Estimating User Affect and Focus of Attention from Physiological Signals in Life-Like Character Based Interaction Scenarios

Helmut Prendinger
National Institute of Informatics (NII)
Department of Informatics, School of Multidisciplinary Sciences
The Graduate University for Advanced Studies
2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan
helmut@nii.ac.jp

Abstract

This paper abstract highlights some of our recent research efforts in designing and evaluating life-like characters that are capable of entertaining affective and social communication with users. The key novelty of our approach is the use of human physiological information: *first*, as a method to evaluate the effect of life-like character behavior on a moment-to-moment basis, and *second*, as an input modality for a new generation of interface agents that we call 'physiologically perceptive' characters. By exploiting the stream of primarily involuntary human responses, such as biometric signals and eye movements, those characters are expected to respond to users' affective and social needs in a truly sensitive, and hence effective, friendly, and beneficial way.

1 Introduction

Life-like character based interfaces are computer interfaces that emulate aspects of human face-to-face communication by using anthropomorphic animated agents. Those character agents do not only display multi-modal behavior in the form of speech, facial expressions, and gestures, but they may also perceive, and to a limited extent, understand the user [Prendinger and Ishizuka, 2004]. A salient feature of life-like character based interfaces is that they support affective communication with users, by expressing emotion verbally and non-verbally, and also recognizing the emotional state of the user. By using deictic gestures, those agents may also provide effective navigational aid.

2 The Impact of Empathic Character Response on Users' Level of Arousal

In this study, a life-like character acts as a quizmaster of a mathematical game [Prendinger *et al.*, 2005c]. This application was chosen as a simple, and for the sake of the experiment, highly controllable, instance of human–computer interaction. As skin conductance increases with a person's overall level of arousal or stress, the impact of empathic response was measured by comparing the skin conductance readings of subjects that received empathic feedback with the readings of subjects that did not.

A quiz game was 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 ($i \leq 4$). Subjects compete for the best score in terms of correct answers and time. They 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 the version of the game using the empathic agent, an apology as depicted in Fig. 1 was shown to subjects, while in the non-affective version the occurrence of the delay was simply ignored by the animated instructor.

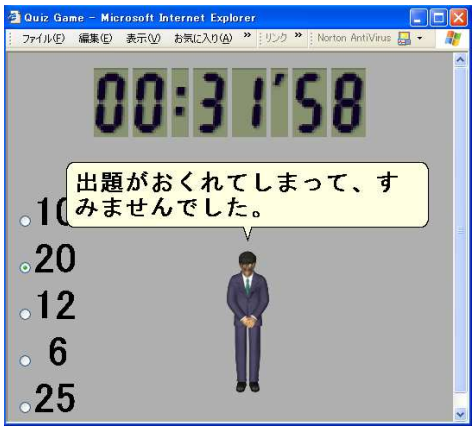

Figure 1: Shima character: "I apologize that there was a delay in posing the question."

In order to demonstrate the effect of the character's empathic behavior on the physiological state of subjects, we have been interested in two specific segments: (i) 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; (ii) the RESPONSE-TO-DELAY segment refers to the period when the character expresses empathy concerning the delay, or ignores the occurrence of the delay.

The main results of the study are as follows:

- If a character shows empathy to a deliberately frustrated user, then skin conductance is significantly lower than when the character does not display empathy, as compared to the induced frustration period.
- A character with affective (including empathic) behavior has a positive effect on users' perception of task difficulty, e.g. users perceive the game as less difficult.

The result indicates that empathic feedback may *undo* some of the user's negative emotions. On the other hand, affective behavior could not be shown to have an impact on users' performance in playing the quiz game.

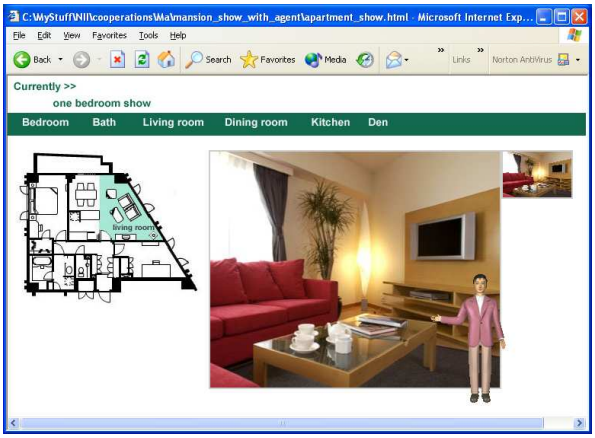

Figure 2: The Kosaku character presents the living room.

3 Physiologically Interactive Gaming

We recently conducted a study that processes affective states (derived from skin conductance and electromyography) in real-time, for users playing the cards game "Skip-Bo" against the 3D character Max (developed at the Univ. of Bielefeld), which extends our previous work on the Empathic Companion [Prendinger *et al.*, 2005b].

Since Skip-Bo is a competitive game, two types of empathy for the user could be implemented.[1] In the *positive empathic* version, the character displays happiness if the user is detected to be in happy or relaxed affective state. In the *negative empathic* version, the character will display e.g. gloating joy if the user is recognized to be negatively aroused. In both cases, the character will also display self-centered emotions, such as being happy about its own successful game move. As control conditions, the character will either display only self-centered emotions or no emotions at all. First results of this study indicate that the absence of negative empathy is conceived as significantly more stressful (derived from skin conductance) and irritating, as it might also be experienced when playing against a human player. A complementary result is that negative emphatic behavior induces negatively valenced emotions (derived from electromyography) in the user.

4 The Effect of a Character's Deictic Gestures on Users' Focus of Attention

For the eye tracking study, a presentation of an apartment located in Tokyo has been prepared using a web page based interface [Prendinger *et al.*, 2005a] (see Fig. 2). Three versions of the apartment show have been implemented:

- *Agent (& Speech)* version: the "Kosaku" character presents the apartment using synthetic speech and deictic facial and hand gestures;
- *Text (& Speech)* version: the presentation content of each scene is displayed by a text box and read out by Microsoft Reader;
- *Voice (only)* version: speech is the only medium used to comment on the apartment.

The same type of (synthetic) voice was used in all versions.

Results were distilled from applying both *spatial* (cumulative) and *spatio-temporal* analysis methods. Spatial analysis counts the gaze points that fall within certain screen areas and hypothesizes users' attentional focus. Our findings of the cumulative analysis include e.g. that users are looking mostly at the character's face, indicating users interact socially with agents. While a spatial analysis can indicate where attention is spent, it cannot reveal how users traverse the interface when watching a presentation. In order to address those more complex aspects of character-based interfaces, we also performed a spatio-temporal analysis.

The main findings are:

- An agent's referential (arm or facial) gestures may direct the user's focus of attention to the intended reference object better than a text box or only voice;
- If the uttered sentence contains a trigger word – a word that has a corresponding semantically related visualization – an agent using gestures helps users to locate the (visual) reference object quickly. By contrast, directional support by a text box or voice often shows considerable latency;
- Users often redirect their attention back and forth between the animated agent and the reference object.

Hence our findings also shed new light on *auditory language processing*, claiming that people who simultaneously listen to speech and are shown a visual object featuring elements that are semantically related to the spoken information, tend to focus on the elements that are most closely related to the meaning of the currently heard spoken language. We observed that auditory language processing has less latency when accompanied by according deictic gestures. Our observations also advance the discussion about the believability of life-like characters in that they contribute quantifiable evidence of the character's ability to direct the user's focus of attention to objects of interest.

[1] Demo at: `http://research.nii.ac.jp/~prendinger/demo/JapanMax_02.wmv`

Acknowledgments

This research was supported by the Research Grant (FY1999–FY2003) for the Future Program of the Japan Society for the Promotion of Science (JSPS), by a JSPS Encouragement of Young Scientists Grant (FY2005–FY2007), and by a NII Joint Research Grant (FY2005).

References

[Prendinger and Ishizuka, 2004] Helmut Prendinger and Mitsuru Ishizuka, editors. *Life-Like Characters. Tools, Affective Functions, and Applications*. Cognitive Technologies. Springer Verlag, Berlin Heidelberg, 2004.

[Prendinger *et al.*, 2005a] Helmut Prendinger, Chunling Ma, Jin Yingzi, Arturo Nakasone, and Mitsuru Ishizuka. Understanding the effect of life-like interface agents through eye users' eye movements. In *Proceedings of Seventh International Conference on Multimodal Interfaces (ICMI-05)*. ACM Press, New York, 2005.

[Prendinger *et al.*, 2005b] Helmut Prendinger, Junichiro Mori, and Mitsuru Ishizuka. Recognizing, modeling, and responding to users' affective states. In *Proceedings of Tenth International Conference on User Modeling (UM-05)*, LNAI 3538, pages 60–69. Springer, Berlin Heidelberg, 2005.

[Prendinger *et al.*, 2005c] Helmut Prendinger, Junichiro Mori, and Mitsuru Ishizuka. Using human physiology to evaluate subtle expressivity of a virtual quizmaster in a mathematical game. *International Journal of Human-Computer Studies*, 62(2):231–245, 2005.