the affective state of users. The character acts as a quizmaster in the context of a mathematical game. This application was chosen as a simple, and for the sake of the experiment, highly controllable, instance of human–computer interfaces and software. Subtle expressivity refers to the character's affective response to the user's performance by emulating multimodal human–human communicative behavior such as different body gestures and varying linguistic style. The impact of em-pathic behavior, which is a special form of affective response, is examined by deliberately frustrating the user during the game progress. There are two novel aspects in this investigation. *First,* we employ an animated interface agent to address the affective state of users rather than a text-based interface, which has been used in related research. *Second,* while previous empirical studies rely on questionnaires to evaluate the effect of life-like characters, we utilize physiological information of users (in addition to questionnaire data) in order to precisely associate the occurrence of interface events with users' autonomic nervous system activity. The results of our study indicate that empathic character response can significantly decrease user stress

☆This article is a significantly revised and extended version of Prendinger et al. (2003).

*Corresponding author. Tel.: +81 3 4212 2650; fax: +81 3 3556 1916.
E-mail addresses: helmut@nii.ac.jp (H. Prendinger), jmori@miv.t.u-tokyo.ac.jp (J. Mori), ishizuka@miv.t.u-tokyo.ac.jp (M. Ishizuka).

1071-5819/$ - see front matter
doi:10.1016/j.ijhcs.2004.11.009

and that affective behavior may have a positive effect on users' perception of the difficulty of a task.

Keywords: Life-like characters; Affective behavior; Empathy; Physiological user information; Evaluation

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 have been shown to respond to computers in an essentially natural way (Reeves and Nass, 1998), we suggest to employ an animated interface agent endowed with subtle expressivity, i.e. affective response in the form of emotional display and the expression of empathy. The agent's expression of affect is communicated by means of linguistic style such as polite speech and conversational gestures including bodily signals to convey sorriness or an apology. In psychology, it argued that humans are affective beings who want their emotional and social needs to be addressed (Myers, 1989). The present study about human–computer interaction will investigate whether animated agents with affective behavior can meet those needs of users (Prendinger et al., 2003).

While animated agents, or life-like characters, start populating the interfaces of numerous computer-based applications (Prendinger and Ishizuka, 2004), their impact on human users is still largely unexplored, or at least very general in its formulation. In this article, we will propose to use affective information about users derived from their physiological signals as an evaluation method. Physiological sensing is becoming an increasingly important technology to monitor users for systems design and evaluation (Allanson, 2000) and to detect emotions from bio-signal changes (Picard, 1997). In our study, the recorded history of users' physiological readings will enable to precisely relate their affective state (emotion) with the user–computer interaction state, and hence track the impact of character behavior. Moreover, 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 rest of this article is organized as follows. The next section reports on related work in using affect in human–computer interaction, animated interface agents, and physiological sensing as an evaluation method. Section 3 introduces the agent we used as a quizmaster for the mathematical game. Sections 4 and 5 detail theory, design, and method of our study, and Section 6 describes its results. In Section 7, some limitations of the experiment are explained. Section 8 concludes the article.

2. Related work

The work presented in this article is strongly influenced by the affective computing paradigm in human–computer interaction (see Picard (2000), Hudlicka (2003) for excellent surveys). Specifically, our approach shares its motivation with the work of Klein et al. (2002) who describe the design and evaluation of an interface implementing strategies aimed at reducing negative affect, such as active listening, empathy, sympathy, and venting. The resulting affect-support agent used in a simulated network game scenario could be shown to undo some of the users' negative feelings after they have been deliberately frustrated by simulated network delays inserted to the course of the game. Our interface differs from the one used in Klein et al. (2002) in the following aspects. *First*, the user is given feedback in a more timely fashion, i.e. shortly after the emotion actually occurs, and not after the interaction session, in response to the user's questionnaire entries. *Second*, affective feedback to the user is communicated by means of a life-like character rather than a text message. Although the study of Klein et al. (2002) supports the argument that embodiment is not necessary to achieve social response (effecting frustration-reduction), it has been shown that embodied characters may boost the tendency of people to interact with computers in a social way, the so-called 'persona effect' (Lester et al., 1997; van Mulken et al., 1998).

In the context of educational software, Lester et al. (1997) 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. (1998) conducted a follow-up study to the investigation of Lester et al. (1997) where a life-like character, the PPP Persona (Andrè et al., 1996), acts as a presenter of technical and non-technical information. In their experiment, a positive effect of the 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. Our work differs from those studies on the persona effect in that we compare "affective persona" vs. "non-affective persona" conditions rather than "persona" vs. "no persona" conditions. Furthermore, 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.

Our work slightly touches on the debate of "direct manipulation vs. interface agents" (Shneiderman and Maes, 1997) in that we propose an anthropomorphic component in the interface, which Shneiderman believes to deceive the users and interferes with the predictability of the interface. He also argues that interface designers should be concerned with user performance rather than properties such as naturalness, friendliness, or intuitiveness. It seems to us, however, that Shneiderman's arguments do not affect our aim of designing empathic embodied interfaces. Animated agents in our approach are used to communicate with the user, as mediators between user and application. If they assume some functionality that is part of the application, it can easily be replaced by other means, such as a textual message. What makes embodied interfaces attractive as a user interface paradigm is their capability to address user emotions in a natural way, with the potential of leaving users less frustrated and more motivated.

The use of physiological information as a method to evaluate human–computer interaction has hitherto not attracted much attention. Previous studies, however, have shown interesting and promising results regarding software usability. In order to assess user cost of different levels of multimedia quality in videoconferencing tools, Wilson and Sasse (2000) monitored users' galvanic skin response, heart rate, and blood volume pulse for different frame rates. While the majority of subjects show an increase in those physiological responses—indicating stress—for low frame rates, interestingly only a small fraction of subjects noticed that the frame rate had changed. In the study of Ward and Marsden (2003), physiological responses of users are taken while performing a task with well-designed vs. poorly designed web pages. The latter are characterized by non-intuitive organization, many pull-down menus, and many pop-ups. By analysing users' skin conductance, heart rate, and finger blood volume, the authors demonstrate that the poorly designed web pages cause more stress. In order to assess task engagement, cognitive load, and stress, physiological monitoring together with eye-tracking and video information has also been extensively studied in NASA systems, e.g. to investigate whether dynamic reconfiguration of the interface can counteract to negative affective states (Pope et al., 1995).

An issue of using bio-sensors is the adequate interpretation of its data as affective states, such as emotions. The most common view is that physiological signals are influenced by two independent dimensions of experience, valence and arousal (Lang, 1995; Feldman–Barrett and Russell, 1999). Valence refers to the degree of pleasantness, and arousal refers to the degree of activation. According to the model of Lang (1995), specific emotions can then be derived from their location in the arousal-valence space. Those two dimensions have also been employed to classify media experiences (Reeves and Nass, 1998). Besides physiological measurement, more direct methods exist, such as (non-verbal) self-report instruments. In the proposal of Desmet et al. (2000), an emotion can be reported by clicking an interface button that graphically depicts one of a variety of different emotions.

For the purpose of our experiment, however, physiological sensor reading has clear advantages. It can be done without interrupting the subject during the interaction, and the method does not rely on the subject's memory of the (interaction) experience. Another reason is that the physiological activity is unconscious and thus deception on the part of the subject is negligible.[1]

3. The virtual quizmaster

The quizmaster of the mathematical game used in our study is a 2D cartoon-style animated agent called "Shima" that was developed by Hottolink (2002) (see Fig. 1).

[1]As known from the area of biofeedback, under certain circumstances, humans *are* able to exert conscious influence over their autonomic nervous system activity (Allanson, 2000).

Fig. 1. "Shima".

The Shima character is designed with gestures of a typical Japanese professional and hence the character's actions are familiar to Japanese people and easily understood. For instance, one animation shows the agent bowing, a gesture which Japanese people readily perceive as a signal of the interlocutor's apology, another animation displays the character with a 'smiling' facial gesture. The experimental work started in McBreen et al. (2000) suggests that synthetic characters are more trustable if their appearance 'matches' to their role in the application. Hence, as a quizmaster has a responsible role, a more formal looking character seemed appropriate. Moreover, since all of our subjects were Japanese, we wanted to employ an agent depicting a Japanese character.

The animated agent is controlled by the Microsoft Agent package (Microsoft, 1998) that provides the following features: (i) controls for facial and body gestures, called "animations" (about sixty); (ii) a Text-to-Speech engine for synthetic speech which is also textually displayed in a balloon adjacent to the character; (iii) a limited form of synchronization between speech and gestures, such as co-occurrence (or overlapping) of speech output and gesture. While more realistic anthropomorphic characters exist, including agents based on 3D models and agent architectures that allow for accurate synchronization of speech and gesture generation (see the proposals in (Prendinger and Ishizuka, 2004)), the rather simple appearance of our character can be shown as sufficient to have users "suspend disbelief" and respond in a social way. However, it might be the case that more sophisticated animations would increase the impact of the character. Notably, Bickmore and Picard (2003) propose a ('relational') animated agent based on the BEAT system (Cassell et al., 2004) that implements a large variety of subtle interactive behaviors, such as varying proximity, varying frequency of head-nods, eyebrow raises, and adjustable gaze behavior.

In the implementation of the quiz game, the actions of the character are specified by using MPML (Multimodal Presentation Markup Language), a scripting language that allows for easy handling of the verbal and non-verbal behavior of characters based on the Microsoft Agent package (Ishizuka et al., 2000). The character is embedded in a JavaScript-based web browser environment where its behavior can be triggered by the user's mouse input.

4. Theory and design

A simple mathematical quiz game has been implemented where subjects are instructed to sum up five consecutively displayed numbers and are then asked to subtract the *i*th number of the sequence ($i \leqslant 4$). The instruction is given by the Shima character. Subjects compete for the best score in terms of correct answers and time. Subjects were told that they would interact with a prototype interface that is not necessarily bug-free. 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, called 'primary frustration' in behavioral psychology (Lawson, 1965).

The mathematical application provides a simple instance of the class of interfaces where the user has to perform some task and may experience frustration due to failures caused by the system (or the user him- or herself). Since our goal is to investigate the effect of particular character behaviors in response to the same type of easily repeatable interface events (e.g. delays) on the affective state of users, we chose not to use a more realistic, i.e. more complicated but presumably more interesting, interface as it would preclude tight experimental control. On the other hand, the mathematical task is potentially challenging and may induce different levels of stress on game players.

The primary hypothesis of our study can be formulated as follows:

> If a life-like character provides affective feedback to the user, it can effectively reduce user frustration and stress.

To our knowledge, this is the first investigation that explores the possibility of employing an affective character 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 (van Mulken et al., 1998) or provided a text-based response as a feedback medium to the (deliberately frustrated) user (Klein et al., 2002).

In order to measure user frustration and stress, two types of physiological user signals have been recorded, galvanic skin response (GSR) and blood volume pulse (BVP). The GSR signal is an indicator of skin conductance. It has been shown that skin conductance varies linearly with the overall level of arousal and increases with anxiety and stress (see Picard, 1997, p. 162; Healey, 2000, pp. 25, 40). The BVP signal is an indicator of blood flow. BVP increases with negatively valenced emotions such as fear and anxiety, and decreases with relaxation (see Picard, 1997, p. 162; Healey, 2000, p. 27).

The original motivation for using those signals was to obtain named emotions by employing the emotion model of Lang (1995), which argues that all emotions (joyful, fearful, enraged, etc.) can be located as coordinates of affective valence and arousal in a two-dimensional space. However, in the experiment, BVP data could be taken reliably in only six out of twenty cases, which effectively precluded mapping the signals onto Lang's dimensions.

5. Method

5.1. Subjects

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. Subjects were given a monetary award of 1000 Yen for participation. An additional amount of 5000 Yen was promised for the best score in the experiment. The experiment was conducted in Japanese, and lasted for about 25 min (15 min for game play, and 10 min for experimenter instructions, attaching the sensors, etc.). The subjects have been randomly assigned to one of the two versions of the game (affective and non-affective).

5.2. Design

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 incorrect answer from the menu displayed in the game window (see the numbers in Fig. 2), the character expresses 'happy for' and 'sorry for' emotions both verbally and non-verbally.
  When a delay in the game play happens, the character expresses empathy for the user after the subject answers the question that was affected by the delay. Note that the apology is given *after* the occurrence of the delay, immediately after the subject's answer, and not during the delay:
- *Non-affective version*: the character does not give any affective feedback to the subjects. It simply replies "right" or "incorrect" 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.

In the two versions of the game, the verbal and non-verbal behaviors of the character differ with respect to two dimensions of subtle expressivity, *linguistic style* and *gesture*. Regarding the first dimension, linguistic style, the non-affective version has the character utter "seikai" or "fu-seikai" (romanized Japanese for "correct" and "incorrect" in English). In the affective version, on the other hand, the character replies by saying "seikai-desu" and "fu-seikai-desu" which contain the postfix "desu". Attaching "desu" indicates a more formal usage of the Japanese language and gives a 'soft' and polite impression to the utterance. Voice parameters associated with different emotions (Murray and Arnott, 1995) have not been used in the experiment.

The second dimension of subtle expressivity concerns facial and body gestures. In the affective version, the character expresses a 'happy for' (the subject's correct answer) emotion by displaying a smiling face, or a 'sorry for' (the subject's incorrect answer) emotion by playing the "sad face" and "hanging shoulders" animations. By

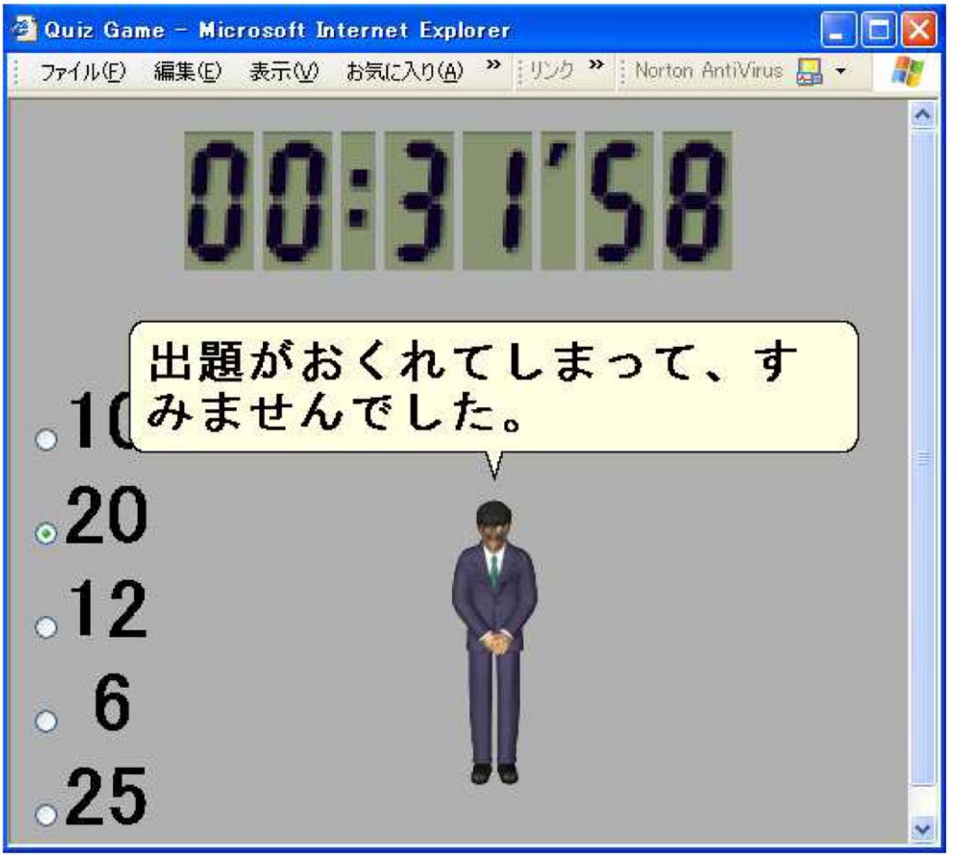

Fig. 2. The ''Shima'' character apologizes: ''I apologize that there was a delay in posing the question''.

contrast, in the non-affective version, the character does not show any affective gestures. If a delay in asking the query happens, the character expresses empathy to the subjects by verbal and non-verbal means. Fig. 2 depicts the character displaying a gesture that Japanese people perceive as a signal of the interlocutor's apology (bowing with both hands in the lap), and saying (in romanized Japanese): ''Shutsudai ga okurete shimatte sumimasen deshita''. The English translation is ''I apologize that there was a delay in posing the question''. The character's wording loosely follows the characterization of ''empathy'' given in Klein et al. (2002, p. 122): ''this is how/ feel *about* what you're going through. Here is *my emotional response* to *your* predicament.'' It is important to note that both the verb conjugations and the words used in the Japanese sentence convey the speaker's politeness and sorriness. Besides the speaker's expressed respect for the addressee, politeness levels in Japanese can be used to emphasize the speaker's emotional state, such as sorriness. In the non-affective version, the agent ignores the occurrence of the delay.

In order to show the effect of the character's behavior on the physiological state of subjects, we have been interested in three specific segments:

- The DELAY segment refers to the period after which the character suddenly stops activity while the question is not completed until the moment when the agent continues with the question.
- The RESPONSE-TO-DELAY segment refers to the period when the character expresses empathy concerning the delay, or ignores the occurrence of the delay, which follows the character's response (regarding the correctness of the answer) to the subject's answer.
- The RESPONSE-TO-QUERY segment refers to the character's response to the subject's correct or incorrect answer to the quiz question.

5.3. Apparatus

The game was displayed on a 20-in color monitor, running Internet Explorer with browsing buttons deactivated. Two flat speakers produced the sound. Physiological signals have been recorded with the ProComp+ unit and visualized with Bio-Graph2.1 software, both available from Thought Technology (2002).

5.4. Procedure

Subjects were recruited directly by the experimenter (the second author). They came to the testing room individually and were seated in front of computer display, keyboard, and mouse. After briefing the subjects about the experiment and asking them to sign the consent form, they were attached to GSR and BVP sensors on the first three fingers of their non-dominant hand (see Fig. 3).

Before subjects actually started to play the game, the character shows some quiz examples and explains 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 between 6 and 14 s (9 s 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. The clocks of the computer providing the game content and the laptop processing the physiological data of the subjects were initially synchronized and then the histories of both the game events and the incoming data were recorded. In this way, we can associate the game state to the subject's physiological state.

After the subjects completed the quiz, the sensors were 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 check a web page that will announce the best score.

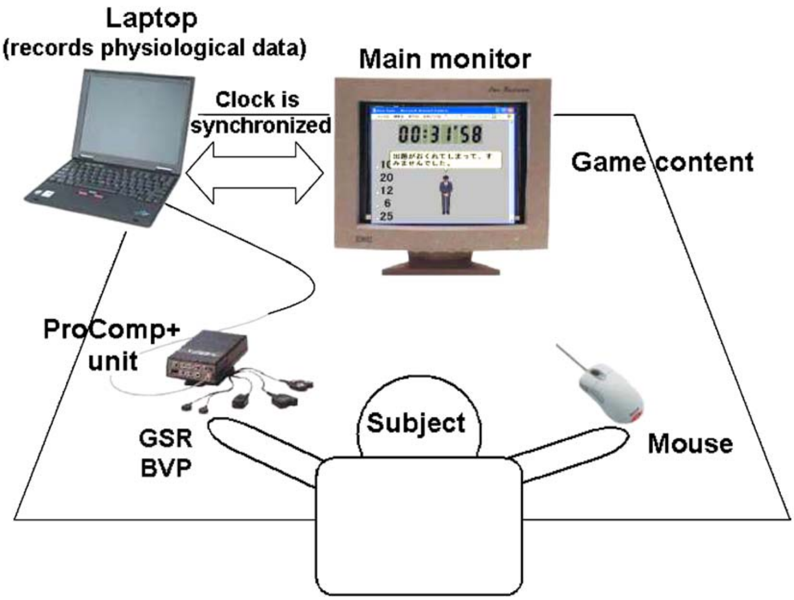

Fig. 3. Schematic of the experimental setup.

6. Results

In the following, the confidence level α is set to 0.05, i.e. we accept a risk of 0.05 of making a false decision as to reject or retain the null hypothesis.[2] The data of two subjects of the non-affective version were discarded because of extremely deviant values. The underlying assumption that delays actually induce stress in users, i.e. effect significantly higher SC levels, has been validated in a pre-experiment.

The general hypothesis about the positive effect of life-like characters with subtle expressivity on a subjective measures (the emotional state of users) can be divided into two specific hypotheses *(Empathy* and *Affective Feedback)*. The third hypothesis *(Score)* relates the affective behavior of the character to an objective measure:

- *Empathy hypothesis*: SC 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, SC is lower in the affective version than in the non-affective version.
- *Score hypothesis*: subjects interacting with the affective version score better in the game than subjects interacting with the non-affective version.

[2]The null hypothesis says that there is in fact no difference between the means of the two groups.

To support the empathy hypothesis, the differences between the mean values of the GSR signal in the DELAY and RESPONSE-TO-DELAY 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 empathic 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.

The affective feedback hypothesis compares the means of GSR values of the RESPONSE-TO-QUERY 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' level of arousal.

The score hypothesis could not be supported in the experiment. The average score in the affective version was 28.5 correct answers (out of 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 et al. (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 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 van Mulken et al. (1998), which show that a character may influence the subjects' perception of difficulty. Recall, however, that van Mulken and co-workers compare "persona" vs. "no-persona" conditions rather than "affective persona" vs. "non-affective persona" conditions. Although not statistically significant, subjects in the affective game version were on average less frustrated with delays and enjoyed the

Table 1
Mean scores for questions about interaction experience in non-affective (NA) and affective (*A*) game version

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

Ratings range from 1 (disagreement) to 10 (agreement).

game more. Subjects' enjoyment of the game suggests that they liked playing the game. In a related study, Brave (2003) investigates the "likeability" dimension (besides competence and trustworthiness) in a card game setting where an agent displays 'happy for' and 'sorry for' emotions depending on the user's performance. He could show that male users preferred the agent, which empathizes with them. Although we did not ask users to rate the quizmaster directly, both studies highlight the importance of interfaces that show concern for the users' affective state.

7. Discussion

While the obtained results are in favor of the claim that life-like characters with subtle expressivity have the potential to alleviate user frustration similar to human interlocutors, some limitations of the present study have to be considered.

First, the study heavily relies on the psychological significance of the galvanic skin response signal as a reliable indicator of user stress and frustration. Although a number of investigations claim that skin conductance positively correlates with arousal and stress (Levenson, 1988; Picard, 1997; Healey, 2000), the study of Pecchinenda and Smith (1996) explains the physiological activity manifested by SC as a result of users' situation appraisal. Specifically, they demonstrate that SC increases with higher levels of engagement in the task of solving a difficult problem-solving task. In the light of their study, the high level of SC during delay period could be interpreted in terms of a high level of users' engagement rather than stress. However, since SC significantly decreased only when the character expressed empathy afterwards (and not when the character ignored the delay), the explanation of relating SC with stress and frustration seems strongly supported in our experiment.

Second, the experiment only considers "affective character" vs. "non-affective character" conditions. In order to directly support the positive effect of an agent with subtle expressivity, a "no persona" condition has to be considered as well (van Mulken et al., 1998), which might be text-based as in Klein et al. (2002) and sub-categorized into non-affective vs. affective conditions. This type of experiment would allow to determine the respective effects of affective text-based and affective embodied interfaces on user emotions.

Third, the positive impact of a life-like character with affective behavior is confined to a situation where users' frustration with the delay is addressed. Affective response related to users' answers to the quiz questions did not show a significant effect. Hence, the question whether life-like characters with affective behavior positively influence user emotions (assessed by their physiological signals) for different types of human–agent interface experiences could not be answered in the present study. However, it can be speculated that affective feedback from a character is most important when users experience disruptive events, such as delays.

Finally, by computing only the mean values of the GSR signal in the specified segments, the analysis of physiological signals was fairly simple. No efforts have been made to distinguish or identify individual emotions, such as surprise, anger, or anxiety. This type of analysis would involve the consideration of signal change, duration, and timing issues (Levenson, 1988).

8. Conclusions

The aim of the experiment described in this article is to investigate the impact of an animated agent with subtle expressivity on the affective state of users playing a mathematical quiz game. The core ideas of this investigation are: (i) the use of an embodied character (rather than textual media) to address the user, and (ii) the utilization of physiological user information to track the effects of the interface on users' affective state. The main results of the study can be summarized as follows:

- A life-like character verbally and non-verbally expressing empathy may significantly decrease user frustration and stress.
- A life-like character with affective behavior (including the display of empathy) may have a positive effect on the users' perception of task difficulty.

In the current experiment, affective behavior of a life-like character could not be shown to have an impact on users' performance in playing the mathematical quiz game. This outcome is in line with comparable research but might also be attributed to the generally low difficulty level of the task.

While the obtained results are promising and a strong indicator of the benefits of employing embodied interface agents with subtle expressivity, further investigations are required. Most importantly, we intend to show that life-like characters are useful for a broader class of interface applications, in a variety of roles as interaction partners for the user, such as educational agents (Conati, 2002), medical advisors (Bickmore and Picard, 2003), and health companions (Lisetti et al., 2003). Furthermore, we plan to analyse physiological user data in interaction settings that are less tightly controlled than in the present study, which was designed to elicit a certain emotion (frustration) at specific moments. First steps in this direction have been taken by Ward and Marsden (2003) who investigate the impact of different web page designs of human physiology. Our initial experiences with a virtual job interview training application, the Empathic Companion, can be found in Prendinger et al. (2004). We hope to address those issues comprehensively in our future work.

Acknowledgements

We would like to thank the reviewers for their helpful remarks and valuable suggestions. This research was supported by the Research Grant (1999–2003) for the Future Program ("Mirai Kaitaku") from the Japan Society for the Promotion of Science (JSPS).

References

Allanson, J., 2000. Supporting the development of electrophysiologically interactive computer systems. Ph.D. Thesis, Lancaster University Computing Department.

Andrè, E., Miiller, J.W., Rist, T., 1996. The PPP Persona: a multipurpose animated presentation agent. In: Proceedings Advanced Visual Interfaces (AVI-96), ACM Press, New York, pp. 245–247.

Bickmore, T., Picard, R.W., 2003. Subtle expressivity by relational agents. In: Suzuki, N., Bartneck, Ch. (Eds.), Proceedings CHI 2003 Workshop on Subtle Expressivity for Characters and Robots, pp. 15–18.

Brave, S., 2003. User responses to emotion in embodied agents. In: Suzuki, N., Bartneck, Ch. (Eds.), Proceedings CHI 2003 Workshop on Subtle Expressivity for Characters and Robots, pp. 25–29.

Cassell, J., Vilhjalmsson, H., Bickmore, T., 2004. BEAT: the behavior expression animation toolkit. In: Prendinger, H., Ishizuka, M. (Eds.), Life-Like Characters. Tools, Affective Functions, and Applications, Cognitive Technologies Series. Springer, Berlin, Heidelberg, pp. 163–185.

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

Desmet, P.M.A., Hekkert, P., Jacobs, J.J., 2000. When a car makes you smile: Development and application of an instrument to measure product emotions. In: Hoch, S.J., Meyer, R.J. (Eds.), Advances in Consumer Research, vol. 27, pp. 111–117.

Feldman-Barrett, L., Russell, J.A., 1999. The structure of current affect: controversies and emerging consensus. Current Directions in Psychological Science 8 (1), 10–14.

Healey, J.A., 2000. Wearable and automotive systems for affect recognition from physiology. Ph.D. Thesis, Massachusetts Institute of Technology.

Hottolink Inc., 2002. http://www.hottolink.com.

Hudlicka, E., 2003. To feel or not to feel: the role of affect in human–computer interaction. International Journal of Human–Computer Studies 59, 1–32.

Ishizuka, M., Tsutsui, T., Saeyor, S., Dohi, H., Zong, Y., Prendinger, H., 2000. MPML: a multimodal presentation markup language with character control functions. In: Proceedings Agents' 2000 Workshop on Achieving Human-like Behavior in Interactive Animated Agents, pp. 50–54.

Klein, J., Moon, Y., Picard, R.W., 2002. This computer responds to user frustration: theory, design, and results. Interacting with Computers 14, 119–140.

Lang, R.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. MacMillan, New York.

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.

Levenson, R.W., 1988. Emotion and the autonomic nervous system: a prospectus for research on autonomic specificity. In: Wagner, H.L. (Ed.), Social Psy-chophysiology and Emotion: Theory and Clinical Applications. Wiley, Hoboken, NJ, pp. 17–42.

Lisetti, C., Nasoz, F., LeRouge, C., Ozyer, O., Alvarez, K., 2003. Developing mul-timodal intelligent affective interfaces for tele-home health care. International Journal of Human–Computer Studies 59 (1-2), 245–255.

McBreen, H., Shade, P., Jack, M., Wyard, P., 2000. Experimental assessment of the effectiveness of synthetic personae for multi-modal e-retail applications. In: Proceedings 4th International Conference on Autonomous Agents (Agents'2000), ACM Press, New York, pp. 39–45.

Microsoft Corp., 1998. Developing for Microsoft Agent. Microsoft Press.

Murray, I.R., Arnott, J.L., 1995. Implementation and testing of a system for producing emotion-by-rule in synthetic speech. Speech Communication 16, 369–390.

Myers, D.G., 1989. Psychology, second ed. Worth Publishers, New York.

Pecchinenda, A., Smith, C.A., 1996. The affective significance of skin conductance activity during a difficult problem-solving task. Cognition and Emotion 10 (5), 481–503.

Picard, R.W., 1997. Affective Computing. The MIT Press, Cambridge, MA.

Picard, R.W., 2000. Towards computers that recognize and respond to user emotions. IBM Systems Journal 39 (3–4), 705–719.

Pope, A.T., Bogart, E.H., Bartolome, D.S., 1995. Biocybernetic system evaluates indices of operator engagement in automated task. Biological Psychology 40 (1–2), 187–195.

Prendinger, H., Ishizuka, M. (Eds.), 2004, Life-Like Characters. Tools, Affective Functions, and Applications, Cognitive Technologies Series. Springer, Berlin, Heidelberg.

Prendinger, H., Mayer, S., Mori, J., Ishizuka, M., 2003. Persona effect revisited. Using bio-signals to measure and reflect the impact of character-based interfaces. In: Proceedings 4th International Workshop on Intelligent Virtual Agents (IVA-03), Springer, Berlin, pp. 283–291.

Prendinger, H., Dohi, H., Wang, H., Mayer, S., Ishizuka, M., 2004. Empathic embodied interfaces: addressing users' affective state. In: Proceedings Tutorial and Research Workshop on Affective Dialogue Systems (ADS-04), Springer, Berlin, pp. 53–64.

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, Cambridge.

Shneiderman, B., Maes, P., 1997. Direct manipulation vs. interface agents. Excerpts from debates at IUI 97 and CHI 97. Interactions, 42–61.

Thought Technology Ltd., 2002. http://www.thoughttechnology.com.

van Mulken, S., Andrè, E., Mïller, J.M., 1998. The Persona effect: how substantial is it? In: Proceedings Human Computer Interaction (HCI-98), Springer, Berlin, pp. 53–66.

Ward, R., Marsden, P., 2003. Physiological responses to different WEB page designs. International Journal of Human–Computer Studies 59, 199–212.

Wilson, G.M., Sasse, M.A., 2000. Listen to your heart rate: counting the cost of media quality. In: Paiva, A. (Ed.), Affective Interactions—Towards a New Generation of Computer Interfaces. Springer, Berlin, Heidelberg, pp. 9–20.